International Journal on Advances in Systems and Measurements

2021 vol. 14 nr. 1&2

The *International Journal on Advances in Systems and Measurements* is published by IARIA.
ISSN: 1942-261x
journals site: http://www.iariajournals.org
contact: petre@iaria.org

Sponsored by IARIA
www.iaria.org

International Journal on Advances in Systems and Measurements
Volume 14, Number 1 & 2, 2021

Javier Calpe, Analog Devices and University of Valencia, Spain
Jaime Calvo-Gallego, University of Salamanca, Spain
Maria-Dolores Cano Baños, Universidad Politécnica de Cartagena,Spain
Juan-Vicente Capella-Hernández, Universitat Politècnica de València, Spain
Vítor Carvalho, Minho University & IPCA, Portugal
Irinela Chilibon, National Institute of Research and Development for Optoelectronics, Romania
Soolyeon Cho, North Carolina State University, USA
Hugo Coll Ferri, Polytechnic University of Valencia, Spain
Denis Collange, Orange Labs, France
Noelia Correia, Universidade do Algarve, Portugal
Pierre-Jean Cottinet, INSA de Lyon - LGEF, France
Paulo Estevao Cruvinel, Embrapa Instrumentation Centre - São Carlos, Brazil
Marc Daumas, University of Perpignan, France
Jianguo Ding, University of Luxembourg, Luxembourg
António Dourado, University of Coimbra, Portugal
Daniela Dragomirescu, LAAS-CNRS / University of Toulouse, France
Matthew Dunlop, Virginia Tech, USA
Mohamed Eltoweissy, Pacific Northwest National Laboratory / Virginia Tech, USA
Paulo Felisberto, LARSyS, University of Algarve, Portugal
Javad Foroughi, University of Wollongong, Australia
Miguel Franklin de Castro, Federal University of Ceará, Brazil
Mounir Gaidi, Centre de Recherches et des Technologies de l'Energie (CRTEn), Tunisie
Eva Gescheidtova, Brno University of Technology, Czech Republic
Tejas R. Gandhi, Virtua Health-Marlton, USA
Teodor Ghetiu, University of York, UK
Franca Giannini, IMATI - Consiglio Nazionale delle Ricerche - Genova, Italy
Gonçalo Gomes, Nokia Siemens Networks, Portugal
Luis Gomes, Universidade Nova Lisboa, Portugal
Antonio Luis Gomes Valente, University of Trás-os-Montes and Alto Douro, Portugal
Diego Gonzalez Aguilera, University of Salamanca - Avila, Spain
Genady Grabarnik,CUNY - New York, USA
Stefanos Gritzalis, University of the Aegean, Greece
Richard Gunstone, Bournemouth University, UK
Jianlin Guo, Mitsubishi Electric Research Laboratories, USA
Mohammad Hammoudeh, Manchester Metropolitan University, UK
Petr Hanáček, Brno University of Technology, Czech Republic
Go Hasegawa, Osaka University, Japan
Henning Heuer, Fraunhofer Institut Zerstörungsfreie Prüfverfahren (FhG-IZFP-D), Germany
Paloma R. Horche, Universidad Politécnica de Madrid, Spain
Vincent Huang, Ericsson Research, Sweden
Friedrich Hülsmann, Gottfried Wilhelm Leibniz Bibliothek - Hannover, Germany
Travis Humble, Oak Ridge National Laboratory, USA
Florentin Ipate, University of Pitesti, Romania
Imad Jawhar, United Arab Emirates University, UAE
Terje Jensen, Telenor Group Industrial Development, Norway
Liudi Jiang, University of Southampton, UK

Kenneth B. Kent, University of New Brunswick, Canada
Fotis Kerasiotis, University of Patras, Greece
Andrei Khrennikov, Linnaeus University, Sweden
Alexander Klaus, Fraunhofer Institute for Experimental Software Engineering (IESE), Germany
Andrew Kusiak, The University of Iowa, USA
Vladimir Laukhin, Institució Catalana de Recerca i Estudis Avançats (ICREA) / Institut de Ciencia de Materials de Barcelona (ICMAB-CSIC), Spain
Kevin Lee, Murdoch University, Australia
Wolfgang Leister, Norsk Regnesentral (Norwegian Computing Center), Norway
Andreas Löf, University of Waikato, New Zealand
Jerzy P. Lukaszewicz, Nicholas Copernicus University - Torun, Poland
Zoubir Mammeri, IRIT - Paul Sabatier University - Toulouse, France
Sathiamoorthy Manoharan, University of Auckland, New Zealand
Stefano Mariani, Politecnico di Milano, Italy
Paulo Martins Pedro, Chaminade University, USA / Unicamp, Brazil
Don McNickle, University of Canterbury, New Zealand
Mahmoud Meribout, The Petroleum Institute - Abu Dhabi, UAE
Luca Mesin, Politecnico di Torino, Italy
Marco Mevius, HTWG Konstanz, Germany
Marek Miskowicz, AGH University of Science and Technology, Poland
Jean-Henry Morin, University of Geneva, Switzerland
Fabrice Mourlin, Paris 12th University, France
Adrian Muscat, University of Malta, Malta
George Oikonomou, University of Bristol, UK
Arnaldo S. R. Oliveira, Universidade de Aveiro-DETI / Instituto de Telecomunicações, Portugal
Aida Omerovic, SINTEF ICT, Norway
Victor Ovchinnikov, Aalto University, Finland
Telhat Özdoğan, Amasya University - Amasya, Turkey
Gurkan Ozhan, Middle East Technical University, Turkey
Constantin Paleologu, University Politehnica of Bucharest, Romania
Matteo G A Paris, Universita` degli Studi di Milano,Italy
Vittorio M.N. Passaro, Politecnico di Bari, Italy
Giuseppe Patanè, CNR-IMATI, Italy
Marek Penhaker, VSB- Technical University of Ostrava, Czech Republic
Juho Perälä, Bitfactor Oy, Finland
Florian Pinel, T.J.Watson Research Center, IBM, USA
Ana-Catalina Plesa, German Aerospace Center, Germany
Miodrag Potkonjak, University of California - Los Angeles, USA
Alessandro Pozzebon, University of Siena, Italy
Vladimir Privman, Clarkson University, USA
Mohammed Rajabali Nejad, Universiteit Twente, the Netherlands
Konandur Rajanna, Indian Institute of Science, India
Nageswara Rao, Oak Ridge National Laboratory, USA
Stefan Rass, Universität Klagenfurt, Austria
Candid Reig, University of Valencia, Spain
Teresa Restivo, University of Porto, Portugal

Leon Reznik, Rochester Institute of Technology, USA
Gerasimos Rigatos, Harper-Adams University College, UK
Luis Roa Oppliger, Universidad de Concepción, Chile
Ivan Rodero, Rutgers University - Piscataway, USA
Lorenzo Rubio Arjona, Universitat Politècnica de València, Spain
Claus-Peter Rückemann, Leibniz Universität Hannover / Westfälische Wilhelms-Universität Münster / North-German Supercomputing Alliance, Germany
Subhash Saini, NASA, USA
Mikko Sallinen, University of Oulu, Finland
Christian Schanes, Vienna University of Technology, Austria
Rainer Schönbein, Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB), Germany
Cristina Seceleanu, Mälardalen University, Sweden
Guodong Shao, National Institute of Standards and Technology (NIST), USA
Dongwan Shin, New Mexico Tech, USA
Larisa Shwartz, T.J. Watson Research Center, IBM, USA
Simone Silvestri, University of Rome "La Sapienza", Italy
Diglio A. Simoni, RTI International, USA
Radosveta Sokullu, Ege University, Turkey
Junho Song, Sunnybrook Health Science Centre - Toronto, Canada
Leonel Sousa, INESC-ID/IST, TU-Lisbon, Portugal
Arvind K. Srivastav, NanoSonix Inc., USA
Grigore Stamatescu, University Politehnica of Bucharest, Romania
Raluca-Ioana Stefan-van Staden, National Institute of Research for Electrochemistry and Condensed Matter, Romania
Pavel Šteffan, Brno University of Technology, Czech Republic
Chelakara S. Subramanian, Florida Institute of Technology, USA
Sofiene Tahar, Concordia University, Canada
Muhammad Tariq, Waseda University, Japan
Roald Taymanov, D.I.Mendeleyev Institute for Metrology, St.Petersburg, Russia
Francesco Tiezzi, IMT Institute for Advanced Studies Lucca, Italy
Wilfried Uhring, University of Strasbourg // CNRS, France
Guillaume Valadon, French Network and Information and Security Agency, France
Eloisa Vargiu, Barcelona Digital - Barcelona, Spain
Miroslav Velev, Aries Design Automation, USA
Dario Vieira, EFREI, France
Stephen White, University of Huddersfield, UK
Shengnan Wu, American Airlines, USA
Qingsong Xu, University of Macau, Macau, China
Xiaodong Xu, Beijing University of Posts & Telecommunications, China
Ravi M. Yadahalli, PES Institute of Technology and Management, India
Yanyan (Linda) Yang, University of Portsmouth, UK
Shigeru Yamashita, Ritsumeikan University, Japan
Patrick Meumeu Yomsi, INRIA Nancy-Grand Est, France
Alberto Yúfera, Centro Nacional de Microelectronica (CNM-CSIC) - Sevilla, Spain
Sergey Y. Yurish, IFSA, Spain
David Zammit-Mangion, University of Malta, Malta

Guigen Zhang, Clemson University, USA
Weiping Zhang, Shanghai Jiao Tong University, P. R. China

International Journal on Advances in Systems and Measurements
Volume 14, Numbers 1 & 2, 2021

CONTENTS

pages: 162 - 177

Asia Techno Farm Initiatives for growing Future Farmers of Asia

Nobutaka Ito, Khon Kaen University, Thailand

Relation Analysis of Investment Behavior and Risk Expectation of Cryptocurrencies

Erik Massarczyk, Peter Winzer, Finn Jakob Müller
Faculty of Design – Computer Science – Media
RheinMain University of Applied Sciences
Wiesbaden, Germany
Email: erik.massarczyk@hs-rm.de, peter.winzer@hs-rm.de

***Abstract*—The technological development of the blockchain technology allows a new way of processing secured transactions and payments between different parties. Therefore, it is not surprising that new virtual currencies are developed to open new payment methods, as well as investment opportunities. To estimate the intention to buy and use cryptocurrencies, an empirical analysis was performed. The question of research is whether an investment in cryptocurrencies is primarily made for speculative reasons or because of a belief in the establishment of a digital currency. Although eight different cryptocurrencies are investigated, most respondents refer to Bitcoin as the one over all cryptocurrencies. The ordinary regression analyses on base of survey data, which was distributed online, outlines that the intention to use cryptocurrencies is mainly driven by investment purposes.**

Keywords-investments; cryptocurrencies; risk; experience; performance expectancy; mean analysis; UTAUT2.

I. INTRODUCTION

The following article is an expanded work presented on the CENTRIC conference [1].

Cryptocurrencies have achieved market capitalization of currently around 250 billion euros due to the strong growth in recent years [2]. On the one hand, investors see cryptocurrencies as an opportunity to reach high revenues accompanied with a specific (potentially high) risk, while on the other hand, researchers and experts see cryptocurrencies as opportunity to create a new known and general accepted currency and payment method [3][4]. Therefore, it will be analyzed what most private customers/users think about cryptocurrencies (e.g., Bitcoin or Ethereum) and how cryptocurrencies are used. To estimate the described customer behavior, quantitative research using an online survey is applied. The resulting database will be analyzed using statistical techniques for data estimation and the statistical program Statistical Package for the Social Sciences (SPSS), which targets on the estimation of results about the later described hypotheses.

In this respect, the paper is structured as follows. In Section II, (a) cryptocurrencies, (b) blockchain, (c) digital versus traditional currencies, as well as (d) challenges will be described. Section III will include the used research model. Following this section, the methodology and the theoretical approach for carrying out the analysis, will be explained. In Section V, the results of the empirical analysis are presented. The paper concludes with a summary of the results in Section VI.

II. TECHNICAL BACKGROUND

The following background section covers the definition of the research objectives cryptocurrencies and blockchain as well as the used research and conceptual model as well as the challenges in the named research field.

A. Cryptocurrencies

Although the first ideas to develop a digital and anonymous currency date back to 1989 [5], the first virtual cryptocurrency was implemented in 2008 [6], when Nakamoto published an approach for an electronic payment system and a new currency "Bitcoin" based on blockchain technology [7]. This approach differed from earlier approaches in particular in that all transfers must be validated by the community. This validation was performed decentral using a synchronized blockchain across multiple users [8]. To this extent, no third party is required as an intermediary to carry out secured transaction. This means that the currency Bitcoin was created primarily with the intention that transmissions can be cryptographically secured and tracked [6][9]. In addition, cryptocurrencies based on blockchain technology are implemented to (a) guarantee fast worldwide money transfers, (b) establish the privacy of the participating parties through anonymity, and (c) advance the development of a payment system independent of the traditional banking system [4].

B. Blockchain

Following Nakamoto [6], a blockchain is a continuously expandable list of data records, called "blocks", which are linked together by cryptographic methods. Each block typically contains a cryptographically secure hash of the previous block, a timestamp and transaction data [10]-[12]. The blockchain allows the linking of transfers within a decentralized platform, which is distributed and publicly assessable [13][14], where through recording of transfers, processes and information are secured by cryptographic techniques [11][15][16]. The fact that a large number of users of the blockchain can access and track the linked blocks within the blockchain creates confidence in the reliability of the digitally applied processes and transfers

[8][11][12][14][16]. Finally, blockchain provides a solution for a trusted, secure, decentralized and (by consensus) peer-validated approach [17]. In general, the entire database is embedded in a peer-to-peer network architecture with equal nodes. Due to the node principle, the system is not dependent on a central location, which could be a single point of failure [4].

Since, the information and data are implemented in the blockchain, which is decentralized distributed, no information can get lost [14]. Any implemented block is irreversibly linked to a previously block and cannot be deleted. Each block contains information about transactions and information of the previous block [6]. A new block is only added in case the verification through the validation and consensus process by the community is done [18]. Any update needs to be performed in a new developed block, which needs to be verified by the described process [19].

The application of blockchains guarantees a technically secure communication on the base of mutual authentication, as well as tamper-resistant asymmetric cryptography, which enables an information exchange by timestamped and logged records [8][13][20][21][22][23]. The blockchain approach implies the irrevocability of changes, i.e., the blocks or information remain permanently in the system and cannot simply be deleted [8][19]. The security mechanisms are implemented to avoid any spam and denial-of-service attacks [24].

The interaction of users within the blockchain takes place by using a related key pair, which comprises a private key and a public key [25]. The latter is publicly visible and comparable to an address that each node has; it can be regenerated for each transaction in order to maintain anonymity. If a node wants to create a transaction and, e.g., add new data, this can be done anytime autonomous by signing it with its (secret) private key [26]. It is then sent to all nodes of the peer-to-peer network to reduce single point of failures [16]. Each node is then able to use the public key to verify the node that created the transaction before a distributed consensus mechanism regulates the addition of the new block [27][28]. A consensus mechanism implemented through the Distributed Ledger Technology ensures that there is only one next block, which is necessary to obtain integrity of the blockchain [16][27][29]. This means that the consensus mechanism ensures that the transactions and blocks are sorted chronologically, which verifies the integrity, coherence and consistency of the blocks sustaining in the blockchain [8][16][20].

A subsequent update process ensures that all participants always have the latest version of the database at their disposal [30]. There are several methods for validating the transaction and reach consensus. The most common of which are currently known as 'Proof of Work' and 'Proof of Stake'. In these two methods, hash values are generated by the network nodes according to a certain pattern. Depending on the length of the blockchain, the degree of difficulty and the computing power required for this increase. In this context, the working nodes are also referred to as 'miners' [28]. The type of the utilized consensus mechanism varies in dependency from the type of network and other factors [26]. In summary, when a transaction is validated, it is stored in the block and chained in the blockchain [16], with the community deciding on the validation. I.e., this validation could only be manipulated by someone who has control over the majority (> 50%) of nodes, which is extremely unlikely due to the worldwide decentralized networks [8]. The timestamp documents (transparently for the whole network) the time of implementation and adjustments [31].

C. Digital versus Traditional Currencies

The main differences between traditional and digital currencies are: (a) The digital currencies are organized decentral using block-chain technology and do not require banks or other intermediaries (unlike traditional currencies). (b) The digital currencies are (uniformly) valid and available worldwide [32], while the traditional currencies are generally specific to individual states or economic areas [33]. The use of traditional currencies (especially for international transactions) results in relatively high transaction costs, whereas digital currencies cause no or only very low transaction costs due to the consensus mechanism and the very fast "automatic" validation of transactions [4][32][34]. (c) Digital currencies offer a high degree of anonymity and protection of personal data, which is not provided by traditional currencies (e.g., credit card payments or money transfers). In traditional currencies, this anonymity could only be achieved through cash payments, but the transaction costs are extremely high. In addition, cash payments are strongly limited or regulated in many countries.

Another central feature of a currency is that it is always available, transportable, and divisible. This is also true for cryptocurrencies [35].

In contrast to the traditional currencies, each cryptocurrency has a fixed limit regarding the maximum currency units that can be issued [35].

D. Challenges

Due to this "gap" regarding the legal and regulatory framework, there are potential uncertainties regarding the clarification of possible conflicts between trading partners [11][36].

In particular, the 'Proof of Work' mechanism causes extremely high-energy consumption, which is a factor of several thousand higher than traditional financial transactions [37].

For a long-term success, a digital currency (using blockchain technology) must achieve the acceptance of the majority of the population. After all, the long-term importance of the digital currency ultimately depends on the number of actual users and the acceptance as a payment system by the trade [38].

III. Research Model – Adjusted Model with Elements of the Unified Theory of Acceptance and Use of Technology 2

In this section, the used research model will be described. The focus in this research paper will be on the relationships between (a) the risks of cryptocurrencies and the intention to use cryptocurrencies, (b) the experiences with cryptocurrencies and the intention to use cryptocurrencies, and (c) the general experiences with investments and the intention to use cryptocurrencies. The analysis of the named research concepts follows the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2), which developed by Venkatesh, Thong, and Xin [39]. The UTAUT2 expands the existing UTAUT by the additional elements of hedonic motivation, price, and habit/experience, which allows a broader consideration of critical influence factors on the user behavior and the behavioral intention to use [39]-[41].

For this reason, to estimate these and further relations, an adjusted approach of the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) will be used, which is displayed in the conceptual model in Figure 1.

In principle, it can be assumed that higher returns or expected returns are generally associated with a higher investment risk. In this respect, it is necessary to examine how much risk they are prepared to take in order to achieve high returns. It can be assumed that investors who have more experience with investments and who have often made these via digital channels (e.g., online banking) are generally more open to the use of cryptocurrencies.

Finally, it should be noted that so far there has been no scientific review of the relationship between (a) performance expectations, (b) experience, (c) perceived risk and behavioral intention to use cryptocurrencies. The variable of perceived risk is treated as external variable in the further analysis. Additionally, the strength of perceived risk and experience will be estimated by linking these variables with the performance expectancy. The performance of the investments in digital currencies is rated by the performance expectancy.

Problematically, (a) the expected performance, (b) experience, and (c) perceptions of risks differ between the individual customers [42]. This means, the user attitudes and beliefs are completely subjective [42]. The experience comes from the fact that users become more and more familiar with a technology or service after it has been used for the first time. [40][43][44]. With the increasing use of a technology or a service, the user gains more and more experience and knowledge and learns with it, whereby the use becomes more and more self-evident and "automatic" [45].

Since habits and experiences allow predictions to be made for later use, it can be predicted that experience positively influences the utilization of cryptocurrencies.

In principle, the existing risks influenced the uses and investment behavior of customers [46]. This is particularly reflected in the fact that the risk has a significant influence on customer acceptance of innovations (e.g., mobile payment, mobile banking, and mobile shopping) [47]-[52].

Previous research identified that the perceived risk is one of the key drivers for the estimation of uncertainties in mobile payments, mobile shopping, mobile banking, and mobile transactions [47][49]-[53], because customers fear a lack of control.

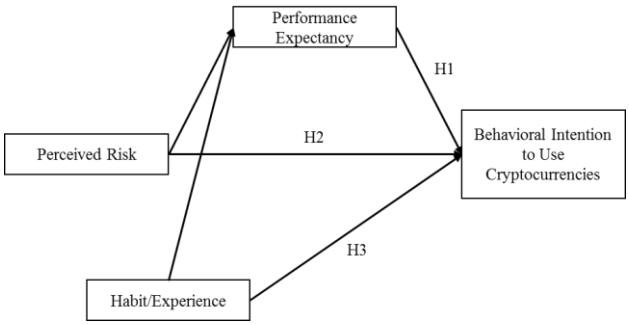


Figure 1. Conceptual Model

Consequently, the literature conveys the feedback that in several cases risks and uncertainties influence the customer user behavior.

Based on the explanations, the hypotheses for this research paper are:

H1: The customer perception of performance expectancy of investments (including digital currencies) has a directly positive effect on the intention to use cryptocurrencies.

H2: Customers' experience with investments (including digital currencies) has a direct positive impact on their intention to use cryptocurrencies.

H3: If customers are generally affine to risk when making investments, this has a direct positive effect on the intention to use cryptocurrencies.

IV. Methodology

In this section, the approach for the verification of the hypotheses will be shortly introduced. Therefore, to test the hypotheses, an online survey was carried out to obtain information on the investment behavior of private individuals. In particular, the survey covered the perception of users regarding cryptocurrencies and the resulting investment behavior. The focus here is on the perceived performance of digital currencies and investments made.

To achieve the needed user information, a cross-sectional online survey ("one-shot survey") had been prepared and distributed through multipliers in social media platforms [54].

As this is an online survey, it cannot be guaranteed (as opposed to a personal survey) that most respondents will fully answer the questions. In addition, the questionnaire was designed in such a way that individual questions could not be skipped without ending the survey. In this respect, a relatively large number of participants prematurely terminated their

responses to the questionnaire. The survey was distributed during the period from May to June 2018. In this period, 155 people have opened and started the questionnaire. However, only 62% (96 out of 155) respondents have finished the questionnaire. For this reason, the sample of the whole analysis will be the data set covering the 96 respondents, which have fully completed the questionnaire.

In the first part, the demographic information (such as age, gender and educational level) of the respondents was collected. In the following part the previous investment behavior and the knowledge of the participants about cryptocurrencies was determined. It should be determined whether the respondents know cryptocurrencies and whether they have already made investments based on cryptocurrencies. A positive answer (= experience with cryptocurrencies) was used to determine in more detail how many transactions, how much with which cryptocurrency the participants have already carried out. Since the third part is in higher importance for the later data analysis, all the implemented questions were coded in the 5-Point-Likert-scale format [55]. The third part covers especially questions regarding the respondent investment intentions in cryptocurrencies. In addition, the risk appetite and expected return (5-Point-Likert-scale: high to low) are important information in this part. The subsequent fourth part of the questionnaire considers questions regarding the user perception about the course of the cryptocurrency investment. As in the part before, the questions are coded in 5-Point-Likert-scale format (very likely to very unlikely). In the last part of the questionnaires, the respondents were queried about the future of cryptocurrencies in general.

The collected data were analyzed using quantitative research methods and the SPSS statistical program. To examine the reliability and validity of the data, the estimation of the Cronbach's Alpha and the Exploratory Factor Analysis were performed.

Only the eight largest cryptocurrencies (measured by market capitalization) were taken into account.

As mentioned above, the used approach only contains elements of the UTAUT2. Therefore, the evaluation is not done by structural equation modeling [39]. Instead, an ordinary least squares regression to test the significance of each hypothesis is used. In the final hypothesis, all the previous considered single variables, like (a) perceived performance, (b) experience, (c) risk appetite, (d) investment and speculation type, (e) regulations, and (f) assessment of the acceptance as alternative payment method will be related to the undertaking of investments in cryptocurrencies.

V. Data Analysis and Results

Following the described approach in conducting the survey, the outcomes for the estimation of the hypotheses will be deeply illustrated.

A. The Difference of Cryptocurrency Applications

Before going in-depth to the analysis of the questionnaire, one aspect is necessary to remind. Although most of the cryptocurrencies base on the blockchain and the wallets of the customers are numeric codes in general, cryptocurrencies are introduced for different purposes. Therefore, e.g., Bitcoin is implemented for solving financial transactions, whereas e.g., Ethereum is commonly used for establishing smart contracts. From this point of view, it would be preferable in differentiating the outcomes in relation to their purpose.

Despite the descriptive results as well as mean considerations will indicate some outcomes differentiated on the respective cryptocurrencies. Most respondents refer to Bitcoin as the one over all cryptocurrencies. In this regard, the outcome is possibly kind of biased. Although further investigations would be preferable, the low number of respondents investing in other cryptocurrencies than Bitcoin and Ethereum does not allow a representative result. Therefore, a greater statistical differentiation cannot be pursued.

B. Descriptive Results

In the following, the descriptive results of the performed survey will be introduced. 53.1% of the respondents are male and the average age of a respondent is between 25 and 29 years.

With 41.7% respectively 24.0%, the group of the 18 to 24 years respectively 25 to 29-year-old persons have the largest proportions of respondents within the survey (see Figure 2). On the base that the age group of the 20- to 29-year-old persons has only a 12.2% share of the total population in Germany, it must be noted that the young persons below the age of 30 years old are overrepresented in the survey by a factor of approx. five [56]. Since cryptocurrencies are virtual goods, their use requires a high Internet affinity. Based on a study of ARD/ZDF from 2015 the age group of the 20 to 29-year-old persons does nearly complete use the internet [57].

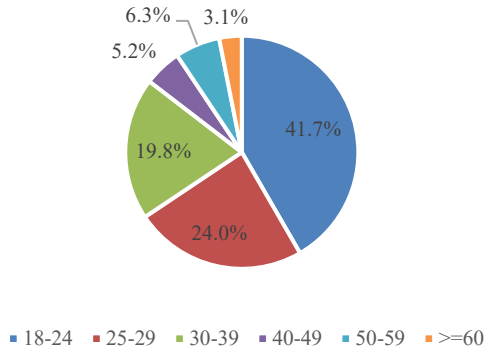


Figure 2. Age Distribution

Since younger people generally use the Internet more often and have a greater interest in virtual goods than older people

have, the previously established overrepresentation of younger age groups is not surprising. With regard to the age, the survey is not representative for the total population of Germany.

Considering the educational background in Figure 3, nearly the half of the respondents (46.9%) state that the school leaving examination is the education degree what they have. A quarter of the respondents have completed the Master degree (25%) from university.

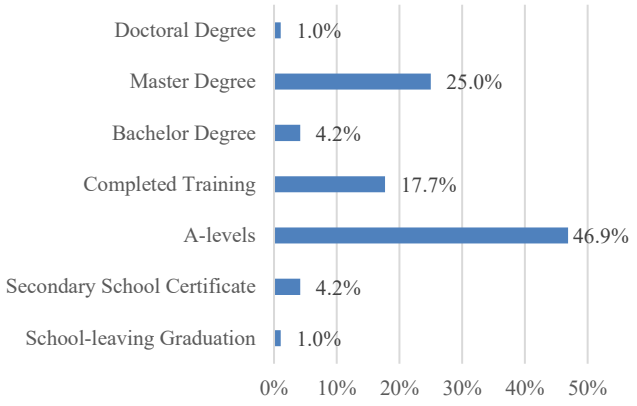


Figure 3. Distribution of Graduation

The average net household income of the respondents is between 1,000 and 1,999 euros per month, with most of the participants (36.3%) having a net (household) income of less than 500 euros per month. In addition, almost three quarters (73.8%) have a net (household) income of less than 2,000 euros. In connection with the level of education and age, it can be assumed that the interviewees are predominantly students.

90.6% (= 87/96) of the respondents know what cryptocurrencies are. These 87 persons are the basic population (= 100%) for questions about cryptocurrencies.

47.9% (= 46/96) of the respondents have already made financial investments. However, only 35.4% (34 of 96 respondents) have already done investments in or with cryptocurrencies. From this point of view, the 34 respondents will be the basic population (= 100%) for all questions regarding the investment behavior with cryptocurrencies, especially number of transactions, amount of invested financial resources and perceptions regarding the development of the invested portfolio.

Firstly, the descriptive results for the respondents, who know cryptocurrencies (=87), will be illustrated. In general, all the respondents know Bitcoin as cryptocurrency, whereas two thirds of the respondents answer to know BitcoinCash and Ethereum, which can be seen in Table I.

Although 36.9% of the respondents are very risk-affine with regard to investments, only 23.4% of the respondents describe themselves as speculators. Contrary, 39.3% of the respondents answer to have a risk-shy nature, which can be also seen in estimation that 37.8% of the respondents estimate to be arbitragers. By regarding the estimation of returns, only 21.7% of the respondents think to get low returns. Although it is well known that higher returns can only be achieved with higher risks, some of the respondents who are risk-averse hope for medium to high returns.

TABLE I. DEGREE OF AWARENESS OF CRYPTOCURRENCIES

Cryptocurrency	Degree of Awareness
Bitcoin	**100.0%**
Bitcoin Cash	**67.1%**
Ethereum	**66.7%**
Litecoin	**61.4%**
Ripple	**58.5%**
EOS	**45.8%**
Neo	**41.0%**
Cardano	**35.4%**

Interestingly, 87.8% of respondents think that the new cryptocurrencies have been brought to life to drive a new form of speculation and investment. This is underlined by the fact that only 22.9% of respondents see Bitcoin as an alternative payment method to credit cards and the like. 43.4% of respondents involved in investment argue for regulatory intervention or restrictions in the cryptocurrency market, while 32.5% reject it.

Now, the results of the respondents using cryptocurrencies are shown. As already mentioned, however, the sample size is very small with 34 respondents, which is why the results cannot be generalized.

Figure 4 shows the distribution of the investment in the eight most important cryptocurrencies. 83.0% of respondents have already invested in Bitcoin. In addition to Bitcoin, the currency Ethereum seems to be of particular interest to investors.

48.5% of the respondents have invested at least 2,000 Euro in cryptocurrencies. 67.6% of these investors state that they make a profit by investing in cryptocurrencies. However, 51.4% of the investors have only short-time experience with cryptocurrencies since they invest in them for the last 2 years. Cryptocurrencies have become much more popular, especially in recent years. It is therefore not surprising that many investors have only recently started to invest in and trade with cryptocurrencies. Due to this short experience time horizon, most investors have an extremely limited ability to assess the long-term performance of cryptocurrencies and the risk of an investment. In this respect, one would normally expect investors to be generally extremely cautious and careful in assessing the development and returns of cryptocurrency investments.

However, the results of the questionnaire, illustrated in Figure 5, show that investors have a vastly different assessment of the performance of their currency investments for each cryptocurrency.

Most of the investors (82%) assumes that a total loss of the investment does not occur. Contrary, over the half of the investors assumes to gain profits (in 6 months: 55%; in 12

months: 71%). Considering, the probability to get a loss in the investments, 58% of the investors estimate this as unlikely within the next 12 months.

Interestingly, in case investors decide to invest in cryptocurrencies, they take a couple of cryptocurrencies. No one of the investors takes only one cryptocurrency. 54.5% of the investors invest in more than 5 different cryptocurrencies. Diversification of financial resources across different investments indicates hedging, if one cryptocurrency fails, other cryptocurrencies can compensate for the loss. Such investor behavior is typical for brokers who trade in different financial securities to make profits. It can therefore be assumed that cryptocurrencies are bought more as a tool to make (speculative) profits than as an alternative means of payment or as a tool to improve trading or production processes.

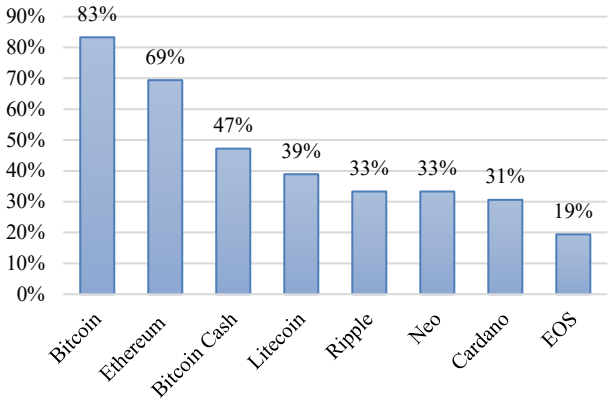


Figure 4. Investments in Cryptocurrencies

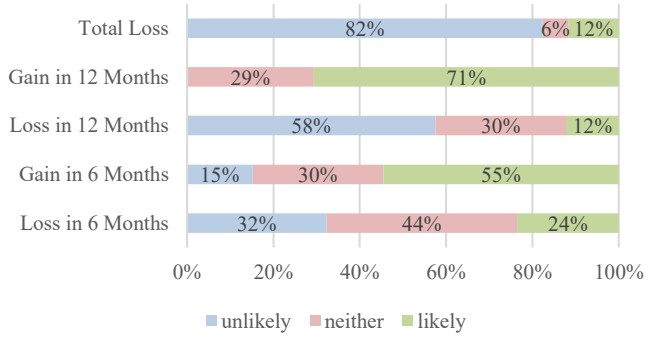


Figure 5. Expectations regard to the Cryptocurrency Investment

Figure 6 shows the objectives of the investments. In general, most investors in cryptocurrencies believe in long-term increases in value. In comparison to the overall group of respondents knowing and using cryptocurrencies, the users of cryptocurrencies believe in a higher degree that Bitcoin could develop to an alternative currency and payment method.

In general, over 50% of the investors have a long-term direction by investing in cryptocurrencies. In this respect, the investment in these currencies usually takes place with a longer time horizon (of several years).

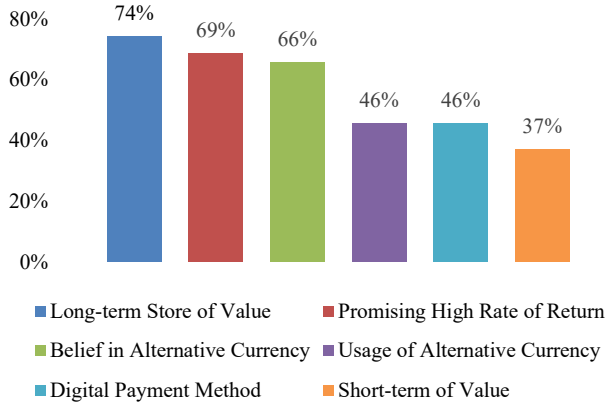


Figure 6. Purpose in Cryptocurrencies

Overall, the investment atmosphere regarding cryptocurrencies is quite positively. Investors perceive high profits by doing the investments and see only minor risks of a loss of their investments.

C. *Reliability and Validity*

The results of the reliability and validity analyses are illustrated in Tables II and III. In general, this study includes the following 7 concepts: (1) performance expectancy, (2) experience, (3) perceived risk, (4) intention to use cryptocurrencies, (5) purposes of investments in cryptocurrencies, (6) usage of cryptocurrencies, and (7) prominence of cryptocurrencies.

TABLE II. RELIABILITY ANALYSIS

Research Concepts	Cronbach's Alpha
Performance Expectancy	**0.335**
Experience	**0.282**
Perceived Risk	**0.779**
Intention to Use Cryptocurrencies	**0.624**
Purposes of Investments in Cryptocurrencies	**0.777**
Usage of Cryptocurrencies	**0.726**
Prominence of Cryptocurrencies	**0.947**

Generally, all named concepts are examined in the terms of reliability and validity. Following Cronbach, Alpha values must be higher than 0.7/0.6 to for a good/sufficient reliability [58]-[60]. Based on the results in Table II, the collected data for 5 of the 7 named aspects are at least sufficiently reliable. Solely, the concepts of performance expectancy and experience seem to be completely not reliable.

After the testing of the reliability, the exploratory factor analysis includes the assessment of Kaiser-Meyer-Olkin criterion (KMO), the significance test from Bartlett, and the examination of the cumulative variance to evaluate the

validity of the collected data [61]-[65]. Validity considers the consistency of an empirical measurement with the based conceptual/logical measurement concept. To reach a good validity, the concepts should reach significant p values ($p<0.05$) in the Bartlett-Test and KMO values above 0.7 (at least higher than 0.5) [61]-[65].

Table III shows a sufficient validity for 6 of the 7 concepts. The validity scores are also supported by the results of the cumulative variances higher than 50% except the concept of experience.

TABLE III. VALIDITY ANALYSIS

Research Concepts	KMO	Bartlett-Test	Cumulative Variance
Performance Expectancy	**0.284**	p < 0.000	**78.844%**
Experience	**0.562**	p < 0.000	**47.657%**
Perceived Risk	**0.637**	p < 0.000	**71.614%**
Intention to Use Cryptocurrencies	**0.640**	p < 0.000	**64.520%**
Purposes of Investments in Cryptocurrencies	**0.686**	p < 0.000	**74.544%**
Usage of Cryptocurrencies	**0.642**	p < 0.000	**69.377%**
Prominence of Cryptocurrencies	**0.911**	p < 0.000	**74.529%**

Despite the mark of 50% is not completely achieved, the explanatory rates of the variances can be rated as sufficiently high [62]-[64]. Consequently, the reliability and validity of the collected data are proved.

D. Mean Analysis

Before the relations between the different concepts will be illustrated in-depth, several mean analyses should give an insight if specific characteristics have an impact on the knowledge about cryptocurrencies as well as on the risk assessment and the development of the investment.

Here, the Analysis of Variances (ANOVA) figures out if there are differences in means between the different groups of a variable (mostly the two specifications of a binary-coded variable) are considered. A difference in means gives a direct suggestion that possibly variables have an impact in changing the variable significantly. In more simplified words, the ANOVA-test reveals in a first stage, if possible, variables relate to the investigated variable. By performing the ANOVA, the F-Test needs to be greater than 3.90 to determine a significant difference in means [64].

Table IV indicates the differences in means in gender regarding the eight mostly used cryptocurrencies. Most importantly, the prominence is coded in a 3-point-scale with one for unknown, two for heard but unknown about the use, and three for known and trusted in the use. Considering the following Tables IV, V and VI, the ANOVA-tests identify significant differences in means between the two regarded groups. Since all F-values are above 3.90, it must be concluded that female and male respondents have different knowledge on average.

TABLE IV. MEAN ANALYSIS GENDER VS INVESTMENT IN SEVERAL CRYPTOCURRENCIES

	Gender	Mean	Difference in Means	ANOVA F
Bitcoin	**Female**	**2.09**	**0.73**	**53.726****
	Male	**2.72**		
Ethereum	**Female**	**1.31**	**1.25**	**104.713****
	Male	**2.56**		
Ripple	**Female**	**1.19**	**1.16**	**82.429****
	Male	**2.35**		
BitcoinCash	**Female**	**1.46**	**0.86**	**34.171****
	Male	**2.32**		
Litecoin	**Female**	**1.31**	**1.05**	**56.690****
	Male	**2.36**		
EOS	**Female**	**1.20**	**0.65**	**22.217****
	Male	**1.85**		
Cardano	**Female**	**1.06**	**0.77**	**31.452****
	Male	**1.83**		
Neo	**Female**	**1.11**	**0.78**	**31.573****
	Male	**1.89**		

* Significant within the error probability of 10%.
** Significant within the error probability of 5%.

TABLE V. MEAN ANALYSIS GENERAL INVESTMENT BEHAVIOR VS INVESTMENT IN SEVERAL CRYPTOCURRENCIES

	General Investment Behavior	Mean	Difference in Means	ANOVA F
Bitcoin	**No**	**2.10**	**0.69**	**72.219****
	Yes	**2.79**		
Ethereum	**No**	**1.41**	**1.19**	**86.333****
	Yes	**2.60**		
Ripple	**No**	**1.27**	**1.14**	**82.889****
	Yes	**2.41**		
BitcoinCash	**No**	**1.48**	**0.92**	**44.147****
	Yes	**2.40**		
Litecoin	**No**	**1.37**	**1.06**	**59.687****
	Yes	**2.43**		
EOS	**No**	**1.12**	**0.90**	**58.496****
	Yes	**2.02**		
Cardano	**No**	**1.05**	**0.90**	**51.721****
	Yes	**1.95**		
Neo	**No**	**1.10**	**0.90**	**49.569****
	Yes	**2.00**		

* Significant within the error probability of 10%.
** Significant within the error probability of 5%.

Also, significant differences in the means can be seen in the ANOVA-tests between the group of investors, who have done general and cryptocurrency investments, and the non-

investors in the knowledge about the eight most common cryptocurrencies.

In Table IV, we see that for the eight cryptocurrencies, male respondents on average have greater knowledge regarding cryptocurrency investments. Bitcoin is the most well-known cryptocurrency, while male respondents generally have at least a general idea about the other currencies. In contrast, female respondents hardly know any cryptocurrencies other than Bitcoin. Table V shows that respondents who invest more frequently (regardless of which areas) are on average knowledgeable about the eight cryptocurrencies.

TABLE VI. MEAN ANALYSIS INVESTMENT BEHAVIOR IN CRYPTOCURRENCIES VS INVESTMENT IN SEVERAL CRYPTOCURRENCIES

	Cryptocurrency Investment Behavior	Mean	Difference in Means	ANOVA F
Bitcoin	No	2.08	0.92	362.195**
	Yes	3.00		
Ethereum	No	1.46	1.39	172.331**
	Yes	2.85		
Ripple	No	1.36	1.23	102.244**
	Yes	2.59		
BitcoinCash	No	1.49	1.15	86.842**
	Yes	2.64		
Litecoin	No	1.38	1.32	134.596**
	Yes	2.70		
EOS	No	1.18	1.00	79.654**
	Yes	2.18		
Cardano	No	1.04	1.19	141.566**
	Yes	2.23		
Neo	No	1.10	1.14	113.228**
	Yes	2.24		

* Significant within the error probability of 10%.
** Significant within the error probability of 5%.

TABLE VII. MEAN ANALYSIS GENDER VS RISK EXPECTATION

	Gender	Mean	Difference in Means	ANOVA F
Risk Appetite	Female	2.39	0.99	15.802**
	Male	3.38		
Expected Return	Female	2.97	0.55	6.431*
	Male	3.52		
Risk Type	Female	1.41	0.76	24.588**
	Male	2.17		
Bitcoin Payment Method	Female	2.26	0.47	4.051*
	Male	2.73		
Cryptoinvestment Speculation	Female	4.00	0.23	2.292
	Male	4.23		
Regulation of Cryptoinvestment	Female	3.14	0.14	0.131
	Male	3.04		

* Significant within the error probability of 10%.
** Significant within the error probability of 5%.

In particular, investors are generally familiar with Bitcoin and Ethereum. Conversely, people who do not invest regularly are, on average, unfamiliar with most of the eight cryptocurrencies surveyed. Only Bitcoin reach a greater prominence. Comparing these results with the responses of investors who regularly invest in in cryptocurrencies (see Table VI), the results are almost the same.

Based on the further tables, the mean analysis should distinguish between the risk and return expectation regarding cryptocurrencies. Besides the risk type, which is coded with 3-point-Likert-scale (1 = arbitrager to 3 = speculator), the other questions are coded in a 5-point-Likert-scale manner. For the risk and return expectation, the 1 stands for a low-level-expectation, whereas the 5 means a high-level expectation. For the other variables, the 5-point-Likert-scale is going from do not agree (1) to agree (5).

TABLE VIII. MEAN ANALYSIS GENERAL INVESTMENT BEHAVIOR VS RISK EXPECTATION

	General Investment Behavior	Mean	Difference in Means	ANOVA F
Risk Appetite	No	2.60	0.71	7.768**
	Yes	3.31		
Expected Return	No	3.02	0.53	5.946*
	Yes	3.55		
Risk Type	No	1.58	0.54	11.502**
	Yes	2.12		
Bitcoin Payment Method	No	2.39	0.28	1.379
	Yes	2.67		
Cryptoinvestment Speculation	No	3.95	0.36	6.075*
	Yes	4.31		
Regulation of Cryptoinvestment	No	3.12	0.07	0.072
	Yes	3.05		

* Significant within the error probability of 10%.
** Significant within the error probability of 5%.

Considering Table VII, the ANOVA-test identifies that regarding the risk expectations, the estimations between male and female respondents differ significantly. For the three questions to risk appetite, expected return, and risk type, the F-scores exceed the mark of 3.90.

In average, the female respondents answer that they take a lower risk if they are doing investments, especially cryptocurrency investments. In this regard, the female respondents seem to be risk-averse, whereas the male respondents seem to be risk neutral. Considering the risk taken, male respondents on average expect a medium-high return, whereas the return expectation of female respondents seems to be lower.

Despite the differences regarding the risk/return expectation, in average, independent from the gender, the respondents report that they see cryptocurrencies just for speculation perspectives. Regarding a possible regulation, the respondents do not give sophisticated feedback, independent from the gender. Accordingly, most respondents (of both genders equally) do not see Bitcoin as an alternative payment method. These results are supported by the ANOVA, as for the three questions a (weakly) significant difference in the means can only be seen for the question about Bitcoin as a means of payment. For the other two variables, the F-scores do not exceed 3.90, indicating that female and male respondents do not answer differently.

TABEL IX. BEHAVIOR IN CRYPTOCURRENCIES VS RISK EXPECTATION

	Crypto-currency Investment Behavior	Mean	Differ-ence in Means	ANOVA F
Risk Appetite	No	2.50	1.19	23.892**
	Yes	3.69		
Expected Return	No	2.90	0.98	24.098**
	Yes	3.88		
Risk Type	No	1.53	0.80	28.604**
	Yes	2.33		
Bitcoin Payment Method	No	2.32	0.53	5.046*
	Yes	2.85		
Cryptoinvestment Speculation	No	4.02	0.28	3.505
	Yes	4.30		
Regulation of Cryptoinvestment	No	3.30	0.54	3.866
	Yes	2.76		

* Significant within the error probability of 10%.
** Significant within the error probability of 5%.

In the following, Table VIII displays that the F-values are only above of the value 3.90 and significant for the relation of the general investment behavior and the risk expectations. Regarding the other parameters in Table VIII, they do not reach a F-value greater than 3.90.

Here, the outcomes of the general investment behaviors regarding the named expectations, do not change. This means, people, who invest in general, they anticipate a greater risk in doing that in comparison to people, who do not invest. However, in comparison with gender perspective, the differences between investors and non-investors are not heavily. Regarding the risk type and the risk appetite, in average both groups seem to be risk neutral.

With Table IX, the cryptocurrency investment behavior is considered regarding the difference in means of the already named expectations. Since the cryptocurrency investment field is a specific part of the general investment, it could be assumed that the general results are not mainly different. The outcomes show that the F-Scores of the risk expectations even exceed the previous consideration. As a result, the risk expectations are more resilient than in the consideration of the general investment behavior. Following this point, investors in cryptocurrencies believe in a greater risk. As the general group, the investment in cryptocurrencies does not trigger the opinion about any regulatory or governmental intervention.

However, overviewing the results of Table VII and Table IX, the outcomes for the questions regarding the usage of Bitcoin as payment method as well as the usage of cryptocurrencies for speculation purposes are differently. Table IX shows an F-value of 5.046 (and thus greater than 3.90) for the use of Bitcoin as a means of payment. Compared to the overall group of all investors, the cryptocurrency investor group is more open to using Bitcoin as an investment option on a larger scale. Cryptocurrency investors have a neutral position regarding the use of Bitcoin as a means of payment, while non-cryptocurrency investors (who are invested in other asset classes) tend to reject the use of Bitcoin as a means of payment on average. However, the group of cryptocurrency investors does not exclusively believe that crypto investments are made only for speculative purposes. In contrast, people who do not invest in cryptocurrencies (but invest in other asset classes) believe that cryptocurrencies are only made for speculative purposes.

Above of Table IX, Tables X and XI go in-depth of the cryptocurrencies Bitcoin and Ethereum in relation to the risk expectations. The outcomes of the performed ANOVA-tests indicate differences in means whether the respondents invest in Bitcoin or Ethereum.

Though, in comparison to the previous explanations, the ANOVA-tests identify that a difference in means between investors and non-investors of Bitcoin and Ethereum have not different expectations regarding risk and return of cryptocurrencies. All the F-tests are insignificantly, which identifies that averagely both considered groups decide similar in the questions. Since averagely there are no differences in the expectations, the in-depth consideration of Bitcoin and Ethereum does not add a new information.

TABLE X. MEAN ANALYSIS INVESTMENT BEHAVIOR IN BITCOIN VS RISK EXPECTATION

	Investment in Bitcoin	Mean	Difference in Means	ANOVA F
Risk Appetite	No	4.17	0.63	1.582
	Yes	3.54		
Expected Return	No	3.60	0.26	0.352
	Yes	3.86		
Risk Type	No	2.60	0.36	1.063
	Yes	2.24		
Bitcoin Payment Method	No	3.00	0.17	0.101
	Yes	2.83		
Cryptoinvestment Speculation	No	4.00	0.31	0.722
	Yes	4.31		
Regulation of Cryptoinvestment	No	3.40	0.71	1.053
	Yes	2.69		

* Significant within the error probability of 10%.
** Significant within the error probability of 5%.

TABLE XI. MEAN ANALYSIS INVESTMENT BEHAVIOR IN ETHEREUM VS RISK EXPECTATION

	Investment in Ethereum	Mean	Difference in Means	ANOVA F
Risk Appetite	No	3.80	0.22	0.256
	Yes	3.58		
Expected Return	No	3.33	0.67	3.922
	Yes	4.00		
Risk Type	No	2.22	0.10	0.119
	Yes	2.32		
Bitcoin Payment Method	No	2.90	0.07	0.025
	Yes	2.83		
Cryptoinvestment Speculation	No	4.00	0.38	1.802
	Yes	4.38		
Regulation of Cryptoinvestment	No	2.70	0.13	0.060
	Yes	2.83		

* Significant within the error probability of 10%.
** Significant within the error probability of 5%.

Considering Tables XII and Table XIII, the relation between the investments in Bitcoin and Ethereum are related to the return development as well as the belief in cryptocurrencies. In average, an investment or non-investment in Bitcoin does not lead to a different value expectation of the investment. In addition, investors or non-investors do not anticipate differently regarding cryptocurrencies. For Bitcoin investment in all cases the F-scores are below the mark of 3.90. For this reason, all the ANOVA analyses are insignificantly and a difference in means regarding the single variables and the investment in Bitcoin can be excluded.

Reviewing the investors in Ethereum in relation to the return expectations, it can be mainly determined that also the most ANOVA-tests have F-scores below the mark of 3.90. This means, the F-Tests are insignificantly. For this reason, there is no difference in means between the investors and non-investors regarding the different variables. However, in comparison to the investors in Bitcoin, one expectation indicates a significant F-test. Regarding the variable, a promising high rate of return, the F-score exceeds the mark of 3.90. On this account, there is a difference in means between the investors and non-investors in Ethereum. The variation is that non-investors have a neutral expectation regarding a promising high rate of return of Ethereum investments. Oppositely, averagely investors in Ethereum believe in a high rate of return.

TABLE XII. MEAN ANALYSIS INVESTMENT BEHAVIOR IN BITCOIN VS PERFORMANCE EXPECTANCY

	Investment in Bitcoin	Mean	Difference in Means	ANOVA F
Long-term Store of Value	No	3.67	0.43	0.474
	Yes	4.10		
Short-term Store of Value	No	2.40	0.80	1.371
	Yes	3.20		
Promising High Rate of Return	No	3.80	0.07	0.009
	Yes	3.87		
Digital Payment Method	No	3.20	0.03	0.002
	Yes	3.23		
Belief in Alter-native Currency	No	4.00	0.50	0.481
	Yes	3.50		
Usage of Alternative Currency	No	3.20	0.10	0.018
	Yes	3.10		

* Significant within the error probability of 10%.
** Significant within the error probability of 5%.

Although the outcomes have mainly no difference in means, it must be concluded that the investors in cryptocurrencies expect in general a long-term store of value by doing cryptoinvestments. Considering the beliefs in the cryptocurrencies as well as the short-term store of value, the investors in cryptocurrencies have generally a neutral attitude.

Concluding the ANOVA-tests by considering, how the investors in Bitcoin and Ethereum averagely estimate how their investments will develop in the upcoming year. Before going in the estimations of the average responses, it needs to be underlined that the investors of cryptocurrencies (independently if Bitcoin, Ethereum or something else), the investors do not anticipate a total loss of their investments. Furthermore, they more likely expect a rise of their investments.

TABLE XIII. MEAN ANALYSIS INVESTMENT BEHAVIOR IN ETHEREUM VS PERFORMANCE EXPECTANCY

	Investment in Ethereum	Mean	Difference in Means	ANOVA F
Long-term Store of Value	No	3.64	0.55	1.262
	Yes	4.21		
Short-term Store of Value	No	2.60	0.68	1.666
	Yes	3.28		
Promising High Rate of Return	No	2.90	1.34	7.917**
	Yes	4.24		
Digital Payment Method	No	2.70	0.58	1.028
	Yes	3.28		
Belief in Alter-native Currency	No	3.30	0.38	0.463
	Yes	3.68		
Usage of Alternative Currency	No	2.90	0.46	0.589
	Yes	3.36		

* Significant within the error probability of 10%.
** Significant within the error probability of 5%.

TABLE XIV. MEAN ANALYSIS INVESTMENT BEHAVIOR IN BITCOIN VS STORAGE DEVELOPMENT

	Investment in Bitcoin	Mean	Difference in Means	ANOVA F
Downtick in 6 Months	No	3.40	0.54	1.482
	Yes	2.86		
Uptick in 6 Months	No	4.40	1.01	5.587*
	Yes	3.39		
Downtick in 12 Months	No	3.60	1.35	9.062**
	Yes	2.25		
Uptick in 12 Months	No	4.20	0.23	0.380
	Yes	3.97		
Total Loss	No	2.00	0.07	0.018
	Yes	1.93		

* Significant within the error probability of 10%.
** Significant within the error probability of 5%.

This is also underlined in Tables XIV and XV with the positive estimation about an uptick of the cryptocurrency investment. Furthermore, investors in cryptocurrencies do not averagely expect a downtick of the investment within an investment period of 12 months. Interestingly, investors, who do not invest in cryptocurrencies but in other kinds of

investments, see cryptocurrency investments more critically and averagely expect a downtick of an investment in Bitcoin or Ethereum. In general, it can be expected that investors expect a greater risk in doing cryptocurrency investments. Oppositely, the non-investors in cryptocurrencies have a more positive opinion about the return progression of the cryptocurrency investments.

TABLE XV. MEAN ANALYSIS INVESTMENT BEHAVIOR IN ETHEREUM VS STORAGE DEVELOPMENT

	Investment in Ethereum	Mean	Difference in Means	ANOVA F
Downtick in 6 Months	No	3.10	0.22	0.415
	Yes	2.88		
Uptick in 6 Months	No	3.78	0.32	0.753
	Yes	3.46		
Downtick in 12 Months	No	3.00	0.75	3.742
	Yes	2.25		
Uptick in 12 Months	No	3.80	0.28	0.933
	Yes	4.08		
Total Loss	No	2.20	0.37	0.869
	Yes	1.83		

* Significant within the error probability of 10%.
** Significant within the error probability of 5%.

However, looking for the F-test, it must be stated that the ANOVA-analysis does not figure out a difference in means in the relation between the return progression and the investment in Ethereum. All the outcomes in Table XV do not reach a F-value greater than 3.90. The F-tests are insignificantly and there is not a difference in means in the expectations about the development of Ethereum between the investors and non-investors in Ethereum.

Considering the ANOVA for the relation between the investment in Bitcoin and the return expectation within the following 12 months. In general, three of five F-tests also identify values below the mark of 3.90, which are insignificantly and do not figure out a difference in means. However, regarding an uptick of Bitcoin within the next six months as well as a downtick in the next twelve months, the F-tests reach scores about the mark of 3.90. For these two questions, the investors and non-investors in Bitcoin have averagely a different expectation about the return development.

Interestingly, for the uptick within the six months, the non-investors in Bitcoin have a more positive expectation about the Bitcoin development than the investors in Bitcoin. Oppositely, within twelve months, the investors in Bitcoin reject a possible expectation about a downtick of the development, whereas the non-investors think about that averagely it would be likely that the Bitcoin investment gets a downtick.

E. Correlation Analysis

The correlation analysis measures the degree of the relationship between two individual variables. It is not, however, the degree of the linear proportionality. A correlation of 1.000 shows a 'perfect' relationship. A correlation coefficient higher than 0.500 is classified as a good correlation. Below 0.300, the correlation coefficients are weak [66][67].

The first correlation analyses build the pre step for the further investigations. For this reason, the correlations will be considered between the intent to use investment (to be an investor) and the variables considering the return and risk.

TABLE XVI. SIGNIFICANT CORRELATIONS FOR THE INTENTION TO USE INVESTMENTS

Variables	Correlation Coefficient
Risk Appetite	0.294
Expected Returns	0.262
Risk Type	0.355
Belief in Alternative Currency	0.349

TABLE XVII. SIGNIFICANT CORRELATIONS FOR THE INTENTION TO USE CRYPTOCURRENCIES

Variables	Correlation Coefficient
Total Loss	-0.349
Risk Appetite	0.475
Expected Returns	0.479
Risk Type	0.513
General Investment	0.728
Investment Duration	0.388
Year of First Investment	0.508
Bitcoin Alternative Payment Method	0.242
Long-term Store of Value	0.539
Short-term Store of Value	0.366
Promising High Rate of Return	0.511
Digital Payment Method	0.350
Belief in Alternative Currency	0.434
Usage of Alternative Currency	0.345

Based on the results in Table XVI, a positive significant relationship can be found between the general investment behavior as well as the risk taken and the expected return. Investors make investments to generate profits or to increase their financial resources. Therefore, the correlation with the expected return can be comprehended. Considering the

relation to the risk expectation, it can be followed those investors, who already have decided to invest, anticipate the possible risk they have to face.

TABLE XVIII. SIGNIFICANT CORRELATIONS FOR THE BITCOIN COURSE DEVELOPMENT

Variables	Correlation Coefficient
Uptick in 6 Months	**-0.391**
Downtick in 12 Months	**-0.476**

TABLE XIX. SIGNIFICANT CORRELATIONS FOR THE ETHEREUM COURSE DEVELOPMENT

Variables	Correlation Coefficient
Expected Returns	**0.440**

This means, investors are normally aware that they that they do not gain a financial profit or could lose their investments. Otherwise, non-investors, who reject to invest, are normally not willing to take the risks to possibly lose the financial resources. When investing, it must be clear that a loss as well as a possible is possible.

Table XVII shows the variables that have a significant correlation with the intention to use cryptocurrencies. In addition to the values shown in Table XVII: (a) There are positive significant correlations for all variables of perceived risk and experience with the intention of using cryptocurrencies. (b) From the concept of performance expectancy, the variable of the expectation regarding the total loss of an investment in cryptocurrencies correlates negatively significant with the intention to use cryptocurrencies. The negative correlation identifies that the investors in cryptocurrencies do not expect a total loss of their taken investments. This result corresponds with the positive correlations between the intention to use cryptocurrencies and the rate of return. On this account, users of cryptocurrencies expect to get a positive return from their investment.

Finally, in Tables XVIII and XIX, it can be remarked that the outcomes of the mean analyses can be supported. For this reason, the negative correlations for the investment in Bitcoin and the development of the investment identify those investors, who do not invest in Bitcoin, have a better expectation about the Bitcoin development than the actual investors. Oppositely, investors in Bitcoin do not expect a reduction of the invested financial resources in the Bitcoin investment. Although they are more carefully in the expectation regarding a positive development of the investment in the short run, in the long run, the investors do not expect a loss of financial resources within the investment.

F. Regression Analyses

As introduced earlier, the regression analysis will be performed on the method of an ordinary least squares regression. The intension is to verify if the dependent variable behavioral intention to use cryptocurrencies is affected by the developed three concepts of independent variables [67]. In this regard, it will be examined, in which degree the predictor variables can explain the generated values of the dependent variable [68].

In the application of the regression analysis, four major indicators need to be considered. Firstly, the r-square will be determined to quantify the explanatory power of the whole regression model. The r-square is the share of the dependent variable, which can be explained by the independent variables. Following Chin and Cohen, the value should be at least 33% [69][70].

Secondly, the analysis of the variances (ANOVA) needs to verify the model fit. The resulting values should be significant ($p<0.05$) and higher than 3 in order to evaluate the model as good, which is the case here.

Thirdly, the regression coefficients of the independent variables need to be significant ($p<0.05$). In particular, the identified estimators must match the expectations in the research hypotheses. Fourthly, the test of multicollinearities by the Variance Inflation Factor (VIF) needs to be performed to find out, whether the variables included in the regression analyses have an identical relation. In the case of existing multicollinearities, i.e., if the VIF values exceed 10 (or in a stricter definition 3), the outcomes of the regression analysis are biased [61][71][72].

TABLE XX. REGRESSION ANALYSIS – PERCEIVED RISK:

Independent variables	Dependent: Intention to Use Cryptocurrencies	
ANOVA = 13.932 p<0.05	**R-Square = 35.2%**	
	Regression Coefficients with Significance	**VIF**
Risk Appetite	**0.052**	**2.656**
Expected Return	**0.143****	**1.328**
Risk Type	**0.182***	**2.524**

* Significant within the error probability of 10%.
** Significant within the error probability of 5%.

TABLE XXI. REGRESSION ANALYSIS – PERFORMANCE EXPECTANCY

Independent variables	Dependent: Intention to Use Cryptocurrencies	
ANOVA = 4.434 p<0.05	**R-Square = 12.2%**	
	Regression Coefficients with Significance	**VIF**
Total Loss	**-0.057****	**1.000**

* Significant within the error probability of 10%.
** Significant within the error probability of 5%.

In performing the regression analysis, the relation between the variables of perceived risk and the intention to use cryptocurrencies are investigated (see Table XX). In general, the r-square achieves a score of 35.2%. Since this value is

slightly above the mark of 33%, there is at least a sufficient explanatory rate of the values of the dependent variable. The ANOVA scores an F-Ratio above the mark of 3.90.

The expected return positively significantly affects the intention to use cryptocurrencies. This means, when investors expect a higher return, they are more open to use cryptocurrencies. In addition, the affinity for risk relates positively significantly but weakly with the intention to use cryptocurrencies. If investors are open to speculate and to take higher risks in investments, they intent to use cryptocurrencies for their investments. The VIF-values are below the mark of 3, so it can be excluded those multicollinearities are within the assumed model.

In Table XXI, the variable of the expectation regarding a total loss of the investment is related to the intention to use cryptocurrencies. The r-square of the regression is 12.2%. Surely, the mark is below 33% and therefore, the explanatory rate seems to be low. In comparison to the other concepts, the expectation of a total loss of an investment in cryptocurrencies reaches a high r-square regarding that only one variable in the regression is considered. The F-Ratio of the ANOVA indicates a value better than the mark of 3.90 and therefore, a model fit is given. The variable total loss is negatively significant related to the intention to use cryptocurrencies.

The negative relationship remarks that investors perceive those investments in cryptocurrencies are very improbable to lead to a full loss of the investment. This induces the openness for and investments in cryptocurrencies. Since there is only one variable, there cannot be any multicollinearities.

In Table XXII, the variables of the concept experience are directly related to the intention to use cryptocurrencies. The r-square of 45.8% describes a low to moderate explanatory rate of the values occurring by the dependent variables. At least two fifths of the values of the dependent variable intention to use cryptocurrencies can be explained by applying the independent variables covering the concept of experience. The F-Ratio of 8.734 remarks an existing model fit.

TABLE XXII. REGRESSION ANALYSIS – EXPERIENCE

Independent variables	Dependent: Intention to Use Cryptocurrencies	
ANOVA = 8.734 p<0.05	R-Square = 45.8%	
	Regression Coefficients with Significance	VIF
General Investment	0.365**	1.090
Investment Duration	0.038**	1.061
Year of First Investment	0.034	1.134

* Significant within the error probability of 10%.
** Significant within the error probability of 5%.

In the concept experience, two variables are positively significant with the intention to use cryptocurrencies. Firstly, the general investment behavior positively affects to the intention to use cryptocurrencies. In general, in case investors do regularly investments (indifferently in which field) they are more open to intent to use investments in cryptocurrencies. Secondly, the variable, which includes the investment duration, is positively significant related to the intention to use cryptocurrencies. This means, investors are more oriented in a long-term store of value. If they behave in this direction, they see cryptocurrencies also as opportunity to invest over a longer time. If investors want to invest for a longer period of time, they are more intent to use cryptocurrencies for their investments. The VIF-scores identify those multicollinearities can be excluded in the model.

Finally, in a combined regression, all independent variables of the three individual regressions are used together. The combination of the independent variables leads to an enhancement to the level of 70.1%. Comparing the resulting r-square to the mark of 33%, the combined approach identifies a high level of explanatory power. In this regard, nearly three quarters of the data points of the dependent variable can be explained by the application of the independent variables. The F-Ratio of 8.041 identifies a good model fit. Through combining all independent variables of the previous regression analyses, only the variable covering the general investment behavior affects positively significant the intention to use cryptocurrencies. When investors have more experience with the application and execution of investments in general, they are more open and willing to use cryptocurrencies. This effect seems to be the most dominant one in the model, since all the other independent variables are getting insignificantly when they are considered in the combined approach. It can be assumed that investors in cryptocurrencies are persons, who have done investments in the past. Therefore, if persons are familiar with investments, they are more willing to do investments in cryptocurrencies.

However, the combined approach identifies two variables (risk appetite and risk type), which have VIF-values above the mark of 3. In this regard, the combined approach cannot fully guarantee that no multicollinearities are within the model. For this reason, the regression coefficients could be biased by the overwhelming effects of the independent variables, which are strongly correlating with each other.

VI. Conclusions And Future Work

The mean analyses underline the importance of Bitcoin as sign for the whole cryptocurrencies. In this regard, the outcomes of the regression and correlation analyses are influenced from the aspect that in the most cases, the investors anticipate Bitcoin with the term cryptocurrency. Although cryptocurrencies are implemented for solving several kinds of transactions, most of the investors see cryptocurrencies with the type of Bitcoin, an alternative payment method with speculation options.

Summarizing the regression analyses, the hypotheses H1 to H3 can be accepted. In general, when investors have made

investments in the past, they are more open to use cryptocurrencies. This result is supported by the fact that how longer the investors do investments and have a long-term store of value, they intent to use cryptocurrencies. In addition, if the investors expect to experience not a total loss of the investment in cryptocurrencies, they have a higher willingness to use cryptocurrencies. Lastly, investors, who do investments with a greater risk, they have also a greater intent to use cryptocurrencies for their investment to reach higher returns.

To sum up, all three concepts identify significant variables, which are influencing the intention to use cryptocurrencies. For this reason, the assumed research model and hypotheses can be fully confirmed. However, as remarked in the beginning, the sample size of the whole analysis is too low. On this account, the achieved results cannot be generalized, and further quantitative analyses and surveys are necessary to deepen the influence factors of cryptocurrencies. As this is a very topical issue, the authors expect that further research works will be performed, which focus on the influence factor for the adoption of Bitcoin, Ethereum and further currencies.

REFERENCES

[1] E. Massarczyk, P. Winzer, and F. J. Müller, "Influence of Performance Expectancy, Experience and Perceived Risk on the Usage of Cryptocurrency Investments," in S. Böhm, L. Berntzen, & F. Volk (Eds.), The Twelfth International Conference on Advances in Human oriented and Personalized Mechanisms, Technologies, and Services (CENTRIC 2019, IARIA) Conference Proceedings and Thinkmind Library, ISSN: 2308-3492, ISBN: 978-1-61208-592-0, pp. 67-76, 2019.

[2] CoinMarketCap, "Top 100 Cryptocurrencies," 2019. [Online]. Available from: https://coinmarketcap.com/de/ [retrieved: 09.2019]

[3] F. Glaser, K. Zimmermann, M. Haferkorn, M. C. Weber, and M. Siering, "Bitcoin - Asset or Currency? Revealing Users," Hidden Intentions, Twenty Second European Conference on Information Systems, Tel Aviv, pp. 1-14, 2014.

[4] E. Sixt, "Bitcoins and further Transaction Systems. Blockchain as Base for Crypto Economy," [German] "Bitcoins und andere dezentrale Transaktionssysteme. Blockchains als Basis einer Kryptoökonomie," Wiesbaden, Springer Gabler, 2017. [Online]. Available from: http://dx.doi.org/10.1007/978-3-658-02844-2 [retrieved: 09.2019]

[5] P. Vigna and M. Casey, "The Age of Cryptocurrency. How Bitcoin and Digital Money are challenging the Global Economic Order," New York, NY, St. Martin's Press, 2015.

[6] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008 [Online]. Available from: https://bitcoin.org/bitcoin.pdf [retrieved: 09.2019]

[7] J. Schütte et al., "Blockchain and Smart Contracts – Technology, Research and Applications," [German] "Blockchain und Smart Contracts – Technologien, Forschungsfragen und Anwendungen," 2017. [Online]. Available from: https://www.fraunhofer.de/content/dam/zv/de/forschung/artikel/2017/FraunhoferPositionspapier_Blockchain-und-Smart-Contracts_v151.pdf [retrieved: 09.2019]

[8] T. Aste, P. Tasca, and T. Di Matteo, "Blockchain Technologies: The Foreseeable Impact on Society and Industry," IEEE. 2017, ISBN: 0018-9162/17, pp. 18-28, 2017.

[9] O. Nica, K. Piotrowska, and K. R. Schenk-Hoppé, "Cryptocurrencies: Economic benefits and risks," FinTech working paper, University of Manchester, Manchester, pp. 1-56, 2017. [Online]. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3059856 [retrieved: 10.2019]

[10] S. R. Basnet and S. Shakya, "BSS: Blockchain Security over Software Defined Network," International Conference on Computing, Communication and Automation (ICCCA2017), IEEE, pp. 720-725, 2017.

[11] D. López and B Farooq, "A Blockchain Framework for Smart Mobility," IEEE Xplore Digital Library, 2018 IEEE International Smart Cities Conference (ISC2), ISBN: 978-1-5386-5959-5/18, 2018.

[12] S. G. Sharma and L. Ahuja, "Building Secure Infrastructure for Cloud Computing using Blockchain," IEEE Xplore Digital Library, 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), ISBN: 978-1-5386-2842-3/18, 2018.

[13] R. Henry, A. Herzberg, and A. Kate, "Blockchain Access Privacy: Challenges and Directions," IEEE Computer and Reliability Societies, *IEEE Security & Privacy*, vol. 16, issue 4, ISBN: 1540-7993/18, pp. 38-45, 2018.

[14] C. Richter, "Blockchain Fog lifts for Assurance," [German] "Der Blockchain-Nebel lichtet sich auch für die Assekuranz - Vom Hype zum Geschäftsmodell für Versicherer," 2017. [Online]. Available from: https://www.accenture.com/de-de/_acnmedia/PDF-55/Accenture-Der-Blockchain-Nebel-Assekuranz-German-2017.pdf [retrieved: 09.2019]

[15] P. Urien, "Blockchain IoT (BIoT): A New Direction for Solving Internet of Things Security and Trust Issues," IEEE, 2018 3rd Cloudification of the Internet of Things (CIoT), 2018.

[16] S. Wang, L. Ouyang, Y. Yuan, X. Ni, X. Han, and F.-Y. Wang, "Blockchain-Enabled Smart Contracts: Architecture, Applications, and Future Trends," IEEE Transactions on Systems, Man, and Cybernetics Systems, ISBN: 2168-2216/2019, pp. 1-12, 2019.

[17] A. M. Antonopoulos, "Mastering Bitcoin: Unlocking Digital Cryptocurrencies," O'Reilly Media, 2014.

[18] G. W. Peters and E. Panayi, "Understanding Modern Banking Ledgers through Blockchain Technologies. Future of Transaction Processing and Smart Contracts on the Internet of Money," University College London, London School of Economics. London, pp. 1-33, 2015.

[19] V. Gatteschi, F. Lamberti, C. Demartini, C. Pranteda, and V. Santamaria, "To Blockchain or Not to Blockchain: That is the Question," IEEE Computer Society, *IT Professional*, vol. 20, issue 2, ISBN: 1520-9202/18, pp. 62-74, 2018.

[20] D. Fakhri and K. Mutijarsa, "Secure IoT Communication using Blockchain Technology," IEEE Xplore Digital Library, 2018 International Symposium on Electronics and Smart Devices (ISESD), ISBN: 978-1-5386-6670-8/18, 2018.

[21] C. O'Flynn, "A Lightbulb Worm? Details of the Philips Hue SmartLighting Design," Black Hat USA 2016.

[22] SRC – SIA/SRC, "Rebooting the IT Revolution: A Call to Action," 2015.

[23] E. Ronen, C. O'Flynn, A. Shamir, and A. Weingarten, "IoT Goes Nuclear:Creating a ZigBee Chain Reaction," IEEE Symposium on Securityand Privacy, *IEEE Security & Privacy*, vol. 16, issue 1, pp. 54-62, 2017.

[24] A. Back, "Hashcash. A Denial of Service Counter-Measure," 2002. [Online]. Available from: http://www.cypherspace.org/hashcash/ [retrieved: 09.2019]

[25] Microsoft, "Understanding Public Key Cryptography," 2014. [Online]. Available from: https://docs.microsoft.com/en-us/previous-versions/tn-archive/aa998077(v=exchg.65) [retrieved: 09.2019]

[26] M. I. Harrigan, L. Shi, and J. Illum, "Airdrops and Privacy: A Case Study in Cross-Blockchain Analysis," IEEE Xplore

Digital Library. 2018 IEEE International Conference on Data Mining Workshops (ICDMW), pp. 63-70, 2018.

[27] K. Christidis and M. Devetsikiotis, "Blockchains and Smart Contracts forthe Internet of Things," IEEE Access, vol. 4, pp. 2292-2303, 2016.

[28] R. Palkovits, N. Pohlmann, and I. Schwedt, "Blockchain Technology revolutionize Digital Business," [German] "Blockchain-Technologie revolutioniert das digitale Business. IT-Sicherheit," pp. 54-60, 2017.

[29] M. Schneekluth, "Blockchain Consensus" [German] "Blockchain Konsens: Welche Konsens-Algorithmen gibt es?," 2018. [Online]. Available from: https://www.wallstreet-online.de/nachricht/10452672-blockchain-konsens-algorithmen-blockchain-konsens-konsens-algorithmen-es/all [retrieved: 09.2019]

[30] J. Altmann, "250 Keywords Bank Economy," [German] "250 Keywords Bankwirtschaft: Grundwissen für Fach- und Führungskräfte," Wiesbaden, Germany, Springer-Verlag, 2018.

[31] S. Haber and W. S. Stornetta, "How to Time-Stamp a Digital Document," *Journal of Cryptology*, vol. 3, pp. 99-111, 1991. [Online]. Available from: https://link.springer.com/article/10.1007%2FBF00196791 [retreived 09.2019]

[32] T. Wu and X. Liang, "Exploration and Practice of Inter-bank Application Based on Blockchain", The 12th International Conference on Computer Science & Education, University of Houston, USA, pp. 219-224, 2017.

[33] M. Rosenberg, "Bitcoin to Smart Contract – Use Cases," [German] "Von Bitcoin zum Smart Contract - Anwendungspotenziale der Blockchain-Technologie," 2016. [Online]. Available from: https://www.bmwi.de/Redaktion/DE/Downloads/Monatsbericht/Monatsbericht-Themen/10-2016-bitcoin.pdf?__blob=publicationFile&v=7 [retrieved: 09.2019]

[34] H. U. Buhl, A. Schweizer, and N. Urbach, "Blockchain Technology as Key for the Future," [German] "Blockchain-Technologie als Schlüssel für die Zukunft?," in: Zeitschrift für das gesamte Kreditwesen : Pflichtblatt der Frankfurter Wertpapierbörse, vol. 70, issue 12, pp. 596-599, 2017.

[35] A. Hayes, "Decentralized Banking: Monetary Technocracy in the Digital Age," in Banking Beyond Banks and Money, pp. 121-131, 2016.

[36] USSEC – U.S. Securities and Exchange Commission, "Investor Bulletin: InitialCoin Offerings," 2018. [Online]. Available from: https://www.sec.gov/oiea/investor-alerts-and-bulletins/ib_coinofferings [retrieved: 09.2019]

[37] M. Conti, S. Kumar, C. Lal, and S. Ruj, "A Survery on Security and Privacy Issues of Bitcoin," *IEEE Communications Surveys & Tutorial*, vol. 20, issue 4, pp. 3416-3452, 2018.

[38] A. J. Schwartz, "Money Supply," in Concise Encyclopedia of Economics, 2008. [Online] Available from: http://www.econlib.org/library/Enc/MoneySupply.html [retrieved: 09.2019]

[39] V. Venkatesh, J. Y. L. Thong, and X. Xin, "Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology," *MIS Quarterly*, vol. 36, issue 1, pp. 157-178, 2012.

[40] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User Acceptance of Information Technology: Toward a Unified View," *MIS Quarterly*, vol. 27, issue 3, pp. 425-478, 2003.

[41] F.-T. Lin, H.-Y. Wu, and T. T. Nguyet Nga, "Adoption of Internet Banking: An Empirical Study in Vietnam," 10th International Conference on e-Business Engnineering, IEEE Xplore Digital Library, pp. 282-287, 2013.

[42] S. Dhawan, K. Singh, and S. Goel, "Impact of Privacy Attitude. Concern and Awareness on Use of Online Social Networking," 5th International Conference - Confluence The Next Generation Information Technology Summit 2013, IEEE Xplore Digital Library, pp. 14-17, 2014.

[43] S. S. Kim and N. K. Malhotra, "A Longitudinal Model of Continued IS Use: An Integrative View of Four Mechanisms Underlying Post-Adoption Phenomena," *Management Science*, vol. 51, issue 5, pp. 741-755, 2005.

[44] S. S. Kim, N. K. Malhotra, and S. Narasimhan, "Two Competing Perspectives on Automatic Use: A Theoretical and Empirical Comparison," *Information Systems Research,* vol. 16 (4), pp. 418-432, 2005.

[45] M. Limayem, S. G. Hirt, and C. M. K. Cheung, "How Habit Limits the Predictive Power of Intentions: The Case of IS Continuance," *MIS Quarterly,* vol. 31, issue 4, pp. 705-737, 2007.

[46] T. Zhou, "Understanding Mobile Internet Continuance Usage from the Perspectives of UTAUT and Flow," *Information Development*, vol. 27, pp. 207-218, 2011.

[47] J. Zhong, A. Dhir, M. Nieminen, M. Hämäläinen, and J. Laine, "Exploring Consumer Adoption of Mobile Payments in China," Academic Mind Trek'13, pp. 318-325, 2013.

[48] Y. S. Wang, Y. M. Wang, H. H. Lin, and T. I. Tang, "Determinants of User Acceptance of Internet Banking: an Empirical Study," *International Journal of Service Industry Management*, vol. 14, pp. 501-519, 2003.

[49] L.-D. Chen, "A Model of Consumer Acceptance of Mobile Payment," *International Journal of Mobile Communications*, vol. 6, issue 1, pp. 32-52, 2008.

[50] M. A. Mahfuz, L. Khanam, and W. Hu, "The Influence of Culture on M-Banking Technology Adoption: An Integrative Approach of UTAUT2 and ITM," 2016 Proceedings of PICMET'16: Technology Management for Social Innovation, pp. 70-88, 2016.

[51] X. Luo, H. Li, J. Zhang, and J. P. Shim, "Examining Multi-dimensional Trust and Multi-faceted Risk in Initial Acceptance of Emerging Technologies: an Empirical Study of Mobile Banking Services," *Decision Support Systems*, vol. 49, issue 2, pp. 222-234, 2010.

[52] H.-P. Lu and P. Y.-J. Su, "Factors Affecting Purchase Intention on Mobile Shopping Websites," *Internet Research*, vol. 19, issue 4, pp. 442-458, 2009.

[53] T. Zhou, "An Empirical Examination of Initial Trust in Mobile Banking," *Information Development*, vol. 21, issue 5, pp. 527-540, 2011.

[54] A. Diekmann, "Empirical Social Research," [German] "Empirische Sozialforschung," Rowohlt-Taschenbuch-Verlag. Reinbek bei Hamburg, vol. 5, 2011.

[55] R. Likert, "A Technique for the Measurement of Attitudes," Archives of Psychology, pp. 199-224, 1932.

[56] Destatis, Statistisches Bundesamt, "Population," [German] "Bevölkerung," 2015. [Online]. Available from: https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Bevoelkerung/Bevoelkerungsstand/Tabellen_/lrbev01.html [retrieved: 09.2019]

[57] Statista, "Internet Users in Germany from 2001 to 2015," [German] "Anteil der Internetnutzer in Deutschland in den Jahren 2001 bis 2015," 2015. [Online]. Available from: http://de.statista.com/statistik/daten/studie/13070/umfrage/entwicklung-der-internetnutzung-in-deutschland-seit-2001/ [retrieved: 09.2019]

[58] L. J. Cronbach, "Coefficient Alpha and the Internal Structure of Tests," Psychometrika, vol. 16, pp. 297-334, 1951.

[59] C. Fornell and D. Larcker, "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error,"

Journal of Marketing Research, vol. 18, issue 1, pp. 39-50, 1981.

[60] R. Hossiep, "Cronbachs Alpha," [German] "Cronbachs Alpha," In Wirtz, M. A. (editor): Dorsch – Lexikon der Psychologie, vol. 17. Verlag Hans Huber, Bern, 2014.

[61] J. F. J. Hair, R. E. Anderson, R. L. Tatham, and W. C. Black, "Multivariate Data Analysis," Macmillan, New York, NY, Macmillan, vol. 3, 1995.

[62] S. Fromm, "Data Analysis with SPSS Part 1," [German] "Datenanalyse mit SPSS für Fortgeschrittene," Arbeitsbuch, vol. 2, VS Verlag für Sozialwissenschaften, GWV Fachverlage, Wiesbaden, 2008.

[63] S. Fromm, "Data Analysis with SPSS Part 2," [German] "Datenanalyse mit SPSS für Fortgeschrittene 2: Multivariate Verfahren für Querschnittsdaten," Lehrbuch, vol. 1, VS Verlag für Sozialwissenschaften, Springer, Wiesbaden, 2010.

[64] N. M. Schöneck and W. Voß, "Research Project," [German] "Das Forschungsprojekt – Planung, Durchführung und Auswertung einer quantitativen Studie," vol. 2. Springer Wiesbaden, 2013.

[65] A. Field, "Discovering Statistics Using SPSS," Sage Publications Ltd., vol. 4, 2013.

[66] S. Hagl, "Statistics," [German] "Schnelleinstieg Statistik," Rudolf Haufe Verlag. München, vol. 1., 2008.

[67] F. Brosius, "SPSS 8 Professionell Statistics in Windows," [German] "SPSS 8 Professionelle Statistik unter Windows," Kapitel 21 Korrelation. International Thomson Publishing. edition. 1, 1998.

[68] T. Schäfer, "Stastics I – Descriptve Results and Explorative Data Analysis," [German] "Statistik I. Deskriptive und Explorative Datenanalyse," VS Verlag für Sozialwissenschaften, 2010.

[69] W. W. Chin, "The Partial Least Squares Approach for Structural Equation Modeling," in G. A. Marcoulides (Ed.). Modern Methods for Business Research. Lawrence Erlbaum Associates. Mahwah, NJ, pp. 295-336, 1998.

[70] J. Cohen, "Statistical Power Analysis for the Behavioral Sciences," Lawrence Erlbaum Associates. Hillsdale, edition 2, 1988.

[71] S. Petter, D. W. Straub, and A. Rai, "Specifying Formative Constructs in Information Systems Research," *MIS Quarterly*, vol. 31, issue 4, pp. 623-656, 2007.

[72] D. Lin, D. P. Foster, L. H. Ungar, "VIF Regression: a Fast Regression Algorithm for Large Data," *Journal of the American Statistical Association*, vol. 106, issue 493, pp. 232-247, 2009.

A Generation Method for the Discussion Process Model during Research Progress Using Transitions of Dialog Acts

Yoko Nishihara
College of Information Science and Engineering
Ritsumeikan University
Shiga, Japan
nisihara@fc.ritsumei.ac.jp

Seiya Tsuji
College of Information Science and Engineering
Ritsumeikan University
Shiga, Japan

Wataru Sunayama
School of Engineering
The University of Shiga Prefecture
Shiga, Japan

Ryosuke Yamanishi
Faculty of Informatics
Kansai University
Osaka, Japan

Shiho Imashiro
Institute for Organizational Behavior Research
Recruit Management Solutions Co., Ltd.
Tokyo, Japan

***Abstract*—People in business and academic fields work in cooperation rather than alone. They may discuss their progress with others, like co-workers and supervisors, to help them obtain the best results, and sometimes, they may feel that such discussions are not conducted well. However, people do not evaluate the quality of each discussion on every occasion because it is tough work for them, and they usually do not have enough time for that. In the process of evaluating discussions, people might look back on their discussions and make a plan to have an improved discussion next time. This paper proposes a generation method for a discussion process model during research progress. First, discussions are recorded to generate transcripts in which each line has a speaker name and his/her utterance. Then, the transcripts are classified manually into high-quality and low-quality discussion groups. Next, dialog acts are assigned to utterances as labels. The labels of dialog acts are originally designed for discussion analysis for research progress. After the labeling, transitions of dialog acts with a high appearance rate are extracted. The transitions are connected if the same dialog act is in both transitions to make a network of dialog acts. The network represents a model of the discussion process. A model for a high-quality discussion group is compared with a model for a low-quality discussion group. By investigating dialog acts and transitions found only in one group, suggestions for low-quality discussions to high-quality discussions would be found. We used discussions between a supervisor and a student who was studying for a degree at a university. We generated models for high-quality and low-quality discussion groups by the proposed method and revealed suggestions for low-quality discussions to high-quality discussions. Our contributions are summarized in two points: (1) we proposed a new method to generate models that represent the discussion process, and (2) we found suggestions for low-quality discussions to high-quality discussions using the models obtained.**

***Index Terms*—Discussion Analysis; Discussion Mining; Discussion Improvement; Transitions of Dialog Acts**

I. Introduction

People in business and academic fields work in cooperation rather than alone. They must discuss their working progress with others, like co-workers and supervisors, to help them obtain the best results. Such people can exchange their opinions and advise each other. These discussions make people not only understand what others think but also ensure that members of the team agree with their work.

Discussions are sometimes not conducted well. This happens because discussions have a time limit, and people often fail to arrive at a common understanding because of a difference in their thinking styles. However, people do not evaluate the quality of each discussion every time because it is tough work for them; they usually do not have enough time for that. In the process of evaluating discussions, people might look back on their discussions and make a plan to have an improved discussion next time. Discussion summarization utilizing the text and acoustic information of the discussion might be a support for the evaluation. However, few of the current summarization methods could detect problems in discussions or suggestions for those.

This paper proposes a generation method for a discussion process model during research progress to find suggestions for low-quality discussions to high-quality discussions. The proposed method divides discussions into two groups: high-quality and low-quality discussion groups. The method makes models for the discussion process in each group using transitions of dialog acts. By comparing the two models, discussants can find suggestions for low-quality discussions to high-quality discussions. The method uses discussions between a supervisor and a student preparing for a graduation thesis. Note that we define a high-quality discussion to be a discussion in which both the student and his/her corresponding supervisor understand their research progress. We believe that transitions of dialog acts represent the quality of discussion. In our previous paper [1], the basic method was proposed and preliminary analysis results were discussed. The contribution of this paper

is summarized in two points. First, we improved the basic method into a new method to generate a model that illustrates the discussion process. Second, we found suggestions for low-quality discussions to high-quality discussions using the models.

II. Related Work

The objectives of conversation and discussion analysis research are classified into two types. One of them is to acquire knowledge to develop conversational agents. The other is to acquire knowledge to improve conversations and discussions. The smooth interaction between conversational agents and users requires that users feel that the agents understand them. Thus, informing methods of information effectively that is understood by conversational agents from users' utterances have been proposed [2]–[4]. Conversational breakdowns between conversational agents and users would decrease users' satisfaction, trust, and willingness to continue using the agents [5], [6]. Methods for addressing conversational breakdowns in task-oriented dialogs have been proposed [7], [8]. These studies analyzed conversations between agents and users to develop conversational agents. Though we analyze discussions between humans, the findings are utilized to improve discussions rather than develop conversational agents.

Some approaches utilize verbal and non-verbal information to analyze conversations between humans. Methods using non-verbal information utilize gaze information, hand gestures, and acoustic information [9]. Gaze information is related to turn-taking in conversations [10]. It can estimate the degree of engagement in a conversation [11]. The effect of hand gestures on understanding a conversation is studied [12]. These studies analyze conversations using non-verbal information to obtain findings. Conversely, the proposed method here analyzes discussions using verbal information to obtain suggestions. Conversations in real life are conducted by using both verbal and non-verbal media. The present study specifically targets language features to find limitations to the suggestions.

The proposed method uses transcripts of discussions during research progress to analyze the discussion process. Zehnalova et al. and Inches et al. postulated methods for analyzing discussion topics and context using words in transcripts [13], [14]. Nishihara et al.'s method utilized dialog acts of utterances in transcripts for the analysis [15]. Since topics and context of discussions during research progress would be changed, it makes little sense to compare topics and context between the same or different discussants. Therefore, we propose a new method to compare discussions using dialog acts; the types, rates and transitions are utilized.

In linguistics and, particularly, in natural language understanding, a dialog act is an utterance, in the context of conversational dialog, that serves a function in the dialog. Conversation analysis often uses the dialog acts to label utterances [16]–[18]. The set of dialog acts is generally composed for every target conversations because frequent dialog acts and the granularity depend on the conversations. Quinn et al. posited their original dialog acts *Declarative, Interrogative, and Imperative* to analyze conversations between nursing support agents and users [19]. Germesin et al. posited two types of dialog acts *Agreement and Non-agreement* to analyze discussions on consensus building [20]. Qadir et al. developed four types of dialog acts *Commissives, Directives, Expressives, and Representatives* [21], and Lampert et al. set two types of dialog acts *Request and Non-request* [22]. According to the previous studies, we also design a new set of dialog acts for analyzing discussions during research progress. The dialog acts are obtained from actual utterances in the discussions we analyze. The new set of dialog acts would be used to generate a model for the discussion process.

As an approach for conversation analysis, both quantitative and qualitative analysis methods have been proposed [23], [24]. He et al. proposed a method for quantitative conversation analysis on consensus building [24]. The present study will also be a quantitative analysis because utterances are transformed into dialog acts, and their rates will be used for the analysis. A quantitative analysis would be influential in supporting discussion improvement because the analysis result will be quantitative so that the suggestions would be posed as concrete actions.

III. Proposed method

Fig. 1 shows the outline of the proposed method. First, discussions are recorded and classified into high-quality and low-quality discussion groups manually. Transcripts of each discussion are prepared manually. One line includes a speaker's name and an utterance. Each utterance is assigned labels for dialog acts. A matrix of the transition of dialog acts for each group is obtained. Each factor of the matrix has a transition rate between two different dialog acts. Transitions with a high rate are extracted from the matrix. The extracted transitions are connected if the same dialog act is included, and then a model for the discussion process is obtained. Two models for high-quality and low-quality discussion groups are obtained. By comparing the two models for high-quality and low-quality discussion groups, common and different points are investigated. It is posited that dialog acts and transitions included in the high-quality discussion group but not present in the low-quality would be suggestions for improving low-quality discussions to high-quality discussions. Discussants refer to the suggestions to have a better discussion next time.

A. Making transcripts of discussions

We posit that a research discussion should be conducted face-to-face. Research discussions are recorded by a voice recorder to generate transcripts. The proposed method uses transcripts of discussions to make a model for the discussion process. One line of a transcript includes a speaker's name and an utterance; an utterance includes fillers. The length of sound in a word (e.g., "weeeeeeeeell" is represented as just "well"), the length of silent time, and the information of laughing (e.g., "ha-ha-ha") are omitted. An utterance includes several sentences before turn-taking occurs. TABLE I exhibits an example of a part of a transcript.

TABLE I
EXAMPLE OF A ONE-TO-ONE DISCUSSION TRANSCRIPT. ONE LINE HAS A SPEAKER NAME AND AN UTTERANCE. THE THIRD COLUMN DISPLAYS ASSIGNED LABELS FOR DIALOG ACTS THAT ARE NOT INCLUDED IN A TRANSCRIPT.

Speaker	Utterance	Dialog Act
T	はい、じゃあ、えっと、よろしくお願いします。(Well, let's start the meeting.)	Greeting_T
T	えーっと、書き起こしが、まあ当然まだだと思ってるんですけども、えっと１人もらった？ (I'm sure that you have not made a transcription naturally. Did you get a recording data?)	Question_T
S	はい。(Yes.)	Answer_S
T	それは誰のん？(Who did you get it from?)	Question_T
S	岡村さんから。(From Riko.)	Answer_S
T	岡村さんからもらった、了解、了解。(You got it from Riko, OK.)	Understanding_T
T	それはなん分ぐらいのデータ？(How long was the data?)	Question_T
S	えっと、４５分ぐらいだったと思います。(Well, about 45minutes)	Answer_S
T	ながっ。(So long.)	Comment_T
S	長かったです。(It's so long.)	Repetition_S
T	すごい。(Awesome.)	Comment_T
S	今回は長かったって言ってたので。(She said the meeting was so long.)	Explanation_S, Report_S
T	あー、大変、大変、わかった、そしたらそれを書き起こしを是非、頑張ってください。(It will be tough work. Please make the transcript though it is tough.)	Request_T, Suggestion_T
S	はい。(Sure.)	Agreement_S
T	で、書き起こすときのポイントがあって、コンピュータで最後言葉を処理することになるので、単語の書き方を揃える。(There are some points in making transcripts. Please use the same words for the same objects because the transcripts would be processed by computer programs.)	Explanation_T, Suggestion_T
S	平仮名とか片仮名とかそういうことですか？(Do you mean to use the same representation for one object?)	Question_S
T	そういうことです、必ず、文の終わりに丸をつける（はい。）っていうのをしてってもらっていいでしょうか？ (That's right. Please add periods every end of a sentence.)	Answer_T

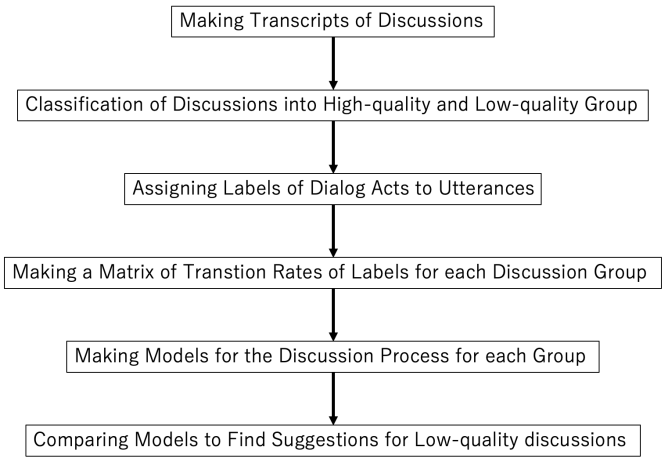


Fig. 1. Flowchart of the proposed method.

TABLE II
LABELS FOR DIALOG ACTS. T DENOTES A SUPERVISOR, WHILE S DENOTES A STUDENT. THERE ARE 62 TYPES OF LABEL; 31 LABELS FOR A SUPERVISOR'S UTTERANCE AND THE REST 31 LABELS FOR A STUDENT'S.

Greeting_T, Confirmation_T, Question_T, Answer_T, Agreement_T, Repetition_T, Explanation_T, Opinion_T, Admiration_T, Suggestion_T, Understanding_T, Topic shifting_T, Report_T, Degression_T, Soliloquy_T, Nodding_T, Request_T, Planning_T, Denial_T, Filler_T, Consultation_T, Response_T, Comment_T, Advice_T, Indication_T, Correction_T, Wondering_T, Surprise_T, Acknowledgement_T, Chatting_T, Additional comment_T
Greeting_S, Confirmation_S, Question_S, Answer_S, Agreement_S, Repetition_S, Explanation_S, Opinion_S, Admiration_S, Suggestion_S, Understanding_S, Topic shifting_S, Report_S, Degression_S, Soliloquy_S, Nodding_S, Request_S, Planning_S, Denial_S, Filler_S, Consultation_S, Response_S, Comment_S, Advice_S, Indication_S, Correction_S, Wondering_S, Surprise_S, Acknowledgement_S, Chatting_S, Additional comment_S

B. Labeling Dialog Acts to Utterances

Each utterance in a transcript is labeled with dialog acts. Though automatic labeling methods have been proposed [15], [25], the labeling accuracy is not enough to generate accurate models. Therefore, each utterance is labeled manually.

Though label sets of dialog acts also have been proposed [19]–[22], [26], most of the sets of labels have their target discussions and conversations. We design a new set of labels for our target discussions, that is, discussions for research progress.

TABLE II shows designed labels for dialog acts. These labels are designed by referring to transcripts of discussions that will be used in experiments in Section IV. An utterance from a supervisor and that from a student will be distinguished by a subscript to a label (T and S). The proposed method uses $31 \times 2 = 62$ types of the label for dialog acts in total.

When labeling dialog acts to an utterance, the already appeared utterances are also considered because one utterance may not have enough information for labeling. If a student says "Yes" to a question from a supervisor, it means "Yes, you are right" labeled as "Answer_S." However, if a student says "Yes" to a proposal from a supervisor, it means "Yes, I will do it" labeled as "Agreement_S." The two instances of "Yes" are different types of utterance. Multiple labels may be put to an utterance because a single utterance may have several roles.

C. Making a matrix of transition rates of dialog acts

A matrix of transition rates of dialog acts is obtained for each group. The transition rate means a rate between labels for dialog acts. Suppose that ith utterance has a vector of labels

TABLE III
EXAMPLE OF A MATRIX OF TRANSITION RATES OF DIALOG ACTS. START DENOTES A DIALOG ACT IN THE STARTING POINT OF TRANSITION AND END REPRESENTS THAT IN THE ENDING POINT.

Start ＼ End	Greeting_T	Question_T	Understanding_T	Suggestion_T	Confirmation_T	Answer_S	Repetition_S	Agreement_S	Question_S	...
Greeting_T	0	0	0	0	0	0	0	0	0	...
Question_T	0	0	0	0	0	8.2%	0	3%	0	...
Understanding_T	0	0	0	0	0	0	0	0	0	...
Suggestion_T	0	0	0	0	0	0	0	1.4%	1.4%	...
Confirmation_T	0	0	0	0	0	3.1%	0	0.5%	0	...
Answer_S	0	5.4%	0	0.3%	0	0	0	0	0	...
Repetition_S	0	0	0	0	0	0	0	0	0	...
Agreement_S	0	3%	0	0.8%	0.7%	0	0	0	0	...
Question_S	0	0	0	0	0	0	0	0	0	...
...	...	...	...	...	...	...	...	...	...	...

$L(i)$ and jth utterance has a vector of labels $L(j)$ ($j = i+1$). $L(i)$ is described by Eq. (1).

$$L(i) = \{l_n | 0 <= n <= 61\}, \quad (1)$$

where n is an index of label l_n which is equal to 0 or 1. If an nth label is put to an utterance, $l_n = 1$. Otherwise, $l_n = 0$. The frequency of transition from the label l_n in $L(i)$ to the label l_m in $L(j)$ is counted up if both l_n and l_m are not equal to zero. Let the frequency be $f_{n,m}$. Let the number of lines of the transcript be NL. $NL-1$ is the number of turn-taking in a discussion. The transition rate $r_{n,m}$ from the label l_n to the label l_m is calculated by Eq. (2).

$$r_{n,m} = \frac{f_{n,m}}{NL-1}, \quad (2)$$

where n and m are indices of labels. By assigning a rate $r_{n,m}$ to each cell, a matrix of transition rates is obtained. TABLE III illustrates an example of a part of the matrix.

D. Making a model for the discussion process

A model of the discussion process for each group is obtained by using the matrix of transition rates. Transitions of dialog acts with a rate more than the threshold TH are extracted from the matrix. The extracted transitions are connected if the same dialog act is included in two different transitions. The network of connected transitions will be a model for the discussion process.

E. Comparing models to find suggestions

The two models for high-quality and low-quality discussion groups are compared. The comparison is conducted by investigating dialog acts and transitions included in the models. First, common points in dialog acts and transitions between the two models are found. Then, different points between the two models are investigated. The different points might be suggestions for low-quality discussions to high-quality discussions.

IV. EXPERIMENT

We investigated whether the proposed method could generate a discussion process model corresponding to the quality of discussion and whether the proposed method could support finding suggestions for low-quality discussions to high-quality discussions. We would verify the former through the experiments in this section, and the latter in Section V.

A. Experimental hypotheses

We proposed two main hypotheses H1 and H2, each of which had two smaller hypotheses; in total there were four small hypotheses.

- H1a If the quality of the discussion is different, the dialog acts are different.
- H1b If the quality of the discussion is different, the transitions are different.
- H2a If the stage of the thesis is different, the dialog acts are different.
- H2b If the stage of the thesis is different, the transitions are different.

H1a hypothesized that if the quality of the discussion differs between the models, the dialog acts in each model are different. H1b was obtained by replacing dialog acts with transitions. If those hypotheses are proved, it will be verified that the proposed method can generate a model for the discussion process that corresponds with the quality of discussion.

H2a hypothesized that if the stage of the thesis differs between the models, dialog acts in each model are different. H2b was obtained by replacing dialog acts with transitions. Discussions during research progress should be held by supervisors and students regularly. The students will acquire the skill of discussion by practicing discussions repeatedly. The supervisors will also change the mode of explanation to match each student. If those hypotheses are proved, dialog acts and transitions only in a model of recent discussions might indicate changes obtained by the practices repeatedly.

B. Experimental procedures

The experiments were conducted as follows:

1) The experimenter collects discussion records and classifies them into high-quality and low-quality discussion groups. The experimenter collects discussions in two stages of the thesis (early and late stages).
2) The models for high-quality and low-quality discussion groups are obtained using the proposed method.
3) The hypotheses are tested from the four models obtained.

The experiments generated four discussion process models for:

1) a **high-quality** discussion group in the **early** stage of the thesis,
2) a **low-quality** discussion group in the **early** stage of the thesis,
3) a **high-quality** discussion group in the **late** stage of the thesis, and
4) a **low-quality** discussion group in the **late** stage of the thesis.

Hypothesis H1a was to be verified by comparing dialog acts between models 1) and 2), and between 3) and 4). Hypothesis H1b was to be verified by comparing transitions of dialog acts between models 1) and 2), and between 3) and 4). Hypothesis H2a was to be verified by comparing dialog acts between models 1) and 3), and between 2) and 4). Hypothesis H2b was to be verified by comparing transitions of dialog acts between models 1) and 3), and between 2) and 4).

C. Used data

We used 16 transcripts of discussions between a supervisor and a student in our laboratory. Half of them (eight out of 16) were collected in the early stage of the thesis (around April in 2019), while the rest was collected in the late stage (around October in 2019, after six months). Two supervisors and eight students (four males and four females) were in our laboratory, and each of the supervisors had four students. The students were 21-22 years old and studying in the College of Information Science and Engineering. The transcripts of discussions were read by the first and second authors. The two authors classified the transcripts into two groups: a high-quality and a low-quality discussion groups.

TABLE IV shows the details of the eight transcripts of discussions held in the early stage of the thesis: the length of discussion, the number of utterances from a supervisor, and that from a student are described. The average length of the discussions was 26min and 58s. The average number of utterances from a supervisor was 81, whereas that from a student was 71.5. The transcripts with IDs 1-4 were in the high-quality discussion group, while those with IDs 5-8 were in the low-quality discussion group. TABLE V shows the details of the eight transcripts of discussions held in the late stage of the thesis. The average length of the discussions was 25min and 17s. The average number of utterances from a supervisor was 86.3, whereas that from a student was 81.0. The transcripts with IDs 9-12 were in the high-quality discussion group, while those with IDs 13-16 were in the opposite group.

Compared with the discussions in the early stage of the thesis, the length was shortened by approximately 1 min, the number of utterances from a supervisor increased (five utterances), and the number from a student also increased (10 utterances). The increase in the number of utterances, even in less discussion time, denotes that the discussions in the late stage were more active. The number of utterances from a student was more significant than that from a supervisor, which indicated that the students might acquire the skill of discussion through repeated practice.

TABLE IV
USED EIGHT TRANSCRIPTS OF THE DISCUSSION HELD IN THE **EARLY** STAGE OF THE THESIS. THE LENGTH OF DISCUSSION, THE NUMBERS OF UTTERANCES FROM SUPERVISORS AND STUDENTS ARE PRESENTED. IDS 1-4 ARE IN HIGH-QUALITY DISCUSSION GROUP, WHEREAS IDS 5-8 ARE IN LOW-QUALITY DISCUSSION GROUP.

Discussion ID	Length (min : sec)	# of utterances from a supervisor	# of utterances from a student
1	48:02	139	127
2	24:29	50	43
3	19:20	50	44
4	44:11	151	145
5	20:38	53	50
6	28:55	100	76
7	17:25	67	58
8	12:42	38	29
Average	26:58	81.0	71.5

TABLE V
USED EIGHT TRANSCRIPTS OF THE DISCUSSION HELD IN THE **LATE** STAGE OF THE THESIS. THE LENGTH OF DISCUSSION, THE NUMBERS OF UTTERANCES FROM SUPERVISORS AND STUDENTS ARE PRESENTED. IDS 9-12 ARE IN HIGH-QUALITY DISCUSSION GROUP, WHEREAS IDS 13-16 ARE IN LOW-QUALITY DISCUSSION GROUP.

Discussion ID	Length (min : sec)	# of utterances from a supervisor	# of utterances from a student
9	32:02	135	130
10	19:57	48	46
11	24:47	89	82
12	16:38	68	65
13	22:05	88	82
14	30:43	94	85
15	24:37	76	73
16	31:23	92	85
Average	25:17	86.3	81.0

D. Experimental results

First, we report frequently appearing labels for dialog act in discussions of high-quality and low-quality groups. TABLE VI and TABLE VII exhibit labels for dialog acts that appeared more than 10 times in each group (high-quality/low-quality in the early/late stage). Though there were common points in appeared dialog acts, the different points were also found.

Next, we report characteristic transitions found in high-quality and low-quality discussion groups. TABLE VIII and TABLE IX exhibit transitions frequently appearing in each group (the appearance rate was more than 1.7% which was the threshold TH [1]). Though there were common points in appearing transitions of dialog acts, different points were also found.

Figs. 2, 3, 4, and 5 illustrate the models obtained for the discussion process. Fig. 2 is for a high-quality discussion group and Fig. 3 is for a low-quality discussion group in the early stage. Fig. 4 is for a high-quality discussion group and Fig. 5 is for a low-quality discussion group in the late stage.

[1]The rate of the used transitions was almost 40% for all transitions found in the transcripts.

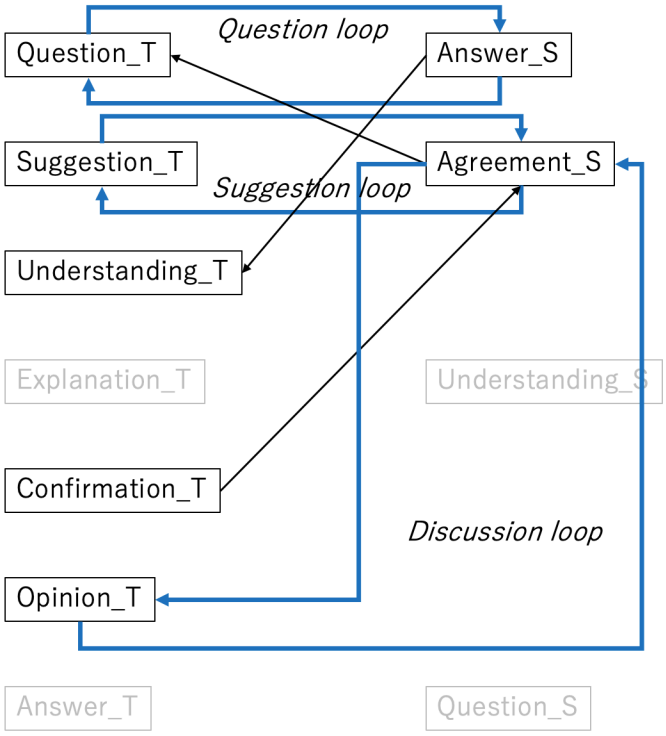


Fig. 2. **High-quality** discussion process model in the **early** stage.

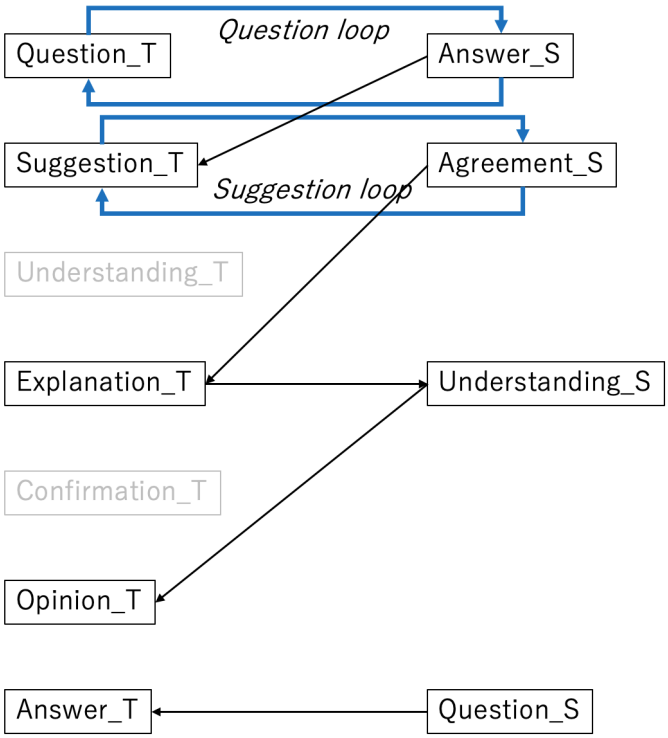


Fig. 3. **Low-quality** discussion process model in the **early** stage.

TABLE VI
LABELS OF DIALOG ACT APPEARING MORE THAN 10 TIMES IN EACH GROUP OF DISCUSSIONS HELD IN THE **EARLY** STAGE. THE LEFT-SIDE IS FOR HIGH-QUALITY DISCUSSION GROUP AND THE RIGHT-SIDE IS FOR LOW-QUALITY DISCUSSION GROUP. THE UPPER-HALF IS FROM SUPERVISORS AND THE BOTTOM-HALF IS FROM STUDENTS.

High-quality		Low-quality	
Dialog act	Frequency	Dialog act	Frequency
Suggestion_T:	53	Question_T	121
Question_T	51	Opinion_T	70
Opinion_T	33	Suggestion_T	34
Explanation_T	29	Confirmation_T	33
Answer_T	24	Explanation_T	26
Confirmation_T	13	Understanding_T	19
Understanding_T	10	Agreement_T	18
Agreement_T	10	Indication_T	17
		Answer_T	16
		Chatting_T	12
		Advice_T	12
Agreement_S	50	Answer_S	121
Answer_S	48	Agreement_S	96
Understanding_S	28	Understanding_S	34
Question_S	25	Opinion_S	20
Report_S	20	Question_S	19
Confirmation_S	12	Report_S	17
Opinion_S	10	Repetition_S	17
		Chatting_S	11
		Wondering_S	11

TABLE VII
LABELS OF DIALOG ACT APPEARING MORE THAN 10 TIMES IN EACH GROUP OF DISCUSSIONS HELD IN THE **LATE** STAGE. THE LEFT-SIDE IS FOR HIGH-QUALITY DISCUSSION GROUP AND THE RIGHT-SIDE IS FOR LOW-QUALITY DISCUSSION GROUP. THE UPPER-HALF IS FROM SUPERVISORS AND THE BOTTOM-HALF IS FROM STUDENTS.

High-quality		Low-quality	
Dialog act	Frequency	Dialog act	Frequency
Question_T	87	Opinion_T	63
Opinion_T	68	Explanation_T	48
Suggestion_T	44	Question_T	46
Explanation_T	41	Understanding_T	42
Confirmation_T	36	Confirmation_T	40
Request_T	18	Suggestion_T	35
Agreement_T	16	Agreement_T	28
Understanding_T	15	Answer_T	14
Wondering_T	12	Indication_T	11
Repetition_T	11	Advice_T	10
Indication_T	10		
Agreement_S	110	Opinion_S	114
Answer_S	84	Answer_S	44
Understanding_S	34	Understanding_S	38
Report_S	31	Report_S	36
Nodding_S	24	Opinion_S	35
Opinion_S	17	Confirmation_S	19
		Nodding_S	15
		Wondering_S	13
		Question_S	12

E. Hypotheses Verification

We investigated the differences of dialog acts between the high-quality and low-quality discussion groups in the early stage to verify H1a. TABLE VI shows 22 types of dialog acts in total. Thirteen of them were common, whereas seven of them were found only in high-quality discussion group or low-quality discussion group. Then, we investigated the differences of dialog acts between the high-quality and low-

TABLE VIII
TRANSITIONS OF DIALOG ACTS APPEARING FREQUENTLY IN EACH GROUP OF DISCUSSIONS HELD IN THE **EARLY** STAGE. THE LEFT-SIDE IS FOR HIGH-QUALITY DISCUSSION GROUP AND THE RIGHT-SIDE IS FOR LOW-QUALITY DISCUSSION GROUP.

High-quality		Low-quality	
Transition	Rate	Transition	Rate
Question_T→ Answer_S	8.2%	Question_T→Answer_S	12.4%
Suggestion_T→Agreement_S	4.4%	Answer_S→Question_T	5.8%
Agreement_S→Suggestion_T	3.8%	Opinion_T→Opinion_S	3.8%
Question_S→Answer_T	3.4%	Agreement_S→Opinion_T	3.7%
Answer_S→Question_T	3.1%	Suggestion_T→Agreement_S	2.7%
Explanation_T→Understanding_S	2.3%	Confirmation_T→Agreement_S	2.2%
Understanding_S→Opinion_T	1.9%	Agreement_S→Question_T	2.0%
Agreement_S→Explanation_T	1.7%	Answer_S→Understanding_T	1.7%
Answer_S→Suggestion_T	1.7%	Agreement_S→Suggestion_T	1.7%

TABLE IX
TRANSITIONS OF DIALOG ACTS APPEARING FREQUENTLY IN EACH GROUP OF DISCUSSIONS HELD IN THE **LATE** STAGE. THE LEFT-SIDE IS FOR HIGH-QUALITY DISCUSSION GROUP AND THE RIGHT-SIDE IS FOR LOW-QUALITY DISCUSSION GROUP.

High-quality		Low-quality	
Transition	Rate	Transition	Rate
Question_T→Anwer_S	9.1%	Question_T→Answer_S	5.2%
Suggestion_T→Agreement_S	4.3%	Confirmation_T→Agreement_S	4.3%
Confirmation_T→Agreement_S	3.4%	Agreement_S→Opinion_T	3.7%
Answer_S→Question_T	3.3%	Opinion_T→Agreement_S	3.6%
Agreement_S→Opinion_T	3.0%	Suggestion_T→Agreement_S	3.1%
Opinion_T→Opinion_S	3.0%	Agreement_S→Suggestion_T	2.6%
Opinion_S→Suggestion_T	2.7%	Explanation_T→Understanding_S	2.5%
Opinion_S→Question_T	2.5%	Agreement_S→Confirmation_T	2.2%
Explanation_T→Understanding_S	2.2%	Report_S→Understanding_T	2.1%
Answer_S→Opinion_T	2.0%	Understanding_S→Explanation_T	2.0%
Answer_S→Confirmation_T	1.8%	Opinion_T→Understanding_S	1.9%
Request_T→Agreement_S	1.8%		
Agreement_S→Explanation_T	1.7%		

quality discussion groups in the late stage. TABLE VII shows 23 types of dialog acts in total. Thirteen of them were common, while six of them were found only in one discussion group. The results indicated that the dialog acts found in the discussion process depend on the quality of discussion. Hypothesis H1a was verified.

We investigated the differences of transitions between high-quality and low-quality discussion groups in the early stage to verify H1b. TABLE VIII shows 14 types of transition in total. Four of them were common, whereas 10 of them were found in only one group. Then, we investigated the differences in transitions between high-quality and low-quality discussion groups in the late stage. TABLE IX shows 16 types of transitions in total. Five of them were common, whereas 14 of them were found only in one group. The results indicated that the transitions found in the discussion process depend on the quality of discussion. Hypothesis H1b was verified.

We investigated the differences of dialog acts found in high-quality discussion groups between the early stage and the late stage to verify H2a. TABLE VI and TABLE VII show 20 types of dialog act in high-quality discussion group. Twelve of them were common, while eight of them were found only in one group. Then, we investigated the differences of dialog acts found in low-quality discussion groups between the early stage and the late stage. TABLE VI and TABLE VII show 23 types of dialog act in low-quality discussions. Sixteen of them were common, while seven of them were found in only one group. The results indicated that the dialog acts found in the discussion process depend on the stage of the thesis. Hypothesis H2a was verified.

We investigated the differences of transitions found in high-quality discussion groups between the early stage and the late stage to verify H2b. TABLE VIII and TABLE IX show 17 types of transition in high-quality discussion groups. Five of them were common, while 12 of them were found in only one group. Then, we investigated the differences of transitions found in low-quality discussion groups between the early stage and the late stage. TABLE VIII and TABLE IX show 15 types of transition in low-quality discussion groups. Six of them were common, while nine of them were found only in one group. The results indicated that the transitions found in the discussion process depend on the stage of the thesis. Hypothesis H2b was verified.

Since all the hypotheses were verified, we found that the proposed method could generate a model for the discussion process that corresponded to the quality of discussion and the stage of the thesis. The next section investigates suggestions for low-quality discussions to high-quality discussions comparing the four models.

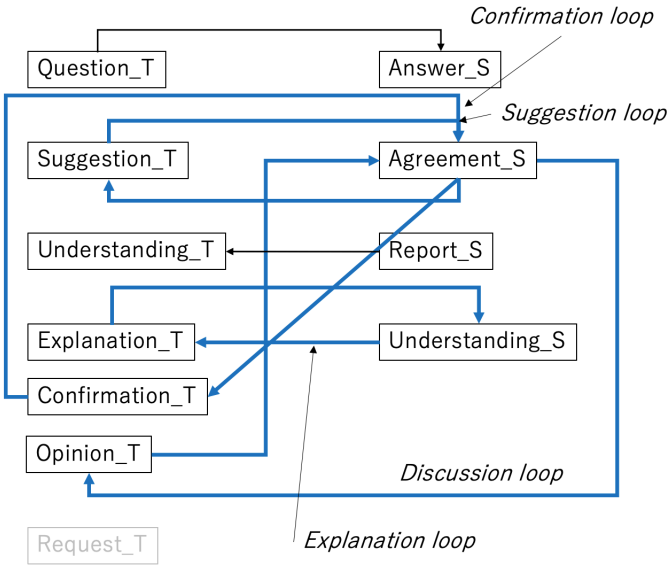


Fig. 4. **High-quality** discussion process model in the **late** stage.

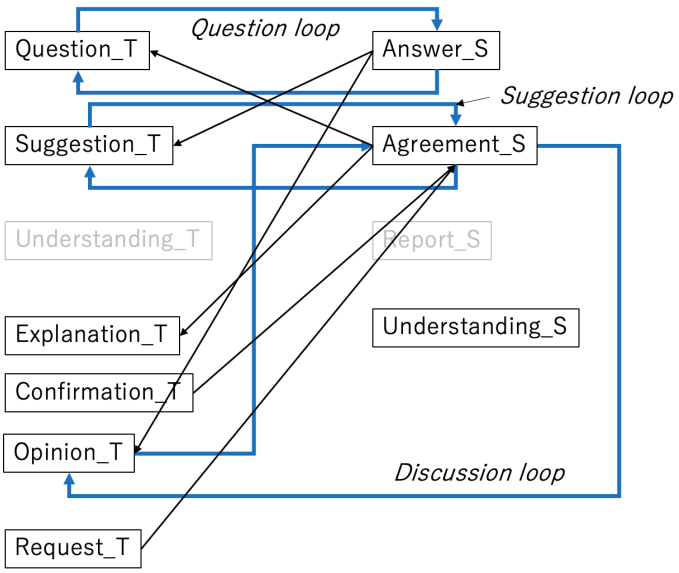


Fig. 5. **Low-quality** discussion process model in the **late** stage.

V. Discussion

We discuss the models obtained for the discussion process and investigate suggestions for low-quality discussions to high-quality discussions.

A. Difference between models in the early stage

First, we discuss the common points between high-quality and low-quality discussion groups in the early stage, referring to Figs. 2 and 3. Both models had the same loop of Question_T $\leftrightarrow$ Answer_S that illustrated a repetition of questioning by supervisors and answering by students. The loop could be interpreted as a process by which supervisors tried to make students understand their thesis and progress (we name it the *Question loop*). Moreover, both models had the same loop of Suggestion_T $\leftrightarrow$ Agreement_S that illustrated a repetition of suggestions by supervisors and agreements by students. The loop could be interpreted as a process by which supervisors tried to make their students agree on the objectives of their thesis (we name it the *Suggestion loop*).

The different points between the high-quality and low-quality discussion groups in the early stage were also found. The model for the high-quality discussion group had a loop of Opinion_T $\leftrightarrow$ Agreement_S that represented a repetition of supervisors giving opinions and students agreeing. The loop could be interpreted as a process by which both supervisors and students discussed their thesis (we name it the *Discussion loop*).

The model for the high-quality discussion group had other significant transitions; Answering_S $\rightarrow$ Understanding_T that represented supervisors' understandings of students' answers and Confirmation_T $\rightarrow$ Agreement_S that illustrated students' agreements on supervisor's confirmations. The significant transitions indicated that both supervisors and students in high-quality discussion group could have constructive discussions based on students' understandings.

Meanwhile, the model for the low-quality discussion group had different transitions: Answer_S $\rightarrow$ Suggestion_T that represented supervisors' suggestions complying with students' progress, Understanding_S $\rightarrow$ Opinion_T that represented supervisors giving opinions for the students' understandings, and Question_S $\rightarrow$ Answer_T that represented supervisors' answering students' questions. Those transitions indicated that supervisors compensated for the lack of students' understanding.

B. Difference between models in the late stage

First, we discuss the common points between the high-quality and low-quality discussion groups in the late stage, referring to Fig. 4 and Fig. 5. Both models had *Discussion loop*s that were included only in the model for the high-quality discussion group in the early stage. The partial change indicated that both supervisors and students in the late stage came to discuss their thesis to some extent.

However, the model for the low-quality discussion group kept in three transitions Answering_S $\rightarrow$ Suggestion_T, Agreement_S $\rightarrow$ Explanation_T, and Explanation_T $\rightarrow$ Understading_S that were included in the model for the low-quality discussion group in the early stage. The remaining of the three transitions indicated that the students in the low-quality discussion group did not understand their thesis perfectly. The model was added to with new transitions Answering_S $\rightarrow$ Confirmation_T and Answering_S $\rightarrow$ Opinion_T that represented supervisors giving confirmations and opinions for students' answers. The two transitions indicated that supervisors could not understand or satisfy with the students' answers. Moreover, a transition of Request_T $\rightarrow$ Agreement_S also emerged that implied supervisors had trouble making their

students understand by explanation and moved to order them specific tasks to finish their thesis.

Meanwhile, the model for the high-quality discussion group had a new loop of Explanation_T ↔ Understanding_S that represented students' understandings of supervisors' detailed explanations (we name it the *Explanation loop*). The model had another new loop of Confirmation_T ↔ Agreement_S that represented a repetition of students' agreements for supervisors' confirmations (we name it the *Confirmation loop*). The model for the high-quality discussion group comprised the four loops of transitions. The four loops illustrated that supervisors and students discussed their specific topics in depth. Since more loops were found in the discussion process, we could estimate that the discussions had various topics with deep insights.

The model for the high-quality discussion group did not include a *Question loop* that was in the model for the high-quality discussion group in the early stage, because a transition of Answer_S → Question_T was disappeared, which meant that students started to answer supervisors' questions appropriately; therefore, it was not necessary to repeat questions. One more newly added transition was Report_S → Understanding_T that represented students reporting their progress and supervisors understanding them. The transitions indicated that students came to understand and conduct their thesis appropriately.

C. Suggestions for low-quality discussions to high-quality discussions

The above findings were summarized as suggestions for low-quality discussions to high-quality discussions. The discussants should drive their discussions to include the following transitions.

1) Supervisors conduct discussions so that students can understand their thesis. The suggestion is realized by removing one-way transitions in Fig. 3 except the transition of Explanation_T → Understanding_S.
2) Supervisors lead students to agree on supervisors' opinions based on the students' understanding. The suggestion is realized by adding a *Discussion loop*.
3) Both supervisors and students have discussions to understand each other's opinions. The suggestion is realized by conducting discussions that the five changes occur:
    a) Students answer supervisors' questions appropriately based on their understandings. The suggestion is realized by removing a *Question loop*.
    b) Supervisors do not explain unilaterally. They should check students' understandings and explain points one by one. The suggestion is realized by adding an *Explanation loop*.
    c) Supervisors do not confirm unilaterally. They should check students' agreements and confirm them one by one. The suggestion is realized by adding a *Confirmation loop*.
    d) Students should come to understand supervisors' opinions. They should not agree on everything without understanding.
    e) Students should report their progress in a way that can be understood by supervisors, which means students must conduct their thesis based on their understandings.

VI. Conclusion

This paper proposed a generation method for the discussion process model during research progress. First, discussions are recorded, then transcripts of them are made that have a speaker name and an utterance in each line. The transcripts are classified into high-quality and low-quality discussion groups. After classification, labels for dialog acts are assigned to each utterance. The labels for dialog acts are originally designed for discussion analysis for research progress. Frequently appearing transitions of dialog acts are extracted in each group. By connecting dialog acts that are common in two transitions, a network of dialog acts is generated. The network illustrates a model for the discussion process. The models for high-quality and low-quality discussion groups are compared. The differences between the two models should be the suggestions for low-quality discussions to high-quality discussions.

We experimented with the proposed method using eight discussions in the early stage of the thesis and another eight discussions in the late stage. We made two main hypotheses: if the quality of discussion is different, the dialog acts and the transitions in the models are also different (H1), and if the stage of the thesis is different, the dialog acts and the transitions in the models are also different (H2). We classified the 16 discussions into four groups considering the quality (high/low) and the stage (early/late), then generated models for each group. By comparing the four models, we proved that all hypotheses were valid. The experimental results verified that the proposed method generated models corresponding to the quality of discussion and the stage of the thesis. We investigated the four models and found the suggestions for low-quality discussions to high-quality discussions. The suggestions were summarized in three points: (1) Supervisors should conduct discussions so that students understand their thesis. (2) Supervisors should lead students to agree on their opinions based on the students' understanding. (3) Both supervisors and students have discussions to try to understand each other's opinions.

Future studies will endeavor to apply the findings to improve discussions. It is assumed that the way of conducting discussions depends on the research field and experiences. There might be limitations in the proposed method because the discussions used were about information science and the students were engaged in their first piece of research. We will make several improvements to the proposed method to cope with the differences in the research field and experiences.

ACKNOWLEDGMENT

This work was supported by Recruit Management Solutions Co., Ltd. Grant. We show our best appreciation.

REFERENCES

[1] S. Tsuji, Y. Nishihara, W. Sunayama, R. Yamanishi, and S. Imashiro, "Analysis method for one-to-one discussion process for research progress using transition probability of utterance types," in The Thirteenth International Conference on Advances in Computer-Human Interactions, 2020, pp. 210–213.

[2] D. J. Litman and S. Silliman, "Itspoke: An intelligent tutoring spoken dialogue system," in Demonstration Papers at HLT-NAACL 2004, ser. HLT-NAACL–Demonstrations '04. USA: Association for Computational Linguistics, 2004, p. 5–8.

[3] A. Raux, B. Langner, D. Bohus, A. Black, and M. Eskenazi, "Let's go public! taking a spoken dialog system to the real world," in 9th European Conference on Speech Communication and Technology, 01 2005, pp. 885–888.

[4] K. Inoue, D. Lala, K. Yamamoto, S. Nakamura, K. Takanashi, and T. Kawahara, "An attentive listening system with android ERICA: Comparison of autonomous and WOZ interactions," in Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue. 1st virtual meeting: Association for Computational Linguistics, Jul. 2020, pp. 118–127. [Online]. Available: https://www.aclweb.org/anthology/2020.sigdial-1.15

[5] F. Bentley, C. Luvogt, M. Silverman, R. Wirasinghe, B. White, and D. Lottridge, "Understanding the long-term use of smart speaker assistants," Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., vol. 2, no. 3, Sep. 2018. [Online]. Available: https://doi.org/10.1145/3264901

[6] Z. Ashktorab, M. Jain, Q. V. Liao, and J. D. Weisz, "Resilient chatbots: Repair strategy preferences for conversational breakdowns," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–12. [Online]. Available: https://doi.org/10.1145/3290605.3300484

[7] C.-H. Li, K. Chen, and Y.-J. Chang, "When there is no progress with a task-oriented chatbot: A conversation analysis," in Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services, ser. MobileHCI '19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3338286.3344407

[8] T. J.-J. Li, J. Chen, H. Xia, T. M. Mitchell, and B. A. Myers, "Multi-modal repairs of conversational breakdowns in task-oriented dialogs," in Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology, ser. UIST '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1094–1107. [Online]. Available: https://doi.org/10.1145/3379337.3415820

[9] X. Liu, A. Xu, V. Sinha, and R. Akkiraju, "Voice of customer: A tone-based analysis system for online user engagement," in Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, ser. CHI EA '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–6. [Online]. Available: https://doi.org/10.1145/3170427.3188454

[10] K. Jokinen, H. Furukawa, M. Nishida, and S. Yamamoto, "Gaze and turn-taking behavior in casual conversational interactions," ACM Trans. Interact. Intell. Syst., vol. 3, no. 2, Aug. 2013. [Online]. Available: https://doi.org/10.1145/2499474.2499481

[11] R. Bednarik, S. Eivazi, and M. Hradis, "Gaze and conversational engagement in multiparty video conversation: An annotation scheme and classification of high and low levels of engagement," in Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction, ser. Gaze-In '12. New York, NY, USA: Association for Computing Machinery, 2012. [Online]. Available: https://doi.org/10.1145/2401836.2401846

[12] Y. Xiong, F. Quek, and D. McNeill, "Hand gesture symmetric behavior detection and analysis in natural conversation," in Proceedings of the 4th IEEE International Conference on Multimodal Interfaces, ser. ICMI '02. USA: IEEE Computer Society, 2002, p. 179. [Online]. Available: https://doi.org/10.1109/ICMI.2002.1166989

[13] S. Zehnalova, Z. Horak, and M. Kudelka, "Email conversation network analysis: Work groups and teams in organizations," in Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, ser. ASONAM '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 1262–1268. [Online]. Available: https://doi.org/10.1145/2808797.2808862

[14] G. Inches and F. Crestani, "Online conversation mining for author characterization and topic identification," in Proceedings of the 4th Workshop on Workshop for Ph.D. Students in Information and Knowledge Management, ser. PIKM '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 19–26. [Online]. Available: https://doi.org/10.1145/2065003.2065009

[15] Y. Nishihara and W. Sunayama, "Estimation of friendship and hierarchy from conversation records," Inf. Sci., vol. 179, no. 11, May 2009, p. 1592–1598. [Online]. Available: https://doi.org/10.1016/j.ins.2008.11.024

[16] G. A. Wang, H. J. Wang, J. Li, A. S. Abrahams, and W. Fan, "An analytical framework for understanding knowledge-sharing processes in online q&a communities," ACM Trans. Manage. Inf. Syst., vol. 5, no. 4, Dec. 2014. [Online]. Available: https://doi.org/10.1145/2629445

[17] S. Oraby, M. Bhuiyan, P. Gundecha, J. Mahmud, and R. Akkiraju, "Modeling and computational characterization of twitter customer service conversations," ACM Trans. Interact. Intell. Syst., vol. 9, no. 2–3, Mar. 2019. [Online]. Available: https://doi.org/10.1145/3213014

[18] A. Ahmadvand, J. I. Choi, and E. Agichtein, "Contextual dialogue act classification for open-domain conversational agents," in Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, ser. SIGIR'19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1273–1276. [Online]. Available: https://doi.org/10.1145/3331184.3331375

[19] K. Quinn and O. Zaïane, "Identifying questions & requests in conversation," ACM International Conference Proceeding Series, 08 2014.

[20] S. Germesin and T. Wilson, "Agreement detection in multiparty conversation," in Proceedings of the 2009 International Conference on Multimodal Interfaces, ser. ICMI-MLMI '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 7–14. [Online]. Available: https://doi.org/10.1145/1647314.1647319

[21] A. Qadir and E. Riloff, "Classifying sentences as speech acts in message board posts," in Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. Edinburgh, Scotland, UK.: Association for Computational Linguistics, Jul. 2011, pp. 748–758. [Online]. Available: https://www.aclweb.org/anthology/D11-1069

[22] A. Lampert, R. Dale, and C. Paris, "Detecting emails containing requests for action," in Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Los Angeles, California: Association for Computational Linguistics, Jun. 2010, pp. 984–992. [Online]. Available: https://www.aclweb.org/anthology/N10-1142

[23] F. Cabrerizo, I. Pérez, and E. Herrera-Viedma, "Managing the consensus in group decision making in an unbalanced fuzzy linguistic context with incomplete information," Knowledge-Based Systems, vol. 23, no. 2, 2010, pp. 169 – 181. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950705109001610

[24] N. He and O. Yoshie, "Conversation analysis based on interpersonal relationship in consensus building," in Proceedings of International Conference on Information Integration and Web-Based Applications & Services, ser. IIWAS '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 26–33. [Online]. Available: https://doi.org/10.1145/2539150.2539178

[25] Y. Hoshikawa and K. Wakabayashi, "Automatic extraction of discussion based on sentence type estimation," in Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, ser. CSCW '17 Companion. New York, NY, USA: Association for Computing Machinery, 2017, p. 203–206. [Online]. Available: https://doi.org/10.1145/3022198.3026336

[26] D. Jurafsky, E. Shriberg, B. Fox, and T. Curl, "Lexical, prosodic, and syntactic cues for dialog acts," in Discourse Relations and Discourse Markers, 1998. [Online]. Available: https://www.aclweb.org/anthology/W98-0319

Edge Computing and Analytics: An Extended Systematic Mapping Study

Andrei-Raoul Morariu*, Kristian Nybom*, Jonathan Shabulinzenze‡, Petteri Multanen†,
Jerker Björkqvist*, Kalevi Huhtala†
*Faculty of Science and Engineering, Åbo Akademi University
Vesilinnantie 3, 20500 Turku, Finland
‡Devecto Oy,
Hermiankatu 12, 33720 Tampere, Finland
†Faculty of Engineering and Natural Sciences, Tampere University
Korkeakoulunkatu 6, 33720 Tampere, Finland
*firstname.lastname@abo.fi ‡jonathan.shabulinzenze@devecto.com †firstname.lastname@tuni.fi

***Abstract*— Connected sensors and devices, the Internet of Things, already today produce more data than data connectivity and cloud services can handle. This gives rise to various forms of distributed sensor data handling, from surveillance cameras with built-in feature detection to alarm functionality in temperature sensors. There is an increased interest in being able to use this sensor data for distributed intelligence. The term "Edge Computing" is often used to denote this distributed computing, performed close to the sensors providing the data. Edge Computing enables services typically provided by cloud services with less communication, lower latency, and independence of the internet infrastructure. In this paper, we present a comprehensive, unbiased overview of state-of-the-art research on edge computing and analytics. From the taxonomy of the 90 identified articles, most articles address task scheduling and operation partitioning while data management and engineering, image and facial recognition, power optimization, and anomaly detection are generally also covered. Simulation remains the most used approach for validation, and research results based on implementations of edge systems in real-life environments are still sparse.**

Index Terms*— *edge computing*; *systematic mapping study*; *taxonomy

I. Introduction

In this paper, we discuss the new trends in terms of scientific publications related to edge computing and analytics. We present the protocol of the Systematic Mapping Study (SMS) that we use to select the scientific publications. This article is an extension of a previous SMS we conducted on the same topic [1], which was published at the Cloud Computing 2020 conference. This paper contains all the new publications that have resulted after proceeding with the SMS methodology in continuation to the previous results from April 10^{th} 2019, up to August 25^{th} 2021.

The number of Internet-connected devices, Internet of Things (IoT), was in the year 2018 estimated to be 23 billion and is estimated to grow to 75 billion by year 2025 [2]. The amount of data transferred over the internet is measured in tens of zettabytes per year [3]. Most of the data on the internet is from people sharing and consuming videos, pictures, and sound, but a substantial and growing part is today's data from sensors. There are estimates that in the future, 75 % [4] of the data generated from sensors will be handled close to the sensors. The concept of handling data close to the sensor is often called Edge Computing.

Edge Computing as a term has been used since 2015 [5], solving the issues with providing the emerging amount of data to end-users by content distribution systems and content caches [6]. Today, Edge Computing commonly symbols the distributed computing performed on sensor data, refining the value and usability of the data. This is also driven by the need for distributed intelligence, where subsystems directly or intermittently connected to the Internet should provide autonomous and independent operations. Edge Computing is also driven by the need for low latency [7] and reduced communication costs [8]. Typical use cases for edge computing varies depending on the domain in where is implemented. For industrial implementations, the objective for edge computing is often reduced maintenance costs, optimized operation performance, predictive maintenance, quality improvement, and safety [9].

Edge computing is a technology that can be applied to numerous different areas and domains. It is possible to also learn between domains, where edge computing architectures, equipment, or infrastructures have been successfully deployed and provided value to the operator. The goal of this paper is to find the state-of-art of experimentation, research, and scientific contributions when it comes to edge computing and related technologies.

Our main contributions are as follows:

- We study in which application domains edge computing is applied. Our results indicate that smart cities and homes are the most common targets for edge computing.
- We identify that algorithms doing task scheduling and operation partitioning are the most common algorithms in edge computing.
- Most of the primary studies contribute to architectural edge computing approaches.
- The most commonly used metric for evaluating edge computing systems in the primary studies is the energy efficiency of the proposals.

The remainder of the paper is organized as follows. Section II describes the protocol used for the SMS used to find and evaluate papers in this study. Section III presents the results of the study according to the research questions from Section II-A. In Section IV we discuss the threats related to the validity of the study. Section V summarizes our work.

II. The Systematic Mapping Study

This section describes the protocol used for the Systematic Mapping Study. The protocol is largely based on the one used in [10], but it has been modified according to the topic of this paper.

A SMS is "a broad review of primary studies in a specific topic area that aims to identify what evidence is available on the topic." [11]. The SMS follows a set of guidelines for articles to include in the primary studies: search for articles, remove duplicates, go through screening phases, perform a study quality assessment checklist and procedure. After the screening phase, researchers extract the most important data from papers for performing the data synthesis.

Subsections II-D – II-G describe how the screening phases and the data extraction and synthesis were performed. Since this paper is a continuation on [1], here we only describe how we worked in this extended version.

A. Research Questions

The research questions (RQ) are as follows:

RQ1: In which fields are edge computing applied?
RQ2: What methods or algorithms are used in edge computing?
RQ3: What proposals exist regarding edge frameworks?
RQ4: What kind of performances do proposed solutions have?
RQ5: What is the standardization level on edge computing?
RQ6: How are the proposals evaluated?

B. Search Strategy for Primary Studies

This section presents our search strategy. It is based on the Systematic Literature Review guidelines from [11] [12].

1) Search Terms: Table I lists the search terms used when searching for original papers for this study. The search terms are derived from the research questions.

TABLE I
Search terms with alternate spellings

Term	Alternate Spelling
Edge	
Comput*	Computing, Compute, Computation
Algorithm*	Algorithms
Analy*	Analytic, Analytics, Analytical, Analysis
Algorithm*	Algorithms
Defect*	Defects
Malfunction	
Anomal*	Anomaly, Anomalies
Performance*	Performances
Complexit*	Complexity, Complexities
Energy	

TABLE II
Search strings

#	Search String
1.	Edge AND (Comput* OR Algorithm OR Analy* OR Defect OR Malfunction OR Anomal*) AND (Performance* OR Complexit* OR Energy)
2.	Edge AND (Comput* OR Algorithm OR Analy*) AND (Defect OR Malfunction OR Anomal*) AND (Performance* OR Complexit* OR Energy)
3.	Edge AND (Comput OR Algorithm OR Analy) AND (Defect OR Malfunction OR Anomal) AND (Performance OR Complexit)

2) Search Strings: The search terms listed in Table I were combined into two search strings for use in the digital libraries. These are shown in Table II.

3) Databases: The search strings shown above was applied in the following digital libraries:

- Institute of Electrical and Electronics Engineers (IEEE) Xplore
- Association for Computing Machinery (ACM) Digital library
- ScienceDirect

The first search string was mainly used for the three databases. For the IEEE Xplore database, we used the second search string when searching in abstracts. This was done to reduce the number of papers found because the first search string resulted in more than 32.000 papers in the abstract search. The third search string was used for the Science Direct database because the maximum number of search strings is limited to 8, and asterisks can not be used. We decided to skip the last search term from the original search string. From the collected results from all databases, duplicates were removed.

C. Study Inclusion Criteria

The inclusion criteria for primary studies were as follows:

- Written in English *AND*
- Published in a peer-reviewed journal, conference, or workshop of computer science, computer engineering, embedded systems, signal processing, or software engineering *AND*
- Describing any of the following:
  - Methods or approaches for edge computing or analytics *OR*
  - Infrastructural or architectural approaches to edge computing and analytics *OR*
  - Performance evaluations of existing edge computing and analytics approaches

If several papers presented the same approach, only the most recent was included, unless the contributions of those papers were different.

D. Title and Abstract Level Screening

In this phase, the inclusion criteria in Section II-C were applied to publication titles and abstracts. One researcher first screened all the titles from the databases. Consequently, two

researchers independently screened the abstracts, excluding all papers that were not relevant for this study. When a researcher was uncertain about including or excluding a particular paper, he discussed it with the other researchers to decision. The results from this phase were used as starting point for the full text screening.

E. Full Text Level Screening

In this phase, the remaining papers were analyzed based on their full text. Four researchers applied the inclusion criteria in Section II-C on the full text. Here, each of the four researchers screened a quarter of the total number of articles. The researchers also documented a reason for each excluded study [13].

F. Study Quality Assessment Checklist and Procedure

The selected papers were assessed based on their quality. Four researchers assessed the quality of the selected papers, each one assessing a quarter of the total number of papers. Any papers not meeting the minimum quality requirements were excluded from the set of primary studies. The output from this phase was the final set of papers listed in Table VI.

Table III presents the checklist for the study quality assessment. For each question in the checklist, a three-level numeric scale was used [13]. The levels were: yes (2 points), partial (1 point), and no (0 points). Based on the checklist and the numeric scale, each study could score a maximum of 34 and a minimum of 0 points. We used the first quartile ($34/4 = 8.5$) as the cutoff point for the inclusion of studies. Therefore, if a study scored 8 points or less, it was excluded due to its lack of quality with respect to this study. The researcher documented the obtained score of each included/excluded study.

TABLE III
STUDY QUALITY ASSESSMENT CHECKLIST, PARTIALLY ADOPTED FROM [10], [13]

#	Question
Theoretical contribution	
1	Is at least one of the research questions addressed?
2	Was the study designed to address some of the research questions?
3	Is a problem description for the research explicitly provided?
4	Is the problem description for the research supported by references to other work?
5	Are the contributions of the research clearly described?
6	Are the assumptions, if any, clearly stated?
7	Is there sufficient evidence to support the claims of the research?
Experimental evaluation	
8	Is the research design, or the way the research was organized, clearly described?
9	Is a prototype, simulation, or empirical study presented?
10	Is the experimental setup clearly described?
11	Are results from multiple different experiments included?
12	Are results from multiple runs of each experiment included?
13	Are the experimental results compared with other approaches?
14	Are negative results, if any, presented?
15	Is the statistical significance of the results assessed?
16	Are the limitations clearly stated?
17	Are the links between data, interpretation and conclusions clear?

G. Data Extraction Strategy

We used the form shown in Table IV to extract data from the primary studies. Four researchers extracted the information from the papers, and each researcher obtained data from one-quarter of the papers. The extracted data was then used for analysis. We extracted such data that it could be used for answering the research questions listed in Section II-A.

TABLE IV
DATA EXTRACTION FORM

Data Item	Value	Notes
General		
Data extractor name		
Data extraction date		
Study identifier (S1, S2, S3, ...)		
Bibliographic reference (title, authors, year, journal/conference/workshop name)		
Publication type (journal, conference, or workshop)		
Edge Computing and Analytics Related		
(RQ1) The domain in which the edge analytics are applied (e.g., smart cities, industry, air industry, shipping, heavy/professional vehicles, health sector)		
(RQ2) Edge computing and analytics method or algorithm		
(RQ3) Edge framework (infrastructure or architecture)		
(RQ4) Performance metrics of proposal (e.g., algorithm complexity, computing, data compression, energy requirements, real-time)		
(RQ5) Mentions of standardization level		
(RQ6) Evaluation method (analytical, empirical, simulation)		

H. Synthesis of the Extracted Data

The extracted data from the papers was used for analysis in order to obtain a high-level view of different aspects related to edge analytics. The papers were categorized in different ways, and collective results were extracted. The results from this phase are presented and discussed in Section III.

III. RESULTS

In this section, we present the main findings of the research. We did the first paper search on April 10^{th} 2019, and the results from that search are reported in [1]. On August 25^{th} 2021, we performed the second paper search. The results presented in this section summarise both of the searches performed.

Several contexts of research contain search terms such as ”edge” and ”algorithm.” Out of the results gathered, some findings did not relate to edge computing. They focused on other examples, such as analysis of image edges or parsing methods for graph edges.

Table V shows the number of papers at the end of each phase of the study. As can be seen, the initial paper search resulted in a large number of papers. From the initial number of papers found using the search strings mentioned in Section

II-B2, we decided to discard papers from earlier than 2016. The discarding of papers was done after the title and abstract screening. To our understanding, the term ”edge computing” was introduced at the end of 2015 [5]. In addition, we also discarded papers related to mobile edge computing, as our research relates to the industrial environment. On the other hand, papers related to fog computing were not discarded, because the technologies used in fog computing are closely related to edge computing.

TABLE V
THE NUMBER OF PAPERS IN EACH PHASE OF THE PROTOCOL

Phase	Number of papers
Initial search results without duplicates	3524
After title and abstract screening	236
After full text screening	123
After quality assessment	90

The initial paper search from the previous study and the search performed in this study resulted in 3524 papers after removing all duplicates. After the title and abstract screening, only 236 papers were included for the next screening phase. Less than half of these papers were included after the full text screening, resulting in 123 papers for the quality assessment. In the quality assessment phase, a few papers were excluded. The number of primary studies included in this research is 90. Of the 90 primary studies, 51 have been published in conference proceedings. The remaining 39 papers have been published in journals.

The term ”edge computing” was firstly mentioned by the ending of 2015, according to a publication of Garcia Lopez et al. [5]. Figure 1 illustrates the publication years of the primary studies. During 2018 many publications related to edge computing appeared. During the following years, there has been a noticeable decrease in the number of publications.

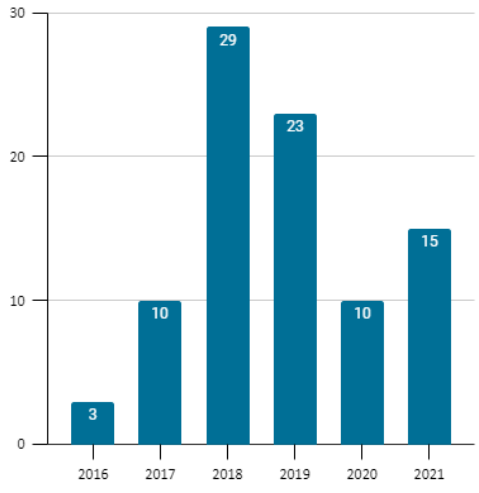


Fig. 1. Reviewed articles sorted by publication years

TABLE VI
PRIMARY STUDIES INCLUDED, WITH CORRESPONDING REFERENCES

ID	Reference	ID	Reference	ID	Reference
S1	[14]	S31	[15]	S61	[16]
S2	[17]	S32	[18]	S62	[19]
S3	[20]	S33	[21]	S63	[22]
S4	[23]	S34	[24]	S64	[25]
S5	[26]	S35	[27]	S65	[28]
S6	[29]	S36	[30]	S66	[31]
S7	[32]	S37	[33]	S67	[34]
S8	[35]	S38	[36]	S68	[37]
S9	[38]	S39	[39]	S69	[40]
S10	[41]	S40	[42]	S70	[43]
S11	[44]	S41	[45]	S71	[46]
S12	[47]	S42	[48]	S72	[49]
S13	[50]	S43	[51]	S73	[52]
S14	[53]	S44	[54]	S74	[55]
S15	[56]	S45	[57]	S75	[58]
S16	[59]	S46	[60]	S76	[61]
S17	[62]	S47	[63]	S77	[64]
S18	[65]	S48	[66]	S78	[67]
S19	[68]	S49	[69]	S79	[70]
S20	[71]	S50	[72]	S80	[73]
S21	[74]	S51	[75]	S81	[76]
S22	[77]	S52	[78]	S82	[79]
S23	[80]	S53	[81]	S83	[82]
S24	[83]	S54	[84]	S84	[85]
S25	[86]	S55	[87]	S85	[88]
S26	[89]	S56	[90]	S86	[91]
S27	[92]	S57	[93]	S87	[94]
S28	[95]	S58	[96]	S88	[97]
S29	[98]	S59	[99]	S89	[100]
S30	[101]	S60	[102]	S90	[103]

A. Application Domains of Edge Computing (RQ1)

Research question 1 strives to identify the domain in which edge computing has been applied in the primary studies. Figure 2 illustrates those domains. Smart cities and homes are the dominating domain of application in the primary studies, and professional vehicles, the health sector, and the industry have also been the application domain in other primary studies. We also included a category named ”Other.” This category covers more domain-specific applications such as tracking systems for drinking activity, micro-services, social media applications, or data centers.

In the majority of the primary studies, however, the domain of application was not specified. Consequently, those papers provided more general contributions that possibly could be applied in several different fields.

B. Edge Computing Method or Algorithm (RQ2)

Table VII shows the purpose of algorithms used in the primary studies.

Approximately one-third of the primary studies relied on algorithms used for task scheduling and operation partitioning, which is understandable since those characteristics are essential when implementing edge systems. One of the second most addressed uses for algorithms was addressing power optimization. It is understandable since often edge computing is applied to small devices with limited resources, most notably computing power and battery. Therefore, power optimization

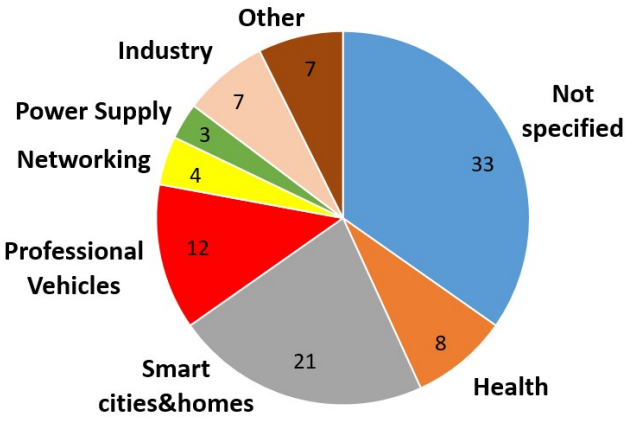


Fig. 2. Edge computing application domains from reviewed studies

is a necessary factor to consider in edge computing. Many primary studies contributed with algorithms related to image and video processing, data transmission, reduction, and mining. A smaller number of primary studies contributed with algorithms related to anomaly detection, audio measurements, or time efficiency. When comparing Table VII with the results we presented in our previous work [1], it is evident that there has been a noticeable increase in publications that contribute with algorithms.

C. Edge Computing Framework (RQ3)

Figure 3 shows the number of papers that contributed to architectures or infrastructures. In some study identifiers, the design was presented as a framework, while others proposed a method. However, the proposals were widely varying, and we were unable to classify the frameworks any further. This research question was consequently challenging to answer. Also, the distinction between architecture and infrastructure may be vague, considering that the infrastructure describes the set of components that make up a system whereas, the architecture represents the design of those components and their relationship. In our rough classification, we considered architecture mostly device-internal and infrastructure on an edge device network level.

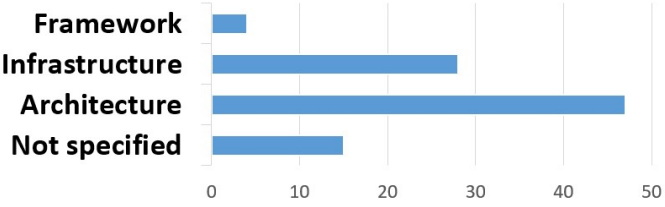


Fig. 3. Articles organized by the type of edge framework proposed

D. Performance of the Proposals (RQ4)

The purpose of RQ4 was to evaluate the performances of the edge systems presented in the primary studies. As can be seen in Table VIII, 40 primary studies provided energy-efficient solutions by reducing the energy requirements for performing tasks. Real-time solutions exist in 23 of the primary studies. With real-time, we mean that results were available with minimal but approximately constant delay, and the papers included in this category were such that it was evident that they were real-time solutions. Some 33 primary studies focused on the improvement computational efficiency of the system by reducing the time required to complete specific tasks and reducing the overall memory usage. In addition, nine primary studies focused on network performance issues. Nine primary studies could not fit into the above classes. These primary studies were on task scheduling, road anomaly detection, and superiority in lane switching scenarios.

E. Edge Analytic Standardization Level (RQ5)

In this research, we analyzed the level of standardization used in edge computing and references to ongoing standardization initiatives for edge computing systems. In our previous study, no publication relied on any edge computing-related standard. The situation is still the same based on our extended search. A few primary studies used standards such as Controller Area Network (CAN), IEEE P1363, and NGSI when implementing edge computing, but these are not strictly edge-related. Corresponding references were made to communication standards such as IEEE 802.11, Wi-Fi, and video codecs MPEG and H.264/H.265 (HEVC) used in edge system implementations. Considering that edge computing standards were not used in the research, a reference to edge level standardization existed that was at the time ongoing within ETSI [104]. It was published in 2017 by ETSI Industry Specification Group with the title: “Mobile Edge Computing (MEC) - Mobile Edge Platform Application Enablement.” Another potential reference that may support the implementation of edge systems in the future was the preparatory work of ISO/IEC JTC 1/SC 41. This Subcommittee 41 is currently preparing standards on the area of the Internet of things and digital twins [105].

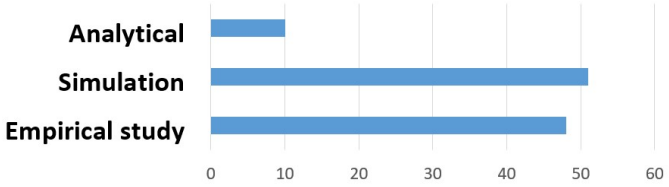


Fig. 4. Evaluation methods

F. Proposal Evaluation Methods (RQ6)

The evaluation of the proposed approach is an essential part of research and scientific papers. The performance and effectiveness of the contribution can be assessed when evaluating the proposal concerning the requirements. The same applies when compared to other approaches. In this study, we analyzed the evaluation methods that were used in the primary studies. Figure 4 illustrates that approaches evaluations were

TABLE VII
TARGETS FOR USING ALGORITHMS IN THE PRIMARY STUDIES

Algorithm Output	Count	Primary Studies	Description
Task Scheduling & Operation Partitioning	29	S7, S11, S13, S16, S20, S23, S26, S27, S31, S34, S40-S42, S44, S45, S47, S50, S57, S68, S69, S70-S72, S74-S77, S80, S81	Decision trees, appliance scheduling, routine handler, offloading algorithm, Markov decision process, sorting, readjustment algorithm
Data Transmission/Reduction/Mining	15	S1, S4, S24, S32, S49, S51, S52, S55, S56, S62, S63, S66, S75, S85, S89	Used for data management
Power optimization	15	S2, S5, S6, S8, S18, S19, S21, S22, S26, S27, S35, S79, S82, S84, S86	Power consumption reduction
Image Classification & Face Recognition & Video Processing & Pattern Recognition	15	S10, S17, S28, S29, S30, S48, S52, S58, S60, S61, S73, S77, S78, S82, S84	Image/video classification recognition, accuracy measurement, fuzzy classification, signal processing in healthcare, lane switching guidance, route planning of autonomous flight devices
Anomaly Detection	12	S12, S15, S37, S53, S54, S64, S80, S83, S87, S88, S89, S90	Vehicle anomaly detection, control loops, digital twin, anomalies in health edge systems, detection of malicious data from edge devices, classifier for predicting component failures
Audio Measurements & Time efficiency & Localization	4	S35, S39, S43, S52	Mosquito wing-beats classification, Bluetooth low energy localization, delay reduction

TABLE VIII
PERFORMANCE METRICS IN THE PRIMARY STUDIES

Performance Metric	Count	Primary Studies	Description
Energy Efficiency	40	S3-S6, S8-S11, S14, S15, S16, S18-S23, S26, S27, S29, S31, S32, S34, S35, S38, S43, S44, S45, S47, S49, S50, S52, S57, S58, S60, S70, S82, S85, S86, S89	Reduced energy requirements for performing computations; power savings; increased battery life of wearable health monitoring devices; early notification from critical health condition
Computational Efficiency	33	S2, S7, S33, S37, S39, S41, S51, S53, S54, S56, S61-S69, S74-S78, S80, S81, S83-S88, S90	Reduced computation time and memory usage; detection of road anomalies; anomaly detection; tracking precision; improved system utility; reduction of operating flight costs; classifiers comparison
Real-time	23	S1, S12, S13, S24, S28, S29, S30, S34, S35, S36, S39, S40, S43, S45, S46, S48, S55, S63, S72, S73, S77, S78, S80	Real-time computation; minimal delay; delay patterns in communication technology; water surface profile predictions; driver notification of critical events
Network performance	9	S17, S25, S30, S36, S45, S59, S71, S76, S79	Network architectures for data transmission; enhancing data availability; efficient bandwidth usage
Other	9	S27, S28, S34, S40, S42, S51, S58, S57, S85	Task scheduling; road anomalies detection; superiority in lane switching scenarios

achieved using analytical, simulation, or empirical studies. In most of the primary studies, simulations were used for evaluation. However, empirical studies were used in almost as many cases. The earlier study shows that there is an increase in the share of empirical evaluations about simulations. Different evaluation methods combinations were used in several studies. In 15 primary studies, empirical evaluation was supported by simulation. In four primary studies, simulation was used along with analytical evaluation. Among the primary studies that were evaluated by empirical studies, case studies were the dominant method chosen. Even though the case studies relied on real-implementations for the evaluations, they used a lab environment for experiments. It means that the conditions for assessment were constructed and controlled by the researchers. In lab test circumstances, there are typically some differences when compared to the actual operating environment. Due to this, some case-specific events that might take place in natural environments may not be admitted in the evaluation phase.

IV. THREATS TO VALIDITY

A threat to the validity of this study is that we dismissed papers related to mobile edge computing since this study focused on edge computing and analytics in non-mobile environments. Consequently, the authors may not have added some relevant papers to this study.

This study also only included papers published from 2016 onward. The reason was that the appearance of the term ”edge” came towards the end of 2015. This way, there may be papers published related to this paper’s topic that was published earlier and subsequently missed.

Another threat to validity is that the screening phases were performed partially by different persons. No researcher followed the entire protocol from beginning to end, but instead, the screening work was divided between the researchers due to time constraints. The researchers may have had different views regarding paper relevancy, potentially excluding relevant papers.

The work-related data extraction was divided between the

researchers. The data extracted in our previous work [1] was double-checked by other researchers, but due to time constraints, we were not able to double-check the data extracted in this extended work. The authors may have missed some of the data during the data extraction phase.

We point out, however, that in our previous work [1], we had consensus discussions in every phase of the protocol. In this extended work, whenever a researcher was uncertain whether to include or exclude a paper, he discussed the matter with the other researchers. Therefore, we believe that the risk of researchers having made mistakes while following the protocol is small.

V. Conclusions

In this paper, we presented a systematic mapping study on edge computing and analytics. The term "edge computing" is moderately new, but it is the same category as other terms such as Internet of Things or fog computing. The term shifts nowadays towards being included in a more widespread section of "distributed intelligence."

We have found an increased number of papers that focused on the application domain of smart homes and cities, professional vehicles, industry, and health. However, among the primary studies we selected, power supply and networking had lower application domains, indicating a clear gap for those fields.

Most of the primary studies we identified focused on task scheduling and operation partitioning. Data management and engineering, image and facial recognition, power optimization, and anomaly detection are other targets for using algorithms within the primary studies. Similar to the previous article, the simulation remains widely used as a tool for validation. The implementation of edge systems is still somewhat sporadic with a few real-life experiments.

Many of the primary studies did not specify the application domain. A similar situation is happening within the specification of the edge framework. Those indicate a lack of strategy for implementing the authors' proposals.

Acknowledgments

This work has been sponsored by Finnish projects funded by Business Finland.

References

[1] A.-R. Morariu, J. Björkqvist, K. Nybom, J. Shabulinzenze, M. Jaurola, P. Multanen, and K. Huhtala, "A systematic mapping study on edge computing and analytics," *CLOUD COMPUTING 2020*, pp. 69–76, 2020.

[2] Statistica, "Internet of things (iot) connected devices installed base worldwide from 2015 to 2025." https://www.statista.com/, Accessed on Nov. 11, 2021.

[3] Bernardmarr, "How much data is there in the world?." https://bernardmarr.com/how-much-data-is-there-in-the-world/, Accessed on Nov. 11, 2021.

[4] Gartner, "Four causes of information governance failures." https://www.gartner.com/smarterwithgartner/, Accessed on Nov. 11, 2021.

[5] P. Garcia Lopez, A. Montresor, D. Epema, A. Datta, T. Higashino, A. Iamnitchi, M. Barcellos, P. Felber, and E. Riviere, "Edge-centric computing: Vision and challenges," *SIGCOMM Comput. Commun. Rev.*, vol. 45, pp. 37–42, Sept. 2015.

[6] L. Ramaswamy and J. Chen, "Efficient delivery of dynamic content: the cooperative ec grid project," in *2005 International Conference on Collaborative Computing: Networking, Applications and Worksharing*, 2005.

[7] J. Luo, X. Deng, H. Zhang, and H. Qi, "Ultra-low latency service provision in edge computing," in *2018 IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE, 2018.

[8] S. Teerapittayanon, B. McDanel, and H.-T. Kung, "Distributed deep neural networks over the cloud, the edge and end devices," in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, pp. 328–339, IEEE, 2017.

[9] T. Qiu, J. Chi, X. Zhou, Z. Ning, M. Atiquzzaman, and D. O. Wu, "Edge computing in industrial internet of things: Architecture, advances and challenges," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2462–2488, 2020.

[10] K. Nybom, A. Ashraf, and I. Porres, "A systematic mapping study on api documentation generation approaches," in *2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, pp. 462–469, Aug 2018.

[11] B. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering (version 2.3)," Tech. Rep. EBSE-2007-01, Keele University and University of Durham, 2007.

[12] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering*. Springer-Verlag Berlin Heidelberg, 1 ed., 2012.

[13] M. Usman, E. Mendes, F. Weidt, and R. Britto, "Effort estimation in agile software development: A systematic literature review," in *Proceedings of the 10th International Conference on Predictive Models in Software Engineering*, PROMISE '14, (New York, NY, USA), pp. 82–91, ACM, 2014.

[14] M. Saez, S. Lengieza, F. Maturana, K. Barton, and D. Tilbury, "A data transformation adapter for smart manufacturing systems with edge and cloud computing capabilities," in *2018 IEEE International Conference on Electro/Information Technology (EIT)*, pp. 0519–0524, May 2018.

[15] M. O. Ozmen and A. A. Yavuz, "Low-cost standard public key cryptography services for wireless iot systems," in *Proceedings of the 2017 Workshop on Internet of Things Security and Privacy*, IoTS&P '17, (New York, NY, USA), pp. 65–70, ACM, 2017.

[16] S. Liu, C. Guo, F. Al-Turjman, K. Muhammad, and V. H. C. de Albuquerque, "Reliability of response region: A novel mechanism in visual tracking by edge computing for iiot environments," *Mechanical Systems and Signal Processing*, vol. 138, p. 106537, 2020.

[17] R. Morabito and N. Beijar, "A framework based on sdn and containers for dynamic service chains on iot gateways," in *Proceedings of the Workshop on Hot Topics in Container Networking and Networked Systems*, HotConNet '17, (New York, NY, USA), pp. 42–47, ACM, 2017.

[18] F. Xiao, L. Yuan, D. Wang, H. Cai, and X. Ma, "Max-fus caching replacement algorithm for edge computing," in *2018 24th Asia-Pacific Conference on Communications (APCC)*, pp. 616–621, Nov 2018.

[19] W. Huang, K. Ota, M. Dong, T. Wang, S. Zhang, and J. Zhang, "Result return aware offloading scheme in vehicular edge networks for iot," *Computer Communications*, vol. 164, pp. 201–214, 2020.

[20] J. Wang, Y. Hu, H. Li, and G. Shou, "A lightweight edge computing platform integration video services," in *2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC)*, pp. 183–187, Aug 2018.

[21] Y. Fukushima, D. Miura, T. Hamatani, H. Yamaguchi, and T. Higashino, "Microdeep: In-network deep learning by micro-sensor coordination for pervasive computing," in *2018 IEEE International Conference on Smart Computing (SMARTCOMP)*, pp. 163–170, June 2018.

[22] O. Gómez-Carmona, D. Casado-Mansilla, D. López-de Ipiña, and J. García-Zubia, "Simplicity is best: Addressing the computational cost of machine learning classifiers in constrained edge devices," in *Proceedings of the 9th International Conference on the Internet of Things*, IoT 2019, (New York, NY, USA), Association for Computing Machinery, 2019.

[23] L. Feng, P. Kortoçi, and Y. Liu, "A multi-tier data reduction mechanism for iot sensors," in *Proceedings of the Seventh International Conference on the Internet of Things*, IoT '17, (New York, NY, USA), pp. 6:1–6:8, ACM, 2017.

[24] G. S. Aujla, N. Kumar, A. Y. Zomaya, and R. Ranjan, "Optimal decision making for big data processing at edge-cloud environment: An

sdn perspective," *IEEE Transactions on Industrial Informatics*, vol. 14, pp. 778–789, Feb 2018.

[25] S. Ding, L. Li, Z. Li, H. Wang, and Y. Zhang, "Smart electronic gastroscope system using a cloud–edge collaborative framework," *Future Generation Computer Systems*, vol. 100, pp. 395–407, 2019.

[26] S. Ci, N. Lin, Y. Zhou, H. Li, and Y. Yang, "A new digital power supply system for fog and edge computing," in *2018 14th International Wireless Communications Mobile Computing Conference (IWCMC)*, pp. 1513–1517, June 2018.

[27] L. Weijian, J. Yingyan, L. Yiwen, C. Yan, and L. Peng, "Optimization method for delay and energy consumption in edge computing micro-cloud system," in *2018 5th International Conference on Systems and Informatics (ICSAI)*, pp. 839–844, Nov 2018.

[28] H. Ren, D. Anicic, and T. A. Runkler, "The synergy of complex event processing and tiny machine learning in industrial iot," in *Proceedings of the 15th ACM International Conference on Distributed and Event-Based Systems*, DEBS '21, (New York, NY, USA), p. 126–135, Association for Computing Machinery, 2021.

[29] D. Rahbari, M. Nickray, and G. Heydari, "A two-stage technique for quick and low power offloading in iot," in *Proceedings of the International Conference on Smart Cities and Internet of Things*, SCIOT '18, (New York, NY, USA), pp. 4:1–4:8, ACM, 2018.

[30] B. Confais, A. Lebre, and B. Parrein, "Performance analysis of object store systems in a fog/edge computing infrastructures," in *2016 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, pp. 294–301, Dec 2016.

[31] D. Liu, Y. Zhang, D. Jia, Q. Zhang, X. Zhao, and H. Rong, "Toward secure distributed data storage with error locating in blockchain enabled edge computing," *Computer Standards & Interfaces*, vol. 79, p. 103560, 2022.

[32] R. Ghosh, S. P. R. Komma, and Y. Simmhan, "Adaptive energy-aware scheduling of dynamic event analytics across edge and cloud resources," in *2018 18th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, pp. 72–82, May 2018.

[33] M. El Chamie, K. G. Lore, D. M. Shila, and A. Surana, "Physics-based features for anomaly detection in power grids with micro-pmus," in *2018 IEEE International Conference on Communications (ICC)*, pp. 1–7, May 2018.

[34] S. Becker, F. Schmidt, A. Gulenko, A. Acker, and O. Kao, "Towards aiops in edge computing environments," in *2020 IEEE International Conference on Big Data (Big Data)*, pp. 3470–3475, Dec 2020.

[35] S. K. Bose, B. Kar, M. Roy, P. K. Gopalakrishnan, and A. Basu, "Adepos: Anomaly detection based power saving for predictive maintenance using edge computing," in *Proceedings of the 24th Asia and South Pacific Design Automation Conference*, ASPDAC '19, (New York, NY, USA), pp. 597–602, ACM, 2019.

[36] T. Rausch, C. Avasalcai, and S. Dustdar, "Portable energy-aware cluster-based edge computers," in *2018 IEEE/ACM Symposium on Edge Computing (SEC)*, pp. 260–272, Oct 2018.

[37] G. Merlino, R. Dautov, S. Distefano, and D. Bruneo, "Enabling workload engineering in edge, fog, and cloud computing through openstack-based middleware," *ACM Trans. Internet Technol.*, vol. 19, Apr. 2019.

[38] Z. Zhou, H. Yu, C. Xu, Z. Chang, S. Mumtaz, and J. Rodriguez, "Begin: Big data enabled energy-efficient vehicular edge computing," *IEEE Communications Magazine*, vol. 56, pp. 82–89, December 2018.

[39] P. Ravi, U. Syam, and N. Kapre, "Preventive detection of mosquito populations using embedded machine learning on low power iot platforms," in *Proceedings of the 7th Annual Symposium on Computing for Development*, ACM DEV '16, (New York, NY, USA), pp. 3:1–3:10, ACM, 2016.

[40] A. Zavodovski, N. Mohan, S. Bayhan, W. Wong, and J. Kangasharju, "Exec: Elastic extensible edge cloud," in *Proceedings of the 2nd International Workshop on Edge Systems, Analytics and Networking*, EdgeSys '19, (New York, NY, USA), p. 24–29, Association for Computing Machinery, 2019.

[41] J. Lim, J. Seo, and Y. Baek, "Camthings: Iot camera with energy-efficient communication by edge computing based on deep learning," in *2018 28th International Telecommunication Networks and Applications Conference (ITNAC)*, pp. 1–6, Nov 2018.

[42] S. Ning, Q. Ge, and H. Jiang, "Research on distributed computing method for coordinated cooperation of distributed energy and multi-devices," in *2018 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pp. 905–910, May 2018.

[43] L. Cai, X. Wei, C. Xing, X. Zou, G. Zhang, and X. Wang, "Failure-resilient dag task scheduling in edge computing," *Computer Networks*, vol. 198, p. 108361, 2021.

[44] L. Pu, X. Chen, G. Mao, Q. Xie, and J. Xu, "Chimera: An energy-efficient and deadline-aware hybrid edge computing framework for vehicular crowdsensing applications," *IEEE Internet of Things Journal*, vol. 6, pp. 84–99, Feb 2019.

[45] S. Dey and A. Mukherjee, "Robotic slam: A review from fog computing and mobile edge computing perspective," in *Adjunct Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing Networking and Services*, MOBIQUITOUS 2016, (New York, NY, USA), pp. 153–158, ACM, 2016.

[46] C. Li, Y. Wang, H. Tang, Y. Zhang, Y. Xin, and Y. Luo, "Flexible replica placement for enhancing the availability in edge computing environment," *Computer Communications*, vol. 146, pp. 1–14, 2019.

[47] Z. Wang, F. Guo, Y. Meng, H. Li, H. Zhu, and Z. Cao, "Detecting vehicle anomaly by sensor consistency: An edge computing based mechanism," in *2018 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–7, Dec 2018.

[48] K. Kolomvatsos and T. Loukopoulos, "Scheduling the execution of tasks at the edge," in *2018 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, pp. 1–8, May 2018.

[49] I. Lujic, V. D. Maio, K. Pollhammer, I. Bodrozic, J. Lasic, and I. Brandic, "Increasing traffic safety with real-time edge analytics and 5g," in *Proceedings of the 4th International Workshop on Edge Systems, Analytics and Networking*, EdgeSys '21, (New York, NY, USA), p. 19–24, Association for Computing Machinery, 2021.

[50] T. Elgamal, A. Sandur, P. Nguyen, K. Nahrstedt, and G. Agha, "Droplet: Distributed operator placement for iot applications spanning edge and cloud resources," in *2018 IEEE 11th International Conference on Cloud Computing (CLOUD)*, pp. 1–8, July 2018.

[51] S. P. Khare, J. Sallai, A. Dubey, and A. Gokhale, "Short paper: Towards low-cost indoor localization using edge computing resources," in *2017 IEEE 20th International Symposium on Real-Time Distributed Computing (ISORC)*, pp. 28–31, May 2017.

[52] S. Y. Nikouei, Y. Chen, A. Aved, E. Blasch, and T. R. Faughnan, "I-safe: Instant suspicious activity identification at the edge using fuzzy decision making," in *Proceedings of the 4th ACM/IEEE Symposium on Edge Computing*, SEC '19, (New York, NY, USA), p. 101–112, Association for Computing Machinery, 2019.

[53] T. Nguyen and E. Huh, "Ecsim++: An inet-based simulation tool for modeling and control in edge cloud computing," in *2018 IEEE International Conference on Edge Computing (EDGE)*, pp. 80–86, July 2018.

[54] C. X. Mavromoustakis, J. M. Batalla, G. Mastorakis, E. Markakis, and E. Pallis, "Socially oriented edge computing for energy awareness in iot architectures," *IEEE Communications Magazine*, vol. 56, pp. 139–145, July 2018.

[55] C. Li, J. Bai, and J. Tang, "Joint optimization of data placement and scheduling for improving user experience in edge computing," *Journal of Parallel and Distributed Computing*, vol. 125, pp. 93–105, 2019.

[56] D. Amiri, A. Anzanpour, I. Azimi, M. Levorato, A. M. Rahmani, P. Liljeberg, and N. Dutt, "Edge-assisted sensor control in healthcare iot," in *2018 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Dec 2018.

[57] P. K. Sharma, S. Rathore, Y. Jeong, and J. H. Park, "Softedgenet: Sdn based energy-efficient distributed network architecture for edge computing," *IEEE Communications Magazine*, vol. 56, pp. 104–111, December 2018.

[58] J. Xue, Q. Hu, Y. An, and L. Wang, "Joint task offloading and resource allocation in vehicle-assisted multi-access edge computing," *Computer Communications*, vol. 177, pp. 77–85, 2021.

[59] C. Xia, W. Li, X. Chang, F. Delicato, T. Yang, and A. Zomaya, "Edge-based energy management for smart homes," in *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, pp. 849–856, Aug 2018.

[60] S. Nousias, C. Tselios, D. Bitzas, A. S. Lalos, K. Moustakas, and I. Chatzigiannakis, "Uncertainty management for wearable iot wristband sensors using laplacian-based matrix completion," in *2018 IEEE 23rd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, pp. 1–6, Sep. 2018.

[61] J. Jang, J. Jung, and J. Hong, "K-lzf : An efficient and fair scheduling for edge computing servers," *Future Generation Computer Systems*, vol. 98, pp. 44–53, 2019.

[62] C. Sonmez, A. Ozgovde, and C. Ersoy, "Edgecloudsim: An environment for performance evaluation of edge computing systems," in *2017 Second International Conference on Fog and Mobile Edge Computing (FMEC)*, pp. 39–44, May 2017.

[63] X. Li, Y. Dang, and T. Chen, "Vehicular edge cloud computing: Depressurize the intelligent vehicles onboard computational power," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 3421–3426, Nov 2018.

[64] C. Xia, X. Jin, L. Kong, C. Xu, and P. Zeng, "Lane scheduling around crossroads for edge computing based autonomous driving," *Journal of Systems Architecture*, vol. 95, pp. 1–8, 2019.

[65] A. M. Khan, I. Umar, and P. H. Ha, "Efficient compute at the edge: Optimizing energy aware data structures for emerging edge hardware," in *2018 International Conference on High Performance Computing Simulation (HPCS)*, pp. 314–321, July 2018.

[66] D. Callegaro, S. Baidya, and M. Levorato, "A measurement study on edge computing for autonomous uavs," in *Proceedings of the ACM SIGCOMM 2019 Workshop on Mobile AirGround Edge Computing, Systems, Networks, and Applications*, MAGESys'19, (New York, NY, USA), p. 29–35, Association for Computing Machinery, 2019.

[67] Y. Liu, L. Kong, M. Hassan, L. Cheng, G. Xue, and G. Chen, "Litedge: Towards light-weight edge computing for efficient wireless surveillance system," in *Proceedings of the International Symposium on Quality of Service*, IWQoS '19, (New York, NY, USA), Association for Computing Machinery, 2019.

[68] T. Mekonnen, M. Komu, R. Morabito, T. Kauppinen, E. Harjula, T. Koskela, and M. Ylianttila, "Energy consumption analysis of edge orchestrated virtualized wireless multimedia sensor networks," *IEEE Access*, vol. 6, pp. 5090–5100, 2018.

[69] S. K. Bose, B. Kar, M. Roy, P. K. Gopalakrishnan, L. Zhang, A. Patil, and A. Basu, "Adepos: A novel approximate computing framework for anomaly detection systems and its implementation in 65-nm cmos," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, pp. 913–926, March 2020.

[70] T. Buddhika, M. Malensek, S. Pallickara, and S. L. Pallickara, "Living on the edge: Data transmission, storage, and analytics in continuous sensing environments," *ACM Trans. Internet Things*, vol. 2, July 2021.

[71] S. Li and J. Huang, "Energy efficient resource management and task scheduling for iot services in edge computing paradigm," in *2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)*, pp. 846–851, Dec 2017.

[72] F. Cicirelli, A. F. Gentile, E. Greco, A. Guerrieri, G. Spezzano, and A. Vinci, "An energy management system at the edge based on reinforcement learning," in *Proceedings of the IEEE/ACM 24th International Symposium on Distributed Simulation and Real Time Applications*, DS-RT '20, p. 155–162, IEEE Press, 2020.

[73] X. Zhao, G. Huang, L. Gao, M. Li, and Q. Gao, "Low load dids task scheduling based on q-learning in edge computing environment," *Journal of Network and Computer Applications*, vol. 188, p. 103095, 2021.

[74] T. Bahreini, M. Brocanelli, and D. Grosu, "Energy-aware speculative execution in vehicular edge computing systems," in *Proceedings of the 2Nd International Workshop on Edge Systems, Analytics and Networking*, EdgeSys '19, (New York, NY, USA), pp. 18–23, ACM, 2019.

[75] P. Andrade, I. Silva, G. Signoretti, M. Silva, J. Dias, L. Marques, and D. G. Costa, "An unsupervised tinyml approach applied for pavement anomalies detection under the internet of intelligent vehicles," in *2021 IEEE International Workshop on Metrology for Industry 4.0 IoT (MetroInd4.0 IoT)*, pp. 642–647, June 2021.

[76] B. Boons, M. Verhelst, and P. Karsmakers, "Low power on-line machine monitoring at the edge," in *2021 International Conference on Applied Artificial Intelligence (ICAPAI)*, pp. 1–8, May 2021.

[77] I. Petri, A. R. Zamani, D. Balouek-Thomert, O. Rana, Y. Rezgui, and M. Parashar, "Ensemble-based network edge processing," in *2018 IEEE/ACM 11th International Conference on Utility and Cloud Computing (UCC)*, pp. 133–142, Dec 2018.

[78] D. Amiri, A. Anzanpour, I. Azimi, M. Levorato, P. Liljeberg, N. Dutt, and A. M. Rahmani, "Context-aware sensing via dynamic programming for edge-assisted wearable systems," *ACM Trans. Comput. Healthcare*, vol. 1, Mar. 2020.

[79] Y. S. Patel, S. Banerjee, R. Misra, and S. K. Das, "Low-latency energy-efficient cyber-physical disaster system using edge deep learning," in *Proceedings of the 21st International Conference on Distributed Computing and Networking*, ICDCN 2020, (New York, NY, USA), Association for Computing Machinery, 2020.

[80] C. Pan, M. Xie, and J. Hu, "Enzyme: An energy-efficient transient computing paradigm for ultralow self-powered iot edge devices," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 37, pp. 2440–2450, Nov 2018.

[81] J. Zhang, C. Deng, P. Zheng, X. Xu, and Z. Ma, "Development of an edge computing-based cyber-physical machine tool," *Robotics and Computer-Integrated Manufacturing*, vol. 67, p. 102042, 2021.

[82] Y. Feng, Z. Liu, J. Chen, H. Lv, J. Wang, and J. Yuan, "Make the rocket intelligent at iot edge: Stepwise gan for anomaly detection of lre with multi-source fusion," *IEEE Internet of Things Journal*, pp. 1–1, 2021.

[83] K. Bhargava, G. McManus, and S. Ivanov, "Fog-centric localization for ambient assisted living," in *2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, pp. 1424–1430, June 2017.

[84] H. Huang, L. Yang, Y. Wang, X. Xu, and Y. Lu, "Digital twin-driven online anomaly detection for an automation system based on edge intelligence," *Journal of Manufacturing Systems*, vol. 59, pp. 138–150, 2021.

[85] J. G. Boubin, N. T. R. Babu, C. Stewart, J. Chumley, and S. Zhang, "Managing edge resources for fully autonomous aerial systems," in *Proceedings of the 4th ACM/IEEE Symposium on Edge Computing*, SEC '19, (New York, NY, USA), p. 74–87, Association for Computing Machinery, 2019.

[86] B. Cheng, G. Solmaz, F. Cirillo, E. Kovacs, K. Terasawa, and A. Kitazawa, "Fogflow: Easy programming of iot services over cloud and edges for smart cities," *IEEE Internet of Things Journal*, vol. 5, pp. 696–707, April 2018.

[87] T.-H. Yang, C.-W. Wang, and S.-J. Lin, "Ecomsnet – an edge computing-based sensory network for real-time water level prediction and correction," *Environmental Modelling & Software*, vol. 131, p. 104771, 2020.

[88] A. Kouloumpris, T. Theocharides, and M. K. Michael, "Metis: Optimal task allocation framework for the edge/hub/cloud paradigm," in *Proceedings of the International Conference on Omni-Layer Intelligent Systems*, COINS '19, (New York, NY, USA), p. 128–133, Association for Computing Machinery, 2019.

[89] X. Chang, W. Li, C. Xia, J. Ma, J. Cao, S. U. Khan, and A. Y. Zomaya, "From insight to impact: Building a sustainable edge computing platform for smart homes," in *2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)*, pp. 928–936, Dec 2018.

[90] X. Jiang, Z. Ma, F. R. Yu, T. Song, and A. Boukerche, "Edge computing for video analytics in the internet of vehicles with blockchain," in *Proceedings of the 10th ACM Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications*, DIVANet '20, (New York, NY, USA), p. 1–7, Association for Computing Machinery, 2020.

[91] C. Pan, M. Xie, S. Han, Z.-H. Mao, and J. Hu, "Modeling and optimization for self-powered non-volatile iot edge devices with ultralow harvesting power," *ACM Trans. Cyber-Phys. Syst.*, vol. 3, Aug. 2019.

[92] D. Y. Zhang, T. Rashid, X. Li, N. Vance, and D. Wang, "Heteroedge: Taming the heterogeneity of edge computing system in social sensing," in *Proceedings of the International Conference on Internet of Things Design and Implementation*, IoTDI '19, (New York, NY, USA), pp. 37–48, ACM, 2019.

[93] A. Anzanpour, D. Amiri, I. Azimi, M. Levorato, N. Dutt, P. Liljeberg, and A. M. Rahmani, "Edge-assisted control for healthcare internet of things: A case study on ppg-based early warning score," *ACM Trans. Internet Things*, vol. 2, Oct. 2020.

[94] Z. Zhu, G. Han, G. Jia, and L. Shu, "Modified densenet for automatic fabric defect detection with edge computing for minimizing latency," *IEEE Internet of Things Journal*, vol. 7, pp. 9623–9636, Oct 2020.

[95] B. Tang, Z. Chen, G. Hefferman, S. Pei, T. Wei, H. He, and Q. Yang, "Incorporating intelligence in fog computing for big data analysis in smart cities," *IEEE Transactions on Industrial Informatics*, vol. 13, pp. 2140–2150, Oct 2017.

[96] A. Awad Abdellatif, A. Emam, C.-F. Chiasserini, A. Mohamed, A. Jaoua, and R. Ward, "Edge-based compression and classification

for smart healthcare systems: Concept, implementation and evaluation," *Expert Systems with Applications*, vol. 117, pp. 1–14, 2019.

[97] H. Wang, L. Muñoz González, D. Eklund, and S. Raza, "Non-iid data re-balancing at iot edge with peer-to-peer federated learning for anomaly detection," in *Proceedings of the 14th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, WiSec '21, (New York, NY, USA), p. 153–163, Association for Computing Machinery, 2021.

[98] G. Gobieski, B. Lucia, and N. Beckmann, "Intelligence beyond the edge: Inference on intermittent embedded systems," in *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*, ASPLOS '19, (New York, NY, USA), pp. 199–213, ACM, 2019.

[99] C. J. L. de Santana, B. de Mello Alencar, and C. V. S. Prazeres, "Reactive microservices for the internet of things: A case study in fog computing," in *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, SAC '19, (New York, NY, USA), p. 1243–1251, Association for Computing Machinery, 2019.

[100] A. Joglekar, G. Gurrala, P. Kumar, F. C. Joseph, T. S. Kiran, K. R. Sahasranand, and H. Tyagi, "Open-source heterogeneous constrained edge-computing platform for smart grid measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.

[101] S. Yi, Z. Hao, Q. Zhang, Q. Zhang, W. Shi, and Q. Li, "Lavea: Latency-aware video analytics on edge computing platform," in *Proceedings of the Second ACM/IEEE Symposium on Edge Computing*, SEC '17, (New York, NY, USA), pp. 15:1–15:13, ACM, 2017.

[102] H. Djelouat, M. Al Disi, I. Boukhenoufa, A. Amira, F. Bensaali, C. Kotronis, E. Politi, M. Nikolaidou, and G. Dimitrakopoulos, "Real-time ecg monitoring using compressive sensing on a heterogeneous multicore edge-device," *Microprocessors and Microsystems*, vol. 72, p. 102839, 2020.

[103] S. Vasavi, K. Aswarth, T. Sai Durga Pavan, and A. Anu Gokhale, "Predictive analytics as a service for vehicle health monitoring using edge computing and ak-nn algorithm," *Materials Today: Proceedings*, vol. 46, pp. 8645–8654, 2021. 3rd International Conference on Materials, Manufacturing and Modelling.

[104] ETSI, "Industry specification group (isg) on multi-access edge computing (mec)." https://www.etsi.org/committee/mec, Accessed on Nov. 11, 2021.

[105] ISO/IEC JTC 1/SC 41, "Iso/iec jtc 1/sc 41 internet of things and digital twin." https://www.iso.org/committee/6483279.html, Accessed on Nov. 11, 2021.

Systemic Modeling and Advance Reengineering of Territory (SMART) as a Path to the Smart Basic Entity (SBE)

Pierre Fournié, Christian Bourret
Laboratoire Dicen-Idf
Université Gustave Eiffel Paris Est Marne la Vallée
Paris, France
pierre.fournie@edu.univ-eiffel.fr
christian.bourret@univ-eiffel.fr

Jean-Pierre Caliste
Université de Technologie de Compiègne
Compiègne, France
jean-pierre.caliste@utc.fr

***Abstract*—Smart cities (SC) became, for a few years, a regular topic in the scientific literature, and both political and economic agendas. Indeed, the connection between urban development and Information and Communication Technologies (ICT) represents a large market. It is presented as a multi-dimension tool to face the challenges of the 21st century. We intend here to demonstrate that developing such a concept only on cities may reinforce the already existing fracture between rural and urban territories. The opportunity exists to bring smart technologies to a lower level, that we call the Smart Basic Entity (SBE). We advocate that, to do so, it could be wise, to experiment a Digital Clone Approach.**

Keywords—Territorial Intelligence; Smart City; Smart Basic Entity; Digital Clone; Rural Territories; Ariège; Angola

I. INTRODUCTION

Smart cities (SC) became, for a few years, a regular topic in the scientific literature, and both political and economic agendas. Former President Bill Clinton is often considered as the father of the concept. Telecommunications and Information Technology (IT) giants, SISCO and IBM, having played the role of initiators. Indeed, the connection between urban development and Information and Communication Technologies (ICT) represents a large market. It is presented as a multi-dimension tool to face the challenges of the 21st century [1].

Between the possible denominations - Eco City, Smart City, Knowledge City, Sustainable City, Resilient City, Low Carbon City, Green City, Ubiquitous City, Intelligent City, Digital City, Information City, Liveable City - Smart City is the most commonly used. More recently, Smart Sustainable City (SSC) and Linking City joined the race.

In 2009, the number of people in urban areas surpassed the number of people living in rural areas. Although, we shall keep in mind that national definition of what is urban is not uniform across the World. For the World Bank, rural population registered a sharp decline during the 1960-2018 period, from 66.4% to 44.7%. Such a figure hides huge disparities between regions and continents. In France, the rural population amounts only at 19.56% in 2018 (divided by 2 since 1960 (38.12%)) whereas it remains at 34.49% in Angola (from 89.56% 58 years earlier) and at 44.68% in Indonesia (85.41% in 1960) [2]. The interpretation of such a worldwide trend shall take into accounts local and regional specificities.

Cities offer multiple advantages: access to electricity, sanitation, water, health and education. Incomes are also higher even they shall be related to living costs. Supported by infrastructures of transportation and communication, high density of individuals and businesses, the city is a territory for serendipity. Urban areas also tempt individuals fleeing war or environment disasters. The attractivity of towns, urban centres or urban clusters, do also impact the structure of employment as there is a shift from agriculture towards manufacturing or services.

Therefore, showing disinterest for the rural population may, on the medium-long term, result in dramatic social, economic and political consequences. At the same time, an opportunity exists, by customizing SC concepts, to create a smart rural development model. We will illustrate the issues and solutions, we currently work at, by taking the examples of France, China and Angola. The interpretation of such a worldwide trend shall take into accounts local and regional specificities. Data related to population, population growth, but also human settlements variations shall be combined. We must consider at the same time the analysis of Cities, and the ones of urban Centres and Urban Clusters.

The change in urbanization ratios varies a lot from one continent to another. Africa (+7.3%) is the fastest urbanizing continent for the period 1990-2015, followed by Northern America (6.4%), Oceania (4.4%), Latin America and the Caribbean (3.3%). The growth is only 1.8% in Asia whereas Europe decreases (-0.8%). Most of all, the surface of urban clusters has nearly doubled in 40 years. From 4% of the global landmass in 1975, they represented 7.6% in 2015. Angola is one of 31 countries for which urban areas have more than double during the period. In some states, often small ones, cities cover more than half of the surface. In 58 others (Russia and Australia, for example) less than 1% [3].

Java, Indonesia hosts two of the 10th most populated urban clusters. In Jakarta, the area of 13 000 square kilometres is home of about 50 million people. The one of Semarang, with a more substantial area (around 19 000 sq/km), shelters 22 million people.

After an introduction, we present, in Section II, the Smart cities as a dominant model. The long term debate about urban concentration and growth will follow (Section III). In a fourth step we will consider what makes the cities so unique (Section IV) before identifying the risks associated with abandoning the rural areas (Section V). We will study in Section VI the case of Ariège before exploring government strategies to transform rural areas into attractive territories (Section VII). We will, by presenting our model, indicate how technology could support such a move (Section VIII) and develop further the digital clone approach (Section IX). Before concluding (Section XI), we will illustrate by our Angola and China experiments, the current status of our research (Section X).

II. SMART CITIES: A DOMINANT MODEL

Smart Data, Smart People, Smart Technology and Smart Governance represent the four pillars of the Smart City. SC has become the dominant model of development for towns in the 21st century. Without any restriction related to traditions, culture and religion, almost all aspects of urban inhabitants' life do enter in the SC scope.

All over the world, political leaders need to answer to the combination of significant issues, namely the explosion of demography and the revolt of Earth. By 2025, the level of urbanization will reach 58.20% (4.7 Bn people) from 44.70% (2.57 Bn), twenty years earlier. The projections show that the two-third of humanity will live in an urban environment by 2050, thus prompting the attention of international agencies such as UNESCO or UN HABITAT.

The explosion of demography and the high concentration of human beings into cities may have devastating effects. Pollution, security concerns, mental and health disorders, increased and concentrated needs for energy and resources are the most intensively documented. Such a massive trend always requires more substantial storage capacities and efficient distribution networks. It also increases the vulnerability of human centres to natural (volcanoes, earthquakes, floods, sea elevation, high tide) and health disasters as well as to terrorist and cyber-attacks.

The revolt of Earth takes multiple forms from climate changes to disappearance of fauna and flora species, from freshwater scarcity to pest invasions. It profoundly affects, together with human-generated conflicts, the living conditions worldwide and creates mutations and transformations at many levels as well as migrations and destruction of human settlements. In such a tense environment, Smart City appears as an "easy to sell" political tool. Expandable and flexible, it is a kind of "swiss knife urban concept" able to resolve all issues mentioned above. It is not a surprise then that it be regularly adopted in the Asia-Pacific, a region facing with Africa the fastest rate of urbanization. There, urbanization would reach 70% by 2050, whereas 90% of the increase will take place in Asia and Africa. Six of the 10th most populated cities in the world are in North, East or South East Asia (Tokyo, Delhi, Shanghai, Mumbai, Beijing, Osaka). The score amounts at 32 for the 50 largest cities of which 16 are megacities (over 10 million people). Some of them, Tokyo, Bangkok, Ho Chi Minh, Jakarta, Manila face, on the same way that New Orleans or West Netherlands, both subsidence and sea-level rise. They enter in the group of "Sinking Cities" and shall partially be below sea level by 2025.

Most of the Asian countries are sensitive to the SC rhetoric. They suffer from environmental threats, are often places of significant disparities (gini ratio), face pollution of air and water, unequal access to education, jobs and services, poor infrastructures for energy, telecommunications and transportations. For many people, the anxiety facing an uncertain future dominates. Also, it is no surprise that considering human gregarious instinct, concentration on cities, also supported by job opportunities, higher standards of living and multiple services, may continue to progress.

The political discourse is supported mainly by the revolution of ICT (Information and Communication Technologies), the one of IoT (Internet of Things) and more recently the fast development of AI (Artificial Intelligence). They allow managing ever-increasing volumes of data that shall grow from 33 Zettabytes (ZB) in 2018 to 175 ZB by 2025 [4].

For a large share of the population, those signs of progress are synonym of job destructions and highly tricky adaptation. Political discourses attempt to calm fears. Thanks to technology, the city would become the place where, through monitoring, air and water purity would remain unchallenged, traffic and transports efficiency would be enhanced, garbage would be invisible and immediately recycled. Criminality and terrorist risks would be assessed, and new services would regularly appear. A smart city is also a place where available jobs would be managed and fulfilled. Still, most of all, its development would require a high number of qualified technicians to support every day appearing positions.

Whatever is the type of decision process (centralized or decentralized) used to drive the development of the Smart City, all insist on people empowerment. Many suggest the promotion of platforms for bottom-up participatory governance [5]. Then, through the access to data, the contribution by forwarding ideas, the development of software and applications by its inhabitants or local companies, the city of the future may appear as a new place for democracy. At least in the political discourse and notwithstanding the increased pressure and control that smart data and IoT may allow on the citizens. By becoming smart, the city will anticipate and satisfy all needs as defined by the Maslow Pyramid from physiological to self-actualization.

Dr. Joan Clos, Secretary-General of the United Conference on Housing and Sustainable Urban Development (Urban III) perfectly summarizes the commonly shared vision: "In this unprecedented era of increasing urbanization, and in the context of the 2030 Agenda for Sustainable Development, the Paris Agreement, and other global development agreements and frameworks, we have reached a critical point in understanding that cities can be the source of solutions to, rather than the cause of, the challenges that our world is facing today. If well-planned and well-managed, urbanization can be a powerful tool for sustainable

development for both developing and developed countries" [6].

III. Urban Concentration and Growth: A Long-term Debate

Does urban concentration support growth? The question remains a place for fierce debates. Jane Jacobs, the well-known author of The death and life of great American cities, appeared as a pathfinder when she claimed that "the understanding of cities, and also of economic development generally, has been distorted by the "dogma of agricultural primacy." From initially being primary organs of cultural development, cities have become primary economic organs [7] and as such a centre of growth. As places of innovation and industrial production, cities offer to the rurality the services and products required to increase the output, to upgrade raw into transformed goods with added value. Polese [8] studies the pro and cons of such a theory extensively. He identifies the laudators of cities whether considering the relationships between per capita income and urbanization levels, the contribution of urban areas to national income and product, the definite link between productivity and the agglomeration of economic activities in cities. But he rejects a direct causal relationship as scientifically impossible to demonstrate. Tolley and Thomas conclude that "Urbanization as such is neither the source nor the enemy of development" [9]

IV. What Makes Cities So Unique ?

Urbanization mainly concerns the rural-urban shift. The denomination is also used when population growth is predominantly urban. As indicated earlier, agricultural productivity directly impacts urbanization pressure. As indicated earlier, agricultural productivity directly impacts urbanization pressure. Reversely effectiveness of agriculture may lower or slower urban growth. Urbanization, part of the farm workforce becoming free, constitutes an inferred-effect of the agricultural revolution [10]. Castells-Quintana and Royuela [11] edulcorate such a proposal by stressing,that rural population are often expelled from the rural areas.

The populations of major cities scarcely decrease: it is the case for Busan and Nagasaki but also in several European towns submitted to low fertility and emigration.

Before 1850, no country was predominantly urbanized. It was the case only for England by 1900. From then, the rate rose dramatically from 14% after World War II, to 55% in 2018 with projections at 68% by 2050. During the period 2018-2050, three countries, namely India (+416 million urban dwellers), China (+255 million) and Nigeria (+189 million), account for about 35% of the projected increase. Northern America (82%), Latin America and the Caribbean (81%), Europe (74%) and Oceania (68%) are highly urbanized. On the other hand, Asia (50%) and Africa (43%) remain mostly rural, but their rate of population growth will automatically generate, as demonstrated by Tolley both urban and non-urban growth.

The genuine attraction for cities shall be analyzed. Cities are the mothers of human progress [12] and, no development may take place without towns [13]. Often the change in status (from city to capital as it was the case for Jakarta in 1961-1964); the efforts from government or municipality; the development of the hinterland, accelerate the urbanization process. Thanks to the concentration of capital and means, the city offers a place for innovation [14]. "The city promotes the monetization of the economy, facilitates social mobility and the adequacy between offer and demand for qualified manpower, expands the markets for industrial and agricultural productions" [15]. Even a specific size shall be reached to boost technological progress [16], dense and marge populations constitute a fertile ground able to welcome the exchange of ideas [17]. City diversity supports employment growth [18]. By limiting the distance, offering efficient transportation and communication networks, they reduce the cost of transmitting information and increase efficiency and productivity [19] [20]. Unsurprisingly the patterns have changed with time also in the 90's "Cities with high levels of human capital did well, and cities with large numbers of the poor did poorly" [21]. Environmental concerns also require adaptative and innovative means of transportation. Intermodal platforms aim at improving efficiency [22].

Cities are intricated into a complex system: the system of cities. For Pumain, cities also have an intrinsic quality to transform themselves: an evolutive capacity or (re)organization. Remembering the "General systems of cities" conceptualized by J.Reynaud in 1841; the works of Berry and its famous "cities as systems within systems of cities" [23] and the ones of Pred ; using analogies with physical systems and synergetics [24]; Pumain compares cities systems to dynamic systems governed by an auto-regulation. She goes even further by defining an evolutional theory centred on the notion of "*system of cities*"; thus, ending the supremacy of a static vision of cities [25].

V. The Risk of Abandoning Rural Areas and "The Catastrophe Scenario"

Could any government decently abandon 60 to 20% of its population? The "yellow vest" movement in France, initiated in October 2018 and still active as of March 2020 in some rural areas, constitutes, for any government, a fierce reminder. The images of violence in Champs Elysees, the groups of rioters spreading in the capital and main cities, made people forget that many Yellow Vest were leaving in rural areas and were protesting against decisions taken by the central government that directly impacted their living conditions.

France is presented regularly as an heir of Jacobinism. Despite some tentative of decentralization, regionalism is refrained, and decisions centres are often far from the countryside. Strong disconnection exists between elites, budget and environmentally oriented, unknowledgeable about local realities, and the rurality. To fight against deficits, dramatic cuts were done, during the last thirty years, in public services. They are now doubled by the disappearing of private services, in particular in the field of finance and communications. An untrained and unprepared population faces a new "virtual world", as well as non-efficient networks.

Conscious that a balance shall exist across the entire national territory, President Charles de Gaulle created in 1963

the Délégation à l'Aménagement du Territoire et à l'Action Régionale (DATAR). Under the Prime Minister, such a structure had to impulse, coordinate, reequilibrate the actions of the state whereas in developing rural areas, reinforcing transportation networks, meshing the country. By using DATAR, the state aimed at organizing and modernizing France; at preserving cohesion, at making territories more attractive.

On 2014, DATAR merged with the Comité Interministériel des Villes and l'Agence Nationale pour la Cohesion Sociale et l'Egalité des Chances into CGET, le Commissariat à l'Egalité des Territoires. Such a combination of agencies was lately replaced by the Agence Nationale pour la Cohésion des Territoires (ANCT) under the ministry of Cohesion of Territories and Relationships with Territorial Collectivities.

Terminology matters here: cohesion, equality constitute the fertile soil of a territory that any disorder may spoil forever.

In the 1970s, DATAR prepared a prospective study forecasting what will be France by the end of the 20th century. One of the output, referred to as the "Catastrophe scenario", concentrating all developments into large metropoles and consequently creating desertification of rural areas had to be avoided. Dou and Fournié [30] have shown that, notwithstanding expert's advices, France has been developed under such a configuration. The sizeable interstitial space created suffers from insufficient means, infrastructures and is dramatically abandoned by its population.

The state agency France Strategy proposed that national investments be channeled to the 15 largest French cities through metropolitan pacts of innovation. Such a policy, justified by a lack of resources, may lead in territories located at the fringe, to the reinforcement of inequalities as regards as public services, access to medical services, connectivity between others.

France suffers from a triple fracture: a Territorial Fracture, a Technological Fracture and "Data-consciousness" fracture.

France is not the sole country to face problems with its rurality. The phenomena concerns almost all countries of the continent. The European Union, now aware of the situation, initiated some action process with the Cork declaration, Ireland, 2016; the UE Action Plan for Smart Villages (11/04/2017) that aims at « investing in the viability and vitality of rural areas »; the Bled declaration, Slovenia, 2018.

VI. The Case of Ariege

Separated from Spain by the Pyrenees mountains, Ariège is one of the 13 departments composing the Occitanie region, second largest province in France (72 724 sq/km). With 5.8 million inhabitants, organized around 18 urban poles, two large metropoles (Toulouse and Montpellier), the region shelters a strong industry recognized at international level (aeronautics, spatial, in-vehicle systems, agro-industries, biotech) supported by 15 poles of competitivity and several large universities.

Ariège is a department limited in size, home of 152 724 inhabitants of which 46.04% are over 50 years old compared to 33.08 % in the neighbour department of Haute Garonne (central city: Toulouse). Largest cities in 2017 were Pamiers (15675 habs), Foix (9 532 habs), Saint Girons (6 383 habs) and Lavelanet (6 137 habs). Only 31.1% of Ariege's inhabitants have a graduate-level, far from the 50.8% of Haute Garonne and 97% of the companies have less than 50 employees.

The territory, called in some media the anti-startup nation [26], has suffered from the closure of few large factories (in particular of paper, an industry-supported by hydroelectric capacities and forests) and mines in the 90's, of "green tourism" being impacted by fierce competition with other regions and international destinations since the year 2000, and more recently of "white tourism" being affected by climatic change.

Still, Ariège benefits of several assets. A vast cultural and historical heritage (between others Cathar castles), an immaculate nature with 55 000 ha of regional Natura 2000 park that welcomes bears and wolves, high peaks over 3000 m, thermalism are only parts of them.

In this French department the disappearing of state presence, mass transportations and private services alter the living conditions and destroy the efforts to promote tourism On the long term, they might be the synonym of exode and increased poverty. A feeling of exclusion may prevail that could be transformed into social unrest and affect social and national cohesion. State presence disappearance takes multiple forms: the closing of taxation and perception offices, of classes in schools and colleges, of tribunals and legal offices, the reorganization of beds in hospitals or health services. In parallel, traditional shops are impacted by slow local and touristic activities. Low traffic and profitability condemn branches of banks and post offices. Bank Automated Distributors are suppressed. As in many rural regions, people have to drive 20 to 40 km to access essential services. Not the least, doctors and specialists abandon those areas, and there is a lack of professionals in both private sector (-26% between 2014 and 2018) and hospitals (-16% for the same period).

Moreover, new regulations made possible the absence of controllers in the trains creating tense situations as regards as security and law enforcement in transports. On many lines, tickets cannot be sold into trains whereas at the same time commercial offices are being suppressed or opened during a short period of the day. Automatic machines are destroyed if located outside of the station and remain not accessible out of office hours or during weekends. The law referred to as LOM (Loi d'Organisation des Mobilités) signed December 24th, 2019, creates a right to mobility. It may also, despite allowing through open data the access to information on transports and reinforcing the role of region, broadly impact rural territories.

As regards as communication networks, 48.3% of housing have access to a high-speed network whereas over 10% still have connection problems. At the same time, only 22.6% are eligible to optic fiber [27].

Thus, at the same time authorities are pushing for more eco-friendly means of transportation and numeric transformation, an ageing population, suffering from a lack of knowledge and adaptative capacities, has difficulties in adapting to tools not fully available and to which it has not been trained. Besides the rupture of equality between

territories, such a situation may endanger - this a guess that shall be assessed on the field - on a large scale, mental health, create deep feelings of exclusion and increase the vulnerability of rural populations. The social patterns that have been prevailing for over a century are destroyed by decisions coming from Paris or regional levels, neither understandable to the majority nor understood.

VII. Could Rural Areas Become Attractive Again?

Could rural areas become attractive again? Or in other words could we, create in those regions, at a time environmental concerns become a priority, the conditions to make the soil fertile again for living, to invert the rural-urban shift, to attract capital and means, to transform rural areas into places of innovation and serendipity.

The report of Cour des Comptes dated March 2019 and entitled "Accessing public services in rural territories" concludes that, despite multiple initiatives, rural areas have required a long-term effort and remain a permanent failure of the central state. Multiplication of policies, overlapping of competences, and lousy coordination represent only few aspects of the problem. Several laws have been voted with little positive impact: the legislation « Montagne » (Mountain) of 1985; the law for the Orientation and Development of Territories dated 04/02/1995; the bill for the Development of Rural Territories (23/02/2005). Circular letters addressed to the prefects by the Prime Minister and inter-ministerial committees for rurality took place between 2015 and 2016 defining a set of 104 measures to promote rural areas and ensure local development. Were considered as priorities: the access to health and the fight against « medical desert »; the access to services through Maisons de Services Publics (MSP) (Public Services Office); « Nomade » Public; and « mobile » Postmen Services; the fight against school weaknesses; the numeric coverage and implementation of networks in rural areas; the execution of « Contracts of Rurality ». Those committees were replaced by CGET soon to merge into ANCT as mentioned earlier.

The strategy of President Macron government for rurality should be read in the continuation of previous initiatives. As an example, Maisons France Services - a system presented as a new device - offers substantial similarities to MSP. In Ariège, 4 units are operational, all located in major cities (Saint Girons, Mirepoix, Tarascon sur Ariège, Ax Les Thermes) but none in remote villages. Most of already existing financial, social and fiscal tools are maintained and strengthened through the continuation of ZRR (Zones de Revitalisation Rurales), of rurality contracts, of reciprocity contracts. Innovation and cooperation shall take place in Territory Workshops (Atelier des Territoires) and Industrial Territories.

Still 18% of the total French population live in medical deserts [28], a figure to compare to the 19.56% representing the total rural population. The same remedies have been proposed over the years without success. Rural territories continue to see their young population disappearing, the closure of public and private services and the constant degradation of their images in a never-ending vicious circle.

VIII. A New Strategy Supported by Technology

Boosting the attractivity of rural territories is related to both a change in image, the development of light infrastructures of utilities and communications, the development of inner innovation capacities and value, the acceptation of the challenge by local people. Technology could be a fantastic chance to invert such an ineluctable destiny. Whatever we call it "Smart Village" or "Smart Basic Entity" (SBE), we advocate that a new frame of organization and development shall be studied. It may become a potential area of growth and allow the inversion of the concentration process into cities. The SBE model shall, of course, and provided customization, benefit of the technical innovations and successful realizations that would be implemented in Smart Cities. SBE and SC models shall not be competing with each other, but live side by side, completing each other through exchanges on data, technologies and experiences, through existing or to be developed networks.

What we foresee today is nothing else than a downsizing process; to go from mainframe (the state or the region) to a connected smart unit, the SBE. Such a reorganization of the territory underpins, on the medium term, a global rethinking of administrative and political organization. The SBE could be defined as an evolutive and complex system, without exact physical limit but characterized by a logic of flows (persons, assets and information) always looking for efficiency improvement. By analogy to the system of cities; SBE will be connected through them as a network: the system of SBEs. Flows exchanges will exist within the SBE and with the exterior.

A way to define Smart Cities is by using a tangible (Hard)/ intangible (Soft) domains approach [29]. By customizing such an approach, we may consider two axes:

- Tangible/Hard: Water, Energy, Land and Environment Resources (Forests, Mines, Parks…), Transportation, Buildings (including health care and education) infrastructures, Security systems
- Intangible/Soft: Governance, Education, Health, Economy, Culture systems and data

And seven related applications: Utility management, Land and Environment Ressources management, Mobility, Buildings, Economy, Security, People (Cultural, social, education and health)

New technology facilitates communication, remote financial operations, distance learning, online medicine. As a consequence, cities are on the way to lose most of their competitive advantages.

IX. The Smart Digital Clone Approach

By collecting data related to the 2 axes, we shall be able to create a digital clone of the SBE. It shall represent its tangible and intangible assets and allow to analyse flows within the SBE as well as exchanges with outside, mainly, within the system of SBE. Such an approach that we call SMART (for Systemic Modeling and Advance Reengineering of Territory) shall facilitate improvements, reinforce governance efficiency

The digital clone would allow, by simulating any structural or logical change, to evaluate it. It shall promote, as

well, the definition and implementation of contingency plans; and make possible dynamic stress tests on all tangible and intangible dimensions. It appears mandatory to respect when implementing such an approach, several rules that we summarized as the HAWKS principles. H: Holistic: the digital clone covers all aspects and enlightens even shadow areas A: Accepted: by the population W: Wise K: Creative S: Secured

The digital clone terminology is used by analogy with recent development in medical technology. If a digital clone can be created to save or cure a human patient, why not applying the same to study, monitor, optimize the living conditions in a rural SBE?

X. Angola and China Cases

In Southern Africa and more specifically in Angola, the question of the economic development based on a systemic (system and system of systems) approach is crucial. The problem is multidimensional from two different points of view. Multidimensional, according to the aspects which constitute the domains where solutions must be designed, developed, implemented and evaluated: production and consumption of energy, wastes treatments, water control (drinking and wastewater), forest preservation and agriculture development, artisanal (handmade) and industrial activities, mobility … . Multidimensional, because at the same time, cultural, societal, technological, collective and individual.

Under the umbrella of the DNDTI (National Direction for the Development of Technology and Innovation) of the MESCTI (Ministry for Higher Education, Sciences, Technology and Innovation) of Angola a "SMART Villages" project is close to being launched. The question is how new technologies can be used for assuring a real, sustainable development of a small city (village)? How could a global approach be defined and implemented? As we noticed at the beginning, the project aims to consider each field of challenge (energy, wastes, water …) as a system and to consider the global interactions between all these systems. The objective is not only to solve problems; it is more the optimization of different solutions under a set of balanced criteria (economic development, alphabetization rate, safety and security …).

This pilot project will consist of two main parallel parts. The first will be focused on scientific and technological studies to propose and implement selected solutions, for example, about production and consumption of energy. But the most original part could be the second one. This part will consist of developing a "digital ghost" (or digital clone) of the small city (village) based on the capture of data concerning all the aspects to take into account even the feelings and the opinions of the citizens, visitors tourists…. Also, if at the origin the "digital ghost" will be more a database than a real digital representation of the city, each development will be the occasion to reinforce the amount of data (information) creating step by step a new digital structure able to facilitate original representations of the city based on the selection of data according to specific criteria (as mobility by example).

On the other hand, to be able to assess the project itself and its results, a model of excellence, as the EFQM model used for evaluating the efficiency of enterprises or public organizations will be used. This approach will allow to assure or to analyze the results for different stakeholders. According to this last point, it will be interesting to compare the evolutions (transformations) of the "digital ghost" of the city with variations of the stakeholder's opinions reinforcing by this way the capacity to understand better how a SMART integrated development impacts positively or negatively the citizens' lives.

Another exploratory work was carried out in the city of Shanghai (China). In partnership with the company Mobike (free bicycle), students worked on the development of an onboard pollution capture system (integrated into the bike frame). The aim was to transmit this information to a platform in order to advise cyclists on "greener" alternative routes. The recording of pollution data, combined with travel data and other information (the digital ghost), resulted in a city (quarter) map of statistically green routes. But this digital ghost (big data) also made possible to study the behavior of the cyclist (deciding whether or not to follow the green route), depending on the level of pollution, distance, time of day, weather conditions, etc. In such a way, it was helpful in supporting the definition of urban policies. The analysis of data made it possible to identify clusters (e.g., long journeys) for which specific actions could be established (fare reduction linked to the use of the green route, longer trip, etc.).

XI. Conclusion

There is no inevitable future for rural territories. We firmly believe that the rural-urban shift is not a fatality and could be inverted provided rural areas modify their image and become land of innovation and investment. Thanks to technology, distance is no more a concern as regards as accessing information, finance and education. Environmental and security issues could reinforce the position of territories towards cities that have lost part of their competitive advantages. Such a revolution could take place at the SBE level and that the development of SMART Digital Clones constitutes a significant leap into the future for rural territories. Its application has no frontier from China to Angola, from rural territories of France to the ones of Indonesia. And the current ongoing pandemic of Corona virus, that constitutes a real stress test for the economy and states structures at the international level may question the future of megapoles. Isn't it time to invest in rural territories, therefore keeping in mind that their specificities and environment shall be preserved ?

References

[1] P. Fournié, C. Bourret, and J.-P. Caliste, "Smart Territories: Advocating for Smart Basic Entity (SBE) and the digital clone approach," in Proc. *The Sixth International Conference on Human and Social Analytics*, HUSO, IARIA, 2020, pp. 9-14.

[2] https://data.worldbank.org/indicator/SP.RUR.TOTL.ZS, retrieved on December 15th, 2021.

[3] M. Pesaresi et al., "Atlas for the Human Planet 2016: Mapping human presence on Earth with the Global Human Settlement Layer", European Commission, 2016

[4] D. Reinsel, J. Gantz and J. Rydning, "The Digitization of the World: From Edge to Core", An IDC White Paper #US44413318, 2018.

[5] UN Economic and Social Council, Smart cities and infrastructures. Geneva, 2016.E/CN.16/2016/2.

[6] New Urban Agenda - Habitat III. United Nations. Quito, Ecuador, 2016. ISBN: 978-92-1-132731-1.

[7] J. Jacobs, “The economy of the cities”, New York, Vintage Books, 1969.

[8] M. Polèse, “Cities and National Economic Growth: A Reappraisal”, Urban Studies, Vol. 42, 8, pp. 1429-1451, July 2005.

[9] G.S. Tolley and V. Thomas, “The economics of urbanization and urban policies in developing countries”, The World Bank,1987, p. 29 and 185.

[10] J. Véron, “Enjeux économiques, sociaux et environnementaux de l'urbanisation du monde”, Mondes en développement, Vol. 2008/2, 142, 2008, pp. 39-52.

[11] D. Castells-Quintana and V. Royuela, “Malthus living in a slum:Urban concentration, infrastructures and economic growth”, St. Petersburg, Russia : European Regional Science Association (ERSA), Louvain-la-Neuve, 26-29 August 2014.

[12] P. Bairoch, “De Jéricho à Mexico, Villes et économie dans l’histoire”, Paris, Gallimard, 1985.

[13] N. Keyfitz, “International Migration and Urbanization”, in P. Demeny and M.F. Perutz, Resources and Population, Clarendon Press, 1996, pp. 269-285.

[14] K.F. Sokoloff, “Inventive Activity in Early Industrial America: Evidence From Patent Records, 1790–1846”, The Journal of Economic History, Vol. 48, 04, 1998, pp. 813-850.

[15] P. Bairoch, “Cinq millénaires de croissance urbaine” in I. Sachs. Quelles villes, pour quel développement ?, Paris, PUF, 1996, pp. 17-60.

[16] J. Henrich, “Demography and cultural evolution: How adaptive cultural processes can produce maladaptive losses-The Tasmanian case”, American Antiquity, Vol. 69, 2, 2004, pp. 197-214.

[17] P.M. Romer, “Endogenous Technological Change”, The Journal of Political Economy, Vol. 98, 5 Part 2, October 1990, pp. S71-S102.

[18] E.L. Glaeser et al., "Growth in Cities”, Journal of Political Economy, Vol. 100, 6, 1992, pp. 1126-1152.

[19] A. Marshall, “Principles of Economics”, Macmillan and Co, New York, 1890.

[20] G. Duranton and M.A. Turner, “Urban growth and transportation”, The review of economic studies, Vol. 79, 4, October 2012, pp. 1407-1440.

[21] E.L. Glaeser and J. Shapiro, “Is there a new urbanism ? The growth of the US cities in the 1990's”, Cambridge, MA, National Bureau of Economic Search, 2001. Working Paper 8357.

[22] V. Vuchic, “Transportation for livable cities”, New York, Routledge, 1999. 378 pages.

[23] B.J.L. Berry, “Cities as systems within systems of cities”, Papers of the Regional Science Association, 1964, Cited by Pumain, 1997.

[24] A. Pred, “City systems in advanced societies”, London, Hutchinson, 1977.

[25] D. Pumain, “Pour une théorie évolutive des villes”, Espace géographique, Vol. 26, 2, 1997, pp. 119-134.

[26] A. Rousseau, “L'Ariège, l'anti-Start-Up Nation”, March 22nd 2019, Les Echos.

[27] ZoneAdsl, Couverture internet de l'Ariège, retrieved 1st March 2020. https://www.zoneadsl.com/couverture/ariege/.

[28] C. Maisonneuve, “Nouvelle carte des déserts médicaux : votre commune est-elle concernée ? ”, La gazette.fr, 2017.

[29] P. Neirotti et al., “Current trends in Smart City initiatives: Some stylised facts”., Cities, June 2014, pp. 25-36.

[30] H. Dou and P. Fournié, “Les Smart Villages dans le contexte européen” ,Val d'Europe, Codata 2019: Data value chain in Science and Territories, 2019, pp. 19-26.

Cooperation Strategies for Swarms of Collaborating Robots: Analysis of Time-Stepped and Multi-Threaded Simulations

Liam McGuigan, Catherine Saunders, Roy Sterritt, George Wilkie
School of Computing, Faculty of Computing, Engineering and the Built Environment
Ulster University
Jordanstown, N. Ireland
email: mcguigan-l8@ulster.ac.uk, c.saunders@ulster.ac.uk, r.sterritt@ulster.ac.uk, fg.wilkie@ulster.ac.uk

***Abstract*— Swarms of robots have been proposed for use in many tasks, such as space exploration, search & rescue operations, and mine clearance. For a robot swarm to be successful, it needs to be self-adaptive, making its own decisions and adjusting its behaviour without relying on human intervention. This paper investigates the potential for using an autonomic system for a robot swarm engaged in a foraging task, capable of adjusting its cooperation strategy based on the ongoing performance in the task, rather than sticking with an initial strategy. The results show that while support for changing the strategy completely is limited, there remains the potential for adjusting the parameters of the given strategy to suit the ongoing situation. In addition, a comparison of two approaches to the implementation of a simulation is also presented. A time-stepped approach is compared with the multi-threaded approach used in previous work, with a view to embedding simulation within the swarm as a means of aiding the autonomic decision-making process through simulation of potential options. It is found that even when the underlying robot behaviour is identical, the time-stepped simulation is faster and more flexible, and is therefore more suitable for embedding.**

Keywords- Swarm robotics; Self-adaptation; Autonomic Computing; Simulation.

I. INTRODUCTION

This paper is an extended version of the work published in [1]. It extends those results with new findings from further research.

The use of robotic swarms consisting of a large number of robots operating in concert can benefit applications, such as space exploration [2][3], search & rescue [4] and mine clearance [4][5] among others, taking advantage of a robot's ability to operate in conditions where human involvement is too dangerous or difficult.

The individual craft in a robotic swarm will need to be capable of managing themselves without requiring constant supervision. They may be required to make quick decisions to protect themselves or to act on opportunities, and will need to adapt to best suit the conditions of the task being carried out [6]. This can be achieved by making the swarms autonomic.

Autonomic computing concepts will embody the swarm with the properties of self-configuration, self-healing, self-optimization and self-protection, ensuring that the swarm [7] is implemented by including an autonomic component running a Monitor, Analyse, Plan, Execute, with a shared Knowledge (MAPE-K) control loop to monitor and analyse the situation, and plan and execute any changes to behaviour aided by a knowledge base of pre-set or previously acquired information [8], as seen in Figure 1. Autonomic robotics combines the concepts of MAPE-K from autonomic computing, with Intelligent Machine Design (IMD) from robotics [9][10].

Due to the cost and impracticality of using real hardware during the development of large-scale swarm behaviour, simulators are often used in the process [11], able to create artificial swarms of hundreds or even thousands of robots engaged in tasks, such as foraging, surveillance and exploration of unknown environments.

The research presented in this paper has two objectives. The first is to investigate the potential for self-adaptation through selection of a cooperation strategy during a foraging task, through analysis of the performance of different strategies over the course of the task. The second objective is to identify which of two simulation implementation approaches used would be most suitable for deploying on an individual agent within the swarm as a means of using simulation-in-the-loop to help with the in-task strategy switching decision.

The rest of this paper is organised as follows. Section II discusses related work on self-adaptation in swarm robotics, and the varying use of simulations in development. Section III describes the foraging task used in the simulation, and the

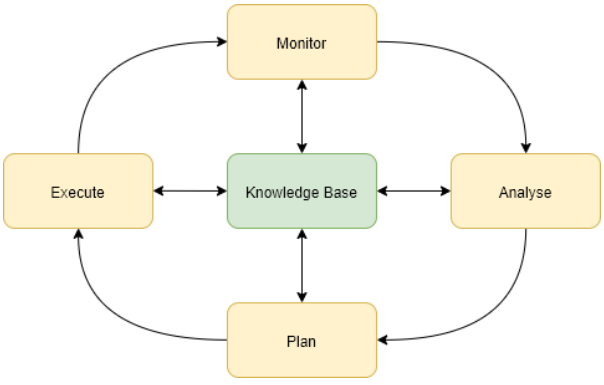


Figure 1. MAPE-K loop, as used by an autonomic component. The loop proceeds through each of the four stages in turn.

general robot behaviour. Section IV describes the cooperation strategies used in the research, while Section V describes the simulators used for comparison. Section VI describes the test scenarios that were run, Section VII presents the results of the cooperation strategy comparison, and Section VIII presents the results of the simulator comparison. Section IX concludes the paper with a summary and presents the future research directions.

II. Related Work

The following subsections discuss current research in swarm self-adaptation, and the use of simulations within swarm development.

A. Swarm Self-Adaptation

Self-adaptation of a robotic swarm concerns the ability of the swarm to adjust its behaviour in response to external or internal conditions, such as a foraging swarm choosing to abandon a depleted deposit in order to find newer deposits, or a surveillance swarm organizing itself so as to provide maximum coverage of the target area.

Swarm self-adaptation can be considered based on two factors – the approach to adaptation, and the location within the swarm where this is applied. Approaches to swarm adaptation include engineering emergent processes where adaptation arises naturally out of the agent behaviour [12], reasoning and learning approaches where the swarm explicitly reasons about the decision being made [13] and may learn from experience [14], and evolutionary approaches which explore alternatives through genetic algorithms [15].

Regarding location, a lot of the research focuses on applying adaptation to individual agent behaviour [16]–[18]. This low-level adaptation results in a bottom-up approach to swarm behaviour, with the resulting performance of the swarm arising from the aggregate performance of the individual agents. This can allow for more specific adaptation, such as balancing an individual's conflicting objectives [19], which may be difficult to apply at the swarm level. Agent behaviour adaptation can have the most direct impact on the swarm's performance, but it is difficult for an agent to make an individual decision on aspects of collaboration or coordination between multiple agents.

Adaptation through the selection of swarm-level cooperation strategies can be used to address the problem of collaboration. In this approach, agents within the swarm can collectively determine an alternative approach which is swapped with the existing agent behaviour either in part or in whole. This selection may be driven by an autonomic component that assesses the suitability of alternative strategies [14][20], and may be employed with in a subset of the entire swarm [21].

Where research has traditionally considered homogeneous swarms with simpler behaviour, a swarm of heterogeneous robots can be more flexible and capable of handling a wider variety of situations [22].

This research is focused on identifying the potential for swarm-level adaptational changes by assessing the performance of a selection of candidate strategies in a set of scenarios. Through noting any effect the scenario has on the performance of a particular strategy, the benefits of the ability to select an alternative strategy will become apparent.

B. Use of Simulations in Swarm Development

Simulation has long been employed as a tool for the development of robotic and swarm simulations, providing the means to test and analyse systems in an artificial environment. Simulators range in complexity, from detailed physical simulations of actual robots [11][23], to abstract approaches where robots move within a grid-based environment. The difficulty of producing an accurate simulation of the real world can manifest as the "reality gap" [24], where results obtained in simulation are not replicated in reality. Nonetheless, it is not necessary for the results of a simulation to be precisely reproducible in the real world for the simulation itself to prove useful.

Simulation need not be restricted to the offline development phase. It may be used to assist the decision making process [25][26], trying out "what-if" scenarios in order to assess the effects of potential actions or strategies ahead of time. For this to be effective, a simulation must be detailed enough to provide useful information, while remaining lightweight enough to be able to run on an individual agent within the swarm.

When choosing or designing a simulator for researching robotic swarms, the accuracy of the physical simulation required will depend on the impact specific hardware has on the research being conducted. Developing a robotic controller without a suitably accurate physical simulation can lead to the robots in the simulation carrying out behaviour that is impossible with the actual robots [27], but when researching purely software based systems, abstractions can be used to trade accuracy for a faster execution time [23].

Further gains in execution time may be made by simplifying the world representation. Grid based approaches need not produce markedly different results to continuous space [28], and can be used in cases where the specific motion of agents can be abstracted.

The majority of multi-robot simulators available make use of discrete time when updating their simulations, in which all agents and physical reactions are updated in sequence with a small time step, rather than independent execution in real time, such as assigning a robot its own execution thread. This ensures synchronous execution of robots [23] and simplifies physical interactions.

The research conducted in this paper abstracts physical movement using cell-based movement within a grid, and the performance is contrasted with a real-time, multi-threaded approach used in previous research [20].

III. FORAGING TASK DESCRIPTION

This research focuses on simulations of a swarm of heterogeneous robots engaged in a foraging task. Each run begins with a world constructed as a rectangular grid of cells, in which a number of target items and robots are initially placed, as shown in Figure 2. Each cell may contain only one item, but any number of robots. It can be assumed that each cell is represents a larger area than the footprint of a single robot, and the simulation may therefore ignore potential collisions.

The robots are tasked with foraging for all the items in the area. Each item is associated with a type, and may only be foraged by a robot with the corresponding type. The simulation proceeds until all the items have been foraged.

Foraging of an item happens at the place it is found, rather than returning to a home base – the simulation can therefore be considered to logically represent applications, such as mine deactivation, analysis of mineral deposits, or environmental cleanup.

Each robot begins in an exploration state, during which they move towards a random location in the world. Each step, if an item is found, the robot will forage if it is able to, removing the item from the world, before resuming exploration.

In the extended research, the behaviour in the exploration state has been modified so that instead of selecting a random location on the map to move towards over time, a robot engages in a random wander, selecting an available direction to move each update.

IV. COOPERATION STRATEGIES

For the initial research, cooperation during the task is determined using one of three strategies, as developed in [20]. The extended research updated this with a set of three configurable strategies, two of which replicate the initial approaches. The following subsections describe each set of strategies.

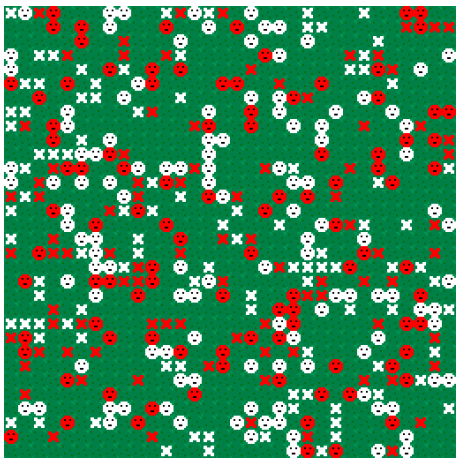


Figure 2. A view of the simulation at the start of a task. The colour of a robot (face) or item (cross) indicates its type.

A. Initial Strategies

When a robot encounters an item that it cannot forage, it broadcasts a help request with a range of 5 cells. The behaviour of the robots is then determined by the strategy:

1) Multiple Responders: If not already engaged in foraging or responding to a previous request, a receiving robot of the correct type will respond to the request by transitioning to a Respond state, in which it moves towards the item. All receivers, whether they can forage the item or not, will rebroadcast the message. In this way, the message will filter throughout the swarm. The robot initiating the help request plays no further part in the cooperation and returns to exploration.

2) Selective Responders: The behaviour here is similar to the Multiple Responders approach, but the message is only rebroadcast if the receiving robot cannot help. This works to reduce the number of robots responding to the request.

3) One Responder: The robot initiating the request transitions to a WaitForOffers state. Offers to help are sent by receiving robots that meet the criteria, who themselves transition to a WaitForAssignment state. No rebroadcasting of the message occurs. If no offers are received after a short delay, the requesting robot returns to its previous behaviour, otherwise it assigns the task to the nearest responding robot and resumes exploration. Robots that do not receive assignment after a period of time return to exploration, while the assigned robot transitions to the Respond state.

Both Multiple and Selective Responder strategies are likely to result in multiple robots moving towards the item. This would provide contingency in the event of robot failure before reaching the target item. Robot failure is not implemented in the current simulation, but will be in a future study.

B. Extended Configurable Strategies

These strategies were used in the extended research. Each strategy offers one or more parameters to fine tune the behaviour.

1) Simple Broadcast: This strategy replicates the behaviour of the Multiple and Selective Responders strategies described above, but parameters allow for the modification of the broadcast range, the number of rebroadcasts, and the likelihood of a robot responding to the request. The parameters are listed in Table I.

2) Help Recruitment: This strategy replicates the One Responder strategy, with parameters to allow adjustment of the broadcast range and maximum number of recruits, as listed in Table II.

3) Blackboard: In this new strategy, each robot maintains a record of items found and items foraged, and periodically synchronises this information with neighbouring robots via broadcasts. During the update loop, a robot will move to the nearest unforaged item of a

matching type that it is aware of, provided it is within the configured range. If a robot subsequently receives information to say that an item has been foraged, it will abandon its efforts to retrieve the item, and resume exploration. The configuration parameters are described in Table III.

This is a fully decentralised implementation in which robots gather and share information. When a robot finds an item, it records this in its memory as unforaged. If it subsequently forages the item, it updates the record. During synchronisation, a robot updates its current memory, adding new items it was not aware of, or updating the forage state of items it previously believed to be unforaged.

Due to the decentralised nature of the blackboard implementation, each robot will have a different interpretation of the current state of the field. However, as the synchronisation process will only move an item from unknown, to found, and then to foraged, it is not possible for the swarm to "forget" an item's state, only to be unaware of any change. As the simulation proceeds, the data will filter throughout the swarm as robots rebroadcast the new details.

TABLE I. SIMPLE BROADCAST PARAMETERS

Parameter	Description	Values
Broadcast Range	Maximum cell distance each broadcast for help can reach	[1 .. ∞]
Max Chain	Number of rebroadcasts for any given message	[0 .. 4]
Rebroadcast Mode	Determines whether or not a robot rebroadcasts	Never Type Mismatch Always
Response Type	How a robot responds to a request	Always Response Curve
Response Distance	The maximum distance at which a robot may respond (Response Curve only)	[0 .. ∞]

TABLE II. HELP RECRUITMENT PARAMETERS

Parameter	Description	Values
Broadcast Range	Maximum cell distance each broadcast for help can reach	[1 .. ∞]
Max Recruits	Maximum number of robots to assign an item to	[1 .. ∞]

TABLE III. BLACKBOARD PARAMETERS

Parameter	Description	Values
Broadcast Range	Maximum cell distance each synchronisation broadcast can reach	[1 .. ∞]
Response Range	Maximum distance of an unforaged item that prompts a response	[1 .. ∞]
Sync Period	Simulation ticks between synchronisations	[1 .. ∞]

V. SIMULATORS

The following sections describe the simulators that were used in this research.

A. Time-Stepped Simulator

In the time-stepped simulator, each robot in the simulation is updated in sequence, with a single tick of the simulation ending once all robots have been updated. Each tick of the simulation can therefore represent a discrete period of time, and the performance of the swarm at completing the task can be measured by the simulation ticks taken to complete the task. Figure 3 shows an example of how the time-stepped simulator updates.

Robots are implemented as a finite state machine (FSM). In each tick of the simulation, the robot will update its current state, and check for state transition conditions. Thus, in a single tick a robot may choose to move one cell, to forage an item underfoot, or to participate in the cooperation strategies described in Section IV.

In this simulation, broadcast messages are first buffered, and only processed after all robots have been updated, in order to maintain synchronisation between independent robot movements, and avoid the possibility of a later-updated robot being able to process a message that was sent in the same tick by an earlier-updated robot. Once the messages are processed and received by reach robot within range, they are further shuffled on the receiving robot in order to remove the effects of update order. If Robot A and Robot B both send a message to Robot C in the same tick, Robot C would always respond to Robot A first. Shuffling the message list before processing removes this problem.

B. Multi-Threaded Simulator

The multi-threaded simulator implementation used here is the same as that presented in [20]. In this implementation, each robot is assigned its own independent CPU thread, complete with pauses to allow for real-time viewing of the simulation.

Each robot executes an infinite loop and carries out a set of steps in sequence: process messages, move one cell towards a target destination, look for an item in the current cell, and forage/request help as required. Figure 4 shows an example of how the multi-threaded simulator operates.

Broadcast messages are sent and buffered by each robot, and processed at the start of their loop. If the robot is using the One Responder strategy, after a help request is sent the robot pauses to wait for an offer, and similarly after an offer is sent, a robot pauses to wait for assignment.

Unlike the time-stepped simulator, this simulator is restricted to a 30x30 map, so all comparisons between the two must be made on that map size.

C. Redesigned Multi-Threaded Simulator

In order to address the fixed map size and differences in robot implementation in the previous multi-threaded simulator, this was rewritten from the ground up in order to

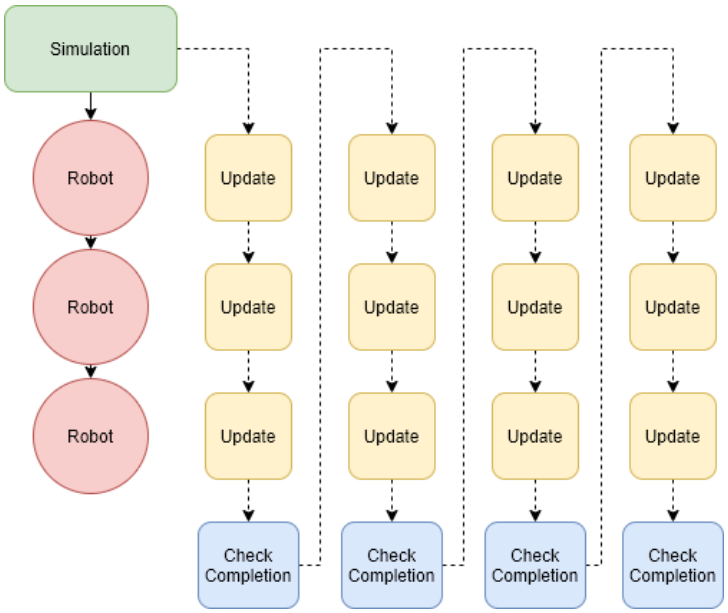


Figure 3. An example of how the time-stepped simulator updates the robots. All robots are updated in sequence on the same thread, represented by the dotted line, before the simulator checks for task completion. If the task is still active, the process repeats.

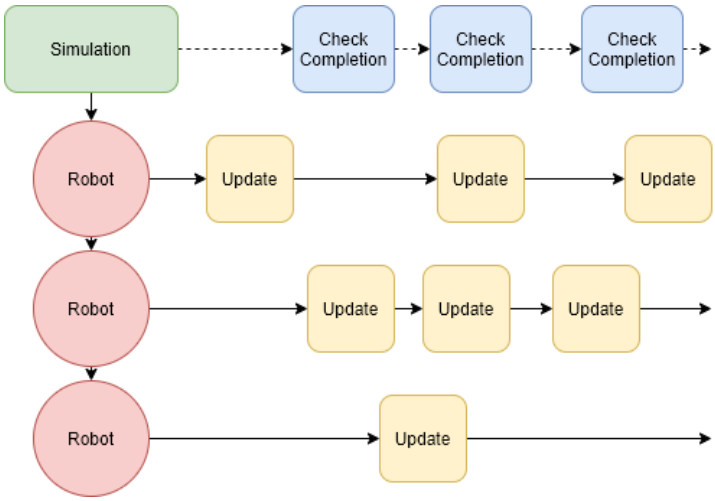


Figure 4. An example of how the multi-threaded simulator updates the robots. Once created, each robot updates on a separate thread to the main simulator and to each other. The frequency and order in which the robots update cannot be controlled. The execution of an update may also be halted at any time – this has not been shown for clarity.

allow a greater variety of tests, and to ensure robot behaviour is identical.

Again, each robot is assigned its own CPU thread and started in a random order, however thread execution is no longer tied to display considerations, and sleeps are limited to 1 millisecond per loop in order to prevent the operating system from locking up during execution. Robots implement the same state machine as in the time-stepped simulator, and messages are also handled in the same way, however no shuffling occurs as the order robots are updated is no longer fixed. If a robot must wait in a particular state, it does so over a number of updates rather than halting execution.

In this implementation, only the Help Recruitment strategy has been implemented, as the intention is to compare performance of the simulators rather than evaluate strategies.

VI. Test Scenarios

The following subsections describe the test scenarios used to evaluate the strategies and simulator approaches.

A. Initial Research

For the initial research, three scenarios were used as described in Table IV.

The Robot Type Imbalance scenario represents a scenario where the swarm configuration deployed is not best suited to the task, and must adapt. The Item Type Imbalance scenario represents one where the reality of the mission differs from that expected, and again the swarm must adapt.

To compare the performance of the cooperation strategies, each scenario is tested on both 30x30 and 90x90 maps. Each simulation is run 30 times, with the initial placement of items and robots randomised at the start of each run.

For simulator comparison, due to limitations of the multi-threaded simulator used, only the Equal Split and Robot Type Imbalance scenarios were able to be run on a 30x30 map. Each scenario was run 30 times on the time-stepped simulator, and 10 times on the multi-threaded simulator.

In assessing the performance of each strategy, the number of simulation ticks until completion of the task is the main metric, as it is a measure of the time taken to forage all items. If the energy cost of actions taken by robots is of interest, then the total number of steps and the number of messages broadcast will also become factors. The simulation does not currently assign an energy cost to individual actions, but the counts may be used as a guide, and for each metric a lower value is considered more efficient.

B. Extended Research

Following on from the above work, two further sets of tests were conducted to compare strategies and simulator approaches. To confirm that no dominant strategy exists and thus support the use of an autonomic system for strategy selection, variants of each of the new set of strategies were created, as described in Table V.

These strategies were each tested in a variety of scenarios created by adjusting the map size, robot counts, item counts, and the ratios of each type, as described in Table VI. A total of 600 scenarios were used to compare the strategies.

To establish any differences in performance between the two simulation implementations, tests were run on four map sizes, each featuring 256 robots and items evenly distributed between the two types, in order to examine the performance of the two simulations in larger environments.

Each scenario was run using the Help Recruitment strategy, and also with no cooperation strategy enabled. To account for the increased distance between robots that would arise in a larger map, the broadcast range for the Help Recruitment strategy was increased in as the map size increased, with 32x32, 64x64, 96x96 and 128x128 maps having ranges of 4, 8, 16 and 24 respectively. Each scenario was run 100 times on each simulator.

For each set of simulations, the average real time to complete the task was recorded, along with the simulation ticks (for the time-stepped simulation) or mean number of robot updates (for the multi-threaded simulation), the total steps taken by all robots in the task, and the number of messages sent, also broken down by message type.

TABLE IV. INITIAL TEST SCENARIOS

Scenario	Items		Robots	
	White	***Red***	***White***	***Red***
Equal Split	100	100	100	100
Robot Type Imbalance	100	100	180	20
Item Type Imbalance	180	20	100	100

TABLE V. STRATEGY VARIANT PARAMETERS

Strategy	Parameter	Values	Variants
Simple Broadcast	Broadcast Range	8, 16	40
	Max Chain	0, 1, 2	
	Rebroadcast Mode	Never, Type Mismatch, Always	
	Response Type	Always, Response Curve	
	Response Distance	32, 64, 128	
Help Recruitment	Broadcast Range	8, 16, 24, 32	12
	Max Recruits	1, 2, 3	
Blackboard	Broadcast Range	8, 16	32
	Response Range	8, 16, 32, 64	
	Sync Period	8, 16, 32, 64	
No Cooperation	N/A		1
Total			**85**

VII. RESULTS – COOPERATION STRATEGY COMPARISONS

The following subsections present the results of the initial research, followed by the extended work which followed.

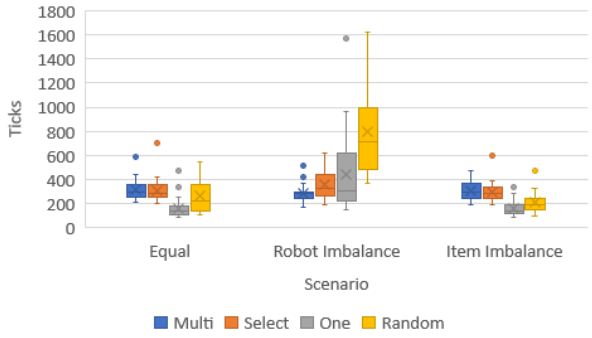
A. Initial Research

Figure 5 shows the ticks to completion, steps taken, and messages sent for each of the test scenarios in a 30x30 grid.

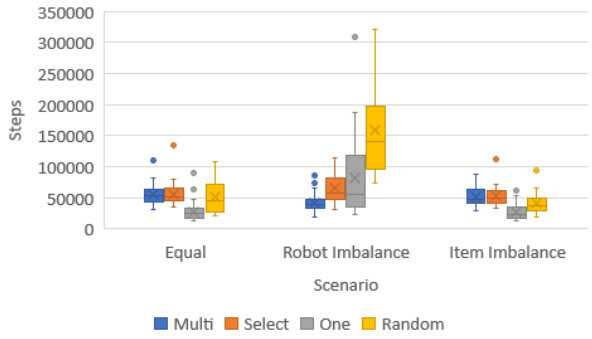
Comparing the results in both the Equal and Item Imbalance strategies, the One Responder strategy is the best performing approach, having the lowest count in each metric. Multiple and Selective Responder strategies can actually perform worse than no cooperation strategy at all, which can be explained by robots that respond to messages halting any exploration while they respond.

TABLE VI. SCENARIO GENERATION PARAMETERS

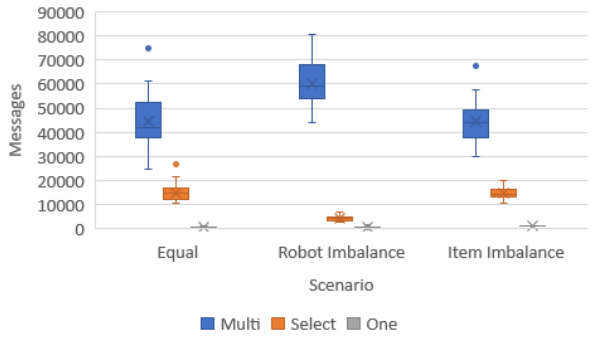
Parameter	Values
Map Size	32x32, 64x64, 96x96, 128x128
Items	64, 128, 256, 384, 512
Item Ratio (White:Red)	1:1, 7:1
Robots	64, 128, 256, 384, 512
Robot Ratio (White:Red)	1:1, 7:1, 1:7



(a) Simulation ticks



(b) Steps taken



(c) Messages sent

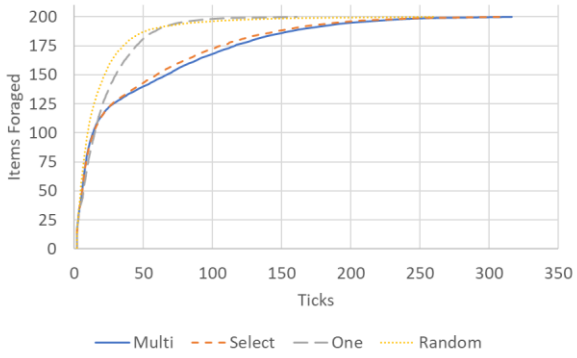
Figure 5. Metrics for each cooperation strategy in a 30x30 map. Circles represent data outliers: (a) simulation ticks, (b) total steps taken, (c) messages sent

In the Robot Imbalance scenario (Figure 5 (b)), however, One Responder does not reliably perform, and is subject to a great deal of variance caused by the initial placement of items and robots, and the subsequent movement of robots within the arena.

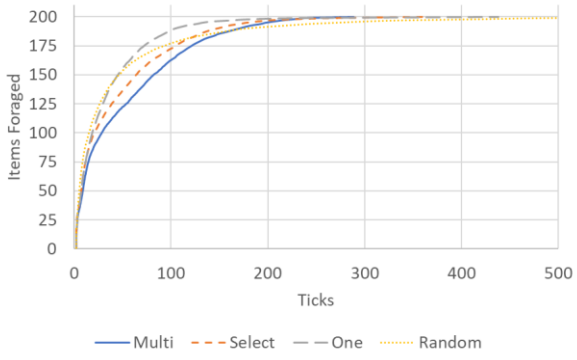
When considering energy costs, Multiple Responders has an extremely high message count setting it apart from Selective Responders, which it otherwise performs very similarly to. A full assessment of the respective efficiency of each would require an assignment of cost to each of the metrics, with the total cost calculated accordingly.

Figure 6 shows the progress of each strategy over time for the three scenarios. In Equal (a) and Item Imbalance (c) scenarios, performance is again similar, however it is notable that using no cooperation strategy at all is the quickest approach until the item count decreases substantially, after which One Responder performs best. This would suggest some system of changing the cooperation strategy used during the test based on the changing situation could lead to stronger overall performance, at least in terms of time taken.

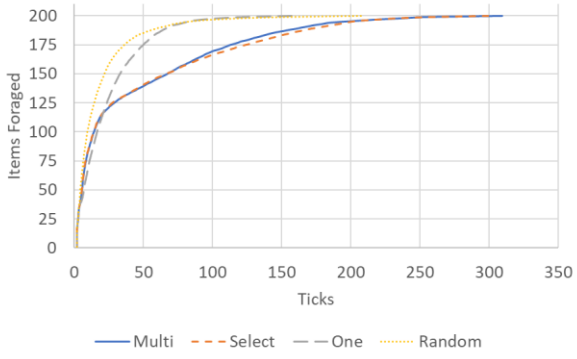
In Figure 6 (b), the Robot Imbalance scenario shows only a slight favouring of Random and One Responder strategies until most items are gathered, but the imbalance of robots then leads to both strategies taking much longer to complete the task than the other approaches. Again, strategy



(a) Equal



(b) Robot Imbalance



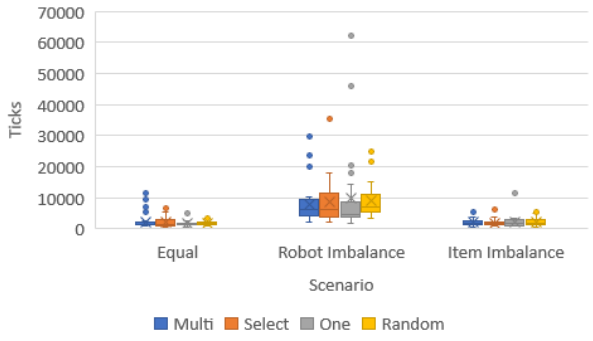
(c) Item Imbalance

Figure 6. Items foraged over simulation ticks for each of the strategies in a 30x30 map: (a) Equal scenario, (b) Robot imbalance scenario, (c) Item imbalance scenario.

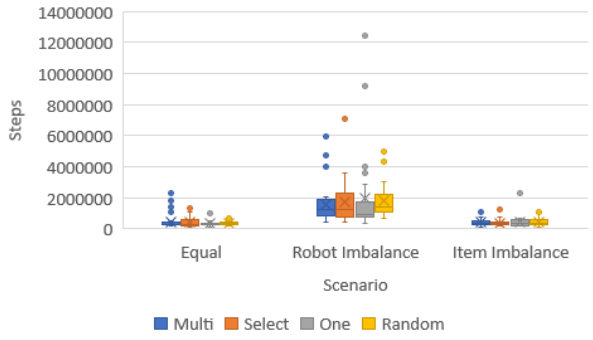
selection during the task could recognise this situation and adopt the strategy most suited.

If individual robot failure is considered, a robot imbalance can occur during the task. A system that can monitor the current swarm composition as well as estimate the progress in the task would therefore be able to adopt a suitable strategy in response to such unpredictable change.

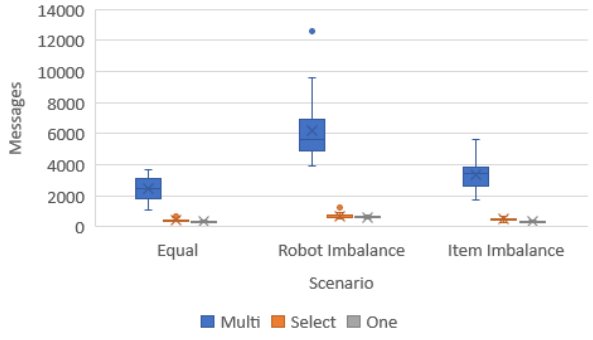
Figure 7 shows the ticks to complete, steps taken, and messages sent for the cooperation strategies in the larger 90x90 grid. Here, it can be seen that the performance of each strategy tends towards that of no cooperation, with large variances in the data and, other than the number of messages sent, similar average values for each metric in each scenario.



(a) Simulation ticks

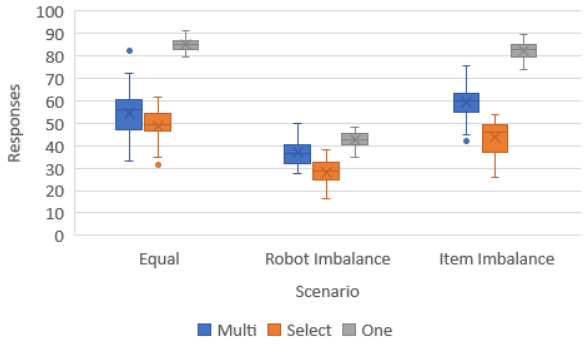


(b) Steps taken

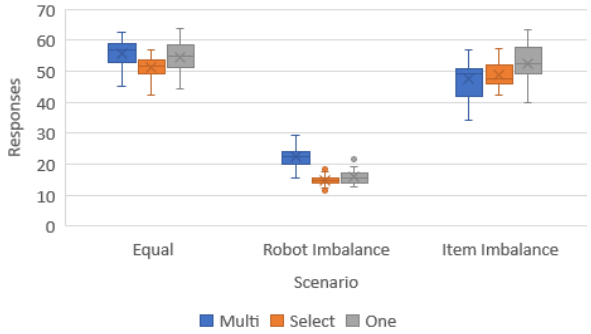


(c) Messages sent

Figure 7. Metrics for each cooperation strategy in a 90x90 map. Circles represent data outliers: (a) simulation ticks, (b) total steps taken, (c) messages sent



(a) 30x30 map



(b) 90x90 map

Figure 8. Percentage of help requests receiving at least one response, for each cooperation strategy and scenario. Circles represent data outliers: (a) 30x30 map, (b) 90x90 map.

It may be expected that the larger map explains the results as messages are not being broadcast far enough in order be received, but a comparison of data in Figure 8 shows that this is not necessarily the case. The proportion of requests receiving a response does not change much between the map sizes for the Multiple and Selective Responders cases, other than when there is a robot imbalance where it can be understood the chances of a robot of the correct type being nearby is significantly lower in a larger area.

The One Responder strategy can be seen to have a much higher percentage of requests receiving a response than the other approaches in a 30x30 map. This is due to the other approaches causing robots who would be able to help to be otherwise engaged in moving to forage an item, and thus unable to respond until they complete that help task. As the One Responder strategy causes only one robot at most to take on a task, other robots remain to be selected. In the 90x90 map, this then drops because of the distance between robots, and more closely matches the other approaches.

The dominant effect in the 90x90 map is the random exploration of the environment, and can be seen in the time taken to complete the task and understood by considering that the number of items remains the same between the two maps. As such, only 2.5% of the cells in a 90x90 map have

an item, compared to 22.2% of the cells in the 30x30 map. It is this decreased chance of stumbling upon an item that has the strongest effect on swarm performance.

The results suggest that allowing a swarm to adjust its cooperation strategy during a task, rather than relying on an initial strategy, could prove beneficial to performance by allowing the swarm to adjust its approach in response to the situation.

B. *Extended Research*

With 600 scenarios and 85 strategy variants, a direct comparison similar to the above is not possible with the extended data. However, as the aim is to show that there is no dominant strategy, a broader overview is sufficient.

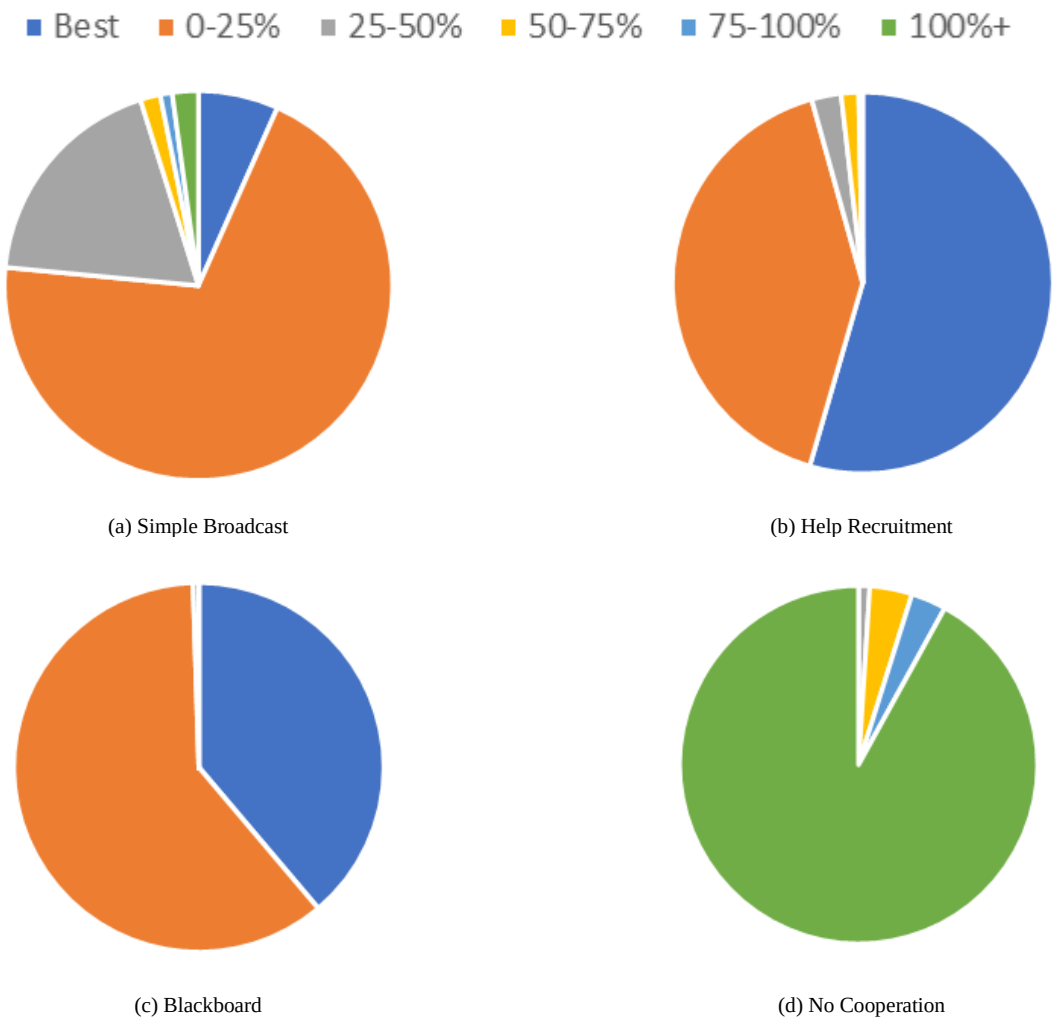
Figure 9 is a summary of the performance of each strategy across all 600 scenarios, as determined by the best performing strategy variant in each scenario, relative to the best performance from all four strategies.

The Help Recruitment and Blackboard strategies perform strongest, in that they each have a high proportion of scenarios for which they are the best performing strategy. At the other end of the scale, the No Cooperation strategy performs poorest, in line with expectations, usually taking more than twice as long to complete as the best performer in a given scenario.

This suggests that the initial hypothesis is confirmed and that no dominant strategy exists, as a scenario best suited to each of the three cooperative strategies may be found. Further, there is no set of parameters for any one strategy that performs well in all scenarios.

Examining the Blackboard strategy further, we see that in all but three of the cases, it performs within 25% of the best performing strategy, and in fact its worst relative performance sees it take just 28% longer than the best strategy, whereas both Simple and Help Recruitment may perform worse in certain scenarios, each having at least one scenario where they take more than twice as long as the best strategy.

Figure 10 compares each strategy according to the



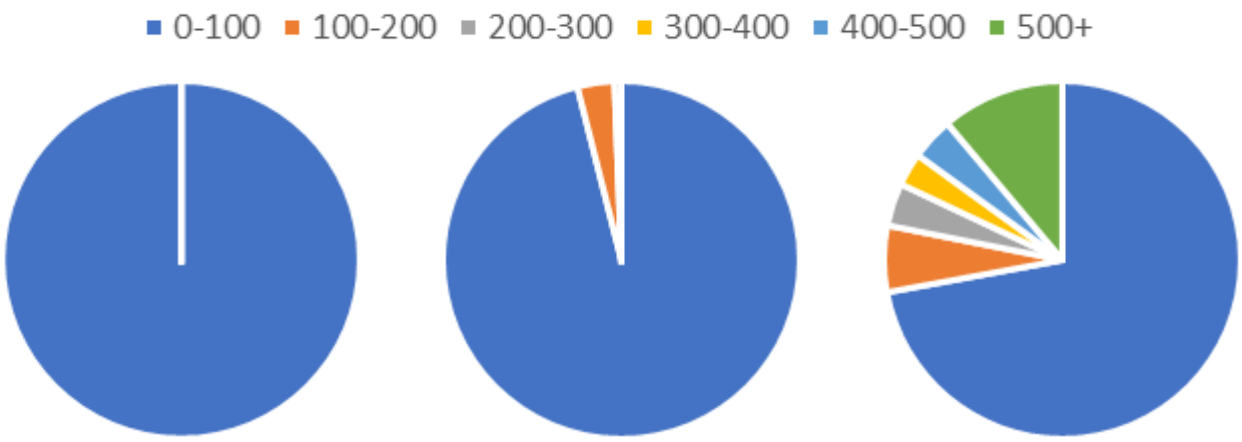
(a) Simple Broadcast

(b) Help Recruitment

(c) Blackboard

(d) No Cooperation

Figure 9. Performance of each strategy across all scenarios. Percentages indicate additional time to complete task relative to best performing strategy.



(a) Simple Broadcast (b) Help Recruitment (c) Blackboard

Figure 10. Performance advantage of each strategy, for scenarios where it is the best performer. Values indicate the difference in simulation ticks to complete the task between the strategy, and the second best option.

performance advantage of each strategy for the scenarios where it is the best performer. The Simple Broadcast strategy offers only a small performance benefit, averaging 6.9 ticks better than the nearest challenger, and with a maximum recorded advantage of 22.5 ticks. Help Recruitment fares little better, averaging 26.8 ticks with a maximum of 208. The Blackboard strategy, on the other hand, shows a clear advantage in many scenarios with an average boost of 264.0 ticks, and a maximum recorded lead of 3515.

This would suggest that the Blackboard strategy, while not dominant in terms of being the best strategy in any given scenario, performs strongly enough in each case that it could be considered de facto dominant, with minimal benefit offered in choosing another strategy and the associated complexity of implementing an autonomous system to make that choice.

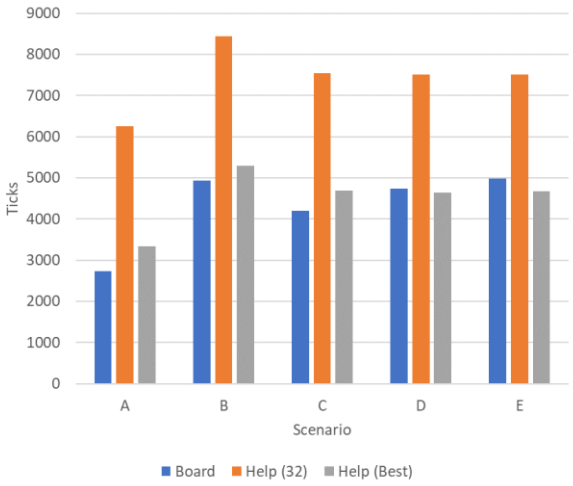


Figure 11. Simulation ticks to complete the five scenarios with the largest performance difference between the Blackboard and Help Recruitment strategies, before and after extending the maximum broadcast range.

However, a closer look at the scenarios in which the Blackboard strategy is dominant reveals a different story. There are a total of 19 scenarios in which the Blackboard outperforms its nearest rival by over 1000 ticks, and each involves the largest map, with the smallest amount of robots with a ratio of 1 white to 7 red robots. As a result, the density of the robot swarm, and in particular the density of the smallest group of robots that would expect to be the target of most requests for help, is at its lowest.

The maximum broadcast range used for the Help Recruitment strategy in this work was 32 cells. Such a range covers less than 20% of the map, targeting one of just eight robots that may be found elsewhere, so it is not surprising that the strategy struggles when the swarm density is reduced. Simulating the 128x128 scenarios for Help Recruitment using increased ranges of 48, 64, 80 and 96 cells produced comparable results to the Blackboard strategy, as seen in Figure 11.

Rather than implementing a complex system to select a strategy based on identifying the scenario a swarm finds itself in, the results indicate that either a Help Recruitment or a Blackboard strategy would provide suitable performance given the right parameters. An autonomous system may still be used in order to adjust those parameters accordingly.

In addition, it may be better to use the simpler of the strategies with the least overhead – a Help Recruitment strategy requires no memory of items found and foraged, and does not require any periodic communication with the other members of the swarm. On the other hand, energy considerations may make large broadcast ranges unsuitable, and the short range communications between robots in the Blackboard strategy may be more efficient. Further research on this will be needed to investigate the relative impact of each strategy.

VIII. Results – Simulator Comparisons

The following subsections present the results of the initial research, followed by the extended work which followed.

A. Initial Research

Table VII compares the time taken to complete the task for each of the simulators in the Equal scenario. The time-stepped simulation presented here is significantly faster than the original multi-threaded simulation. This can largely be accounted for by the deliberate delays introduced previously to allow for visualization, with some impact of the reliance on real-time delays for communication, which makes a true comparison difficult. The table also shows the execution time for the One Responder protocol as implemented in the updated multi-threaded simulator, showing it is still slower than the time-stepped approach.

Figure 12 shows that the time-stepped simulation takes a much larger number of steps in the Multiple and Selective Responder strategies, and also shows an increase under One Responder. This unexpected result may be explained by the specific behaviour of the robots in each simulation. In the multi-threaded approach, robots pause frequently, the effect of which is that fewer robots will move in each step. For example, on deciding to respond to a help request a robot pauses for three seconds. Further, if another help request is received during that pause, that too may be processed and the robot may choose to act on that, with a further pause.

The effect of these pauses is to reduce the number of robots moving at any given time. In the time-stepped simulation, a robot will only pause when waiting for help responses or assignments in the One Responder strategy.

It is notable that despite these pauses, robots in the multi-threaded approach take fewer steps overall, rather than taking the same number of steps over a longer period. This suggests there may be a benefit to reducing the number of robots engaged in random exploration, but this will need to be investigated further.

The impact of the two simulations on the host platform was compared and Table VIII displays the approximate processor and memory usage of the two platforms when running simulations.

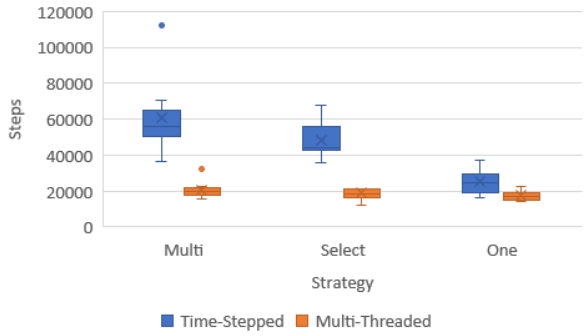


Figure 12. Total steps for each simulator using the three strategies in the Equal scenario. Circles represent data outliers.

TABLE VII. Simulation Duration (Equal Scenario)

Simulator	Time (s)		
	Multiple Responders	***Selective Responders***	***One Responder***
Time-Stepped	2.11	0.89	0.17
Multi-Threaded v1	155.14	142.01	130.86
Multi-Threaded v2			1.35

TABLE VIII. Simulator CPU and Memory Usage

Simulator	CPU Usage (%)	Memory Usage (MB)
Time-Stepped (Display)	5-6	30-35
Time-Stepped (No Display)	55-60	30-35
Multi-Threaded	35-40	700-750

Overall, the time-stepped approach will put less strain on the CPU, as despite its higher usage during execution without a display, it will run for a fraction of the time. With a display, the execution is halted between ticks to update the display at a framerate of the user's choosing, and so CPU usage drops. The multi-threaded simulation has no option to disable display updating, but the use of a separate thread for each robot results in a moderate level of CPU usage for a longer period of time.

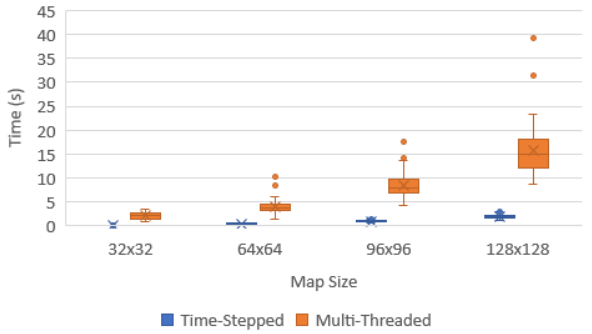
The lower memory footprint of the time-stepped simulation is most likely due to specific implementation differences. Each robot in the multi-threaded simulation contains a copy of the world map and lists of robots and items, whereas the time-stepped simulation uses a shared resource. While requiring local copies is a factor in any real scenario, it is not required to simulate that unless it is expected that robots will have different local data. If this is a requirement, the memory usage would increase accordingly.

B. Extended Research

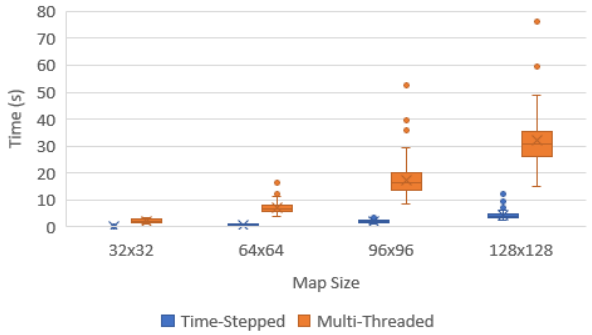
The implementation differences between the time-stepped and multi-threaded simulator used in the initial research have made a performance comparison difficult to achieve, prompting the rewrite of the multi-threaded simulator to use identical robot behaviour. The following results are based on comparisons run between the time-stepped simulator and the re-written multi-threaded simulator.

Figure 13 shows the mean time taken for each simulator to complete the task on each map size, for the Help Recruitment and No Cooperation strategies respectively. It is clear from the results that the multi-threaded simulator is slower at completing each simulation, and it can also be seen that the duration increases more quickly with an increase in map size for the multi-threaded implementation.

Figures 14 and 15 show the simulation ticks and total steps taken within the simulator for each implementation and map size. As discussed in Section V, the multi-threaded simulator uses the mean number of updates executed by the robots as an estimate of the ticks taken by the simulation to complete. The results here show that in both simulators, the

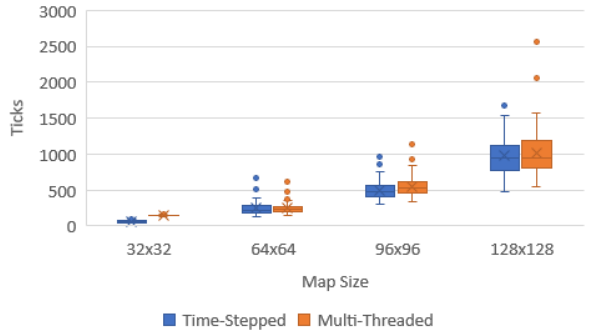


(a) Help Recruitment

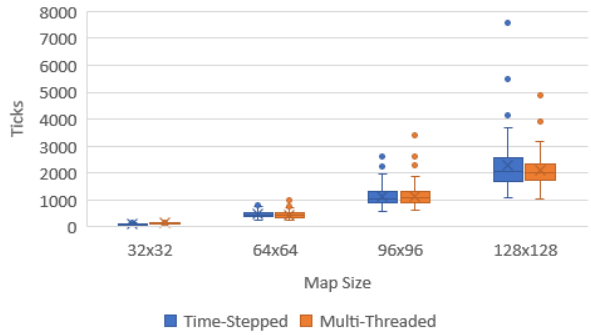


(b) No Cooperation

Figure 13. Average simulation execution time across all four maps. Circles represent data outliers: (a) Help Recruitment strategy, (b) No cooperation strategy.



(a) Help Recruitment



(b) No Cooperation

Figure 14. Average simulation ticks (time-stepped) or robot updates (multi-threaded) across all four maps. Circles represent data outliers: (a) Help Recruitment strategy, (b) No cooperation strategy.

use of the Help Recruitment strategy is able to reduce the number of ticks, and by extension the number of steps taken by the swarm.

It can also be seen that the results are broadly similar between the two simulations, except for the 32x32 map, where the multi-threaded implementation takes more ticks on average, and more steps, during the Help Recruitment strategy. A further investigation into the behaviour of the multi-threaded simulator shows that the robots each receive a different number of updates, as might be expected, but also that some robots do not receive any updates at all throughout the course of the simulation.

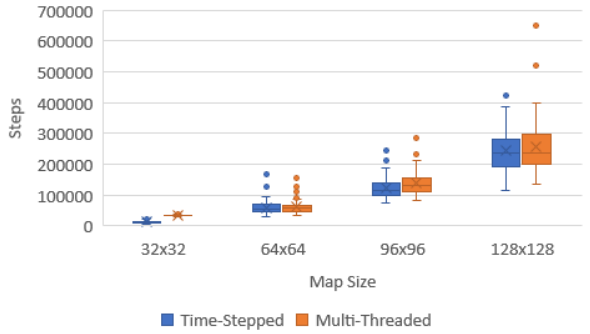
Figure 16 shows the average number of help offers sent in response to each help request when using the Help Recruitment strategy on each map size. It can be seen that the multi-threaded approach results in fewer offers generated per request. This may be due to the fact that robots are not always given processing time, however as the messages are buffered, a response will be sent provided the robot is updated at least once since receiving the request.

The average number of help requests sent on each map size is seen in Figure 17, where it can be seen that the multi-threaded simulator sends more requests than in the time-stepped approach. Again, this may be because of the processing time imbalance. If fewer robots are able to respond to a request, there remain more items for the robots to encounter, and thus more occasions where help requests will be sent.

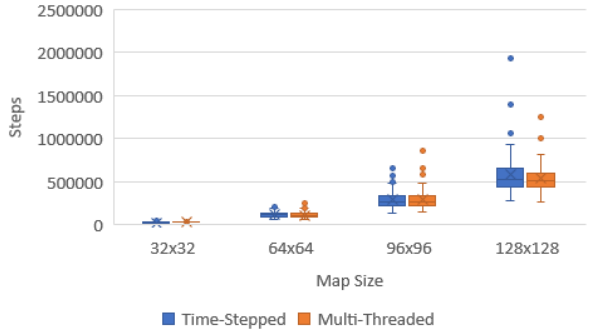
Investigations into the memory and CPU usage of the two simulators showed no appreciable difference, suggesting that the differences shown for the previous multi-threaded approach were down to implementation differences rather than inherent in the time-stepped or multi-threaded approach chosen.

Overall, these results show that despite the underlying implementation of the robot behaviour being the same in each simulator, the scheduling approach taken by each can result in different behaviour. As the simulation duration increases, these effects have less of an outcome on the final result.

Delegating the execution of the robots to the CPU scheduler can result in unpredictable behaviour by the robots, and results in a reduced degree of control in the simulation. A time-stepped approach ensures that each robot receives an identical amount of execution time in a predictable way. Furthermore, the time-stepped approach is significantly faster than the multi-threaded simulator, allowing more simulations to be run in a shorter period of time.



(a) Help Recruitment



(b) No Cooperation

Figure 15. Average steps taken by all robots, across all four maps. Circles represent data outliers: (a) Help Recruitment strategy, (b) No cooperation strategy.

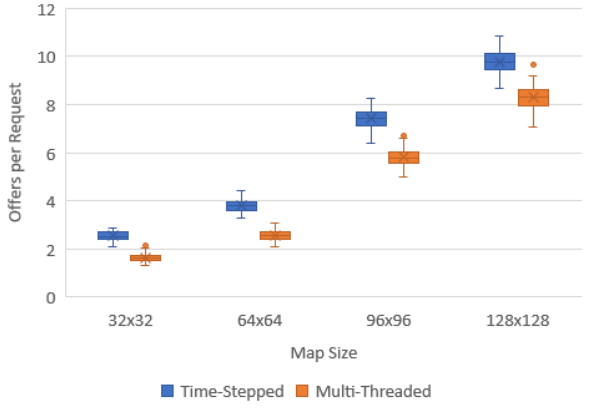


Figure 16. Help offers per request sent, across all four map sizes. Circles represent data outliers.

IX. Conclusion and Future Work

The presented research used a time-stepped simulation to investigate the effects of different cooperation strategies for a swarm carrying out a foraging task. It was shown that the performance of a strategy is sensitive to the scenario in which the swarm operates, and so sticking with a single strategy may lead to suboptimal performance. However, while the initial findings leaned towards some potential for changing the underlying strategy mid-mission, the extended research presented here suggests similar performance gains

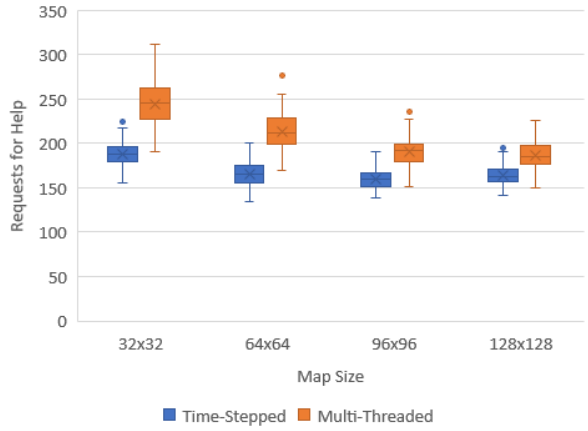


Figure 17. Help requests sent, across all four map sizes. Circles represent data outliers.

can be obtained simply by adjusting the parameters of an existing strategy.

Further, different stages of the task appear to favour different approaches. During the initial phase where large numbers of items remain to be discovered, random exploration with no cooperation strategy produces the best results. Only when a small proportion of the items remain does the adoption of a cooperation strategy start to benefit the performance of a swarm.

There is therefore a potential benefit to be gained by implementing a self-adaptive system that can modify these parameters based on the swarm's understanding of the current scenario, allowing for faster completion of the task. If the factors that affect selection of suitable parameters are simple, such as using swarm density alone to determine a broadcast range, then this may be achieved using a simple calculation. However, if the factors are sufficiently numerous and complex, there may still be a benefit to using simulation within the swarm to explore the possibility space without the risks associated with executing inefficient behaviours in reality.

These results are based on a simplified simulation of the foraging task. As noted in Section II, simulation can be a useful means of testing complex scenarios in situations where making use of large numbers of robots may be considered infeasible. Nonetheless, corroboration of the data with results from other simulations, and possibly physical robotic tests, would be useful.

The time-stepped simulation was compared against a real-time, multi-threaded approach and found to execute faster and with more reliable results, thanks to each robot being guaranteed an equal amount of processing time throughout the task. This would make the time-stepped simulation more suitable for use as part of the MAPE-K loop for a foraging swarm, forming MAPSE-K (Figure 18) [2]. This could be achieved by embedding the simulator on one or multiple robots within the swarm, to analyse and adjust the strategy without risking reduced performance.

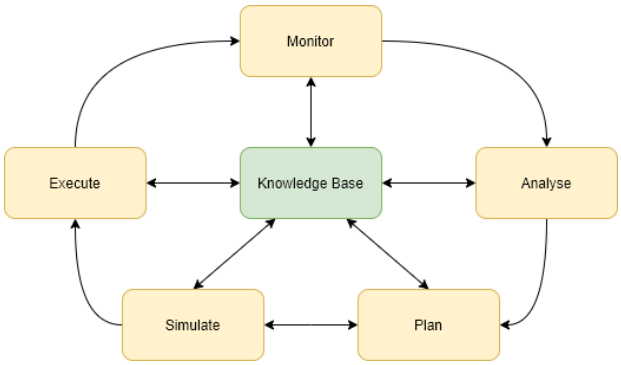


Figure 18. MAPSE-K loop. Simulation is used to test the plans before executing the best performing option, or returning to the planning stage.

The expected limited processing capabilities of the host robot mean managing the overhead that simulation entails will become a major factor. A time-stepped approach allows execution to be stopped and restarted with predictable results, as well as executing quicker overall, so can be less demanding on the host platform.

Future work will explore the methods of giving the swarm an autonomic ability to adjust the Help Recruitment broadcast range in response to the scenario faced. Other factors that can affect the performance will also be investigated, such as the impact of energy costs, as well as changes to the scenarios that may prompt the swarm to adjust their strategy accordingly.

REFERENCES

[1] L. McGuigan, C. Saunders, R. Sterritt, and G. Wilkie, 'Cooperation Strategies in a Time-Stepped Simulation of Foraging Robots', in *The Twelfth International Conference on Adaptive and Self-Adaptive Systems and Applications (ADAPTIVE 2020) IARIA*, Oct. 2020, pp. 135-142.

[2] R. Sterritt et al., 'Inspiration for Space 2.0 from Autonomic-ANTS (Autonomous NanoTechnology Swarms) Concept missions', presented at the Reinventing Space Conference, 2019.

[3] A. Farahani, G. Cabri, and E. Nazemi, 'Self-* properties in collective adaptive systems', in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, Heidelberg, Germany, Sep. 2016, pp. 1309–1314.

[4] L. Bayindir, 'A review of swarm robotics tasks', *Neurocomputing*, vol. 172, pp. 292–321, Jan. 2016.

[5] I. Navarro and F. Matía, 'An Introduction to Swarm Robotics', *ISRN Robot.*, vol. 2013, pp. 1–10, 2013.

[6] G. Beni, 'From Swarm Intelligence to Swarm Robotics', in *Swarm Robotics*, Berlin, Heidelberg, 2005, pp. 1–9.

[7] IBM, 'An architectural blueprint for autonomic computing, 4th ed.' IBM White Paper, 2006.

[8] J. O. Kephart and D. M. Chess, 'The vision of autonomic computing', *Computer*, vol. 36, no. 1, pp. 41–50, Jan. 2003.

[9] R. Sterritt, G. Wilkie, G. Brady, C. Saunders, and M. Doran, 'Autonomic robotics for future space missions', 13th Symposium on Advanced Space Technologies in Robotics and Automation (ASTRA 2015) - ESA/ESTEC, Noordwijk, Netherlands, Sept. 2015.

[10] M. Doran, R. Sterritt, and G. Wilkie, 'Autonomic architecture for fault handling in mobile robots', *Innov. Syst. Softw. Eng.*, Apr. 2020, doi:10.1007/s11334-020-00361-8

[11] M. Torres-Torriti, T. Arredondo, and P. Castillo-Pizarro, 'Survey and comparative study of free simulation software for mobile robots', *Robotica*, vol. 34, no. 4, pp. 791–822, Apr. 2016.

[12] J. Prasetyo, G. De Masi, and E. Ferrante, 'Collective decision making in dynamic environments', *Swarm Intell.*, vol. 13, no. 3, pp. 217–243, Dec. 2019.

[13] J. Zelenka, T. Kasanický, and I. Budinská, 'A Self-adapting Method for 3D Environment Exploration Inspired by Swarm Behaviour', in *Advances in Service and Industrial Robotics*, Cham, 2018, pp. 493–502.

[14] N. Capodieci, E. Hart, and G. Cabri, 'An Artificial Immunology Inspired Approach to Achieving Self-Expression in Collective Adaptive Systems', *ACM Trans. Auton. Adapt. Syst. TAAS*, vol. 11, no. 2, p. 6:1-6:25, Jun. 2016.

[15] N. Bredeche, E. Haasdijk, and A. Prieto, 'Embodied evolution in collective robotics: A review', *Front. Robot. AI*, vol. 5, p. 12, 2018.

[16] K. S. Kappel, T. M. Cabreira, J. L. Marins, L. B. de Brisolara, and P. R. Ferreira, 'Strategies for Patrolling Missions with Multiple UAVs', *J. Intell. Robot. Syst.*, vol. 99, pp. 499–515, Sep. 2019.

[17] G. Leu and J. Tang, 'Survivable Networks via UAV Swarms Guided by Decentralized Real-Time Evolutionary Computation', in *2019 IEEE Congress on Evolutionary Computation (CEC)*, Jun. 2019, pp. 1945–1952.

[18] M. Frasheri, B. Cürüklü, M. Esktröm, and A. V. Papadopoulos, 'Adaptive Autonomy in a Search and Rescue Scenario', in *2018 IEEE 12th International Conference on Self-Adaptive and Self-Organizing Systems (SASO)*, Sep. 2018, pp. 150–155.

[19] F. Yan, K. Di, J. Jiang, Y. Jiang, and H. Fan, 'Efficient decision-making for multiagent target searching and occupancy in an unknown environment', *Robot. Auton. Syst.*, vol. 114, pp. 41–56, Apr. 2019.

[20] C. Saunders, R. Sterritt, and G. Wilkie, 'Collective Communication Strategies for Space Exploration', *J. Br. Interplanet. Soc.*, vol. 72, no. 12, pp. 416–430, 2019.

[21] M. Puviani, G. Cabri, and L. Leonardi, 'Enabling Self-Expression: The Use of Roles to Dynamically Change Adaptation Patterns', in *2014 IEEE Eighth International Conference on Self-Adaptive and Self-Organizing Systems Workshops*, Imperial College, London, United Kingdom, Sep. 2014, pp. 14–19.

[22] M. Dorigo, G. Theraulaz, and V. Trianni, 'Reflections on the future of swarm robotics', *Sci. Robot.*, vol. 5, no. 49, eabe4385, Dec. 2020.

[23] C. Pinciroli *et al.*, 'ARGoS: a modular, parallel, multi-engine simulator for multi-robot systems', *Swarm Intell.*, vol. 6, no. 4, pp. 271–295, Dec. 2012.

[24] N. Jakobi, P. Husbands, and I. Harvey, 'Noise and the reality gap: The use of simulation in evolutionary robotics', in *Advances in Artificial Life*, vol. 929, F. Morán, A.

Moreno, J. J. Merelo, and P. Chacón, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1995, pp. 704–720.

[25] F. Kamrani and R. Ayani, 'Using On-line Simulation for Adaptive Path Planning of UAVs', in *11th IEEE International Symposium on Distributed Simulation and Real-Time Applications (DS-RT'07)*, Oct. 2007, pp. 167–174.

[26] N. Keivan and G. Sibley, 'Realtime simulation-in-the-loop control for agile ground vehicles', *Lect. Notes Comput. Sci. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma.*, vol. 8069 LNAI, pp. 276–287, 2014.

[27] C. Pepper, S. Balakirsky, and C. Scrapper, 'Robot simulation physics validation', in *Proceedings of the 2007 Workshop on Performance Metrics for Intelligent Systems*, Washington, D.C., Aug. 2007, pp. 97–104.

[28] C. J. E. Castle, N. P. Waterson, E. Pellissier, and S. Le Bail, 'A Comparison of Grid-based and Continuous Space Pedestrian Modelling Software: Analysis of Two UK Train Stations', in *Pedestrian and Evacuation Dynamics*, Boston, MA, 2011, pp. 433–446.

Performance Evaluation on Indoor Positioning System Using SS Ultrasonic Waves for Drone Applications

Tatsuki Okada, Akimasa Suzuki
Graduate school of Software and
Information Science
Iwate Prefectural University
020-0693, Sugo Japan
email g231s004@s.iwate-pu.ac.jp
suzuki_a@iwate-pu.ac.jp

Souichirou Masuda
Faculty of Software and
Information Science
Iwate Prefectural University
020-0693, Sugo Japan
email g031q141@s.iwate-pu.ac.jp

***Abstract*—This study develops a drone positioning system for use in indoor environments, including dark places, inaccessible areas, and ordinary living environments that do not accommodate conventional methods. Various indoor drone applications have been developed, such as drone communication systems and wall surface inspection, which require remote estimation of drone position. For outdoor applications, a Global Navigation Satellite System (GNSS) is generally used to obtain the drone position. However, as the GNSS radio waves cannot reach indoors or between buildings, camera-based methods, such as Simultaneous Localization and Mapping (SLAM), are applied to estimate the drone's position. The system uses noise-resistant, code-division-multiplexed spread spectrum (SS) ultrasonic waves for three-dimensional positioning. Transmitter and receiver hardware is developed using SS ultrasonic waves and the effect of wind and sound of the positioning system during drone operations on the SS ultrasonic positioning is evaluated. Transmitter and receiver hardware is developed using SS ultrasonic waves and the effect of wind and the sound of the positioning system during drone operations on the SS ultrasonic positioning is evaluated. The accuracy of the positioning system is verified through experiments, and the results indicate that a positioning accuracy within 15 cm is possible despite the effects of downwash generated by the drone's wings, and there was no effect of multipath on the positioning error.**

***Index Terms*—Drone; Indoor Positioning System; SS Ultrasonic Waves; Downwash; IoT Devices.**

I. Introduction

This paper is an extension of the paper initially presented at the VEHICULAR 2020 Ninth International Conference on Advances in Vehicular Systems Technologies and Applications [1]. In this paper, we additionally describe a detailed system configuration to store positioning data in cloud storage using Internet of Things (IoT) technology. We studied the robustness of the system to drone noise when using Gold code with the conventional M-sequence. We also compared the experimental results in an anechoic chamber with those in a reverberant experimental environment to study the effect of the reverberation in the tunnel.

Because they can take off and land vertically in small spaces, drones can be used to perform various activities in unstable places where people and vehicles cannot access. Previous studies have investigated the use of drones for many uses, including autonomous search and rescue operations for victims following a disaster [2], meteorological observations [3], and logistics such as home delivery [4].

When used indoors, drones act as communication robots [5]. However, an appropriate distance is required to allow natural and smooth communication between a human and an autonomous mobile robot. To ensure the appropriate positioning in indoor spaces, the drone's coordinates can be used to develop real-time centimeter-order positioning. A relevant study investigated the use of drones for periodic inspection to detect aging degradation of locations where staff are unable to work, such as high walls of tanks and industrial chimneys [6]. Using drones for this purpose is expected to reduce the high cost of these inspections.

Use of drones indoors is more dangerous than outdoors because the drone can easily crash into obstacles such as humans and walls. Thus, it is essential to determine the position of the drone in relation to other objects. As horizontal and vertical relationships are important in these applications, it is essential to obtain absolute coordinates in space. While a Global Navigation Satellite System (GNSS) is generally used to obtain the absolute coordinates of a drone, the GNSS signal is difficult to detect indoors. Simultaneous Localization and Mapping (SLAM) is often used in non-GNSS environments. However, the flight path of a routine inspection is often in a dark place and the walls do not always follow a uniform pattern, causing large errors in SLAM's self-position estimation.

We therefore propose an indoor positioning system for drones using spread spectrum (SS) ultrasonic waves [7]. This system is expected to obtain three-dimensional (3D) coordinates with an accuracy of 10cm. However, the accuracy may be negatively impacted by noise from the propellers or downwash of a drone. Downwash is the wind created by

the drone's propellers. Therefore, in this study, we conduct an experiment to evaluate the positioning accuracy of drone flights during a periodic inspection.

Section II presents related research. Section III provides an overview of indoor positioning systems using SS ultrasonic waves. Section IV examines the drone's robustness against noise for positioning in a no-multi-path situation through an experiment conducted in an anechoic chamber with no reverberation. Section V evaluates the positioning error when the experiment is conducted in an echoing environment with the same system configuration as in Section IV. Section VI compares the results of Sections IV and V. Section VII provides a summary and future perspectives.

II. Related and previous works

There does not almost exist positioning method with drones for indoor multi environments, including dark environments, with accuracy under $10\mathrm{cm}$. Various sensor systems have been investigated for indoor positioning purposes, including pseudolites [8] and BLE beacons [9]. Of these, ultrasonic-wave-based systems using the time of flight (TOF) between the time of transmission and the time of reception systems have the lowest cost and greatest accuracy because of their slow propagation speed and ease of improving resolution. However, because these systems use the time-division multiplexing method with on-off keying, which grows increasingly cumbersome as the number of objects to be measured increases, they generally have weak noise resistance and are slow to acquire data. Systems using SS ultrasonic signals have therefore been investigated to overcome these drawbacks [10] [11].

Analogous to SS radiowave systems (e.g., global positioning system (GPS)), we have proposed a real-time 3D positioning system using SS ultrasonic signals with a band-limited transducer, a low-power field programmable gate array (FPGA), and a small microprocessor [12]. In previous studies, we discussed factors such as positioning errors in indoor environments [7] and showed the measurement accuracy of the positioning system using SS ultrasonic signals. We also proposed a calculation algorithm based on the Newton-Raphson method for continuous signals, rather than conventional pulse signals. As a result, 3D coordinates can be obtained every $80\mathrm{ms}$ using Code Division Multiple Access (CDMA) with continuous signals [13].

We evaluated the positioning accuracy of SS ultrasonic waves using a ground-based mobile robot [14]. Other studies have proposed using not only SS ultrasonic waves but also image sensors for drone positioning [15] and applying drones to limited situations such as greenhouses [16]. Indoor positioning accuracy has been discussed using the Kinect camera, with an average positioning error of $48\mathrm{mm}$ [17]. However, the Kinect is difficult to use in the dark. This study develops an indoor positioning system using only SS ultrasonic waves that can be used in dark places where image sensors are ineffective.

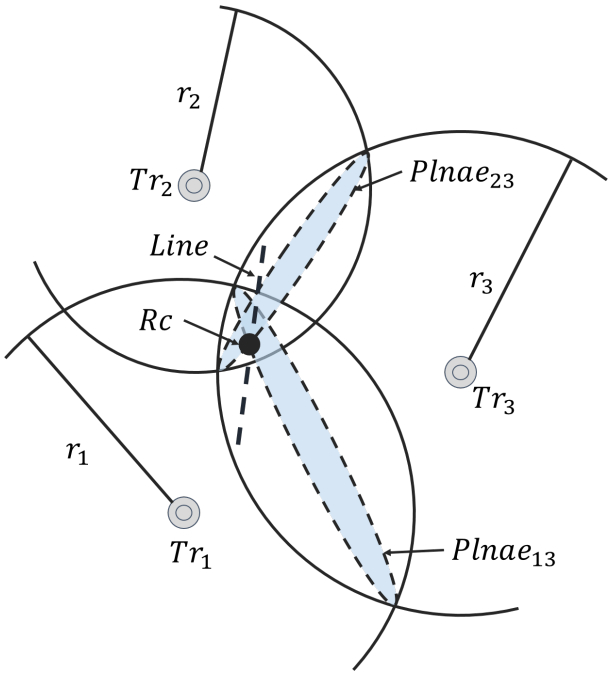


Figure 1. Positioning calculations for the indoor positioning system.

III. Indoor drone positioning system using SS ultrasonic signals

This section describes the indoor positioning method using SS ultrasonic waves and our proposed system.

A. A method for positional calculation

Figure 1 presents the positioning calculations for the indoor positioning system using SS ultrasonic waves. Spheres are drawn to determine the center point on the radius between a receiver Rc and each transmitter. Two pairs of spheres are selected centering on Tr_1 and Tr_3 and Tr_2 and Tr_3, respectively. From these pairs of spheres, Plane_{13} and Plane_{23} are solved simultaneous equations and a line of intersection is obtained from the two planes. Finally, the points at the intersection of the line with an equation of an arbitrary sphere are solved. Figure 2 shows a flowchart of the algorithm for the positioning calculation in Figure 1. Two intersection points are obtained as two transmitters are installed along the bottom or sides of the positioning range. Therefore, one solution is outside the room and the other solution becomes the position of the receiver Rc. When using four transmitters, four position results are obtained. Thus, the measurement position is defined as an average of these results.

B. Hardware structure of a positioning system using SS ultrasonic waves

A 3D position can be calculated on the basis of three or more TOF between the transmitters and the receiver. Figure 3 shows the system architecture of the TOF measurement for the positioning system. The transmitting hardware contains a digital to analog (D/A) converter and an FPGA to generate carrier

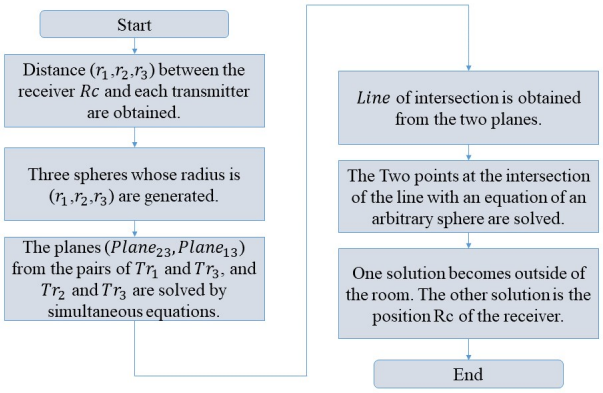


Figure 2. Flowchart of position calculation.

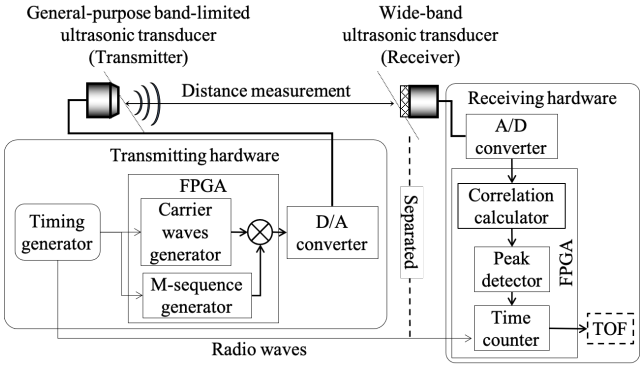


Figure 3. System architecture of the TOF measurement.

waves and M-sequences. The receiving hardware includes an A/D converter and an FPGA for correlation calculation, peak detection, and time measurement.

An SS signal is generated by the transmitting hardware to multiple carrier waves by M-sequences and is output from a transducer after D/A conversion. At the start of the transmission, a time counter is started to measure the TOF and correlation values are calculated from the sound data via the A/D converter as online and real-time hardware processing. The time counter measures the TOF by counting the sampling times until arriving at peak correlation values obtained by the peak detector. Then, the 3D position of the receiver can be calculated based on three or more TOFs between the transmitters and receiver. The correlation calculator component is installed in the hardware as shown in Figure 3. Distance is calculated from the TOF obtained from the hardware and the dimensional position is measured. Real-time positioning is sufficiently available because this processing can be calculated with low cost using optimized expressions.

C. SS signal

In our indoor positioning system, SS signals are modulated by binary phase shift keying using an M-sequence with a direct sequence method. Figure 4 shows the M-sequence generator. ‘0’ or ‘1’ sequence is generated by the shift register.

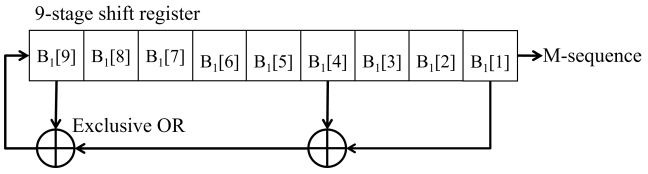


Figure 4. M-sequence generator with tap 4,9.

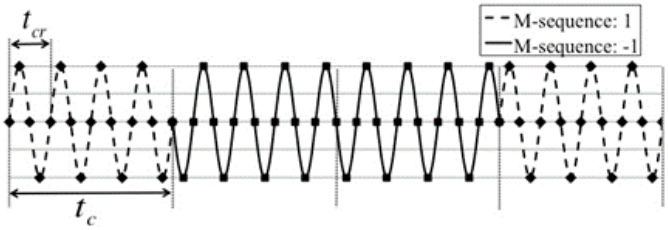


Figure 5. Spread spectrum (SS) ultrasonic signal.

Each number in [] in B_1 of Figure 4 means the number of bits. Tap is defined as a position where exclusive OR calculation is performed. In the case of Figure 4, tap position is shown as {4,9}, which named by the shift register where connected to exclusive OR gates. By changing the position of the tap, several M-sequences with low cross-correlation can be generated.

Figure 5 shows a received SS signal, where the signals corresponding to ‘1’ and ‘-1’ are plotted solid and dashed lines, respectively. We replace values of ‘-1’ with ‘0’ from generated M-sequence for easy signal processing. Each dot in Figure 5 is a sample to convert to digital signals. The amount of sample including one period of carrier waves is decided on the basis of four samples. Here, chip length t_c is defined as the time required to describe a 1-chip of the M-Sequence. The chip length can also be described as $t_c = 4/f$ using carrier frequency f. The length of SS ultrasonic signals becomes $2^9 - 1 = 511$ [chip] owing to a 9-stage shift register for the M-sequence in our system. These four channels of the transmitters are generated by the following tap positions: {4,9}; {3,4,6,9}; {4,5,8,9}; and {1,4,8,9}.

The M-sequence has few combinatorial channels and limited cross-correlation. Therefore, to increase the number of channels, Gold codes obtained by multiplied M-sequences are considered. As shown in Figure 6, the Gold code is generated by using two shift registers, B_1 and B_2 to generate the M-sequence and combining the two registers by an exclusive OR operation. The following tap positions are used to generate the Gold code: combination of{4,9} and {4,5,8,9}, combination of {3,4,6,9} and {4,5,8,9}, combination of {4,5,8,9} and {1,4,8,9} and combination of {1,4,8,9} and {3,4,8,9}. In this system, the frequency of the carrier waves is 40.2kHz.

D. Our proposed indoor positioning system using SS ultrasonic for drones

In this system, we use a transmitter with a closed-type aperture (PC40-18S, Nippon Ceramic Co., Ltd.) and a “Mini”

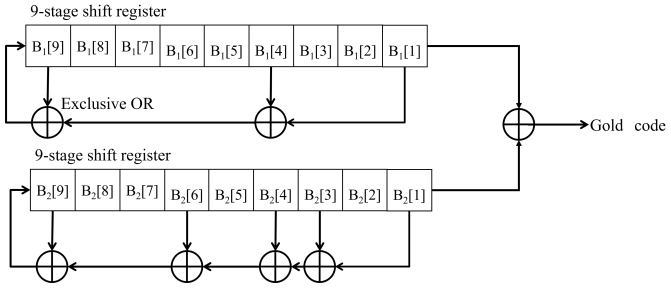


Figure 6. Gold code generated by the M-sequence with tap 9,4 and tap 9,6,4,3.

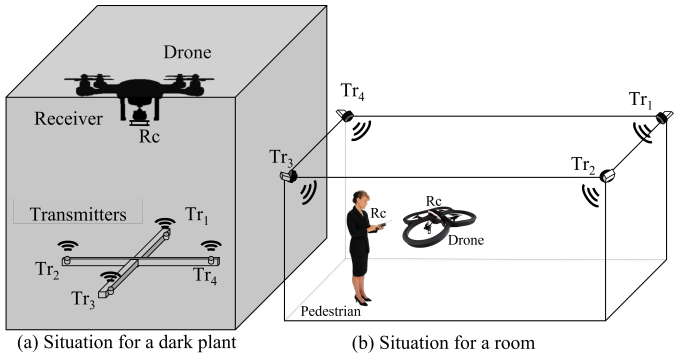


Figure 7. Measurement layout for the proposed system for (a) a dark plant and (b) a room.

Si Sonic™ultrasonic receiver (SPM0404UD5, Knowles) as general-purpose ultrasonic transducers.

Figure 7 shows the layout of the transmitters and receiver for our proposed system. Two example situations, a dark plant and a room as shown in Figures 7(a) and 7(b), respectively, are used. Figure 7(a) represents a periodical inspection at a plant where it is difficult to install infrastructure, such as transmitters, in the building. For convenient mounting, transmitters are therefore set on a cross-shaped mount, as shown in Figure 7(a). Considering the dilution of precision (DOP) [18] [19], especially horizontal DOP, the larger the mount size, the more accurate the expected positioning accuracy. However, a larger size limits the installation position options and is inconvenient to carry. Figure 7(b) represents a communication drone. Transmitters are mounted in four corners of a room. In this situation, the transmitters are more difficult to install, but the DOP is better than in Figure 7(a). In this study, we conduct experiments using the layout shown in Figure 7(a),because this configuration in Figure 7(a) has not been considered in previous studies.

The sequence to send positioning data to the cloud system is shown in Figure 8. A microphone, receiving hardware, and a Raspberry Pi are mounted on the drone. The positioning data will be sent to Azure CosmosDB using the Microsoft Azure IoT Hub. The data are sent from the receiving hardware to the Azure IoT hub via the Raspberry Pi. In the Azure, the received data are stored in Azure Cosmos DB through Azure Stream Analytics jobs that process and analyze the events in

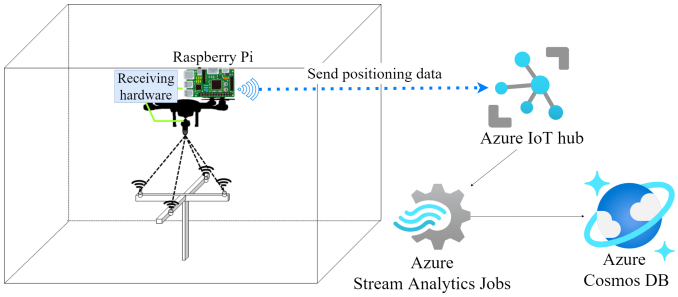


Figure 8. Overview of the sequence to send positioning data from Raspberry Pi to Azure Cosmos DB.

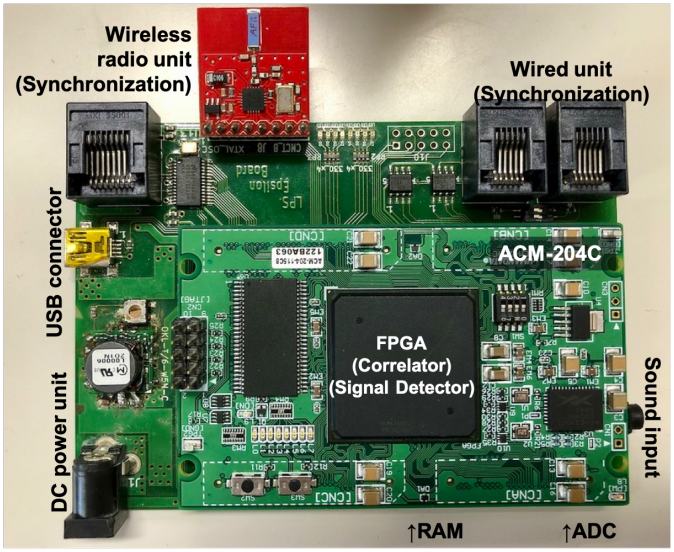


Figure 9. Receiving hardware for drone measurements.

real time. The positioning data can be remotely checked by accessing Azure Cosmos DB online.

Figure 9 illustrates the receiving hardware schematic mounted on the drone, as shown in Figure 8. The hardware consists of an Evaluation Board (ACM204-1158C) installed in the FPGA (Intel Cyclone IV); a transceiver for timing the synchronization of the ultrasonic transmitter unit; a receiver unit; a universal serial bus (USB) interface for output coordinates; an input part to receive the ultrasonic waves, including an A/D converter and amplifier; and synchronous dynamic random-access memory (SDRAM) for the real-time correlation calculations.

The USB interface and the input component are connected to a computer and a microphone, respectively. Ultrasonic waves received by the microphone are converted to A/D and input into the FPGA, where the correlation calculations, peak detection, and TOF calculations are performed. The SDRAM processes the real-time correlation calculations, and the transceiver measures the TOF based on the transmission timing received from the ultrasonic transmitter.

In the Raspberry Pi, the positioning data are calculated by TOF, as obtained by the receiving hardware. Figure 10 shows

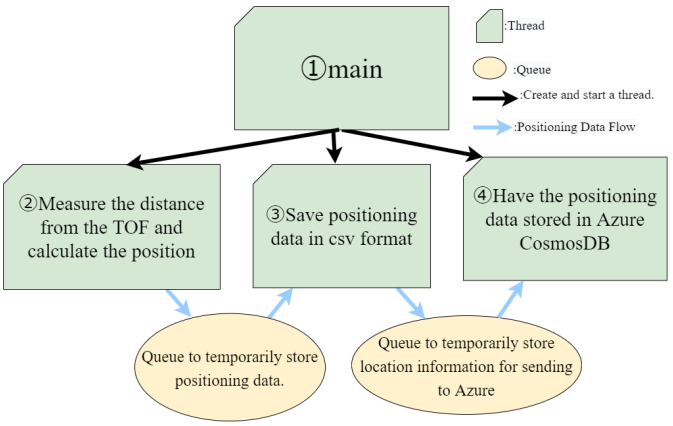


Figure 10. Schematic diagram of the program installed in Raspberry Pi.

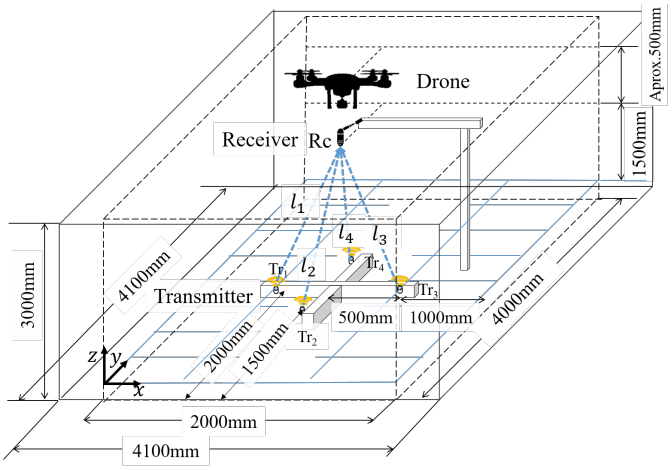


Figure 11. Schematic of the experimental setting in the anechoic chamber.

an overview of the program running on the Raspberry Pi. This program generates four threads for parallel processing. After the initial setup, we create three threads to run methods, calculates positioning, performs backup, and sending to Azure CosmosDB in parallel. The main thread ① is configured to connect to Azure and USB, and three threads, ②, ③, and ④, are created. In thread ②, the distance d between the receiver and the respective transmitter is measured by the TOF sent from the receiver hardware. The distance is obtained by

$$d = v \times \gamma \tag{1}$$

Here, γ and the v are, respectively, the TOF and the speed of sound calculated by the following approximation formula:

$$v = 331.5 + 0.60714 \times T \tag{2}$$

The temperature T°C is measured by SENSIRION's SHT31 module [20]. A receiver's position is calculated by four distances from the four transmitters using the method described in Section III-A. At thread ③, the positioning data are saved in CSV format. At thread ④, positioning data are sent to the Azure IoT hub that manages and monitors IoT devices and their communications.

A constant error could occur in the positioning system. More accurate TOF measurement is expected by calibration, which removes this type of error. Thus, in this system, constant errors are removed from measurement results.

IV. Positioning Experiment Without Sound Reflection.

Before the measurement in a real environment, positioning experiments were conducted in anechoic chamber without sound reflection to study the effect of noise and wind generated by the drone on the distance measurement of SS ultrasonic waves. The experimental environment is shown in Figure 11. The solid line in Figure 11 indicates the environment of the anechoic chamber used, which is $4000\mathrm{mm}$ long, $4100\mathrm{mm}$ wide, and $3000\mathrm{mm}$ high. The area in the dotted line in Figure 11, which is $4000\mathrm{mm}$ long and $2000\mathrm{mm}$ wide, is defined as the comparison area against normal sound reverberation area described in the next section. The lower left corner in this area is defined as the origin of the coordinates. Four transmitters Tr_1-Tr_4 were placed near the center of the room. The coordinates of the speakers were $\mathrm{Tr}_1[\mathrm{mm}] = (500, 2000, 0)$,$\mathrm{Tr}_2[\mathrm{mm}] = (1000, 1500, 0)$,$\mathrm{Tr}_3[\mathrm{mm}] = (1500, 2000, 0)$,$\mathrm{Tr}_4[\mathrm{mm}] = (500, 2000, 0)$,and $\mathrm{Tr}_4[\mathrm{mm}] = (1000, 2500, 0)$.

Figure 12 illustrates the anechoic chamber experiment environment. Figure 13 shows the measurement point of the receiver for this experiment. The measurement points to be obtained are Rc_a, which is $1500\mathrm{mm}$ directly above the center of the $x-y$ plane; Rc_b, which is $1500\mathrm{mm}$ directly above Tr_1; Rc_c, which is $1500\mathrm{mm}$ directly above Tr_2; and Rc_d, which is $1500\mathrm{mm}$ directly above Tr_4. The transmitter and receiver distances l_1-l_4 for each measurement point were obtained. The drone was a Mavic 2 zoom by DJI™. The drone hovered approximately $500\mathrm{mm}$ over the receiver. The position was obtained five times at each measurement point. The transmitting waves were $48\mathrm{V_{p-p}}$, and both the M-sequence and Gold codes were used in the experiment. For the positioning calculation, the average of the difference of the ultrasonic measured distances was subtracted from the true distance at each point as the constant error from the $l_1 - l_4$ measured distance at each point.

A. RMS positioning error of M-sequence

Figure 14 shows the Root Mean Square (RMS) of the difference between the results and the installed distances. The vertical and horizontal axes of Figure 14 show the RMS positioning error and the coordinates of the measurement points, respectively. em_{rms} is defined as

$$em_{\mathrm{rms}} = \sqrt{(dm_i - d_i)^2} \tag{3}$$

where d_i and dm_i are the measured distance and the true distance between a receiver and i-th transmitter, respectively. The line on the bar in Figure 14 has its upper and lower ends at the maximum and the minimum RMS positioning errors, respectively.

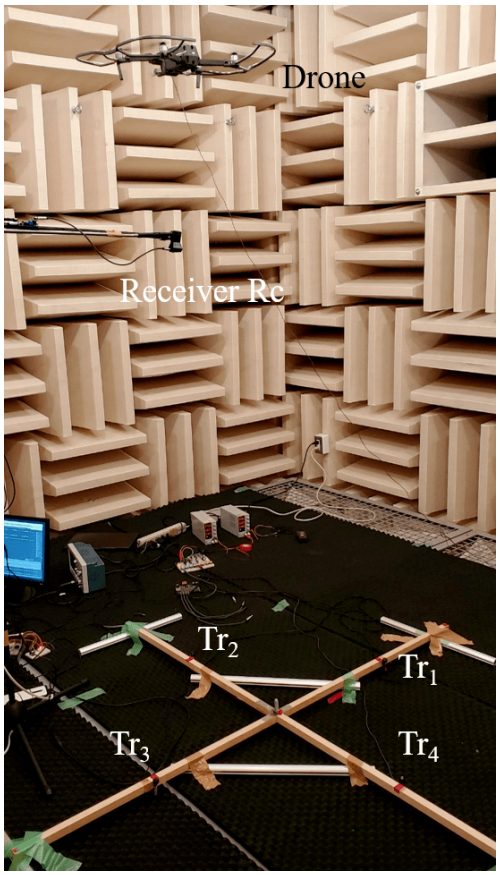


Figure 12. Experimental environment in the anechoic chamber.

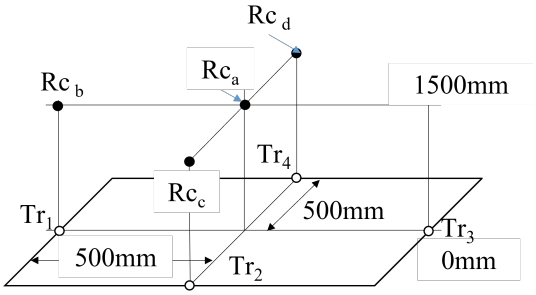


Figure 13. Measurement points of the receivers in Figure 12.

The results show that the range of minimum and maximum RMS positioning errors was wider at all measurement points with the drone because of the sound and wind effects from the drone; however, no other significant trend was identified. The RMS positioning error for all measurement points was within 150mm.

B. Horizontal and vertical errors by M-sequence code

The measurement error on each measurement position was considered for the $x-y$ horizontal error and the z direction error. The horizontal error was calculated as the RMS positioning error of x and y directions on each measurement point. The error in the vertical direction is the absolute value of the

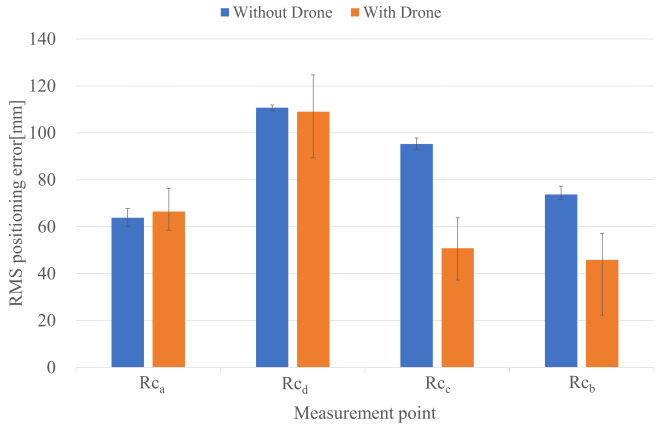


Figure 14. RMS positioning error by M-sequence in an anechoic chamber experiment.

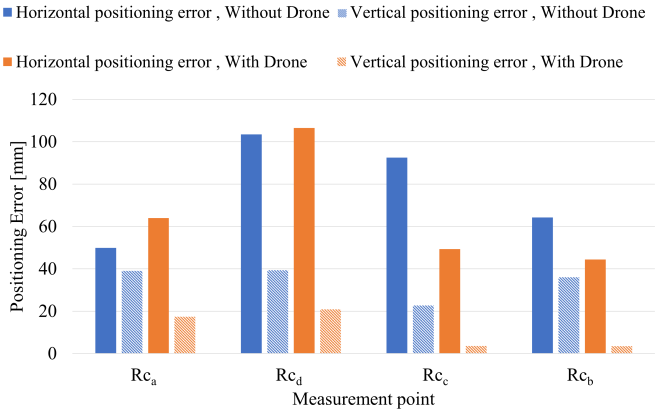


Figure 15. Horizontal and vertical positioning errors by M-sequence code in the anechoic chamber.

error in the z direction. Figure 15 shows the horizontal and vertical errors at each location with and without the drone. The vertical and horizontal axes show the positioning error and measurement point, respectively. Darker and lighter bars in the graph indicate vertical positioning and horizontal positioning errors, respectively.

Figure 15 shows that the horizontal error is larger than the vertical error at all locations. The error in the plane direction is greatly affected by the DOP because the transmitters are placed at a distance of 500mm from the center of the $x-y$ plane. Increasing the interval between the transmitters is expected to improve the accuracy.

C. Comparison of positioning error between Gold code and M-sequence

Figure 16 compares the RMS positioning error after subtracting the constant error of the position using M-sequence and Gold codes. The vertical axis of the graph shows the same RMS positioning error as in Figure 14. The horizontal axis is the measurement point. Similar to Figure 14, the line on the bar in Figure 16 has its upper and lower ends at the maximum

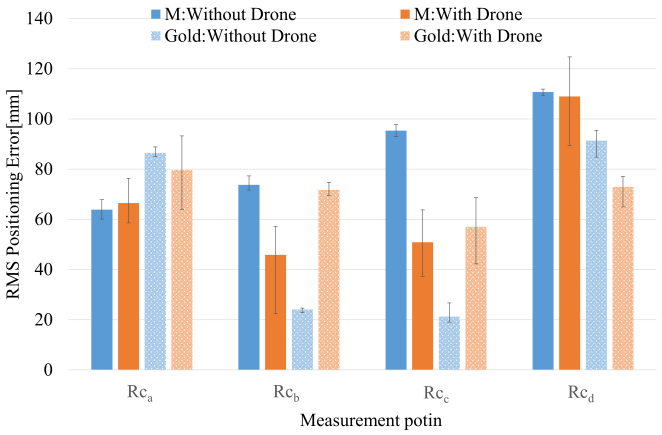


Figure 16. RMS positioning error of Gold code and M-sequence in an anechoic chamber.

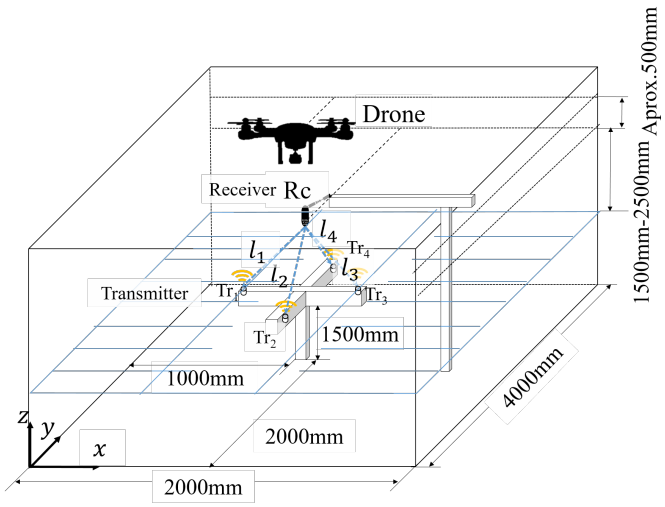


Figure 17. Layout for the positioning experiment in reverberating indoor environment.

and the minimum RMS positioning errors, respectively. The experiment demonstrated that positioning with Gold codes was possible at all points as well as in the M-sequence. Flying the drone increased the range of maximum and minimum RMS positioning errors in the Gold code. In terms of the RMS positioning error between the M-sequence and the Gold code, there was no dominant difference. The RMS positioning errors were 30mm without the drone and 2mm with the drone.

V. Positioning Error by Drone Noise and Downwash in an Indoor Environment with Normal Sound Reflection

In structures such as tunnels, sound reflects off walls, and multi-path effects may occur. To evaluate the effect of motor noise, experiments were conducted in an environment with sound reflection, wind noise, and downwash where reflected waves are generated and comparison is performed with the experiment in the anechoic chamber.

Figure 17 shows the environment used for this experiment, which was a room 2000mm long and 4000mm wide. A Mavic 2 zoom by DJI™drone as in the anechoic chamber experiment was used. The m-sequence was used as pseudo-random code in this experiment. Four transmitters Tr_1-Tr_4 were placed near the center of the room. As shown in Figure 17, the drone's starting point was the floor at the left front edge of the room. To prevent the ultrasonic waves from reflecting off the floor, the transmitter was placed at a height of 1500mm above the floor. The coordinates of the transmitters were $\text{Tr}_1[\text{mm}] = (500, 2000, 1500)$, $\text{Tr}_2[\text{mm}] = (1000, 1500, 1500)$, $\text{Tr}_3[\text{mm}] = (1500, 2000, 1500)$, and $\text{Tr}_4[\text{mm}] = (1000, 2500, 1500)$. The transmitting SS signal was amplified to $50\text{V}_{\text{p-p}}$.

Figure 18 shows the experimental environment. SS ultrasonic waves were transmitted upward from Tr_1-Tr_4, mounted on a tripod, and received by Rc, mounted on a bridge of wood. Figure 19 shows the measurement point. The white and black circles in Figure 19 denote the transmit and receive points, respectively. Rc_A, $\text{Rc}_{\text{A}'}$, and $\text{Rc}_{\text{A}''}$ are the measurement points at the center coordinates $(1000, 2000, 1500)$ of the $x-y$ plane of the four transmitters, located 1500mm, 2000mm, and 2500mm above the transmitters, respectively. Rc_C and $\text{Rc}_{\text{C}''}$ are above transmitter Tr_2 and $\text{Rc}_{\text{B}''}$ is above transmitter Tr_1.

The distances, l_1, l_2, l_3, and l_4, between the transmitters and a receiver (Figure 17) were measured for each measurement point. The drone hovered at a position approximately 500mm above the receiver. The accuracy was examined when the drone was and was not in flight in the environment. Five trials were conducted for each measurement point. As in the Section IV, all experimental results are shown after subtracting the average of the difference from the true value from ultrasonic measured distances.

A. Measurement error in distance in the reverberating Indoor environment

Figure 20 shows the average differences in the distances from the hovering drone for five trials. The vertical and horizontal axes on each graph denote the difference in distance from the drone compared to the measured distance from l_1 to l_4 to Tr_1 to Tr_4, respectively. The differences in distances are shown as absolute values, and the average difference in the distance is shown as a black line.

The results show that all measured distances could be obtained when the drone was flying, but the measurement distance was affected by the drone's flight. Figure 20(a) shows the drone's distance for the four transmitters, where the measurement point is on the center of the $x-y$ plane. A greater distance between the transmitter and receiver indicates larger measurement distance. Figure 20(b) compares the accuracy of the distance measurement at the center position $(1000, 2000)$ with that when the drone is above Tr_2 at heights of 3000mm and 4000mm. The difference in distance measurement above Tr_2 is the same as that shown in Figure 20(a). The difference in the distance between Tr_2 and $\text{Rc}_{\text{C}''}$ is increased by the drone hovering. Figure 20(c) shows the measurement distance

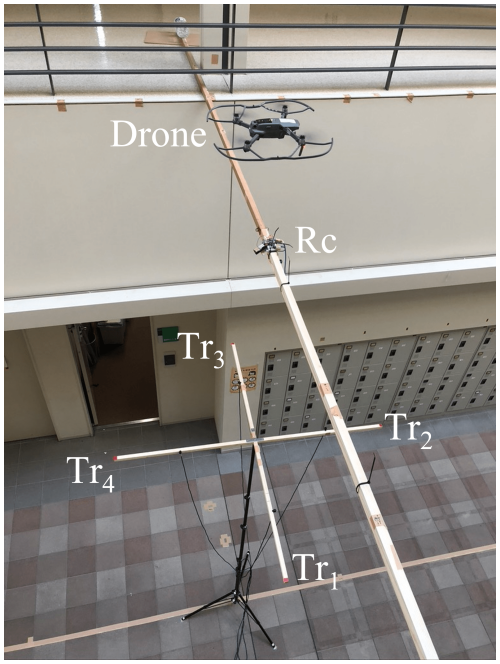


Figure 18. A view of the experiment environment shown from above.

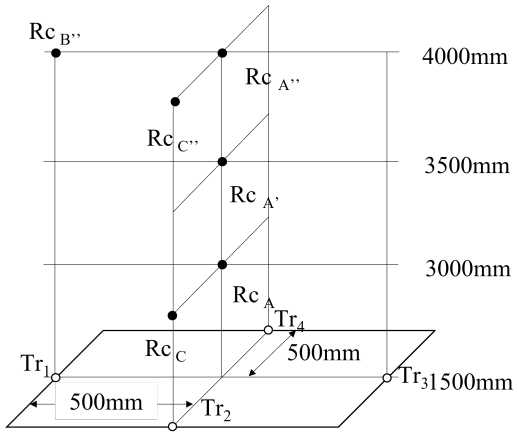


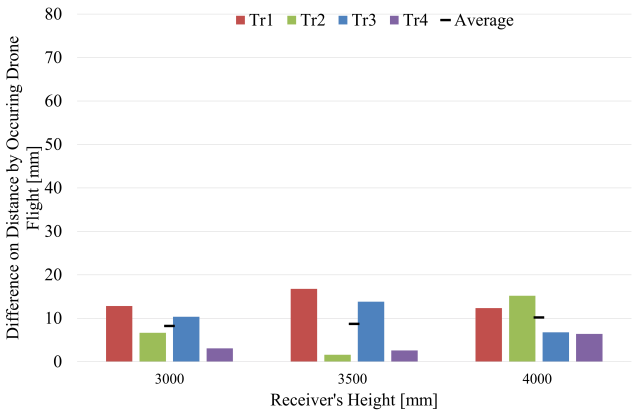
Figure 19. Measurement points of the receivers in Figure 18.

at the height of 4000mm, indicating that the difference in the distance between Tr_1 and $\mathrm{Rc}_{B''}$ increases. Compared to $\mathrm{Rc}_{A''}$, however, the average difference almost the same.

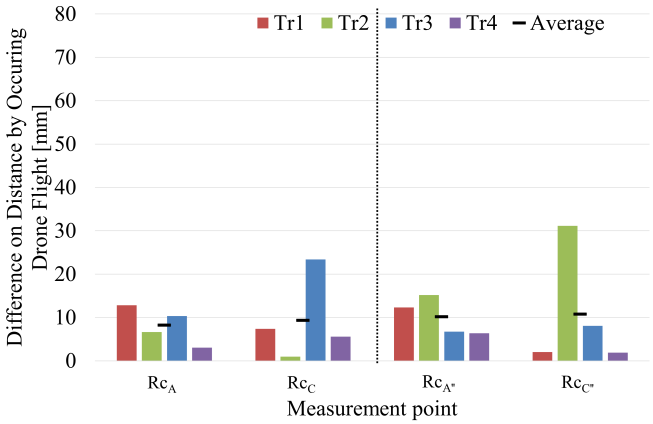
These graphs indicate that a drone's downwash and noise significantly affect the measurement distance when the transmitter and receiver are facing each other. The difference in the measured distance with and without drones is within 4cm.

B. Positioning error in the reverberating indoor environment

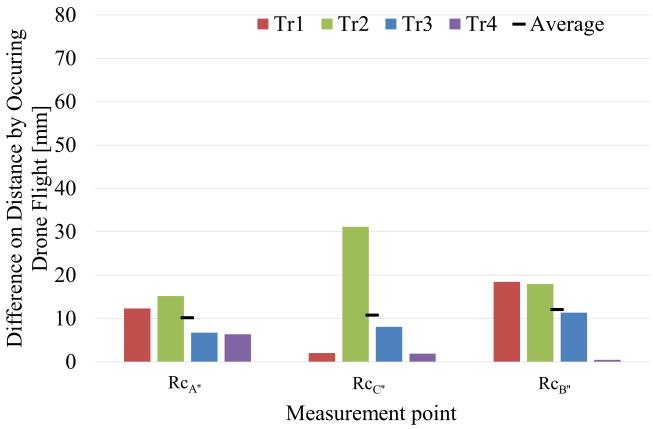
The experimental results were evaluated using the RMS of the difference between the results and the installed distances. Figure 21 shows the RMS positioning errors at the same receivers shown in Figure 20 and the maximum and minimum positioning errors as an expression of variance. The vertical and horizontal axes of Figure 21 denote the RMS positioning



(a) at each receiver's height on the center position



(b) at the center position vs above Tr_2



(c) at a height of 4000mm

Figure 20. Difference in the measured distance due to drone flight for Figure 18 environment.

error and the measurement point, respectively. The positioning errors are an average of five trials.

These results indicate that the positioning error increased when the drone was flying because of downwash and flight noise; however, the average errors were less than 15cm. The

results of Figures 21(a) and (b) confirm that the greater the distance between the transmitter and receiver, the larger the average RMS positioning error and variance when the drone was being flown. Figure 21(c) shows that the most variance was observed at the center of $\mathrm{Rc_{A''}}$.

These results indicate that the transmission is sufficiently accurate enough to measure a drone for a periodic inner wall inspection. More accurate positioning is expected to be achievable by compensating for errors caused by the angle of the transmitter and receiver and by the measurement distance [21].

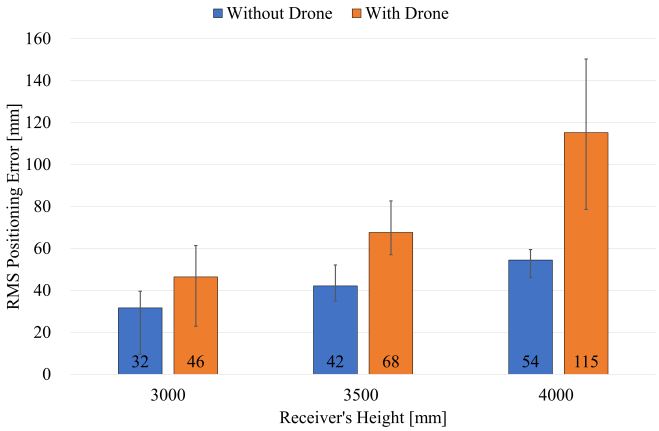
VI. Comparison of RMS positioning errors in an anechoic chamber and a typical acclimation environment

Figure 22 shows the average of all RMS positioning errors acquired in the anechoic chamber and the reverberating indoor environment, respectively, with and without the drone using the M-sequence code. The vertical and horizontal axes in Figure 22, respectively, show the average value of the positioning error for all results and the experimental environment. As with Sections IV and V, the positioning calculation is done after subtracting the average difference from the true value from the ultrasonic measured distances. The line on the bar in Figures 22 and 23 has its upper and lower ends at the maximum and the minimum RMS positioning errors, respectively. As shown in Figure 22, the RMS positioning error in the anechoic chamber and reverberant environment was $32\mathrm{mm}$ with the drone and $10\mathrm{mm}$ without the drone.

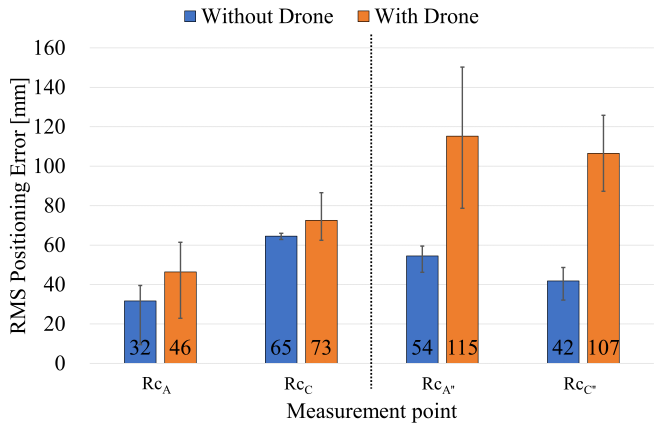
The two positions of Rc_a and Rc_A as well as Rc_c and Rc_C have respectively the same transmitter-receiver distance. Figure 23 compares the difference in errors between the anechoic chamber and the reverberating indoor environment. The vertical and horizontal axes show the RMS positioning error and measurement point, respectively. From Figure 23, no trend of the difference in errors could be confirmed. These results show that multi-path has little effect on the positioning results.

VII. Conclusions and Future Work

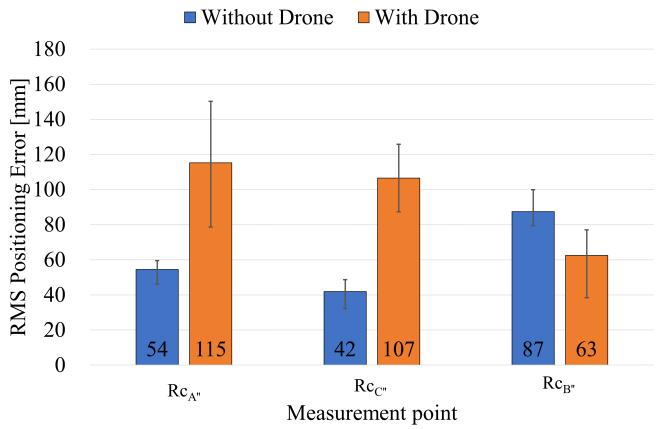
This study proposed a positioning system using SS ultrasonic waves for indoor applications, such as drone communication and wall surface inspection and evaluated the effects of the system against drone downwash and noise. The proposed SS ultrasonic positioning system transmits and receives SS signals using M-sequence, and the distance is measured using the TOF method. From the experiment conducted in an anechoic chamber with no reverberation, the positioning error was within $15\mathrm{cm}$, although the variance of the error increased due to downwash. The results were compared with the experimental results in a reverberating indoor environment, and the effect of multi-path on the positioning error was not observed. Greater accuracy in the layout of a communication robot is expected because of low DOPs. In terms of DOP, a large error was obtained in the horizontal direction because of the layout of the transmitters in the experimental environment.



(a) at the center position



(b) at the center position vs above $\mathrm{Tr_2}$



(c) at a height of $4000\mathrm{mm}$

Figure 21. RMS positioning error for Figure 18 environment.

For the positioning of communication robots, higher positioning accuracy is expected owing to the different transmitter layout. Therefore, our positioning system using SS ultrasonic waves can be applied for drone application. We will conduct positioning experiments by SS ultrasonic waves with a drone in an actual tunnel to investigate the usefulness of the proposed

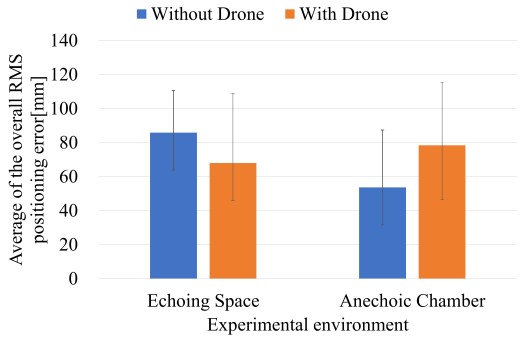


Figure 22. Comparison of average RMS positioning error for all measurement points with anechoic chamber and Figure 18 environment.

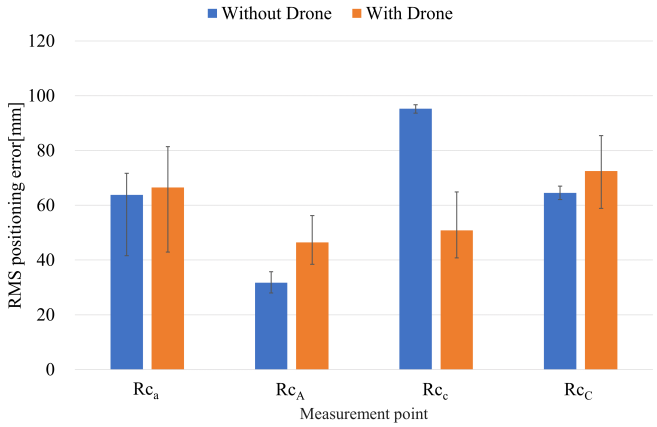


Figure 23. Comparison of RMS positioning error in the same positional relations from the transmitter to receiver with the anechoic chamber and the Figure 18 environment.

system.

ACKNOWLEDGEMENTS

This work was supported by MEXT KAKENHI Grant Number JP 10609423 and Foster Electric Company, Limited. We would like to thank Uni-edit (https://uni-edit.net/) for editing and proofreading this manuscript.

REFERENCES

[1] T. Okada and A. Suzuki, "Measurement accuracy on indoor positioning system using SS ultrasonic waves for drone applications," in VEHICULAR 2020 The Ninth International Conference on Advances in Vehicular Systems,Technologies and Applications, Porto, Portugal, October 2020, pp. 66–71.

[2] L. Apvrille, T. Tanzi, and J.-L. Dugelay, "Autonomous drones for assisting rescue services within the context of natural disasters," in 31st URSI General Assembly and Scientific Symposium, Beijing, China, Aug 2014, pp. 1–4.

[3] A. E. MacDonald, "A global profiling system for improved weather and climate prediction," Bulletin of the American Meteorological Society, vol. 86, no. 12, Dec. 2005, pp. 1747–1764. [Online]. Available: https://doi.org/10.1175/bams-86-12-1747

[4] W. Yoo, E. Yu, and J. Jung, "Drone delivery: Factors affecting the public's attitude and intention to adopt," Telematics and Informatics, vol. 35, no. 6, 2018, pp. 1687 – 1700. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0736585318300388

[5] U. Nagarajan, G. Kantor, and R. L. Hollis, "Human-robot physical interaction with dynamically stable mobile robots," Proceedings of the 4th ACM/IEEE international conference on Human robot interaction - HRI '09, Mar 2009, p. 281–282. [Online]. Available: http://dx.doi.org/10.1145/1514095.1514176

[6] F. Cunha and K. Youcef-Toumi, "Ultra-wideband radar for robust inspection drone in underground coal mines," in 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane,Australia, Mar 2018, pp. 86–92.

[7] A. Suzuki, T. Iyota, Y. Choi, Y. Kubota, K. Watanabe, and A. Yamane, "Measurement accuracy on indoor positioning system using spread spectrum ultrasonic waves," 4th International Conference on Autonomous Robots and Agents, Feb 2009, pp. 294–297.

[8] X. Gan et al., "Doppler differential positioning technology using the BDS/GPS indoor array pseudolite system," Sensors, vol. 19, no. 20. [Online]. Available: http://dx.doi.org/10.3390/s19204580

[9] H. Park, J. Noh, and S. Cho, "Three-dimensional positioning system using bluetooth low-energy beacons," International Journal of Distributed Sensor Networks, vol. 12, no. 10, 2016, p. 1550147716671720. [Online]. Available: https://doi.org/10.1177/1550147716671720

[10] M. Hazas and A. Hopper, "Broadband ultrasonic location systems for improved indoor positioning," IEEE Transactions on Mobile Computing, vol. 5, no. 5, 2006, pp. 536–547.

[11] L. Segers, A. Braeken, and A. Touhafi, "Optimizations for FPGA-based ultrasound multiple-access spread spectrum ranging," Journal of Sensors, vol. 2020, Jun. 2020, pp. 1–26. [Online]. Available: https://doi.org/10.1155/2020/4697345

[12] A. Suzuki, T. Iyota, and K. Watanabe, "Real-time distance measurement for indoor positioning system using spread spectrum ultrasonic waves," in Ultrasonic Waves. InTech, Mar 2012. [Online]. Available: https://doi.org/10.5772/30215

[13] K. Kumakura, A. Suzuki, and T. Iyota, "Code division positioning by continuous signals using spread spectrum ultrasonic waves," International Conference on Indoor Positioning and Indoor Navigation (IPIN 2013), Oct 2013, pp. 1–8.

[14] A. Suzuki, K. Kumakura, D. Tomizuka, Y. Hagiwara, Y. Kim, and Y. Choi, "Positioning accuracy on robot self-localization by real-time indoor positioning system with SS ultrasonic waves," Journal of the Korea Society for Power System Engineering, vol. 17, no. 5, Oct. 2013, pp. 100–111. [Online]. Available: https://doi.org/10.9726/kspse.2013.17.5.100

[15] J. A. Paredes, F. J. Álvarez, T. Aguilera, and J. M. Villadangos, "3D indoor positioning of UAVs with spread spectrum ultrasound and time-of-flight cameras," Sensors, vol. 18, no. 2, Dec. 2018, p. 89. [Online]. Available: https://doi.org/10.3390/s18010089

[16] Z. Huang et al., "A noise tolerant spread spectrum sound-based local positioning system for operating a quadcopter in a greenhouse," Sensors, vol. 20, no. 7, Apr. 2020, p. 1981. [Online]. Available: https://doi.org/10.3390/s20071981

[17] H. Zhang, G. Chen, Z. Wang, Z. Wang, and L. Sun, "Dense 3D mapping for indoor environment based on feature-point SLAM method," in Proceedings of the 2020 the 4th International Conference on Innovation in Artificial Intelligence, ser. ICIAI 2020. New York, NY, USA: Association for Computing Machinery, 2020, p. 42–46. [Online]. Available: https://doi.org/10.1145/3390557.3394301

[18] P. K. Enge, "The global positioning system: Signals, measurements, and performance," International Journal of Wireless Information Networks, vol. 1, no. 2, Apr. 1994, pp. 83–105. [Online]. Available: https://doi.org/10.1007/bf02106512

[19] A. Suzuki, T. Iyota, C. Yongwoon, Y. Kubota, K. Watanabe, and A. Yamane, "Measurement accuracy on indoor positioning system using spread spectrum ultrasonic waves," in ICARA 2009 - Proceedings of the 4th International Conference on Autonomous Robots and Agents, 03 2009, pp. 294 – 297.

[20] "Sht3x (rh/t) - digital humidity sensor," 2021, [accessed: 2021-11-29]. [Online]. Available: https://www.sensirion.com/en/environmental-sensors/humidity-sensors/digital-humidity-sensors-for-various-applications/

[21] A. Suzuki and T. Iyota, "Angular dependence of transducers for indoor positioning system using SS ultrasonic waves," 2012 International Conference on Indoor Positioning and Indoor Navigation, Nov 2012, pp. 1–6.

An Interactive AR-Based Virtual Try-on System Using Personalized Avatars: Augmented Walking and Social Fitme

Yuhan Liu[1], Yuzhao Liu[1], Shihui Xu[1], Kelvin Cheng[2], Soh Masuko[2] and Jiro Tanaka[1]

[1] Waseda University, Kitakyushu, Japan

[2] Rakuten Institute of Technology, Rakuten, Inc., Tokyo, Japan

e-mail: liuyuhan-op@akane.waseda.jp, liuyuzhao131@akane.waseda.jp, shxu@toki.waseda.jp, kelvin.cheng@rakuten.com, so.masuko@rakuten.com, jiro@aoni.waseda.jp

***Abstract* — E-commerce websites offer the convenience to consumers to purchase clothes online. However, they have difficulties imagining how they will look like. To address this problem, we propose a holographic 3D virtual try-on system that provides users a novel experience where they can view garments fitted onto their own personalized virtual body. The garment models are generated from garment images obtained from online shopping websites. Users can animate their dressed virtual body in a real-life scene in augmented reality. We also propose Social Fitme, which provides multiple users an intuitive sharing experience with others. Users can try on clothes with their friends and communicate with each other, thereby giving them the opportunity to make more confident decisions. We conducted a user study to compare our proposed system with an image-only shopping system and validate its effectiveness. We found that Social Fitme can greatly improve the shopping pleasure of users and provide them a more engaging and effective shopping experience. Interactions between users can help them explore new styles, enhance their relationships, and strengthen their social connections.**

Keywords* — *augmented reality; virtual try-on; personalized avatar; virtual fitting; garment modeling; social interaction; shared experience; augmented walking.

I. INTRODUCTION

With the continuous development of the e-commerce technology, the number of consumers purchasing clothes online is increasing. Consumers usually desire to try on garments to assess if the clothes are suitable before purchasing. However, when shopping online, they cannot try them on. Consequently, they may worry how well the clothes will fit on their own body. Furthermore, it is difficult for them to imagine how they will look like with various postures (e.g., standing, walking and posing) or in different settings.

To address these problems, we have proposed a 3D virtual try-on (VTO) system using personalized models [1] (Figure 1).

(a) We generate virtual models of users based on their own body and face information.

(b) We gather some garment information and achieve 3D garment visualization using Cloth-Weaver [2].

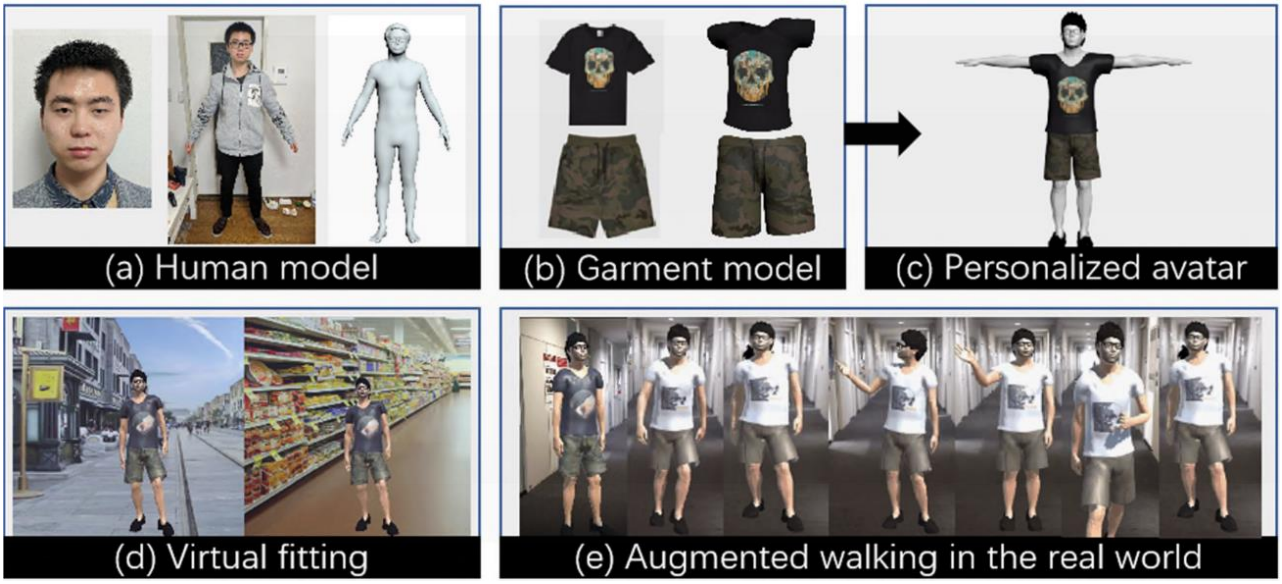


Figure 1. Our system allows users to view virtual garment fitted onto personalized body models and animate them in real-life scene

(c) We customize the garment model for each user and match it to their personalized virtual models.

(d) We enable users to view their own personalized body model fitted with virtual garment.

(e) We enable these models be visualized in a real-life scene together with animated motions.

To determine the user's acceptability, we conducted a user study to evaluate the value and convenience of our system. We also propose Social Fitme, which allows multiple users to virtually fit garments together (Figure 2). Our system reduces the gap between offline shopping and online shopping. It enhances the social experience of online shopping by allowing users to experience the pleasure of physical shopping with friends, choose clothes for each other, and get quick feedback in real-time. The aim of the system is to improve the user experience of online shopping by enabling social interaction between users while engaging with a more personalized VTO system.

We introduce two advanced features of online shopping in our proposed VTO system as the main contributions of this study:

1) A fully personalized VTO system that enables users to view virtual garments interactively and immersively in 360° as well as check their garment on a personalized virtual body augmented with human-like motion in the real-world. Users can try on clothes directly and quickly view the appearance of their dressed body.

2) A socially interactive VTO system that supports an intuitive sharing experience for users. Users can experience social interactions during the VTO process, such as sharing their dynamic dressed virtual body with each other. Social Fitme strengthens the social connections between users when shopping online. Interactions between users can help them make more confident decisions and explore new clothing styles.

The rest of the paper is organized as follows. In Section Ⅱ, a brief review of previous research on VTO, garment modeling and virtual avatar is presented. In Section Ⅲ, we describe the system design, including human model personalization, garment model generation and 3D VTO system. In Section Ⅳ, Social Fitme, which is a socially interactive VTO system is described. In Section Ⅴ, we present our evaluation result. In Section Ⅵ, conclusion and future works are described.

II. Related Work

In this section, we will describe the related work. Related work on VTO, Garment modeling, Virtual avatar, Personalization of VTOs, and Virtual avatar in social augmentation is described in this order.

A. VTO

Earlier works on VTO are mostly conducted in computer graphics [3][4][5]. Previous works focused on two types of VTO: 2D overlay VTO and 3D VTO.

2D overlay VTO:

Hilsmann et al. [6] retextured garment overlay for real-time visualization of garments in a virtual mirror environment. Yamada et al. [7] proposed a method of reshaping the garment image based on human body shapes to make fitting more realistic. However, similar to many other retexturing approaches, it operates only in 2D without using 3D information in any way, which does not allow users to view their virtual self from arbitrary viewpoints.

3D VTO:

3D garment models perform precise garment simulation rather than just 2D overlays. Protopsaltou et al. [8] created a virtual dressing room, where customers can view garments fitted onto their virtual body. Li et al. [9] proposed a multi-part 3D garment model reconstruction method to generate virtual garments for virtual fitting on virtual avatars.

Recently, VTO has been combined with augmented reality (AR) or virtual reality (VR) technologies to provide consumers a more realistic try-on experience. Consumers can get a better sense of how they will look like when wearing the products. Several fashion firms, including Uniqlo

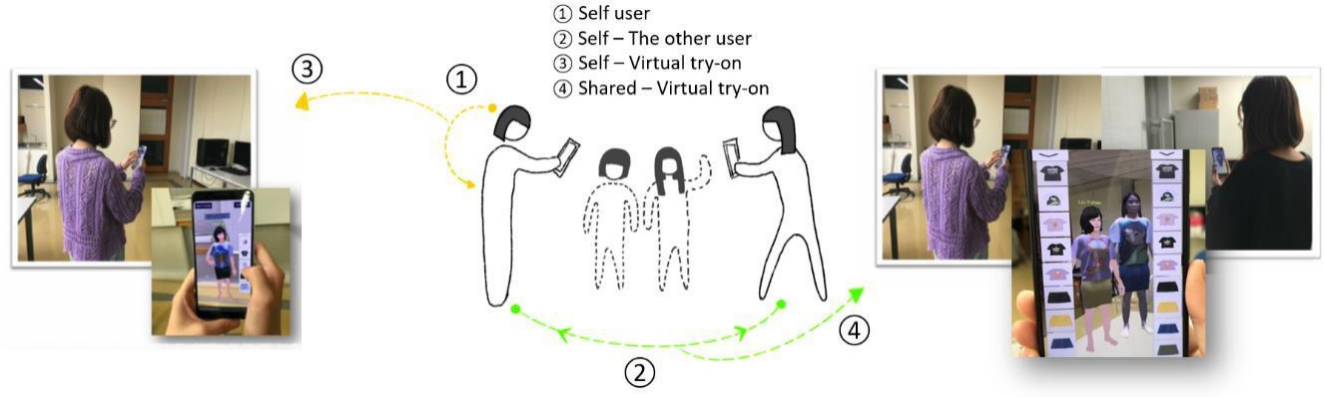


Figure 2. Social Fitme, an AR-based try-on system for multiple users to share their fitting experience together. Users can virtually try- on clothes using a mobile device and view their personalized virtual avatar in motion in the real world. They can speak to each other and give fashion advice in real-time when using the system

and Gap, have utilized the AR technology in the form of a mobile applications [10]. Using the VR technology, consumers can feel like they are physically in a virtual fitting room. Several fashion retailers have provided this kind of shopping experience, such as Alibaba and Dior.

B. Garment modeling

Unlike 2D images, a 3D garment model can be used to perform precise garment simulation. Most garment modeling research focuses on 3D garments for a virtual character. Some garment-retargeting methods transform garment designs from one character to another. For example, Pons-Moll et al. introduced a system using a multi-part 3D model of clothed bodies for clothing extraction and retargeting the clothing to new body shapes [11]. Pattern-based methods simulate the garment creation process in real life, while garment modeling tools, such as Marvelous Designer [12], offer garment modeling and editing of pattern design. Pattern-based methods require professional knowledge of garment design and are difficult for non-experts. To address the problem of digitizing garments, Zhou et al. created virtual garments from a single image [13]. Chen et al. captured real garments using a depth camera and built a coarse shape from its raw RGBD sequence using the RGB color information and depth information [14].

C. Virtual avatar

Most VTO systems provide a virtual fitting experience on a default virtual avatar, rather than one generated from users' own body [15]. The default virtual avatar can be modified by users based on individual preferences and can be personalized by uploading their facial images [16][17]. This type of virtual avatar cannot reflect the true body shape of consumers.

The absence of "true fit" may disappoint customers when shopping online. For our proposed system, we create virtual a personalized model for each user, which can reflect their body shape and facial appearance. This will make their try-on experience more accurate and engaging, thus increasing their confidence when making purchasing decisions on garments online.

D. Personalization of VTOs

Depending on the avatar's level of personalization, the avatar representing the user may or may not provide a real sense of self. According to the avatar's similarity to the user, virtual try-on systems can be divided into two levels [18]:

1) ***Non-personalized VTO:***

Some VTO experiences are based on a default virtual avatar, not generated from the user's own body [19][20]. The lack of precision in describing users and products reduces VTO experience of users.

2) ***Personalized VTO:***

Personalized VTO enables a more realistic user experience where the virtual avatars can mirror their actual looks and fit the clothes on the virtual self [21][22][23]. The virtual avatar is customized with personal features (face, height, weight, and body shape). Kim et al. researched the perceived usefulness, enjoyment, and ease of use of the personalized VTO [20]; Merle et al. argued that the personalized VTO can lead to more positive consumer perceptions than the non-personalized VTO and provide a higher self-congruity with the virtual model [18].

E. Virtual avatar in social augmentation

Early research has shown that a virtual avatar represented by a graphical persona in a virtual social environment can create an illusion for users that they are in the environment and co-present with their companions [24][25][26]. Roth et al. found that virtual avatars that can perform daily social interactions can increase social presence in a multi-user environment and enhance user experience [27]. For co-located social experience, Lankes et al. investigated the effects of avatars and verbal communication on co-presence in a cooperative game [28].

In summary, we found that there is a lack of research exploring the dynamic VTO experience with personalized motions. In addition, the social feature is an important factor of dress fitting that has been ignored in previous VTO systems. Our proposed AR-based try-on system provides multiple users an intuitive sharing experience, allowing them to view their dressed body with personalized human-like motions augmented in the real world and share their outfits with others.

III. 3D VTO

Our 3D VTO system involves human model personalization, garment model generation, and 3D VTO.

Figure 3 presents an overview of our proposed system. This system uses three elements as input: a single face image with a full-frontal face, a short video of the user's full body, and a 2D garment image from online shopping websites. The procedure can be described as follows:

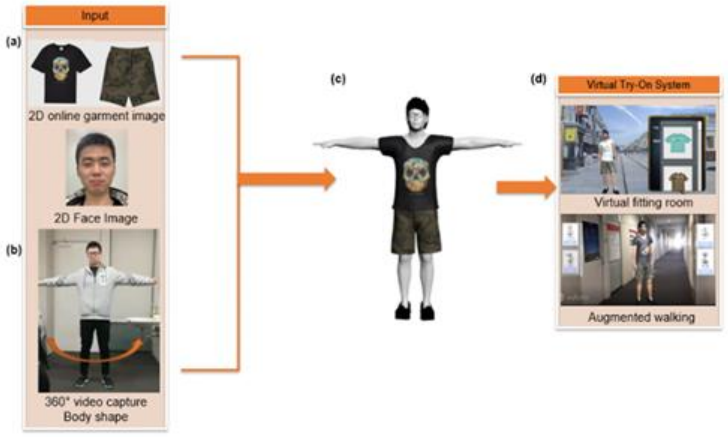


Figure 3. System Overview

Figure 4. Generate 3D garment model based on the information from shopping website

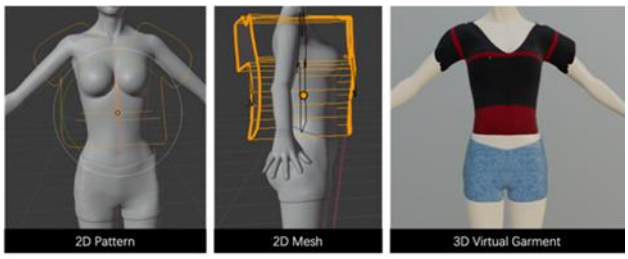


Figure 5. 2D patterns creation and positioning around generic body

(a) Mapping the 2D garment image into the 3D garment model templates and generating the 3D garment model based on online images.

(b) Generating the 3D human model based on the face image and recorded video.

(c) Matching the 3D garment model to the human model.

(d) Users choose different clothes to try on. Users can change the pose in VR fitting room. They can also see their augmented walking in AR environment.

A. Human model personalization

Owing to the lack of physical fitting in online shopping, a gap between the actual and perceived body size can exist, which may make it difficult for consumers to examine the true fit on their own body and influence their purchase selection while shopping online. Therefore, the virtual human body should have an appropriate 3D representation corresponding to the real user's human body shape and facial features.

This will give a better representation of the user and allows for a more accurate clothes fitting, as well as for virtual human body animation. We create human body models based on Alldieck's study [29] and generate face models based on Deng's method [30]. Moreover, a hair model library is prepared, and the most similar hair model is matched to the face model we generated.

B. Garment model generation

To provide users with better garment product visualization, we allow users to view garments from various angles and directions when users are shopping online. Our approach uses garment image information from existing shopping websites (e.g., H&M [31] and Zara [32]) to create a virtual garment library. Textures are extracted from the garment image and mapped onto the 3D garment model. The final 3D garment is shown in Figure 4. We mainly focus on these two parts: garment model templates used in our system and 3D modeling and texturing approaches.

Figure 6. Some 3D garment templates provided for users

3D garment model templates

Garments are created using the traditional 2D pattern approach. We build several 3D templates of virtual garment models for the personalized human model using Cloth Weaver, which is a Blender template library. It allows simulating the methods of traditional garment designs. The 2D pattern is discretized into a triangular mesh. Next, we design and modify the 2D pattern, and then use the reference line to automatically fit the flat pattern to the corresponding part of the body (Figure 5). These are used as the basis for creating a variety of garment models (Figure 6). We simulated several types of clothing for female and male bodies. For females, we prepared them with long sleeves, T-shirt, long pants, dress and skirt for fitting. For males, we prepared them with T-shirts, long sleeves and half pants for fitting.

Texture mapping

We collected garment images from existing shopping websites (H&M, ZARA, etc.) and mapped these clothes images to generated 3D garment model templates (Figure 7). We segmented different parts of the garment from a single garment image. The segmented clothes can be divided into three main parts: left sleeves, right sleeves, and the front of clothes. The 3D mesh of a generated garment template can be extended into a 2D reference mesh in 3ds Max [33]. To map the web garment image into a 3D virtual garment template, we

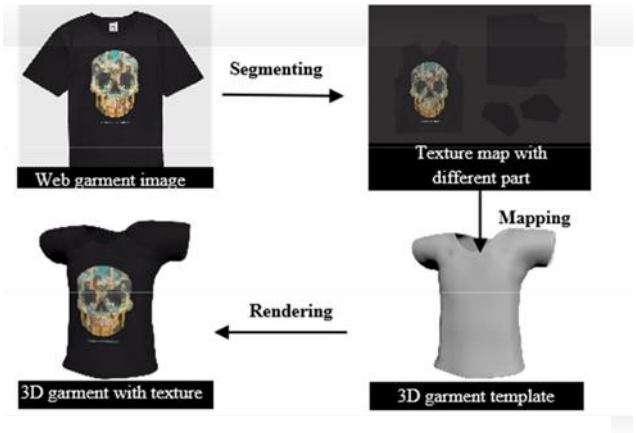


Figure 7. Mapping Web garment image to generated 3D garment templates

Figure 8. Garment model library for female and male

map the different segmentation parts from the garment image to its corresponding parts on the garment template.

In this way, we can generate a 3D garment model with texture. The garment can be customized in various ways to match the desired design. The most obvious change is the customization of appearance and color, which is achieved by modifying the texture of the cloth. Therefore, we collected garment images from online shopping websites as textures and created a garment model library for users (Figure 8).

C. 3D VTO

We gathered various garment information from online websites to enhance the online shopping experience of users. Our system was developed using Unity3D [34] on Windows 10 and we deployed our system on Android smartphones. The 3D VTO system consists of two parts: virtual fitting and augmented walking.

Virtual fitting

Virtual reality relies on an entirely digital environment that can provide an immersive and interactive shopping experience for users. We prepared a variety of fitting scenes for users, such as on the street, in the office and at the supermarket. Users can view the virtual garment based on the different virtual scenes, giving them an idea of how they will look like for various occasions or purposes (Figure 9).

Augmented walking

Normally, when users shop at physical (offline) stores, they often check the attributes of clothes through various motions, such as twisting the body or raising the arm to help confirm the suitability of the clothes. However, when shopping online, users cannot visualize the details of the garment. Compared to the offline try-on experience, the traditional online shopping purchasing environment lacks the capability for users to try-on garments on their own body and check whether the clothes fit on them under various postures. Therefore, we propose a dynamically interactive method that allows users to animate their dressed human body in 360°and

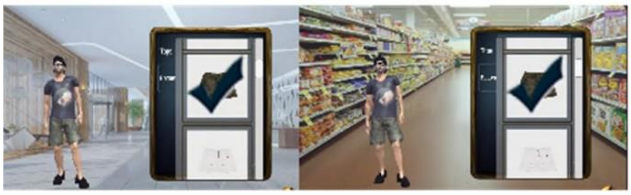


Figure 9. Fitting scenes

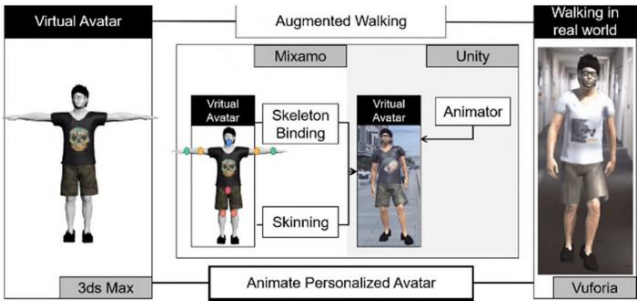


Figure 10. Implementation of augmented walking

enables them to view their virtual body walking in the real-life scene.

The implementation of the augmented walking framework aims to animate the personalized avatar of users in the real world. Figure 10 depicts the workflow of animating the personalized virtual avatar in the real world, which can be described as follows.

1) **Personalized virtual avatar.** We integrate the virtual human model and clothes model in 3ds Max and export the virtual avatar as .fbx file.
2) **Skeleton binding and Skinning.** We upload the personalized virtual avatar to Mixamo [35], which is a web-based service for creating 3D animation of human models. We bind the skeleton to the virtual avatar and skin it using Mixamo.
3) **Animate virtual avatar.** To attach the animation to a personalized avatar, we use the animator controller in Unity [36] to control the virtual avatar and perform various animation.
4) **Augmented walking in the real world.** We realize the augmented walking using Vuforia Augmented Reality SDK [37].

We fitted our generated human model with garment models and created walking and posing animations, as illustrated in Figure 11. Using Mixamo, the motions we generated are very lifelike actions, such as waving/shaking hands, walking, sitting, and turning around.

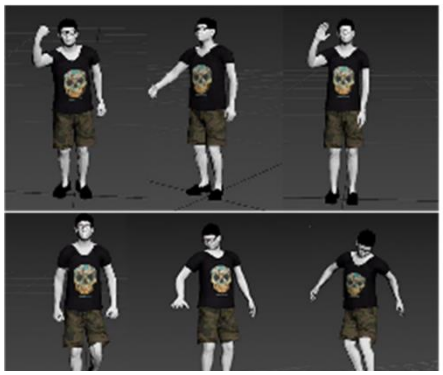


Figure 11. Postures of personalized human model

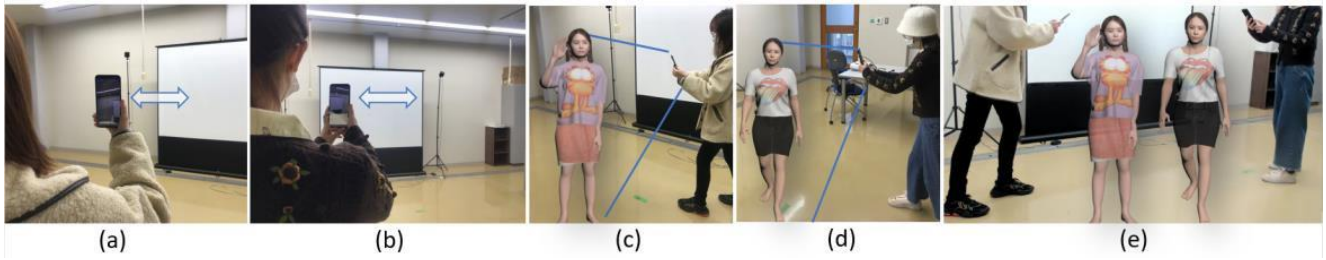


Figure 12. The workflow of Social Fitme

IV. SOCIAL FITME: SHARED VTO EXPERIENCE

AR applications can enable social interactions among users. They can interact with the virtual content while engaging in normal communication in the real world. So far, there is a lack of research on applying AR for VTO.

Social Fitme is an AR mobile application that supports interactive social sharing of VTO experience for multiple online shopping users (Figure 12). Based on online shopping websites, 3D garments are generated, which provide users a more realistic try-on.

Users can first create a fitting room by scanning the ground (a, b), and animate their virtual avatars in the real environment (c). Other co-located users can join in the same fitting room and can be represented as virtual avatars located in the same position and orientation in the same physical environment (d). Users can interact with the virtual avatar, virtually try garments on, communicate with each other and give fashion advice (e).

To provide users a more engaging VTO experience, Social Fitme allows users to socially share their intuitive VTO experience with others, enabling them to get quick feedback as well. Users can both try on clothes on their own body or on other user's body. They can also help each other make decisions and find the best match for each other (Figure 13).

Our system was developed using Unity3D on Windows 10 and deployed on Android smartphones. The implementation of our system consists of the following three parts:

Human model personalization. We personalized theappearance of users based on their face image and full-body images. The 3D face model was generated using 3D avatar SDK [38], whereas the body model was built by 3DLook [39]. They were then combined into a 3D human model.

Motion capture. To gather individual movements of users, we used Microsoft Kinect V2 depth sensor [40] to record postures and user movements and create their own animation library for our system. To convert the captured motion captured data into animation, we used a Unity plugin, Cinema-Mocap [41], which is a marker-less motion capture solution for Unity to create customized animations for users. As the movement captured by Kinect V2 depth sensor is somewhat jittery, we edited and smoothed the animation frame-by-frame using Maya software [42], which is a 3D computer animation, modeling, simulation, and rendering program.

Shared VTO experience. To share the VTO experience with co-located users, we used ARCloud Anchors API from Google. ARCore [43] connects to its ARCore cloud anchor service to host and resolve anchors. The hosting and resolution of an anchor can be deployed on multiple devices through an effective network connection.

Figure 13. Our system provides an intuitive sharing virtual try-on experience for users

TABLE I. QUSTIONNAIRE AND MEASRUEMENT ITEMS

Items	Statements
Enjoyment	**a.** Using the system, shopping experience was enjoyable for me.
Convenience	**b.** I can get a sense of how the outfit might look for the various occasions. **c.** I can get a sense of how I look wearing these clothes.
Augmented Walking	**d.** Seeing a model of me walking in the real-world enhanced my shopping experience. **e.** Having a model walking in a real environment helps me understand more about the appearance of the clothes.
User Behavior	**f.** I want to use this system when I buy some clothes online in the future.

V. EVALUATION

We carried out two user studies. i.e., user study 1 and user study 2, to evaluate our system. The objective of user study 1 is to assess whether our 3D VTO system can enhance the online shopping experience of users. The objective of user study 2 is to evaluate the effectiveness of the system and interactions designed to support the sharing of VTO experience in AR.

A. User study 1

1) Participants

A total of 10 college students participated in both condition 1 (VTO condition) and condition 2 (image-only condition). College students aged 18–30 years are usually targeted by AR/VR applications, as they are more likely to try new technologies and they are proactive in online shopping for fashion products. Hence, we invited N=10 participants (7 males and 3 females) with an average age of 22.5 years to complete our evaluation.

2) Evaluation design

We conducted an initial experiment to evaluate our system. The objective was to assess whether our 3D VTO system could enhance the online shopping experience of users, thereby helping them make better purchasing decisions. To investigate the users' attitudes toward the typical online shopping and 3D VTO with augmented walking, we conducted an experimental study with the following two conditions.

a) VTO condition: simulation of shopping experience using our 3D VTO system.
b) Image-only condition: simulation of typical online shopping experience using only images of garments online.

We hypothesized that the former condition would lead to a higher rating than the latter.

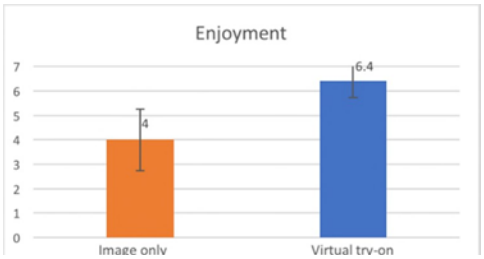


Figure 14. Participants rated their Experiences more enjoyable in the virtual try-on condition

We personalized the human model of each participant based on their 2D face image and 360° body videos. Each participant simulated the shopping experience with two different conditions. The order of the conditions was randomized. After each task, the participants were asked to rate their experience (from 1 "strongly disagree" to 7 "strongly agree") in our questionnaire, indicated on a 7-point Likert scale. At the end of the experiment, we interviewed the participants to gather their preferences and open-ended feedback.

We measured enjoyment, convenience, and user behavior for the two conditions. We also analyzed whether the augmented motion in the real-life scene enhances the users' shopping experience. The questionnaire and measurement items are presented in Table I.

3) Results

We separated the result into two sections: analysis of the rating from questionnaires and thematic analysis of the participants' comments.

We analyzed the result in terms of user enjoyment, convenience, augmented walking, and user behavior.

Enjoyment. As Figure 14 shows, we found a significant effect on the participants' shopping enjoyment. A repeated measures t-test revealed a statistically significant difference between the various conditions ($p<0.01$). The participants rated the enjoyment significantly higher in the VTO condition.

Convenience. We analyzed the user convenience through the two statements below:

A. "I can get a sense of how the outfit may look for various occasions." We found that the participants rated the VTO ($p<0.01$) significantly higher than the other condition (Figure 15).

B. "I can get a sense of how I look wearing the clothes." We also found that participants rated the virtual try-on condition ($p<0.01$) significantly higher, meaning that it gave users a better feeling for how these clothes might look like on their body (Figure 16).

Figure 15. Participants rated their Experiences more enjoyable in the virtual try-on condition

Figure 16. Participants rated that virtual try-on condition gave users a better feel for how these clothes look like on their body

Augmented walking. To understand whether the 3D VTO system within the AR scene enhances the user experience, we prepared two statements:

a. "Seeing a model of myself walking in the real-world enhanced my shopping experience." Figure 17 summarizes the of participants' opinions of VTO condition.

b. "Having a model walking in a real environment helps me understand more about the appearance of the clothes." Figure 18 summarizes the participants' opinions of the VTO condition.

In summary, all participants have the same opinion that augmented walking can enhance their shopping experience. Of the 10 participants, 9 rated that the virtual model walking in the real environment helps them understand more about the appearance of the clothes. One of the main reasons given is that the real environment is very realistic, which helps them to view the appearance of the garment model. Moreover, the virtual models walking in the real-life scene are very interesting and can improve the participants' enjoyment of the online shopping process. At the same time, augmented walking can also provide a better 3D visualization for users.

The dynamic fitting display can show the shape of the clothes when they are in motion and increases the number of clothes attributes that can be observed.

User behavior. All participants preferred the VTO condition for both enjoyment and convenience. We also analyzed the

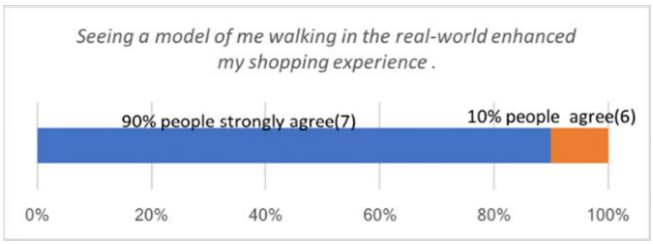


Figure 17. All participants agree that seeing own model in the real-world enhanced their shopping experience. 9 out of 10 participants strongly agree with it

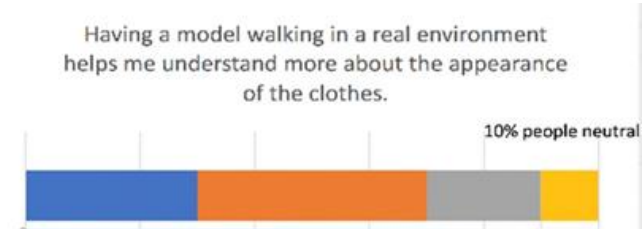


Figure 18. Most participants agree that the model walking in the real environment helps them understand more about the appearance of the clothes

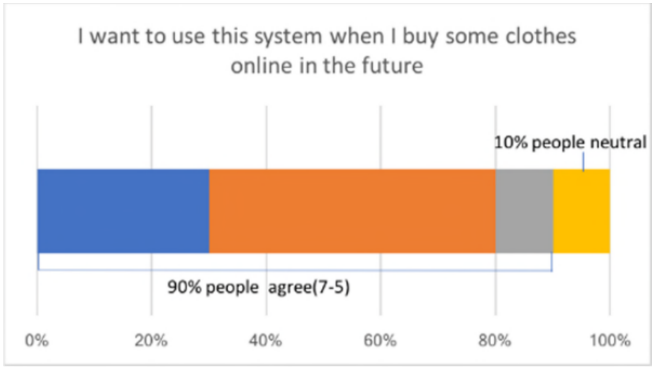


Figure 19. Most participants want to use this system when they buy some clothes online in future

user preference on whether they want to use the VTO system in the future or not. The results indicated that 9 out of 10 participants wanted to use this system in the future (Figure 19).

4) Qualitative results

At the end of the experiment, open-ended feedback was sought from the participants, and a thematic analysis was performed on their responses and their impression of using our 3D VTO system.

Most participants thought that the virtual avatar with augmented walking in the real world offers them a sense of wearing clothes on their own body, which can provide them with a better understanding of the detail of the clothes. Moreover, augmented walking allows them to visualize their personalized model in the real world, increasing their shopping enjoyment. P6 mentioned that the augmented walking makes them feel like they are looking into a mirror. P7 said that augmented walking in the real world can help

them observe more details of clothes.

Most participants thought that our system is interesting. Our system can enhance their experience and narrow their selections when shopping online. For instance, P1 mentioned, "Shopping online is difficult because the model's body shape is pretty, while people in real life don't have such a perfect body. This system shows how the clothes look like on my body in the real world, which gives me confidence when buying clothes." Similar comments were received from P3 and P4.

Furthermore, the 3D VTO system provides users outfit ideas and more clothing information to users. P3 thinks that the 3D virtual system provides various virtual scenes to help users in selecting clothes, especially for special occasions. P6 mentioned that the 3D garment model allows him to see himself wearing clothes in 360° and obtain additional clothing information than just looking in a mirror. P7 said that the virtual model walking in the real world may help them in checking how they will look like in the real wearing conditions.

We also received comments about future improvement. P3 suggested that the material of clothes could be improved to look more like a real fabric, and P9 thought that it would be better to use motion capture to simulate real movements of users moving in the real world. The free comments from participants are summarized in Table Ⅱ.

B. User study 2

1) Participants

We recruited 12 participants (6 females and 6 males) aged between 20–25 years old. We selected participants from this demographic because they are usually the target users of AR/VR applications, since they are more likely to try new technologies and are proactive in online shopping for fashion products. We gathered their relationship with their partner by the inclusion of community in self-scale (IOS) [44].

TABLE II. SUMMARY OF FREE COMMENT

Keyword	Conclusion andComments
Augmented walking	**Judging of fitting:** Wearing clothes doing some activities in the real world provides users with better understanding of the detail of clothes, which allows to better judge of fitting. **Humanoid motion:** Using motion capture technology to capture user's movement may offer users a better sense of "real me".
Garment model	**Information visualization:** The 3D virtual try-on system gives users outfit ideas and provides more clothing information to users (muti-direction and muti-angle). **Realistic garment:** Garment material can be more like real fabric.
Shopping experience	The 3D virtual try-on system can narrow users' selections of clothes and increase their purchase confidence. Increases the enjoyment of shopping experience.

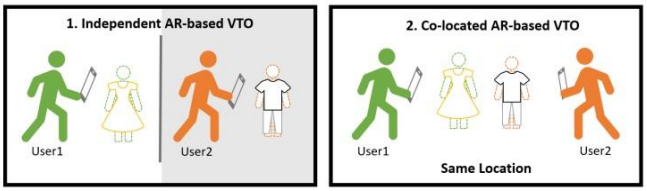


Figure 20. The experiment conditions

2) Evaluation design

Our study used a within-subject design in which two participants used the VTO system with two different modes of independent and co-located AR-based VTO in random order. The two conditions are as follows (Figure 20):

1. **Independent AR-based VTO.** Participants place their virtual avatar separately in their own physical environment and view their own personalized virtual body with garment models and posing or walking in the real world.
2. **Co-located AR-based VTO.** Participants can share their VTO experience with their partners at the same time and in the same place. They can communicate verbally and visually and give advice on each other's fitting effects.

We separated the participants into six pairs, where people of the same sex are in the same group. We asked each pair of participants to virtually try the clothes on and simulate the shopping experience using the independent AR-based try-on condition and the co-located AR-based try-on condition. Afterward, the participants were asked to complete a 7- Likert scale questionnaire to rate their shopping experience. At the end of the experiment, we interviewed the participants and gathered some open-ended feedback.

3) Results

We analyzed the result in terms of body similarity, enjoyment, usefulness, closeness, and user behavior. Paired t-test was performed using SPSS [45] to assess whether there were statistically significant differences between the means of the two conditions (Figure 21).

Body similarity. We analyzed the user body similarity through the statement, "I feel that the virtual body I saw was my own body." Although there is no significant difference found in body similarity ($p = 0.59$) between the two conditions, we found that their friends' evaluations of their virtual bodies often affected their perception of their own virtual avatar. Several participants mentioned that when they heard positive feedback from friends, for example, "This virtual avatar really looks like you," they felt that the virtual body became more similar to themselves. On the other hand, when participants heard negative comments from friends, they reduced the ratings of their virtual avatar.

Enjoyment. We analyzed user enjoyment through the statement, "This VTO system was enjoyable to use." We found that the participants rated the co-located AR condition ($p < 0.05$) significantly higher, meaning that it can provide users with a more enjoyable experience. The participants tended to spend more time and were more engaged in

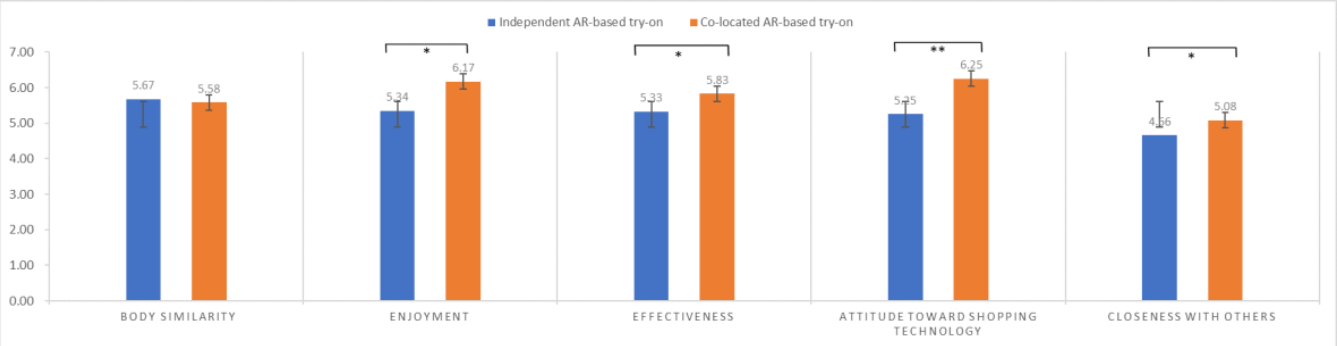


Figure 21. Perceived levels of body similarity, enjoyment, usefulness, confidence in fit problem, attitude toward shopping technology, and closeness. Note: *p < 0.05, **p < 0.01

shopping together with their partners in the co-located AR-based try-on condition. Compared with the independent VTO, they paid more attention to help friends choose clothes and give their personal opinions. P11 mentioned, "It's fun to try on clothes for my partner. I feel more engaged when we can talk to each other and try-on together; it is just like we are shopping together."

Effectiveness. We analyzed the system's effectiveness through the statement, "Using this VTO system would enhance my effectiveness in shopping." We found a significant effect on users' shopping effectiveness ($p < 0.05$). Most participants mentioned that using the co-located AR system allows them to receive recommendations from others. For example, P2 commented, "When I shared my VTO experience to my partner, I could get more helpful advice from my partner."

Several participants thought that comments from others can help them explore a completely new clothing style. P6 mentioned, "When I shop online alone, the styles of clothes that I choose are very similar to what I already have. When using this system, friends will often give me suggestions, encouraging me to try on new styles."

Many participants said that our system offers an intuitive try-on effect, which can help save time when shopping. For instance, P10 stated, "Friends can quickly help us eliminate clothes that are not suitable, which is a great way to save time on shopping."

Compared to just sharing links or outfit photos with friends, sharing an intuitive fitting experience with others can obtain a more accurate recommendation. P2 remarked, "When I shop online, I can only share links and images of clothes, and my friends can only give advice based on their imagination. It is not helpful for me."

User behavior. We analyzed the users' attitude toward our shopping technology through the statement, "I want to use this system when I buy some clothes online in the future." We found that the participants rated the co-located AR condition ($p < 0.01$) significantly higher than the independent AR condition. We also asked the participants' preferences for each condition at the end of the study. Consequently, 11 participants preferred the co-located AR-based try-on condition. Only one participant favored the AR-based independent try-on because he was worried that sharing his body with others would reveal his privacy.

Happiness and closeness with others. We analyzed the users' happiness when using the co-located AR system through the statement, "I feel happy when I am shopping with my partner." The participants strongly agreed that they feel happy ($M = 6.58$, $SD = 0.64$) when shopping with their partners. Most participants reported feeling happy and closer with their partners after using our system. P1 mentioned, "It allows me to experience the pleasure of shopping with friends." Furthermore, P12 noted, "Seeing our avatars doing real-life motions in the real world, similar to this kind of game-like try-on experience, makes me very happy."

We analyzed the closeness between participants by the inclusion of community in self-scale (IOS). The results showed that closeness with their partners increased from an average IOS score of 4.66 ($SD = 1.15$) to 5.08 ($SD = 1.24$) after the experiment ($p < 0.05$). Some participants thought that using our system can enhance their relationship and strengthen their social connections. P2 stated, "It's fun to try on clothes in the same space with my friend's avatars. We can talk to each other, see each other's avatars, and then interact with each other. I think our relationship has become closer."

We summarized the main results and presented explanatory analysis based on our observation during the study as well as findings from the interviews. The results indicated that Social Fitme can greatly improve the shopping pleasure of users and provide them a more engaging and effective shopping experience. Our system provides users an intuitive sharing VTO experience. Comments from others can not only help users make confident decisions, but also allows them to explore a completely new clothing style. Users can enhance their relationship and strengthen their social connection by using our system. We summarized the results in Table Ⅲ.

TABLE III. SUMMARY OF RESULT

Theme	Conclusion
Effectiveness	**Helpful comments:** comments from others can help users make a more confident decision. **Diversified styles:**users can explore a completely new style based on advice given by others. **Intuitive sharing experience:** personalized avatars in AR provide a more intuitive way to share their outfits with others when online shopping, which can offer users a better judge of fit. **Filter mismatched selection:** users can eliminate the unsuitable selections quickly. **Detailed garment visualization:** users can view garments from multiple angles and in more details.
Happiness Closeness with others	**Enhance relationship:** using Social Fitme strengthen social connection and make users feel closer together. **Game-like shopping experience:** users feel happier when simulate try-on experience with others.
Enjoyment	**More engaged:** users are more enjoyable and engaged in sharing outfits with others, they tend to ask for other's recommendations and make a more confident choice.

VI CONCLUSION AND FUTURE WORK

In this paper, we presented a 3D VTO system to help consumers obtain a better sense of how they will look when purchasing clothes online. To allow users to assess how well the displayed products match their actual body, we personalized users' own virtual avatars corresponding to real users' human body shape and face features. Based on online garment images, we generated 3D virtual garments to personalize the human body. Users can fit their 3D user models with a selection of virtual garments, and view the animated body in the real-life scene, as well as various virtual scenes, to get a better sense of the dynamic effects of the clothes. An initial evaluation reveals that the VTO system is more enjoyable and convenient than that of using images only.

Augmented walking provides an interactive dynamic VTO experience for users, which gives them better understanding of the details of clothes. The virtual avatar wearing clothes in the real world can provide a better sense of "true fit," which helps users judge their fitting more effectively. Furthermore, most of the participants prefer using this system for online shopping in the future. They think that this system can increase their purchase confidence and solve the fit problem when shopping online.

However, our system still has certain limitations that can be improved. In the future, we plan to enhance our clothing animations and cloth simulation methods to provide users with a more realistic VTO effect. Motion capture can also be used to better simulate user's walking motion, to provide a more realistic and interactive fitting experience.

We also presented Social Fitme, a platform and concept for co-located social try-on system that supports personalized interactions through smartphones. The system implements interactive technology combined with AR and cloud technologies and provides users with a novel shopping experience. We conducted a user study to explore the effectiveness of our system and interactions designed to support the sharing of VTO experience in AR. We found some possibilities and advantages of multi-user online social shopping:

- **Try clothes on users' friends directly.** Rather than just imagine how the clothes look like on their friend's body, users can try clothes on the personalized avatar of their friend directly and view the effect on the dressed virtual body in the real-life scene quickly.
- **Explore a completely new style of clothing.** Our system allows users to experience the pleasure of physical shopping with friends, allowing them to choose clothes for each other, or even explore a completely new style of clothing that they would otherwise not try on by themselves.
- **More confident decisions.** The sharing VTO experience in AR can provide users with a more effective shopping experience. Sharing the intuitive dressed body allows users to obtain more accurate comments from others, thereby helping them make more confident purchase decisions.
- **Strengthen social connection.** The rating of IOS showed that the shared VTO system can strengthen social relationship between friends and make users feel closer to each other.

In the future, we will improve our work from these two perspectives:

1) **Improve sense of personalization with facial expression.**
Our personalization research result in VTO shows that the level of personalization can be improved to a certain extent, such that the realism and interactivity of VTO may be increased (e.g., facial expression of virtual avatar). Several participants mentioned that they hope that our system will add real facial expressions to make the virtual avatar more natural/humanlike and share the emotional attitude toward virtual clothes with others. Thus, future work may explore expression design of the personalized avatars to provide a more natural and interactive sharing VTO experience for users.

2) **Enrich social interactions with VTO.**
Our system allows users to share their virtual avatar with others from a social interaction perspective. We presented a new way for multiple users to experience a VTO experience in the local space. In the future, we hope to apply the shared AR-based try-on system to remote users, allowing multiple users to cross distance barriers and share fittings through natural communication.

REFERENCES

[1] Y. Liu, Y. Liu, S. Xu, J. Yuan, X. Sun, K. Cheng, et al., "3D Virtual Try-On System Using Personalized Avatars: Augmented Walking in the Real World," in Proceedings of the Thirteenth International Conference on Advances in Computer-Human Interactions (ACHI 2020), pp. 391-398, 2020.

[2] Cloth-Weaver. https://clothweaver.com/. [retrieved: Nov. 2021]

[3] X. Han, Z. Wu, Z. Wu, R. Yu, and L. S. Davis. "Viton: An image-based virtual try-on network," in Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7543-7552, 2018. DOI: https://doi.org/10.1109/CVPR.2018.00787

[4] M. Sekine, K. Sugita, F. Perbet2, B. Stenger, and M. Nishiyama, "Virtual fitting by single-shot body shape estimation," In Proceedings of International Conference on 3D Body Scanning Technologies. pp. 406-413, 2014.

[5] P. Decaudin, D. Julius, J. Wither, L. Boissieux, A. Sheffer, and M.-P. Cani, "Virtual garments: A fully geometric approach for clothing design," In Computer Graphics Forum, pp. 625-634, 2006. DOI: https://doi.org/10.1111/j.1467-8659.2006.00982.x

[6] A. Hilsmann and P. Eisert, "Tracking and Retexturing Cloth for Real-Time Virtual Clothing Applications," in Proceedings of the 4th International Conference on Computer Vision/Computer Graphics Collaboration Techniques Springer-Verlag, Berlin, Heidelberg, pp. 94–105, 2009. DOI: https://doi.org/10.1007/978-3-642-01811-4_9

[7] H. Yamada, M. Hirose, Y. Kanamori, J. Mitani, and Y. Fukui, "Image-based virtual fitting system with garment image reshaping," in Proceedings of 2014 International Conference on Cyberworlds," IEEE, pp. 47-54, 2014. DOI: https://doi.org/ 10.1109/CW.2014.15

[8] D. Protopsaltou, C. Luible, M. Arevalo, and N. Magnenat-Thalmann, "A body and garment creation method for an Internet based virtual fitting room," in Advances in modelling, animation and rendering, pp. 105-122, 2002. DOI: https://doi.org/10.1007/978-1-4471-0103-1_7

[9] D. Li, Y. Zhong, G. Wu, and P. Hu, "Automatic three-dimensional-scanned garment fitting based on virtual tailoring and geometric sewing," Journal of Engineered Fibers and Fabrics, vol. 14, 16 pages, 2019. DOI: Https://Doi.Org/10.1177/1558925018825319

[10] H. Lee and K. Leonas, "Consumer experiences, the key to survive in an omni-channel environment: use of virtual technology," Journal of Textile and Apparel, Technology and Management, vol. 10, pp. 1-23, 2018.

[11] G. Pons-moll, S. Pujades, S. Hu, and Michael J. Black, "ClothCap: seamless 4D clothing capture and retargeting," ACM Transactions on Graphics, vol. 36, 15pages, 2017. DOI: https://dl.acm.org/doi/10.1145/3072959.3073711

[12] Marvelous Designer, https://www.marvelousdesigner.com/. [retrieved: Nov. 2021]

[13] B. Zhou, X. Chen, Q. Fu, K. Guo, and P. Tan, "Garment modeling from a single image," in Computer graphics forum, pp. 85-91, 2013. DOI: https://doi.org/10.1111/cgf.12215

[14] X. Chen, B. Zhou, F. Lu, L. Wang, L. Bi and P. Tan, "Garment modeling with a depth camera," ACM Transactions on Graphics vol. 34, 12 pages. 2015. DOI: https://doi.org/10.1145/2816795.2818059

[15] Warehouse, https://www.warehouselondon.com/. [retrieved: Nov. 2021]

[16] M. Yuan, I. R. Khan, F. Farbiz, S. Yao, A. Niswar and M. H. Foo, "A mixed reality virtual clothes try-on system," IEEE Transactions on Multimedia, vol. 15, pp. 1958-1968, 2013. DOI: https://doi.org/10.1109/TMM.2013.2280560

[17] N. Magnenat-Thalmann, B. Kevelham, P. Volino, M. Kasap, and E. Lyard, "3d web-based virtual try on of physically simulated clothes," Computer-Aided Design and Applications, vol. 8, pp. 163-174, 2011. DOI: https://doi.org/10.3722/cadaps.2011.163-174

[18] A. Merle, S. Senecal, and A. St-Onge, "Whether and how virtual try-on influences consumer responses to an apparel web site," International Journal of Electronic Commerce, vol.16, no.3, pp. 41-64. 2012. DOI: https://doi.org/10.2753/JEC1086-4415160302

[19] A. M. Fiore, J. Kim, and H. Lee, "Effect of image interactivity technology on consumer responses toward the online retailer," Journal of Interactive Marketing, vol.19, no.3, pp. 38-53, 2005. DOI: https://doi.org/10.1002/dir.20042

[20] A. M. Fiore, H. J. Jin, and J. Kim, "For fun and profit: Hedonic value from image interactivity and responses toward an online store," Psychology & Marketing vol.22, no.8, pp. 669-694, 2005. DOI: https://doi.org/10.1002/mar.20079

[21] J. Kim and S. Forsythe, "Sensory enabling technology acceptance model (SE-TAM): A multiple-group structural model comparison," Psychology and Marketing 25, 9, 901-922, 2008. DOI: https://doi.org/10.1002/mar.20245

[22] C. Li and F. Cohen, "In-home application (App) for 3D virtual garment fitting dressing room," Multimedia Tools and Applications, vol.80, pp. 5203–5224 (2021). DOI: https://doi.org/10.1007/s11042-020-09989-x

[23] K. W. Lau and P. Y. Lee, "The Role of Stereoscopic 3D Virtual Reality in Fashion Advertising and Consumer Learning," Advances in Advertising Research (Vol. VI), pp. 75-83, 2016. DOI: https://doi.org/10.1007/978-3-658-10558-7_7

[24] K. L. Nowak and F. Biocca, "The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence," in Virtual Environments. Presence: Teleoperators and Virtual Environments vol.12, no.5, pp.481-494, 2003. DOI: https://doi.org/10.1162/105474603322761289

[25] M. E. Latoschik, F. Kern, J. Stauffert, A. Bartl, M. Botsch, and J. Lugrin, "Not Alone Here?! Scalability and User Experience of Embodied Ambient Crowds in Distributed Social Virtual Reality," IEEE transactions on visualization and computer graphics, vol 25, no. 5, pp. 2134–2144, 2019. DOI: https://doi.org/10.1109/TVCG.2019.2899250

[26] M. E. Latoschik, D. Roth, D. Gall, J.Achenbach, T. Waltemate, and M. Botsch," The Effect of Avatar Realism in Immersive Social Virtual Realities," in Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST '17), no.39, pp. 1-10, 2017. DOI: https://doi.org/10.1145/3139131.3139156

[27] D. Roth, C. Klelnbeck, T. Feigl, C. Mutschler, and M. E. Latoschik, "Beyond Replication: Augmenting Social Behaviors in Multi-User Virtual Realities," in Proceedings of 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 215-222, 2018. DOI: https://doi.org/10.1109/VR.2018.8447550

[28] M. Lankes, J. Hagler, G. Kostov, and J. Diephuis, "Invisible Walls: Co-Presence in a Co-located Augmented Virtuality Installation," in Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '17). Association for Computing Machinery, New York, NY, USA, pp. 553–560, 2017. DOI: https://doi.org/10.1145/3116595.3116609

[29] T. Alldieck, M. Magnor, W. Xu, C. Theobalt, and G. Pons-Moll, "Video based reconstruction of 3d people models," in Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8387-8397, 2018. DOI: https://doi.org/10.1109/CVPR.2018.00875

[30] Y. Deng, J. Yang, S. Xu, D. Chen, Y. Jia, and X.Tong, "Accurate 3D Face Reconstruction with Weakly-Supervised Learning: From Single Image to Image Set," in Proceedings

of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 11pages, 2019.
Retrieved from https://arxiv.org/abs/1903.08527
[31] H&M. https://www.hm.com/. [retrieved: Nov. 2021]
[32] ZARA. https://www.zara.com/. [retrieved: Nov. 2021]
[33] 3ds Max.
https://www.autodesk.co.jp/products/3ds-max/overview/.
[retrieved: Nov. 2021]
[34] Unity3D. https://unity.com/. [retrieved: Nov. 2021]
[35] Maximo. https://www.mixamo.com/. [retrieved: Nov. 2021]
[36] Animator Controller. https://docs.unity3d.com/Manual/class-AnimatorController.html. [retrieved: Nov. 2021]
[37] Vuforia Engine. https://developer.vuforia.com/.
[retrieved: Nov. 2021]
[38] AI-powered 3D Avatars, Avatar SDK. https://avatarsdk.com/.
[39] Body Data Platform, 3D LOOK.
https://3dlook.me. [retrieved: Nov. 2021]
[40] Kinect. https://en.wikipedia.org/wiki/Kinect.
[retrieved: Nov. 2021]
[41] CinemaMocap.
https://assetstore.unity.com/packages/tools/animation/cinema-mocap2-markerless-motion-capture-56576.
[retrieved: Nov. 2021]
[42] Maya. https://www.autodesk.co.jp/products/maya/overview.
[retrieved: Nov. 2021]
[43] ARCore. https://developers.google.com/ar.
[retrieved: Nov. 2021]
[44] D. Mashek, L. W. Cannaday, and J. P. Tangney. “Inclusion of community in self scale: A single-item pictorial measure of community connectedness,” Journal of Community Psychology, 35, 257–275, 2007.
DOI: https://doi.org/10.1002/jcop.20146
[45] SPSS. https://www.ibm.com/analytics/spss-statistics-software.
[retrieved: Nov. 2021]

Vibro-tactile Hazard Notification for Motorcyclists

Akimasa Suzuki

Dept. of Information Science
Iwate prefectural University
020-0693, 152-52, Sugo, Takizawa, Iwate, Japan
Email:suzuki_a@iwate-pu.ac.jp

Yuta Yamauchi

Ad-Sol Nissin Corporation
108-0075, Rivage Shinagawa, 4-1-8,
Konan, Minato, Tokyo, Japan
Email:Yamauchi.Yuta@adniss.jp

***Abstract*—This paper evaluates the effectiveness of vibro-tactile notifications for motorcyclists under external factors. Although many car manufacturers provide side and rear collision warning systems with auditory or visual alarms, such notifications may confuse a motorcyclist because they already need to be aware of many visual targets, such as mirrors and monitors, and of environmental sounds. We propose a vibro-tactile notification system using a vibration speaker installed in a motorcycle helmet between the outer shell and the cushion. The proposed system should enable motorcyclists to correctly identify the directions of five vibrating motors, three levels of risk, and three obstacle types (i.e., pedestrians, vehicles, and motorcycles). We evaluate the system under windy and engine vibration conditions and examine notification accuracy via experiment. Our results indicate that motorcyclists can correctly detect four directions and three threat levels using this system.**

***Index Terms*—vibro-tactile notifications; helmet actuators; vibration speakers; motorcyclists; noise tolerance.**

I. Introduction

This paper is an extension of our paper initially presented at the Ninth International Conference on Advances in Vehicular Systems, Technologies and Applications [1]. In this paper, we conduct additional experiments on several kinds of vibration pattern, and further demonstrate the effectiveness of our proposed system.

There has been considerable research into accident prevention systems for vehicles, particularly in relation to developing driving support systems that will transition to autonomous driving systems. Autonomous vehicles, such as Tesla [2], have achieved practical four-wheel vehicles with Level 2 autopiloting. Furthermore, the vehicles are collecting training data for realization of Level 4 autopiloting. However, to realize autonomous driving systems, we must overcome problems related to cyber-security measures and traffic laws (e.g., responsibility for accidents by autonomous cars [3]), which could take time. Additionally, as many people enjoy driving, the demand for manual driving as a hobby is unlikely to fade. Driving support systems will thus remain an important feature. Moreover, despite the high number of driving support systems used in Japan, many fatal vehicle accidents are caused by violations of safe driving practices, such as failing to keep eyes on the road, careless driving, and failing to make safety checks [4]. This highlights the need to develop more techniques that support drivers. Of the 288,995 road accidents in 2020 in Japan, 208,983 were caused by the violation of safe driving practices.

A support system for motorcycles is more important for driving protection than for other vehicles because when a motorcyclist is in a crash, they are more likely to die, owing to the reduce protection from the vehicle. Hazard notification is vital for motorcyclists because of their limited visibility, the diverse sounds they may hear, and their very high risk of accident. The fatality rate in crashes for motorcyclists is 1.22%, while it is only 0.35% for drivers of four-wheel vehicles [5]. Furthermore, motorcycles are small and difficult for other drivers to recognize. Motorcyclists therefore need to be highly aware of their surroundings, but this is difficult because of the blind spots caused by their helmet and small mirrors. To avoid incidents, an intuitive notification system that can specify direction and threat level would be beneficial. Therefore, we propose a system that uses haptic sensations to quickly notify a motorcyclist of possible hazards or obstacles around the vehicle.

Our proposed hazard notification system uses vibro-tactile actuators installed in a motorcycle helmet. The system notifications indicate the type of object, direction, and threat level surrounding the vehicle. We evaluate robustness against wind and motorcycle engine vibration. We also perform experiments to test the effectiveness of our proposed system. Section II reviews related research. Section III describes system architecture considerations. Section IV presents the system architecture. SectionV discusses the vibration intensity issue. Section VI presents the experimental methods and results under the influence of motorcycle engine vibration. Section VII describes the experiment methods and results under the influence of driving wind. Section VIII gives the conclusions.

II. Previous and related works

For vehicles such as four-wheel vehicles, many practical driving support systems employ image [6], radar [7], and ultrasonic sensors [8] to detect pedestrians and other vehicles with high accuracy. Around-view monitors are increasingly being used for automatic parking [9] and lane-detection systems are being applied using three-dimensional (3-D) laser imaging detection and ranging (LIDAR) [10]. Despite their

weakness to noise, ultrasonic sensors can now be installed in driving support systems for a low cost, and the cost of 3-D LIDAR is also dropping. These devices can be used to detect not only the presence of an obstacle, but also the type of obstacle (e.g., pedestrian, vehicle, or motorcycle). In the current study, we focus on creating a notification method to alert drivers to potential hazards, rather than the development of a sensor system.

For preventing accidents on motorcycles, there are two main approaches: motorcyclists are assisted in checking their surroundings or drivers around a motorcycle are assisted in recognizing motorcyclist locations. For the latter approach, a helmet with brake lights has been investigated for practical use [11]. However, we focus on the former approach in this study. Hazard notification systems have been proposed for four-wheel vehicles [12] [13].

In addition, sensor systems, such as collision [14] or ground [15] detection methods, have been investigated for detection around a motorcycle. Many systems have focused on the sensor, however, we aim to not only sense issues, but also method for providing information more correctly to the motorcyclist. One study [16] proposed a smart helmet using a multimedia Internet of Things (IoT) sensor device and visual notifications. Many conventional notification methods for motorcycles rely on visual images in the motorcyclist's view [17] [18], such as front view, mirror, tachometer, speedometer, navigation system, and indicators. Therefore, there is the potential that excessive visual information may instead impact the motorcyclist's capacity to adhere to safe driving practices.

We focus on vibro-tactile notification as an option for non-visual information. Vibro-tactile notification is a possible candidate for providing information because early motorcycles conveyed engine failure to riders via irregular vibration. Some systems vibrate a motorcycle's steering, but this approach is limited to notifications in front of or behind the motorcycle [19]. We previously proposed a system for four-wheel vehicles that used haptic sensations to quickly notify drivers of possible hazards or obstacles surrounding the vehicle [20]. We examined the system's robustness against different material types and layers used for the driving seat cushions of four-wheel vehicles [21]. In this study, we also discussed vibrating waveforms by using a real 4-wheel vehicle, and illuminated higher accuracy on notification by using category-deformed and normalized waves between each actuator than real and certain vibration waves, respectively. Moreover, we discussed notifying accuracies of our system from an experiment combined with conventional notification system with a display and a sound speaker. From this experiment, we also discussed whether our system is obstructive or useful to driving.

The motorcycle seat is vibrated by the engine, although this may make it more difficult to notify the rider through vibration. In this study, we perform experiments under wind and engine vibration conditions to consider the viability of a highly intuitive notification system for motorcycles.

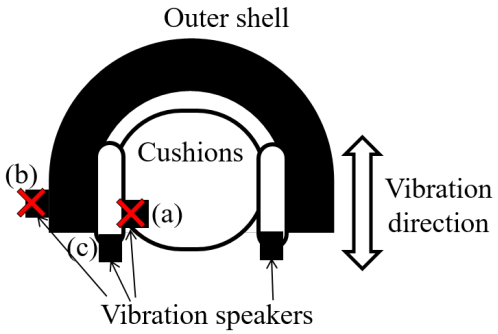


Figure 1. Proposed installation positions for actuators in a motorcycle helmet.

III. Mounting positions of actuators for motorcycles

In this section, we detail the system architecture.

A. Locations of vibro-tactile actuator for notifications

The proposed system aims to help motorcyclists correctly identify the directions of five vibrating motors, three intensity settings to indicate the level of risk, and three obstacle types (i.e., pedestrians, vehicles, and motorcycles). The position of vibration is important for giving critical notifications. We considered vibro-tactile actuators on the motorcyclist's arms, shoulders, or waist within a motorcyclist's suit or on the motorcyclist's head within a helmet. For locations on the body, the motorcyclist must use a wearable device. There are different types of suits for different motorcycles types (e.g., Cruiser or Sport) because the riding posture is different. Drivers may also experience limited mobility in a suit, which would impact accurate notification.

In contrast, helmets are usually fitted to the motorcyclist's head, and vibration positions are not affected by varying postures, although they are limited to facing forward. Helmets are also required in many countries. Therefore, we choose helmets for our proposed system. Figure 1 shows the possible installation positions within a motorcycle helmet. We cannot mount an actuator inside the helmet itself, as at (a), because direct contact with the driver's head would be unsafe. We considered the helmet's outer surface, as at (b), but our attempts showed that a very strong vibration would be needed. Therefore, we mount the actuators at the bottom of the cushion in the helmet, as at (c), so that the actuators vibrate vertically to each cushion.

We considered three types of vibration mechanisms: vibration motors, haptic reactors, and vibration speakers. Vibration motors can only produce sine waves and it is difficult to distinguish different categories, although they can achieve strong vibration. Although haptic reactors can realize a variety of vibrations, such as clicking, their vibration is too weak to produce notifications. We therefore propose a vibro-tactile notification system using a vibration speaker, which can realize the vibration with strong and varied expression. For our proposed system, we utilized a vibrating speaker with an Acoustic

TABLE I
SPECIFICATIONS OF THE ACOUSTIC HAPTIC™ACTUATOR

Shape inch method	$\varnothing 20x27.6$ mm
Weight	39.29 g
Acceleration	6.1 m/s^2
Rise time	34.5 msec (0 to 90 %)
Drop time	34.5 msec (100 to 10 %)
Resonance Frequency	55 Hz (700 g, 2.83 V)
Impedance	8 Ω
Inductance L_e	0.66 mH
Power consumption	0.144 W (700 g, 2.83 V, 55 Hz)

Haptic™actuator developed by Foster Electric Company. The actuator, whose specifications are shown in Table I, is a type of woofer that comes into direct contact with the driver's helmet. The resonance frequency is 55 Hz, which makes it possible to generate strong vibrations in the low frequency range. To achieve uniform signal strength in any waveform, it is important to focus on the resonance frequency.

B. Mounting positions of actuators on a helmet

Motorcycle helmets have shields that can reduce visibility. They may also become fogged with weather and temperature, which further limits the motorcyclist's visibility. Motorcyclists also have blind spots behind the motorcycle where the mirror's field of view does not reach. Therefore, we consider the need for both backward and forward alerts. Figure 2 shows the planned layout of actuators in the helmet. This helmet has four cushions (i.e., front, rear, right, and left), as shown in sections (1) to (4) in Figure 2. To explore the vibro-tactile directional sense at the motorcyclist's head, we installed eight actuators as shown in Figures 2 (a) to (h). We mounted actuators (a), (b), and (h) on cushion (1), (c) and (d) on cushion (2), (e) on cushion (3), and (f) and (g) on cushion (4).

We conducted an experiment to evaluate the resolution of vibration on the human head. By determining the resolution of human-perceivable locations, we aimed to determine which directions are identifiable. We performed the experiment with the engine in idle at 1500 ± 300 rpm, which is frequently used as a typical speed range, on a Yamaha MT-01 motorcycle equipped with a V-twin cylinder 1670 cc engine. We had five participants (students), aged between 19 and 22 years old. We conducted four trials with each participant. We randomly induced vibrations at each position with strengths ranging from -2 dB to -12 dB and participants estimated the position in the helmet.

Figure 3 shows the correct answer rates for direction using all eight or only four actuators (one on each cushion: (a), (c), (e), and (g) in Figure 2). The vertical and horizontal axes denote correct answer rates and installation positions, respectively, and the green and orange bars indicate results for respectively all eight or only four actuators. The results demonstrate that there was confusion when multiple actuators were installed on a single cushion, leading to lower accuracy.

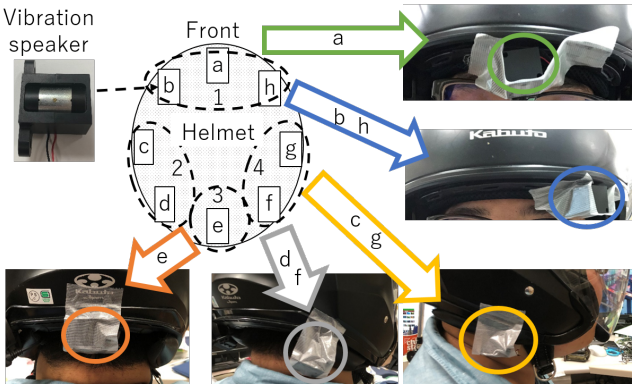


Figure 2. Experimental arrangement of mounted actuators.

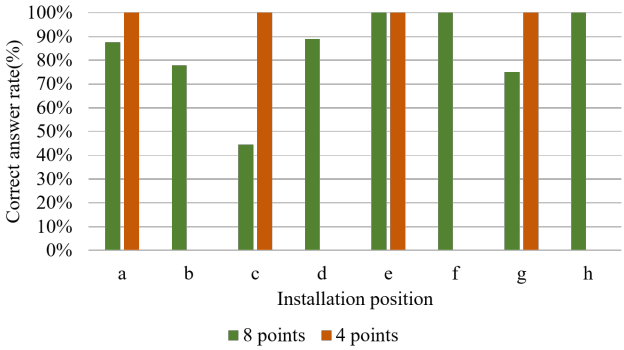


Figure 3. Correct answer rates according to the number and position of actuators.

By contrast, all participants had a 100% correct answer rate when four actuators were used. We therefore decided to utilize only one actuator per cushion.

IV. SYSTEM ARCHITECTURE FOR OUR PROPOSED SYSTEM

In line with the aforementioned experiment shown in Figures 2 and 3, we propose a vibro-tactile helmet as illustrated in Figure 4. Figure 4 (A) shows the positions of the actuators (or vibration speakers) on the helmet. Figure 4 (B) shows pictures of the different actuators in place; note that (b) and (d) are in the same position on the left and right sides of the helmet, so only one is shown.

The actuators shown in Figure 4 vibrate by transmitting sound data via the amplifiers. The sound data are deformed waves for three categories (i.e., pedestrian, four-wheel vehicle, and motorcycle) the same as in our previous study for four-wheel vehicles [21].

A. Waveforms for Categories

From our previous study [21], pedestrian notifications achieved higher correct answer rates than other obstacle types, owing to the more easily recognized notifications of the obstacle type than for other vibration waves. Thus, we aim to improve accuracy for the motorcycle and four-wheel vehicle obstacle types by making their waveforms more contrasting. Therefore, we utilize deformed waves for these obstacle types.

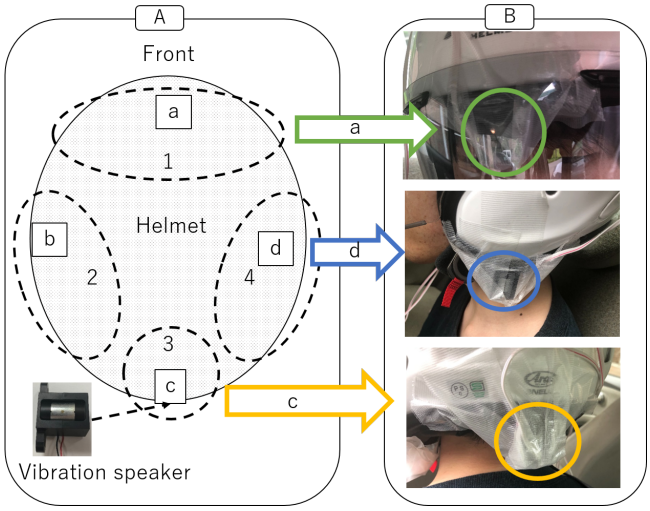


Figure 4. Overview of our proposed system.

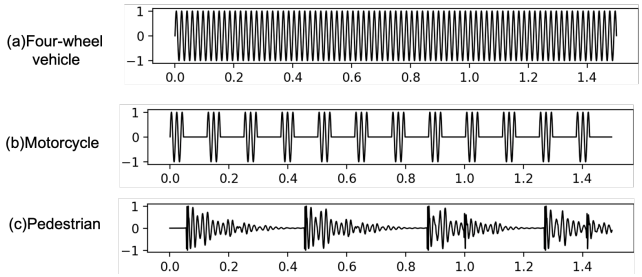


Figure 5. Waveforms for each category of object.

Figures 5(a), (b), and (c) show deformed waves based on the sound of a V6 engine revving up [22], the idling sound of a Suzuki GSX-S1000 motorcycle [23], and a real recording of footsteps with leather shoes [24], respectively. Note that, while a motorcycle is expressed by an intermittent deformed waveform, as in Figure 5(b), the interval is shorter than the waveform of the pedestrian shown in Figure 5(c). By contrast, the four-wheel vehicle is given a constant waveform to increase the contrast with the motorcycle waveform.

V. Vibration Strength

In this section, we discuss the strength of the vibrations for our proposed system.

A. Determination of optimum base sound level of each category

Our previous study in four-wheel vehicles [20] had three vibration strength levels (i.e., large, medium, and small) to indicate the level of risk. However, when riding a motorcycle, there is additional noise, such as from engine vibration or wind. Furthermore, we have not yet applied the vibration strength level concept to a helmet. Therefore, we consider the vibration strengths of the three levels.

We conducted an experiment with five participants aged between 19 and 22 years old, using an idle Yamaha MT-01.

TABLE II
Normalized vibration strength levels on Pattern B.

Category	large	medium	small
Four-wheel vehicle	−3 dB	−8 dB	−12 dB
Motorcycle	−2 dB	−8 dB	−12 dB
Pedestrian	−2 dB	−8 dB	−12 dB

The strength pattern is defined by the difference in the sound pressure. We considered two strength patterns. Pattern A has a small difference between the three levels. We can utilize a fourth level if we can detect the differences in Pattern A. Pattern B has larger differences than Pattern A, with the larger vibration for four-wheel vehicles adjusted so as not to prevent the motorcyclist from driving.

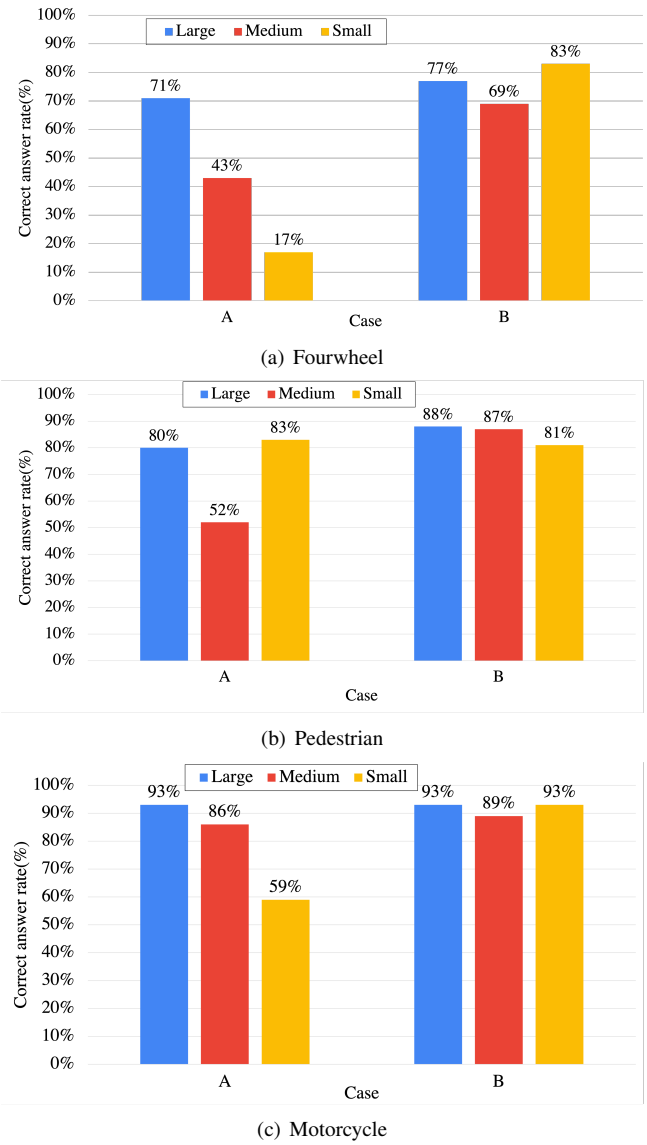
The strength levels of Pattern A, denoted large, medium, and small, were respectively −6 dB, −10 dB, and −12 dB from the original sound data based on the previous experiment [20], for all categories. The strength levels of Pattern B were different according to the category based on the normalization method [21], as shown in Table II.

We conducted this experiment using random directions, categories, and vibration strengths. Test participants answered ”large”, ”medium”, ”small”, or ”insensitive” as the levels of strength. In this experiment, we conducted the test in a stable room, and large, medium, and small strengths were randomly vibrated 51, 64, and 64 times, respectively, for 5 to 10 seconds. The experiment was conducted with the helmet shield closed. The test participants were taught three levels of vibration intensity (large, medium, and small) beforehand at actuator (d) in Figure 4, and then the experiment was conducted.

Figure 6 shows the experimental results for both patterns. The vertical and horizontal axes denote correct answer rates and strength levels, respectively. The blue and brown bars respectively indicate Patterns A and B. Pattern A had low correct answer rates for medium and small, and as well as instances of ”insensitive” shown as 0 percent in6(a). Pattern B had correct answer rates of over 69% for each strength level. Figure 6(b) and (c) show the percentage of correct answers for each pattern A and B in the comparison experiments for pedestrians and motorcycles, respectively. Figure 6(b) shows that 80%, 52%, and 83% of the participants rated the vibration intensity as ”large”, ”medium”, and ”small”, respectively, for Pattern A. Figure 6(b) shows that the percentages of correct responses for large and medium were much higher for Pattern B than for Pattern A. Figure 6(c) also shows that the percentages of correct answers for medium and small were much higher for Pattern B than for Pattern A. Thus, we adopt a notification method of three strength levels with large intervals like Pattern B.

B. Normalization by head sense

In this study, we apply normalized and exaggerated waves to improve notification accuracy because many motorcyclists pointed out that the vibration strength felt uneven depending



(a) Fourwheel

(b) Pedestrian

(c) Motorcycle

Figure 6. Correct answer answer rates for strength patterns A and B.

on the installation position. Here, we normalize the vibration strength for the different parts of the head via an experiment on the motorcycle, in addition to normalization by category with Pattern B in Figure 6.

For normalization, we utilized three actuators, (a), (b), and (c), on the front, rear, and left cushions, respectively, in Figure 4. Actuator (d) in Figure 4 is considered to have the same tendency as actuator (b). First, via a questionnaire, we determined the maximum and minimum strengths motorcyclists can detect with no effort. The results indicated that the gap between the maximum and minimum strengths for actuator (b) was greater than that for the other positions. Thus, we used the maximum and minimum strengths that participants felt for (b) for all the actuators. We denoted the maximum and minimum strengths as large and small, respectively. We defined the medium strength level not as the midpoint value

TABLE III
VIBRATION STRENGTH LEVELS AT EACH POSITION.

position	large	medium	small
a	−6 dB	−10 dB	−14 dB
b	−8 dB	−12 dB	−19 dB
c	−6 dB	−10 dB	−13 dB

(in decibels) between large and small, but instead according to how participants identified “medium” between large and small vibrations. In this normalization, we utilized the vibration of the four-wheel vehicle category [21]. Finally, we adjusted for strength as shown in Table III, which presents the strength levels for each actuator position, noting that the side position is more sensitive than the front and back positions.

VI. Experiment of Idling Scenario

We conducted an experiment to verify the correct answer rates when using actuators mounted on a helmet when there are external factors, including wind and engine noise.

A. Experimental Trials by t-test

We used t-tests to determine statistically significant results. We conducted an independent t-test for each strength level on actuators (a), (b), and (c) to compare the differences in correct answer rates under different wind and idling noises. We adopted a significance level of 5%, a moderate effect size of 0.5, and a detection rate of 80%. Possible answers were ”large”, ”medium”, ”small”, and ”insensitive”. Sample sizes for the answer of ”large” and the other strength levels were determined from one-sided and two-sided tests, respectively. From the t-test, sample sizes from which we could obtain a significant difference were 51, 64, and 64 samples for ”large”, ”medium”, and ”small” answers, respectively.

B. Correct answer rates during idling

We conducted an experiment in the idling state to evaluate the robustness under engine vibration, as illustrated in Figure 7. Six test participants on the motorcycle answered when they felt a vibration. This experiment was conducted with the engine off, at 0 rpm, and rotating at 1000 rpm, 1500 rpm, and 2000 rpm. In the case of 0 rpm, the experiment was considered a situation of engine stopped. For each engine speed, including 0 rpm, we performed 51, 64, and 64 trials for large, medium, and small, respectively, at random. The vibration categories used were those applied to four-wheel vehicles in our previous study [21].

Figure 8 shows the correct answer rates during idling rpm at 0 rpm, 1000 rpm, 1500 rpm, and 2000 rpm, respectively. The blue, red, and yellow bars respectively indicate ”large”, ”medium”, and ”small” as answered by the participants. The vertical and horizontal axes of each figure denote the correct answer rate and the three levels of signal strength, respectively. For example, 10% of participants answered ”medium” for the ”large” strength level given by an actuator at 1000 rpm, as shown in Figure 8(b).

Figure 7. Overview of the experiment during idling.

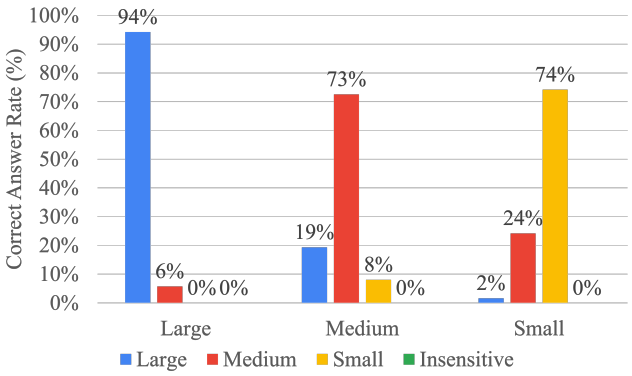
We defined the correct answer rate as the percentage of matches between the answers of participants and the actual level of vibration strength. For example, in Figure 8(b), the correct answer rates for ”large”, ”medium”, and ”small” were 90%, 72%, and 83%, respectively. Let us focus on the medium strength level in Figure 8(b). We confirmed that participants experienced stronger vibration at this level because they answered ”large” more often than ”small”. The correct answer rates were lowest in the case of Figure 8(d). This may be due to the high engine rotation causing stronger vibration and noise from the motorcycle, obscuring the vibration from the actuators.

VII. Evaluation of Driving Scenario

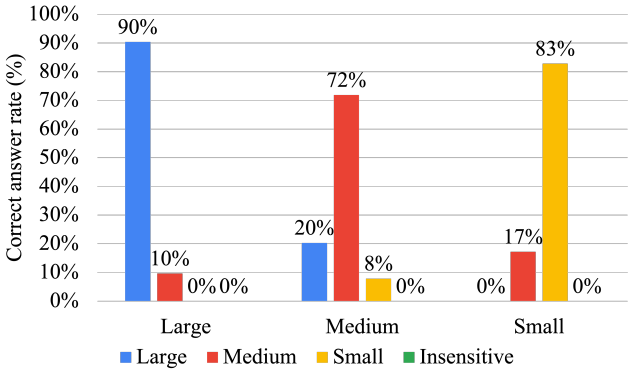
We conducted an experiment with six participants using a car to evaluate accuracy degradation due to wind. We used a car for safety and because of the difficulty in collecting answers. Although the strength of the traveling wind is little different between cars and motorcycles, the effect of the wind can be measured. We compared the wind noise between the car and the motorcycle, and found that the wind noise on the motorcycle was almost the same as on a car with all windows open.

A. Experiment on the vibration categories of a four-wheel vehicle

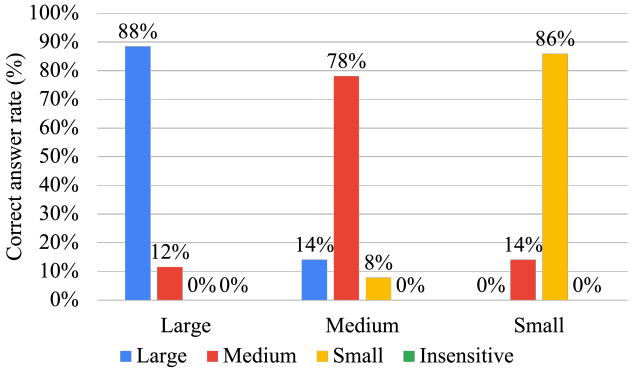
First, we experimented with the vibration categories of a four-wheel vehicle. Each participant performed the experiment at the four speeds of 0 km/h, 60 km/h, 80 km/h, and 100 km/h. At each speed, we performed 51, 64, and 64 trials with large, medium, and small strength levels, respectively, at random. The actuator vibrated for 5-10 s at random for each trial. In the case of 0 km/h, we conducted the experiment in a situation with engine stopped, as in Section VI-B. In the other cases, we used a highway. Figure 9 shows the experimental highway route. This route has two lanes in each direction limited to 100 km/h, and three entrances and exits shown



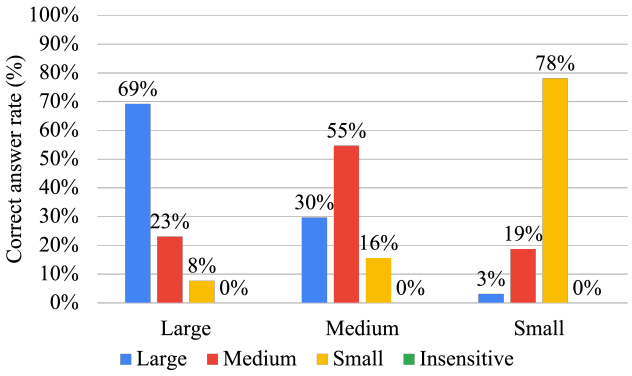
(a) 0 rpm



(b) 1000 rpm



(c) 1500 rpm



(d) 2000 rpm

Figure 8. Correct answer rates for different strength levels at (a) 0 rpm, (b) 1000 rpm, (c) 1500 rpm, and (d) 2000 rpm.

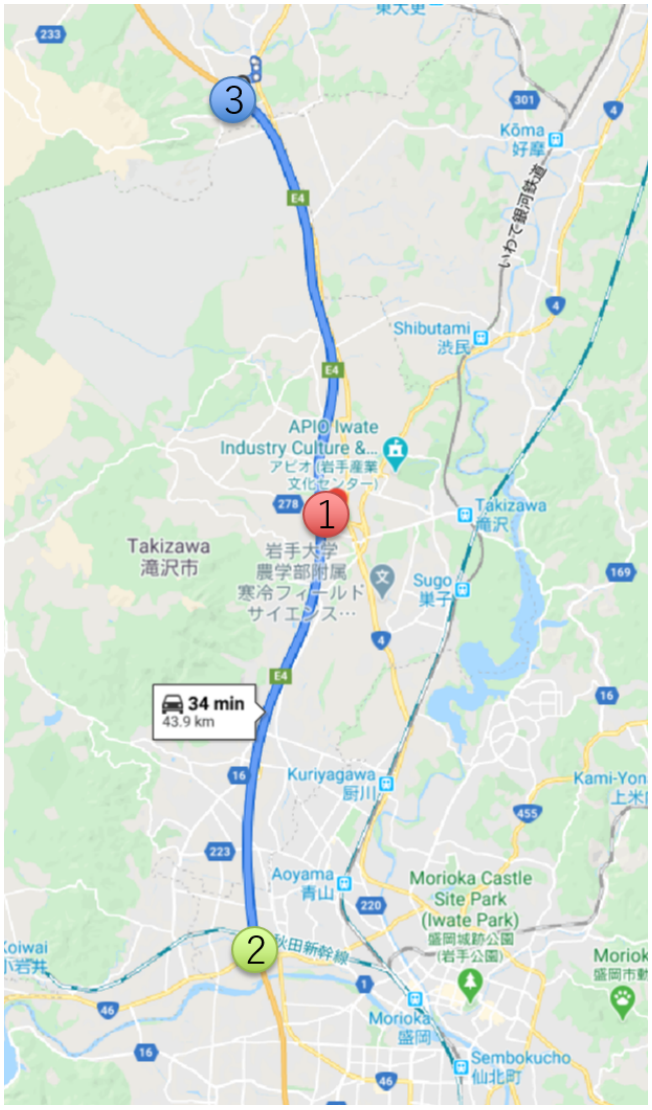


Figure 9. The course of the experiment on a highway.

Figure 10. Seating position of participants in the vehicle.

as ①, ②, and ③. We set up sections of 11.1 km between ① and ②, 20.3 km between ② and ③, and 9.2 km between ③ and ①.

The experiment was conducted at speeds of 80 km/h, 100 km/h, and 60 km/h between ① and ②, ② and ③, and ③ and ①, respectively. Figure 10 shows the seating positions of the participant in this experiment. All windows of the car were open and the helmet shields were closed. Before the experiment, participants were provided with examples of the three strength levels at position (b) of Figure 4. We limited the experiment to the four-wheel vehicle category. The experimental results were saved as movie files and evaluated via post-processing.

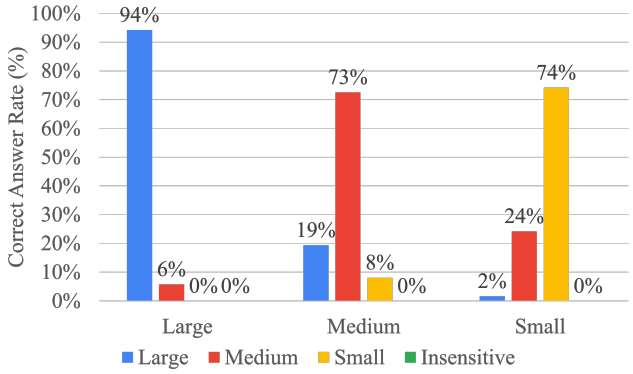
Figure 11 shows the correct answer rates at driving speeds of 0 km/h, 60 km/h, 80 km/h, and 100 km/h. The vertical and horizontal axes of each figure denote the correct answer rate and the three levels of signal strength, respectively. In Figure 11(a), the correct answer rates were in 94%, 73%, and 74% for large, medium, and small vibration strengths, respectively. Figure 11(b) indicates a high correct answer rate for large vibration. Test participants tended to be more likely to select strong vibrations. In the case of Figure 11(c), the correct answer rates were increased and decreased, respectively, for medium and large vibrations as compared to Figure 11(b). In Figure 11(d), a high correct answer rate was obtained even for small vibrations. From Figures 11(c) and (d), the medium strength level showed only small differences in the incorrect answer rate compared with large and small. Therefore, we consider the medium strength level to be appropriate in the high-speed case. We expect a higher notification accuracy can be achieved by adjusting the strength automatically depending on outside noise. We also found that notification accuracy was more degraded by engine rotation than by wind noise, which should be a consideration for practical implementation.

B. Experiment on the other categories

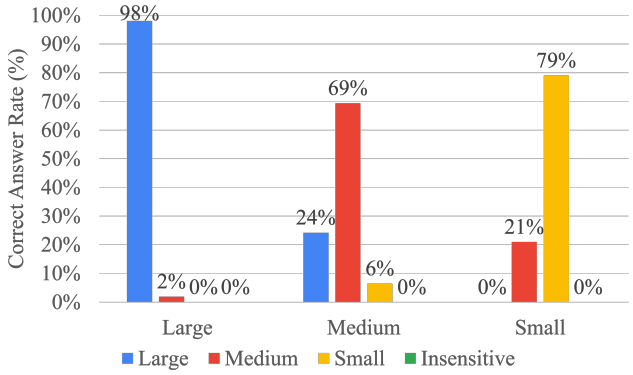
Next, we performed the experiment for other vibrations for motorcycle and pedestrian categories. The experimental settings were the same as Figures 9 and 10. However, each participant evaluated the experiment at the speeds of 0 km/h, 60 km/h, and 80 km/h, because the speed limit had been changed on the target highway by the time of the experiment.

Figures 12 and 13 show the correct answer rates in the vibration patterns for pedestrians and motorcycles, respectively, at the different driving speeds. In the case of pedestrians, the percentage of correct answers for medium and small increased as the wind speed increased from Figures 12(a) to (c). On the other hand, the percentage of correct answers for large was the lowest in Figure 12(c). A lower medium strength might be required, because the number of incorrect "large" answers was higher than the number of "small" answers for the medium vibration, as shown in Figure 12(a). However, Figures 12(b) and (c) show that the difference in the number of "large" and "small" answers in the evaluation of the medium vibration was slight. Thus, the medium intensity is suitable for 60 km/h and 80 km/h as vibrations for the pedestrian category.

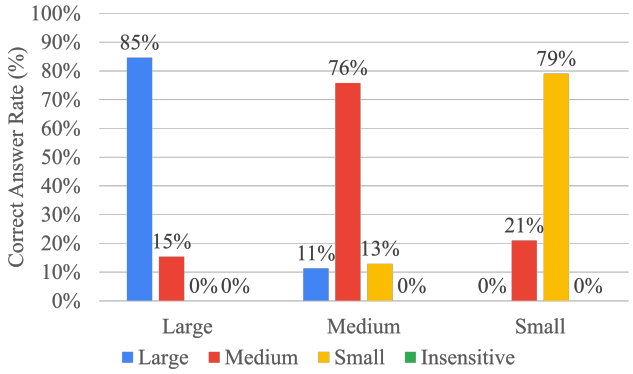
For the motorcycle category, the percentage of correct answers for small increased as the traveling wind became stronger, as shown in Figures 13(a), (b), and (c). These results



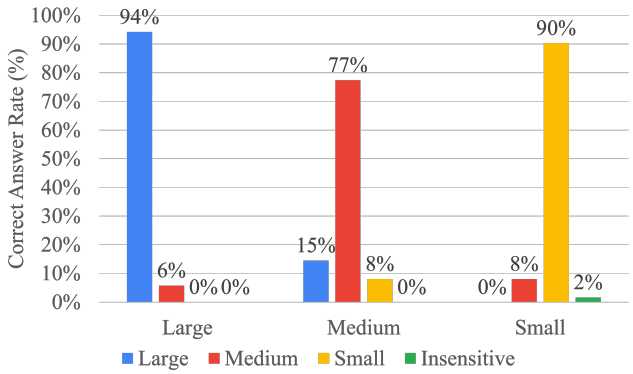
(a) 0 km/h



(b) 60 km/h

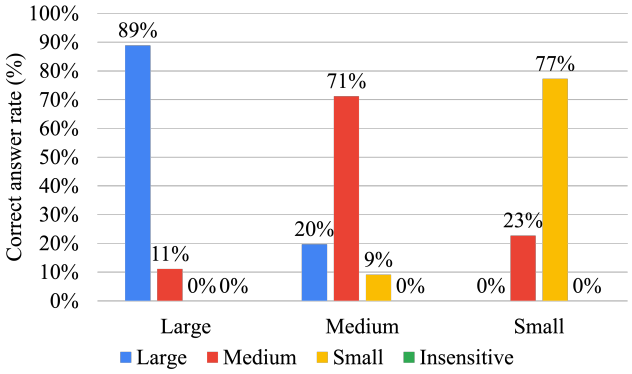


(c) 80 km/h

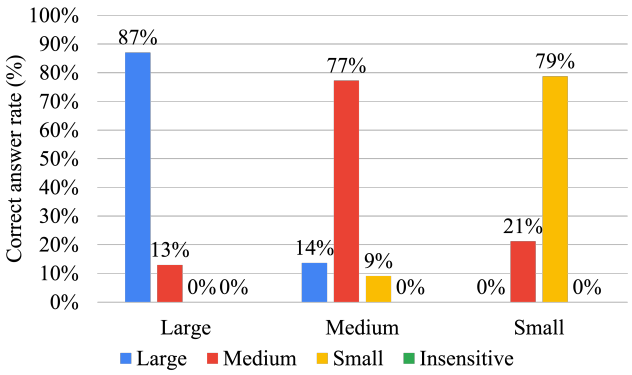


(d) 100 km/h

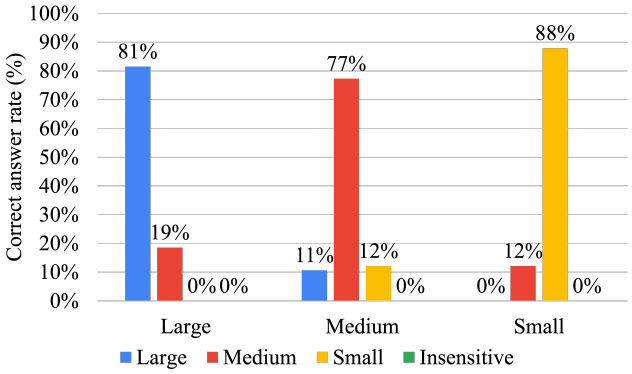
Figure 11. Correct answer rates for strength levels at each speed (four-wheel vehicle).



(a) 0 km/h



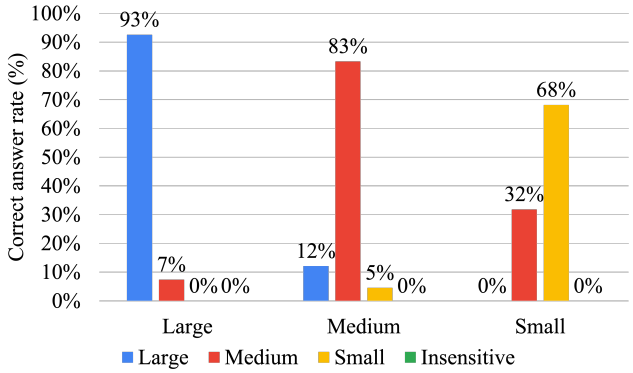
(b) 60 km/h



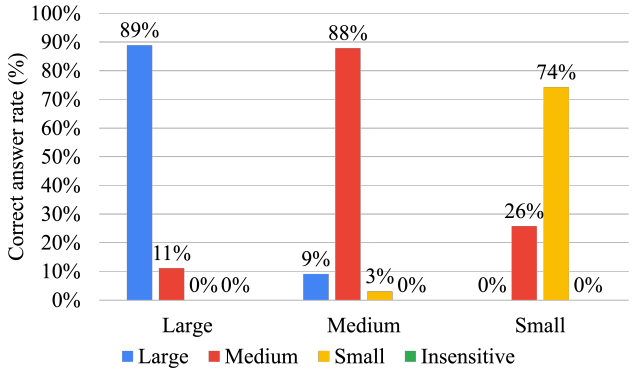
(c) 80 km/h

Figure 12. Correct answer rates for strength levels at each speed (pedestrian).

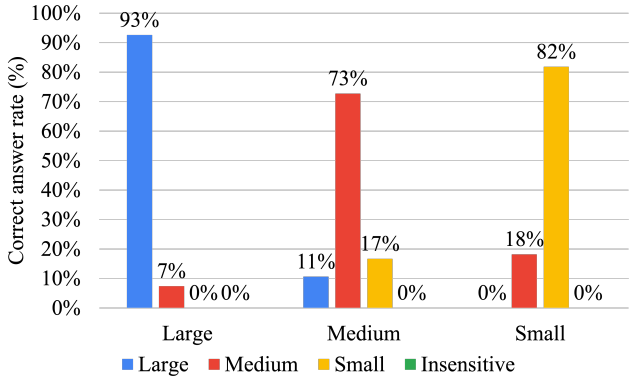
also show that the percentage of correct answers for large remained high in all speed ranges. Figure 13(c) has the lowest percentage of correct answers for medium. The difference in the number of ”large” and ”small” answers for the medium vibration was smaller than that for the small vibration in the case of Figures 13(a) and (b). Furthermore, in Figure 13(c), the difference between the number of ”large” and ”small” answers in the evaluation of the medium vibration was small, indicating that the medium intensity is suitable only for 80 km/h. We expect a higher notification accuracy can be achieved by adjusting the strength automatically depending on outside



(a) 0 km/h



(b) 60 km/h



(c) 80 km/h

Figure 13. Correct answer rates for strength levels at each speed (motorcycle).

noise, as in the case of a four-wheeled vehicle.

VIII. Conclusions

It is difficult for motorcyclists to recognize objects in their surroundings because of the many blind spots from their helmets and their smaller mirrors. Furthermore, accidents are more serious and have higher mortality rates for motorcycles than for four-wheel vehicles. We therefore proposed a notification system for motorcycles based on previous works for four-wheel vehicles. In our system, parts of the helmet vibrate corresponding to the direction of a hazard, the category of an object, and the level of risk. We used the strength of vibration to indicate three strength levels. We evaluated the accuracy of our proposed notification method for motorcycles using haptic actuators in windy and idling situations. We demonstrated the effectiveness of our notification method even for winds of 80 km/h on each waveform category. We expect improved notification accuracy can be achieved by adjusting vibration strength according to the motorcycle's speed. Various helmet types will be studied by quantitative assessment in future works.

Acknowledgment

This work was supported by MEXT KAKENHI Grant Number JP 16723884 and Foster Electric Company, Limited. We would like to thank Uni-edit (https://uni-edit.net/) for editing and proofreading this manuscript.

References

[1] Y. Yamauchi and A. Suzuki, "Vibro-tactile notification in different environments for motorcyclists," in The Ninth International Conference on Advances in Vehicular Systems, Technologies and Applications (VEHICULAR 2020), 2020.

[2] M. Dikmen and C. Burns, "Trust in autonomous vehicles: The case of tesla autopilot and summon," in 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2017, pp. 1093–1098.

[3] A. Hevelke and J. Nida-Rümelin, "Responsibility for crashes of autonomous vehicles: an ethical analysis," Science and Engineering Ethics, 2015.

[4] e-Stat (Japanese Government Statistics), "Statistics about road traffic / annual report," Feb. 2020, [accessed: 2021-07-13]. [Online]. Available: https://www.e-stat.go.jp/en/stat-search/files?page=1&layout=datalist&toukei=00130002&tstat=000001027457&cycle=7&year=20200&month=0

[5] ——, "Statistics about road traffic / annual report," Feb. 2018, [accessed: 2021-07-13]. [Online]. Available: https://www.e-stat.go.jp/en/stat-search/files?page=1&layout=datalist&toukei=00130002&tstat=000001027457&cycle=7&year=20180&month=0&stat_infid=000031800894

[6] Subaru of America, "Subaru eyesight," Jun. 2019, [accessed: 2021-07-13]. [Online]. Available: https://www.subaru.com/engineering/eyesight.html

[7] B. Major, D. Fontijne, A. Ansari, R. Teja Sukhavasi, R. Gowaikar, M. Hamilton, S. Lee, S. Grzechnik, and S. Subramanian, "Vehicle detection with automotive radar using deep learning on range-azimuth-doppler tensors," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, Oct. 2019.

[8] L. Alonso, V. Milanés, C. Torre-Ferrero, J. Godoy, J. P. Oria, and T. De Pedro, "Ultrasonic sensors in urban traffic driving-aid systems," Sensors, vol. 11, no. 1, 2011, pp. 661–673. [Online]. Available: https://www.mdpi.com/1424-8220/11/1/661

[9] J. K. Suhr and H. G. Jung, "Fully-automatic recognition of various parking slot markings in around view monitor (avm) image sequences," in 2012 15th International IEEE Conference on Intelligent Transportation Systems, 2012, pp. 1294–1299.

[10] Q. Li, L. Chen, M. Li, S.-L. Shaw, and A. Nüchter, "A sensor-fusion drivable-region and lane-detection system for autonomous vehicle navigation in challenging road scenarios," IEEE Transactions on Vehicular Technology, vol. 63, no. 2, 2014, pp. 540–555.

[11] Indiegogo, "Brake free helmet light. digital image. indiegogo," 2020, [accessed: 2021-07-13]. [Online]. Available: https://www.indiegogo.com/projects/brakefree-the-smart-brake-light-for-motorcyclists#/

[12] W. Chang, W. Hwang, and Y. G. Ji, "Haptic seat interfaces for driver information and warning systems," International Journal of Human-Computer Interaction, vol. 27, no. 12, Dec. 2011, pp. 1119–1132. [Online]. Available: https://doi.org/10.1080/10447318.2011.555321

[13] Y. Gaffary and A. Lécuyer, "The use of haptic and tactile information in the car to improve driving safety: A review of current technologies," Frontiers in ICT, vol. 5, Mar. 2018, p. 5. [Online]. Available: https://doi.org/10.3389/fict.2018.00005

[14] M. Muzammel, M. Z. Yusoff, and F. Meriaudeau, "Rear-end vision-based collision detection system for motorcyclists," Journal of Electronic Imaging, vol. 26, no. 3, May. 2017, p. 033002. [Online]. Available: https://doi.org/10.1117/1.jei.26.3.033002

[15] G. Gil, G. Savino, S. Piantini, and M. Pierini, "Motorcycles that see: Multifocal stereo vision sensor for advanced safety systems in tilting vehicles," Sensors, vol. 18, no. 2, Jan. 2018, p. 295. [Online]. Available: https://doi.org/10.3390/s18010295

[16] M. I. Tayag and M. E. A. D. V. Capuno, "Smart motorcycle helmet: Real-time crash detection with emergency notification, tracker and anti-theft system using internet-of-things cloud based technology," International Journal of Computer Science and Information Technology, vol. 11, no. 03, Jun. 2019, pp. 81–94. [Online]. Available: https://doi.org/10.5121/ijcsit.2019.11307

[17] V. H. Tran, D. Muscat, J. R. Cruz, A. M. Carrillo, D. A. A. Alonso, and E. Chavez, "Motorcycle and helmet providing advance driver assistance," U.S. Patent Application No. 13/897,570, 2013.

[18] Crosshelmet, "Cross helmet x1 - transform your riding experience," 2020, [accessed: 2021-07-13]. [Online]. Available: https://crosshelmet.com/

[19] C. Inc., "Blind spot detection & lane change assist," 2020, [accessed: 2021-07-13]. [Online]. Available: https://www.continental-automotive.com/en-gl/2-Wheeler/Safe-Mobility/ARAS/Blind-Spot-Detection

[20] A. Suzuki, Y. Murata, and M. Hayashi, "Notification of hazards around a vehicle using seat actuators," in 25th ITS World Congress. Copenhagen, Denmark: ERTICO, Sep. 2018, pp. AP–TP1335.

[21] A. Suzuki, K. Horie, S. Otobe, Y. Murata, and S. Fujimura, "Hazard notifications around a vehicle using seat actuators," The International Journal on Advances in Intelligent Systems, vol. 13, no. 1 and 2, 2020, pp. 59–68, iSBN:978-1-61208-720-7,ISSN:1942-2679.

[22] GYANTARO, "Vellfire v6 3.5l 0-138km/h full acceleration test (alphard 30 series)," Jan. 2016, [accessed: 2021-07-13]. [Online]. Available: https://www.youtube.com/watch?v=WI4voS7YF6Y

[23] ahotakuro, "Suzuki gsx-s1000 normal idling sound, motorcycle," Dec. 2016, [accessed: 2021-07-13]. [Online]. Available: https://www.youtube.com/watch?v=c6pK2z9ANRk

[24] M. studio official Channel, "Recommended sound effects, human sound foot and leg, footstep, leather shoes 04," Apr. 2014, [accessed: 2021-07-13]. [Online]. Available: https://www.youtube.com/watch?v=A5fGVbsHOfI

Estimation of Floor Reaction Force During Walking Using Frequency Analysis

Shohei Hontama, Kyoko Shibata, Yoshio Inoue
Kochi University of Technology
Miyanokuchi 185, Tosayamada, Kami, Kochi, Japan
e-mail: hontama.kut@gmail.com
e-mail: shibata.kyoko@kochi-tech.ac.jp
e-mail: inoue.yoshio@kochi-tech.ac.jp

Hironobu Satoh
National Institute of Information and Communications Technology
Nukui-kitamachi, Koganei, Tokyo, 187-8795, Japan
e-mail: satoh.hironobu@nict.go.jp

Abstract - **This research group aims to develop a gait analysis system that can be easily used by individuals with little user burden. This research group has proposed a method to estimate the floor reaction force, which is an important parameter in analyzing walking, from the force balance equation using the acceleration measured by wearable sensors, thus obtaining its usefulness. Previously, we reduced the number of sensors from 15 for the whole body part to 5 selected for user burden, but the estimation accuracy decreased. Therefore, we consider how to reduce the number of sensors without sacrificing accuracy. In the proposed method, the acceleration of each part is discrete Fourier transformed and expressed in the frequency domain, a function representing the relationship of acceleration between each part is derived. This function is called the motion mode function, which is used to derive the acceleration in the frequency domain from the measured part to the unmeasured part. The inverse discrete Fourier transform was used to derive the acceleration in the time domain. As a result, the acceleration was estimated with high accuracy. In addition, in order to make the system more versatile, the acceleration of the unmeasured part is estimated using the average motion mode function, which is the average of the motion mode functions obtained from the acceleration of multiple trials by multiple persons and averaged for each order. The estimated acceleration was then used to estimate the composite floor reaction force in two directions. Correlation coefficients were used to check the accuracy of the estimates and demonstrate the usefulness of the proposed method.**

Keywords- Discrete Fourier transform; Gait analysis; Average motion mode function; Floor reaction force.

I. INTRODUCTION

Walking is one of the most familiar activities performed by most people. Gait analysis data are important information in clinical medicine and sports for evaluating the effectiveness of rehabilitation therapy and for teaching athletes to improve their normal function. Gait analysis is performed using the widely known optical motion capture (MC) and installed force plate [2]. These systems are capable of deriving floor reaction forces and joint moments, which allow for detailed and accurate analysis of walking [3]. However, these systems are very expensive and have a limited measurement range, so they are limited to use in facilities such as research institutes and hospitals. In addition, electromyography can be used to determine muscle tension, but surface electromyography can only measure a limited number of muscles, while needle electrodes can only be used in specialized institutions. One method for estimating exertional capacity during walking is to perform simulation analysis using a musculoskeletal model [4]. Joint moment and muscle tension can be calculated as the load on the intervertebral discs from the balance of forces, but it involves many assumptions and constraints due to the setting of unclear individual parameters. To address these problems, gait analysis systems using wearable sensors have been proposed. Lei Wang et al. [5] attached four wearable sensors to the lower limbs to estimate step length and left-right asymmetry. Shaghayegh Zihajehzadeh et al. [6] used a wearable sensor attached to the wrist to estimate the walking speed in the walking direction. However, these systems only estimate the most common walking parameters for gait analysis. Also, our research group has developed a wearable motion capture system that combines a sandal-type force plate, which is a wearable version of an installed force plate and embedded with a 3-axis force sensor with motion sensors in six parts throughout the body [7]. However, this system was still expensive, burdensome to the user, causing problems with walking. Moreover, to develop an inexpensive and user-friendly gait analysis system that can be used easily by individuals, this research group proposed a method that can easily estimate the left and right composite floor reaction forces (ground reaction forces), which are important parameters for analyzing walking, using a wearable inertial sensor. This method has no limitations on the measurement range and is inexpensive to use [8]. This method uses the balanced relationship between three kinds of forces acting on a human while walking: inertia, gravity, and floor reaction force. The inertial force of the whole body was derived from the acceleration of the 15 parts and the mass of each part, based on the idea that the whole body can be treated as 15 rigid bodies since the human body can be composed of 14 rigid bodies and the trunk can be divided into the upper and lower trunk at the lower end of the ribs, as described by Ae et al. [9]. The estimated floor reaction force derived from the proposed method showed waveforms comparable to those measured by the force plates. Therefore, to accurately estimate the floor reaction force, each inertial force can be derived from the acceleration measured from 15 parts of the body. In the previous paper [10], we thought that by reducing

the number of sensors used, it would be possible to realize a system with less user burden. we tried to derive the floor reaction force by substituting the acceleration of the unmeasured part with the acceleration of the other measured part, but the estimation accuracy decreased with the reduction of the number of sensors. Therefore, we consider how to reduce the number of sensors without sacrificing accuracy. As one method, we can reduce the number of sensors and the resulting loss of accuracy by selecting measured parts with a limited number of sensors and estimating the acceleration of unmeasured parts from the measured parts. In order to estimate with less accuracy degradation, the relationship between the motion of the measured and unmeasured parts must be understood. However, the quantitative description of the motion relationship in terms of time waveforms is not easy. Because of this, we considered the frequency domain because the waveform can be broken down into its components and analyzed quantitatively [11]. By considering walking as a periodic motion and expressing the acceleration in the frequency domain through a discrete Fourier transform of the acceleration of each part, the characteristics of each component were extracted when the acceleration was decomposed into its constituent components. From the results, the characteristics of the motion of each part were clarified. For the left and right limbs such as the thigh and upper arm, we used the average of the left and right accelerations to align the walking frequencies and make it easier to grasp the characteristics. Based on the results of the discrete Fourier transform, a function was derived by dividing the unmeasured part by the measured part for each order component in order to quantitatively express the relationship between the two parts. This function is called the motion mode function in this study, which quantitatively shows the relationship of acceleration between two parts of the target. By using the obtained motion mode function, the acceleration of the unmeasured part could be estimated from the measured part in the frequency domain. The acceleration in the frequency domain can be derived by multiplying the motion mode function by the acceleration in the frequency domain of the measured part. The acceleration data in the time domain can be obtained by inverse discrete Fourier transforming the acceleration expressed in the frequency domain. In the previous paper [1], we compared the estimated acceleration of the unmeasured part with the measured value of MC, using the correlation coefficient as a measure of the estimation accuracy to show the usefulness of the method in two directions.

In the aforementioned previous report, the motion mode function, which represents the relationship between the accelerations of two parts, was created for each subject. In the method using the motion mode function of one subject, if the walking does not change significantly from trial to trial, the deviation of the motion mode function becomes small, allowing the acceleration of the unmeasured part to be estimated accurately. In this paper, we propose an average motion mode function to enhance the versatility of the system. The average motion mode function is a function that is derived from the motion mode function for each trial of multiple subjects and averaged for each order. The purpose of this paper is to accurately estimate the composite floor reaction force in two directions with a minimum number of measured parts by using this average motion mode function. First, the accelerations at each site for multiple trials of multiple subjects are obtained from the MC to derive the average motion mode function. These are discrete Fourier transformed to extract the features that are important for acceleration estimation. The average motion mode function is derived using the result of the discrete Fourier transform. The next step is to estimate the acceleration of the unmeasured part using the average motion mode function. To estimate the acceleration of the unmeasured part, the acceleration of the unmeasured part in the frequency domain is obtained by multiplying the average motion mode function by the acceleration of the measured part in the frequency domain of the trial that was not used in deriving the average motion mode function. The acceleration in the time domain is estimated by inverse discrete Fourier transforming the acceleration of the unmeasured part in the frequency domain. The accuracy of acceleration estimation for unmeasured parts depends largely on where the measured part is selected. Therefore, the selection of the measured part is discussed. There are two methods for selecting the optimal part. The first method is to derive the correlation coefficient between the actual measured value of the unmeasured part and the value estimated using the average motion mode function to determine the optimal part based on the estimation accuracy. The second selection criterion is the ability to accurately estimate the acceleration of a large mass part, which is important for accurately estimating the floor reaction force. Taking these two points into consideration, we select the optimal site for the acceleration estimation method of the unmeasured part. Finally, the composite floor reaction force in two directions is derived using the acceleration of the unmeasured part estimated using the average motion mode function. Correlation coefficients are again used to examine the accuracy of acceleration and floor reaction force estimates.

This research group aims to develop a system using wearable sensors that can be easily used in daily life. However, compared to MC, current wearable inertial sensors are susceptible to noise and it is difficult to measure motion accurately. Therefore, in this paper, the acceleration data used in the analysis will be obtained using MC to examine the validity of the methodology. In this paper, the vertical direction and the walking direction are focused on because the floor reaction force in the left-right direction is small. The floor reaction force is derived as a composite of the floor reaction forces applied to the left and right feet, because it is derived from the balance between the inertia of the whole body and the sum of gravity.

II. METHOD

A. Method for obtaining acceleration data used in Fourier analysis

The acceleration data used in the analysis is for two steps of one walking cycle, which is a steady-state walk from the

fifth and sixth steps after the beginning of the walk, and the walk is accompanied by a metronome in order to unify the steps. Subjects practice walking to the rhythm of a metronome sufficiently before taking measurements.

The recursive marker is attached at the center of gravity of each segment of the whole body, which is divided into 15 parts as shown in Figure 1.

Figure 1. Attaching position of recursive marker for MC

The division of the whole body is based on the idea that the human body can be composed of 14 rigid bodies and that the trunk can be divided into two parts at the lower end of the ribs, the upper trunk and the lower trunk, so that the whole body can be treated as 15 rigid bodies, as described by Ae et al. [9]. In this study, we use the average of the left and right accelerations for the left and right limbs, such as the thigh and the upper arm, referring to them as the average thigh and the average upper arm, respectively. As a result, 9 parts will be considered instead of 15 parts in the whole body. The acceleration data used in the analysis need to have equal values at both ends to improve the accuracy of the Discrete Fourier Transform results. This time, we will measure 15 trials for each subject and obtain acceleration data for each part.

B. Estimation of acceleration at the unmeasured part

The method of estimating the acceleration of an unmeasured part using the acceleration information of a measured part is described.

First, by performing a discrete Fourier transform of the acceleration data at each part and obtaining the magnitude and phase for each order, the magnitude of the acceleration at each part can be quantitatively evaluated for each order, and the dominant frequency and its band can be determined when estimating the acceleration. In this paper, we focus on the magnitude at which the motion characteristics of each part become prominent in the discrete Fourier transform results of each part. Next, the motion mode function used to estimate the acceleration of the unmeasured part from the measured part is obtained. The motion mode function can be derived by dividing the unmeasured part by the measured part for each order using the result of the discrete Fourier transform of each part. For example, when the upper trunk is the measured part and the foot is the unmeasured part, by dividing the foot by the upper trunk, the motion mode function that can estimate the acceleration of the foot from the upper trunk can be derived. In the previous report, we acquired walking data for 15 trials for each subject and used the motion mode function that averaged the motion mode functions of the data for 14 trials. The reason why one trial is omitted is that it is to be used for verification to evaluate the usefulness of the motion mode function. In this paper, we derive the motion mode function for each trial of multiple subjects and average them for each order to derive the average motion mode function. The number of motion mode functions is 8 for the remaining parts for each measured part and 9 for the possible measured parts, so a total of 72 functions per direction are derived for each subject. The average motion mode function, which is the average of the motion mode functions for each of these parts for each order, is derived as a function that reflects the acceleration relationship between the parts for all subjects. The acceleration of the unmeasured part is estimated in the frequency domain by multiplying the obtained average motion mode function by the result of the discrete Fourier transform of the measured part for the remaining one trial of the subject. For example, when the upper trunk is the measured part and the foot is the unmeasured part, the acceleration in the frequency domain of the foot can be derived by multiplying the average motion mode function of the upper trunk and foot by the acceleration in the frequency domain of the upper trunk. Finally, the acceleration in the time domain is obtained by inverse discrete Fourier transforming the acceleration of the unmeasured part in the frequency domain derived using the average motion mode function.

In this study, we used the average of the left and right acceleration values for the left and right limbs, such as the thighs and upper arms, so that the walking frequency of each part of the body can be aligned. Therefore, it is easier to understand the dominant frequencies and their bands when estimating the acceleration of unmeasured parts as well as the relationship between the head, trunk, and limb movements.

III. EXPERIMENTAL DESIGN

In the experiment, MC with 12 cameras (manufactured by Motion Analysis Co., Ltd., MAC 3D System), force plate 3 units (manufactured by Tec Gihan Co., Ltd., TF-6090-C 1 unit, TF-4060-D 2 units), and a metronome were used. The force plates are a strain gauge type transducer that can measure force, moment, pressure center, etc. The MC and the force plate are synchronized to accurately measure the ground contact timing in walking.

Five healthy subjects (male: age 21±2, height 1.70±0.10 [m], weight 65±5 [kg]) were explained the purpose and contents of this study and gave oral and written consent. In addition, this study was approved by the Ethical Review Committee.

10 steps are measured from the beginning of walking, and the force plates are placed at the 5th to 7th step from the beginning of walking. To measure the ground contact timing required by the discrete Fourier transform, three force plates are set up so that one foot contacts one force plate. The sampling frequency is set to 100 Hz, and low-pass processing with a cutoff frequency of 9 Hz is used for smoothing.

IV. RESULT&DISCUSSION

A. *Estimate the acceleration of the unmeasured part from the measured part*

In this section, we showed whether the acceleration of the non-measured part can be estimated accurately from the measured part using the average motion mode function derived from the five subjects who participated in this experiment.

The number of motion mode functions is 8 for the remaining parts for each measured part and 9 for the possible measured parts, so a total of 72 functions per direction are derived for each subject. However, due to the limitation of space, the acceleration estimation results for the case where the measured part is the upper trunk and the unmeasured parts are the average thigh and the average upper arm are shown as an example. In this paper, the subjects walked to a metronome at a pace of 100 BPM (one walking cycle is 0.6 seconds), and the acceleration of each part was taken out for two steps (1.2 seconds).

First, the measured accelerations in the time domain of the upper trunk, average thigh, and average upper arm of one subject obtained from the experiment were shown. Figure 2 shows the acceleration of the upper trunk, the average thigh, and the average upper arm in the vertical direction, and Figure 3 shows the acceleration of the upper trunk, the average thigh, and the average upper arm in the walking direction. Figure 2 shows that in the vertical direction, the upper trunk and average upper arm move similarly, but the average thigh moves with more high-frequency waves than the upper trunk and average upper arm. Figure 3 shows that even in the walking direction, the average thigh moves with more high-frequency waves than the upper trunk and average upper arm. The acceleration waveforms in the time domain shown in Figure 2 and Figure 3 indicated that it is difficult to quantitatively show the relationship of motion between each part of the body in both the vertical and walking directions and to estimate the acceleration of the average thigh and the average upper arm from the upper trunk simply by making corrections such as the constant correction. In addition, we found that it was difficult to estimate the acceleration of other parts of the body from one part, not only the upper trunk.

Next, the acceleration data is discrete Fourier transformed and decomposed into frequency components, and the magnitude and phase are determined for each direction. From the results, the dominant frequencies and their bands in estimating the acceleration of the unmeasured part were identified. The important components are the walking frequency (The walking frequency is 1.667 Hz, and the order of the graph is an integer.) and it's integer multiples while 0.5 times the walking frequency (The frequency of the 0. 5 times component is 0.833 Hz, and the order of the graph is the value of the half component of the integer.) and its integer multiples are not focused on in this report because they are indicators of the degree of left-right balance. Figure 4 shows the magnitude and phase in the vertical direction of the upper trunk, Figure 5 shows the average thigh, and Figure 6 shows the average upper arm. Figure 7 shows the magnitude and phase in the walking direction of the upper trunk, Figure 8 shows the average thigh, and Figure 9 shows the average upper arm. The magnitude and phase were shown as the average and standard deviation calculated from the walking data of five subjects of 15 trials each (for a total of 75 trials). The result of the discrete Fourier transform has a magnitude in the components up to 9 Hz because of the low-pass processing with a cutoff frequency of 9 Hz.

From Figure 4 to Figure 9, the amplitudes of the walking frequency and its integer multiple components (The order of the graph is an integer value) were large in both the vertical and traveling directions. Since the walking was controlled to some extent, except for the higher-order components of the average thigh, it can be seen that the movements of each part were similar from trial to trial and the standard deviation was small. In the vertical direction, the motion of the upper trunk was highly dependent on the walking frequency component, while the average thigh and the average upper arm contain not only the walking frequency component but also many higher-order components. For the average thigh shown in Figure 5, the reason for the large standard deviation in the third order or higher of the walking frequency component was that the walking of one subject differed greatly from trial to trial. In the direction of travel, the upper trunk and average upper arm were highly dependent on the walking frequency component, while the average thigh not only has a relatively large walking frequency component but also moves with many higher-order components. Therefore, the walking frequency and its integer multiple components are important for the acceleration estimation of the unmeasured part, and higher-order components are also necessary.

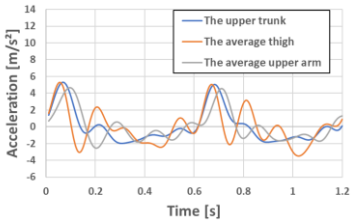


Figure 2. Acceleration data of upper trunk, average thigh, and average upper arm in the vertical direction. The blue line is the upper trunk acceleration, the orange line is the average thigh acceleration, and the gray line is the average upper arm acceleration. Acceleration data were measured by MC.

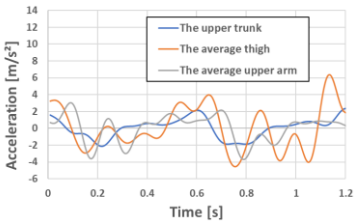


Figure 3. Acceleration data of upper trunk, average thigh, and average upper arm in the walking direction. The blue line is the upper trunk acceleration, the orange line is the average thigh acceleration, and the gray line is the average upper arm acceleration. Acceleration data was measured by MC.

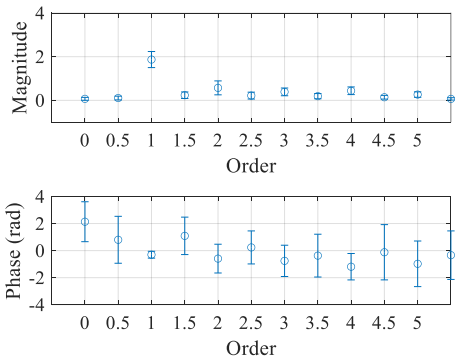


Figure 4. Result of the discrete Fourier transform of the upper trunk acceleration in the vertical direction. The magnitude and phase are shown as the average and standard deviation calculated from 75 walking data(5 subjects, 15 trials each). The upper trunk acceleration in the vertical direction is highly dependent on the walking frequency component.

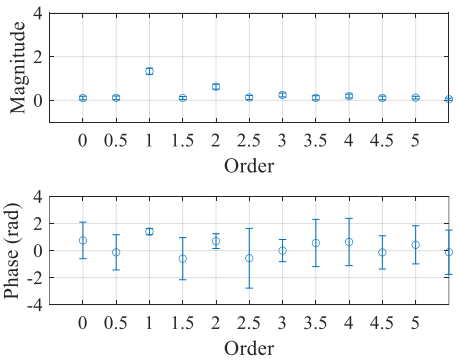


Figure 7. Result of the discrete Fourier transform of the upper trunk acceleration in the walking direction. The magnitude and phase are shown as the average and standard deviation calculated from 75 walking data(5 subjects, 15 trials each). The upper trunk acceleration in the walkingl direction is highly dependent on the walking frequency component.

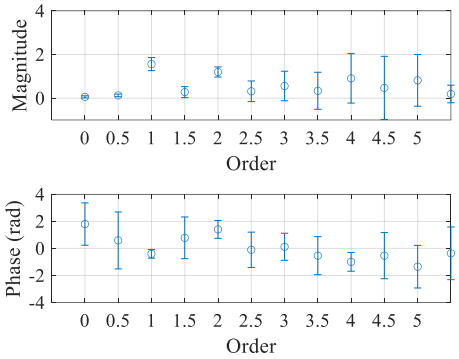


Figure 5. Result of the discrete Fourier transform of the average thigh acceleration in the vertical direction. The magnitude and phase are shown as the average and standard deviation calculated from 75 walking data(5 subjects, 15 trials each). The average thigh acceleration in the vertical direction contains not only the walking frequency component but also many higher-order components.

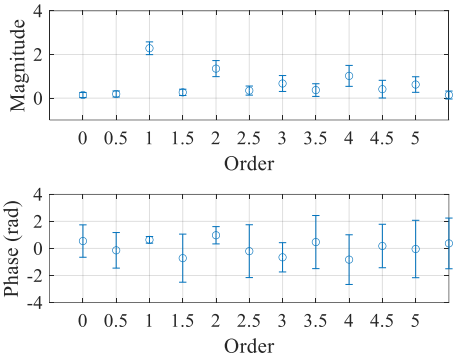


Figure 8. Result of the discrete Fourier transform of the average thigh acceleration in the walking direction. The magnitude and phase are shown as the average and standard deviation calculated from 75 walking data(5 subjects, 15 trials each). The average thigh acceleration in the walking direction contains not only the walking frequency component but also many higher-order components.

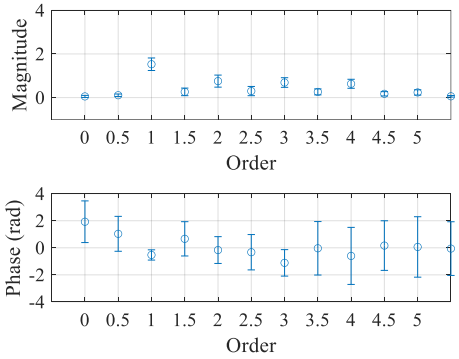


Figure 6. Result of the discrete Fourier transform of the average upper arm acceleration in the vertical direction. The magnitude and phase are shown as the average and standard deviation calculated from 75 walking data(5 subjects, 15 trials each). The average upper arm acceleration in the vertical direction contains not only the walking frequency component but also many higher-order components.

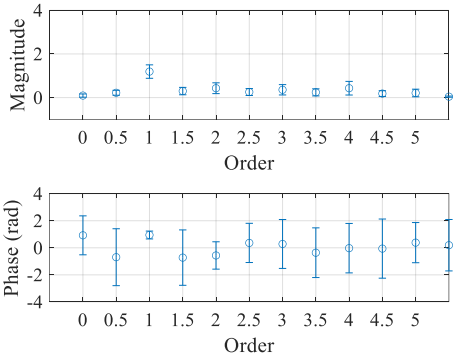


Figure 9. Result of the discrete Fourier transform of the average upper arm acceleration in the walking direction. The magnitude and phase are shown as the average and standard deviation calculated from 75 walking data(5 subjects, 15 trials each). The average upper arm acceleration in the walking direction is highly dependent on the walking frequency component.

Next, the average motion mode function is shown. As an example, the average motion mode function was derived with the upper trunk as the measured part and the average thigh and upper arm as the unmeasured parts. A motion mode function was derived for each of the 14 trials for each subject, and the average motion mode function was averaged for each order of the motion mode function for each trial for five subjects (Average of motion mode functions for 70 trials). Figure 10 shows the results for the average thigh with the upper trunk as the measured part, and Figure 11 shows the results for the average upper arm with the upper trunk as the measured part in the vertical direction. Figure 12 shows the results for the average thigh with the upper trunk as the measured part, and Figure 13 shows the results for the average upper arm with the upper trunk as the measured part in the walking direction. The gain and phase difference of the results were shown as the average and standard deviation calculated from the walking data of 70 trials.

In this paper, we focused on the walking frequency and its integer multiple components, which are important for estimating the acceleration of unmeasured parts. In the vertical direction, Figure 10 and Figure 11 show that the most important walking frequency component has a gain of near 1 and a phase difference near 0, indicating that the motion is similar in the vertical direction. For the integer multiple components larger than the quadratic component, the gain was larger than 1, and the average thigh and the average upper arm moved more than the upper trunk. In the walking direction, Figure 12 shows that the average thigh moves more than the upper trunk because the gain of the walking frequency component is near 2. Figure 13 shows that the upper trunk and average upper arm move similarly because the gain of the most important walking frequency component is near 1 and the phase difference is near 0. For the integer multiple components larger than the quadratic component, the gain was larger than 1 except for the quadratic components of the upper trunk and average upper arm, indicating that the average thigh and average upper arm moved more than the upper trunk. As for the quadratic components of the upper trunk and the average upper arm, Figure 9 shows that the gain of the average upper arm is smaller than 1 due to its smaller magnitude in the integer multiple components of the walking frequency component.
From Figure 10 to Figure 13, we consider that the reason for the large standard deviation in the integer components of the walking frequency is that the magnitude of each part is small in the integer components above the second-order component of the upper trunk. The large standard deviation in the integer multiple components of the walking frequency may affect the error between the measured and estimated values in high-frequency waves when the acceleration at the unmeasured part is estimated.

Next, the acceleration in the time domain of the unmeasured part is shown. As an example, the average thigh and upper arm accelerations were estimated from the upper trunk acceleration of trial 15 for one subject using the average motion mode function obtained from 70 trials of walking data shown in Figure 10 to Figure 13. The acceleration of the average thigh and the average upper arm was estimated in the frequency domain using a total of five points of the walking frequency and its integer multiple components of the average kinematic mode function and an inverse discrete Fourier transform is performed to obtain the acceleration data in the time domain.

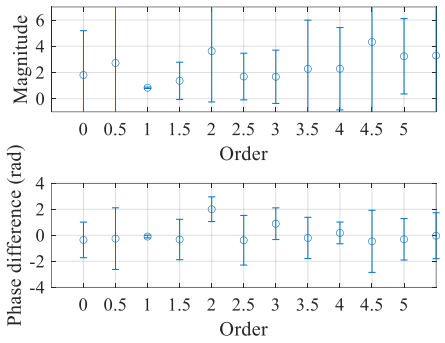


Figure 10. The average motion mode function in the vertical direction with the input as upper trunk and the output as average thigh. The magnitude and phase difference are shown as the average and standard deviation calculated from 70 walking data(5 subjects, 14 trials each).

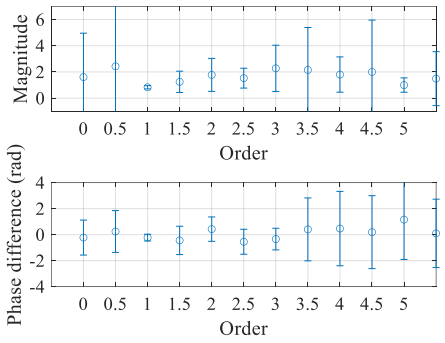


Figure 11. The average motion mode function in the vertical direction with the input as upper trunk and the output as average upper arm. The magnitude and phase difference are shown as the average and standard deviation calculated from 70 walking data(5 subjects, 14 trials each).

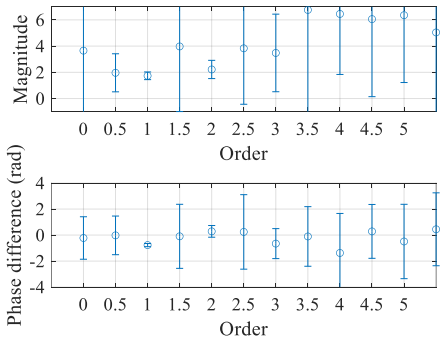


Figure 12. The average motion mode function in the walking direction with the input as upper trunk and the output as average thigh. The magnitude and phase difference are shown as the average and standard deviation calculated from 70 walking data(5 subjects, 14 trials each).

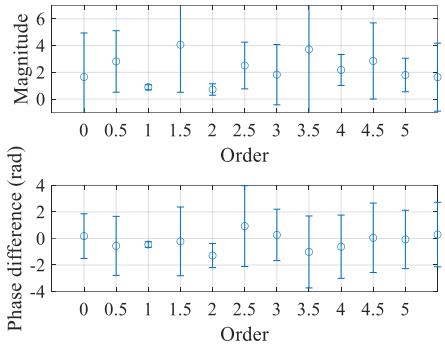


Figure 13. The average motion mode function in the walking direction with the input as upper trunk and the output as average upper arm. The magnitude and phase difference are shown as the average and standard deviation calculated from 70 walking data(5 subjects, 14 trials each).

The results of the comparison between the measured values of MC, the acceleration estimated using the average motion mode function, and the acceleration estimated from the motion mode function derived from 14 trials of one subject were shown in Figure 14 and Figure 15 for the vertical direction of the average thigh and the average upper arm, respectively, and in Figure 16 and Figure 17 for the walking direction of the average thigh and the average upper arm, respectively. In the method using the average motion mode function, the acceleration of the average thigh and the average upper arm can be estimated, using the measured part as the upper trunk, similar to the method using the motion mode function of one subject. The respective correlation coefficients between the measured values of MC and the acceleration estimated using the average motion mode function, and between the measured values of MC and the acceleration estimated from the motion mode function of one subject are shown in Table 1. Compare the correlation coefficients when using the average motion mode function and when using the motion mode function of one subject. In the vertical direction and the walking direction, the motion mode function of one subject is more accurately estimated in both cases where the unmeasured part is the average thigh and the average upper arm. This is because the deviation of the motion mode function becomes small in the method using the motion mode function of one subject, if the walking does not change significantly from trial to trial, and the acceleration can be estimated accurately. However, in the vertical direction, the correlation coefficients were high even when the average operating mode function was used, and accuracy was ensured. Figure 16 and Figure 17 show that in the walking direction, the average thigh captures the characteristics of the waveform, while the average upper arm cannot reproduce the higher-order components, but can estimate the phase and amplitude of the lower-order components.

The results on the accuracy of the acceleration estimation of the unmeasured part when the measured part is the other part are described in the next section.

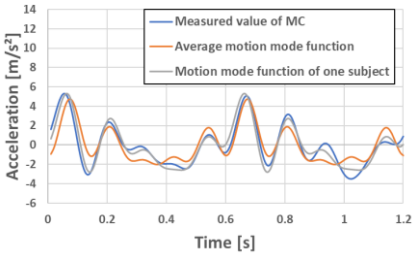


Figure 14. Comparison of acceleration estimated using measured value of MC and the average motion mode function, and acceleration estimated from the motion mode function derived from 14 trials of one subject, using the upper trunk as the measured part and the average thigh as the unmeasured part in the vertical direction.

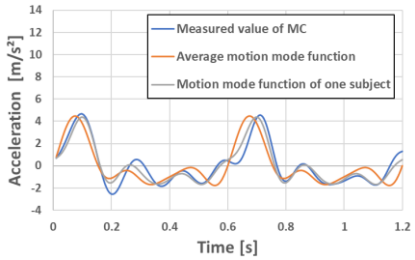


Figure 15. Comparison of acceleration estimated using measured value of MC and the average motion mode function, and acceleration estimated from the motion mode function derived from 14 trials of one subject, using the upper trunk as the measured part and the average upper arm as the unmeasured part in the vertical direction.

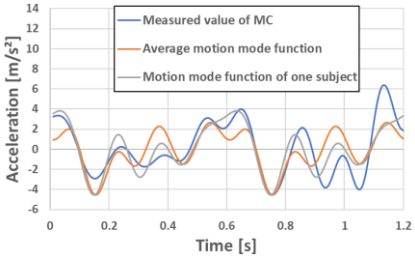


Figure 16. Comparison of acceleration estimated using measured value of MC and the average motion mode function, and acceleration estimated from the motion mode function derived from 14 trials of one subject, using the upper trunk as the measurement part and the average thigh as the unmeasured part in the walking direction.

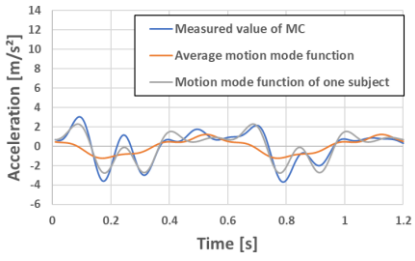


Figure 17. Comparison of acceleration estimated using measured value of MC and the average motion mode function, and acceleration estimated from the motion mode function derived from 14 trials of one subject, using the upper trunk as the measurement part and the average upper arm as the unmeasured part in the walking direction.

TABLE I. CORRELATION COEFFICIENTS OF THE ESTIMATED ACCELERATION OF THE UNMEASURED PART WITH MEASURED VALUES OF MC. CORRELATION COEFFICIENTS WITH MEASURED VALUE OF MC ARE SHOWN FOR THE CASE WHERE THE AVERAGE MOTION MODE FUNCTION OF FIVE PEOPLE IS USED AND THE CASE WHERE THE MOTION MODE FUNCTION OF ONE SUBJECT IS USED.

Vertical direction		
Function	Average thigh	Average upper arm
Average motion mode function	0.825	0.816
Motion mode function of one subject	0.962	0.959
Walking direction		
Function	Average thigh	Average upper arm
Average motion mode function	0.693	0.642
Motion mode function of one subject	0.848	0.926

B. *Selection of the optimal part for estimation of acceleration*

In this section, we consider which part of the body can be selected as the measured part to estimate the acceleration of the unmeasured part with good accuracy.

In the past, the research group has considered that accurate estimation of the acceleration at the part with large mass is important for accurate estimation of the floor reaction force because the higher the accuracy of the part with large mass, the higher the accuracy of the floor reaction force. Considering these factors, the optimal site was selected. First, the correlation coefficient between the measured values of MC and the values estimated from each measured part was derived. Next, the optimal part was selected from the derived correlation coefficients, focusing on the measured part that can accurately estimate the acceleration of the part with a large mass.

In this paper, the whole body was divided into three groups as a large block (A: head, upper trunk, lower trunk B: upper limbs C: lower limbs) and the part with the largest mass in each group is used as the measured part to estimate the acceleration of the unmeasured part. The mass of each body part was derived by multiplying the total body mass measured with a scale by the body part coefficients of Japanese athletes shown in Table 2, which were calculated by Ae et al. using the method of Yokoi et al. [12], who applied the model of Jensen et al. [13]. Body part coefficients for upper and lower limbs were shown as composite values for left and right.

The measurement sites were the upper trunk, the average upper arm, and the average thigh from each of the three groups, and the non-measurement sites were the upper trunk, the average thigh, the lower trunk, the average lower leg, and the head, which are the sites with large mass, respectively.

As an example, when the unmeasured part is the lower trunk, the estimated results from each measured part are shown in Figure 18 for the vertical direction and Figure 19 for the walking direction. In this paper, the average values of the correlation coefficients for each measured part were obtained for each of the two subjects, and are shown in Table 3 for the vertical direction and Table 4 for the walking direction.

TABLE II. BODY PART COEFFICIENTS OF JAPANESE ATHLETES(UPPER AND LOWER LIMBS ARE COMPOSITE VALUES OF LEFT AND RIGHT)

Group	Body part	Mass ration [%]
A	Head	6.9
	Upper tunk	30.2
	Lower trunk	18.7
B	Upper arm	5.4
	Forearm	3.2
	Hand	1.2
C	Thigh	22.0
	Lower thigh	10.2
	Foot	2.2

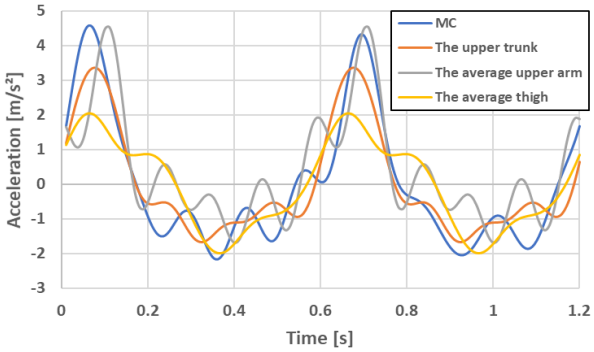


Figure 18. In the vertical direction, the unmeasured part is the lower trunk, and the measured value of MC are compared with the acceleration estimated with the upper trunk as the measured part, the acceleration estimated with the average upper arm as the measured part, and the acceleration estimated with the average thigh as the measured part.

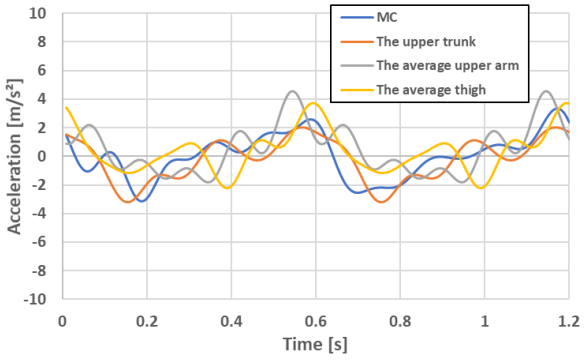


Figure 19. In the walking direction, the unmeasured part is the lower trunk, and the measured value of MC are compared with the acceleration estimated with the upper trunk as the measured part, the acceleration estimated with the average upper arm as the measured part, and the acceleration estimated with the average thigh as the measured part.

In Figure 18 and Figure 19, the characteristics of the measured waveforms in both the vertical and walking directions can be captured in terms of their magnitude and phase. To evaluate the results, the correlation coefficient was again used to quantitatively show the estimation accuracy. Table 3 shows that in the vertical direction, the average values of the correlation coefficients were the highest when the measured part was the upper trunk for subject A and the

average thigh for subject B, and the values were comparable. Considering that the higher accuracy of the estimation of the large mass part is more accurate when estimating the floor reaction force, we decided to select the upper trunk as the measured part, which shows a relatively high correlation coefficient when the unmeasured parts are the upper trunk, average thigh, and lower trunk in the vertical direction.

Table 4 shows that in the walking direction, the average values of the correlation coefficients were the highest when the measured part was the average thigh for subject A and the upper trunk for subject B, and the values were comparable. In the walking direction, the correlation coefficients were similar when the measured parts were the upper trunk, average thigh, and average upper arm. However, for the upper and lower limbs, the average acceleration of the left and right limbs is used, requiring measurements at two parts each. Therefore, the upper trunk was selected as the measured part in order to reduce the number of sensors and to improve the estimation accuracy.

TABLE III. CORRELATION COEFFICIENT BETWEEN THE MEASURED VALUES OF MC AND THE ESTIMATED VALUES FROM EACH MEASURED PART IN THE VERTICAL DIRECTION. IN CASE THE MEASURED AND UNMEASURED PARTS ARE THE SAME, THE CORRELATION COEFFICIENT IS EXPRESSED AS 1.

Subject A						
Measured part \ Unmeasured part	Uppert trunk	Average thigh	Lower trunk	Average lower thigh	Head	Average
Upper trunk	1	0.825	0.929	0.843	0.927	0.905
Average thigh	0.853	1	0.824	0.901	0.846	0.885
Average upper arm	0.900	0.668	0.828	0.545	0.949	0.778
Subject B						
Upper trunk	1	0.661	0.812	0.723	0.867	0.813
Average thigh	0.954	1	0.792	0.651	0.833	0.846
Average upper arm	0.919	0.627	0.880	0.498	0.961	0.777

TABLE IV. CORRELATION COEFFICIENT BETWEEN THE MEASURED VALUES OF MC AND THE ESTIMATED VALUES FROM EACH MEASURED PART IN THE WALKING DIRECTION. IN CASE THE MEASURED AND UNMEASURED PARTS ARE THE SAME, THE CORRELATION COEFFICIENT IS EXPRESSED AS 1.

Subject A						
Measured part \ Unmeasured part	Uppert trunk	Average thigh	Lower trunk	Average lower thigh	Head	Average
Upper trunk	1	0.693	0.781	0.260	0.170	0.581
Average thigh	0.676	1	0.600	0.214	0.552	0.608
Average upper arm	0.721	0.648	0.538	0.173	0.308	0.478
Subject B						
Upper trunk	1	0.736	0.942	0.376	0.188	0.648
Average thigh	0.740	1	0.655	0.430	0.250	0.615
Average upper arm	0.733	0.780	0.707	0.0650	0.100	0.477

V. APPLICATION TO ESTIMATION OF FLOOR REACTION FORCE

A. Floor reaction force estimation method using acceleration of each part

We use the fact that the sum of the inertia force derived from the acceleration measured from 15 parts of the body and the mass of each part, and the gravity force is balanced by the floor reaction force [8]. Equation (1) is an equation for estimating floor reaction forces in the vertical direction. Where F_z is the vertical floor reaction force, m is the mass of each body part, a_z is the vertical acceleration of each body part, i is the number of body parts, M is the total body mass, and g is the acceleration of gravity.

$$F_z = \sum_i m_i a_{zi} + Mg. \quad (1)$$

Equation (2) is an equation for estimating the floor reaction force in the walking direction. The floor reaction force in the walking direction can be derived by removing the gravity term from (1) and using the acceleration in the walking direction for the acceleration at each site. Where F_y is the floor reaction force in the walking direction and a_y is the acceleration in the walking direction.

$$F_y = \sum_i m_i a_{yi}. \quad (2)$$

In this paper, the number of parts i is nine because the average of the acceleration of the left and right limbs, such as the thigh and upper arm, is used. The mass of each body part is derived by multiplying the total body mass measured by a scale by the body part coefficients in Table 2.

B. Estimation Result

The floor reaction force was derived using the acceleration derived from the acceleration estimation method for the unmeasured part shown in Section Ⅳ.A. The floor reaction force in two directions was derived by estimating the acceleration of the remaining eight parts, using the trunk as the measured part, which was selected as the optimal part in the acceleration estimation method for unmeasured parts described in Section Ⅳ.B. The correlation coefficient was used to judge the accuracy of the estimation.

The results of the comparison of the floor reaction force derived from the force plate measurements and the average motion mode function, the floor reaction force derived from the motion mode function of one subject, and the floor reaction force derived using nine measured parts were shown in Figure 20 for the vertical direction and Figure 21 for the walking direction. The measured values from the force plates were smoothed by low-pass processing with a cutoff frequency of 9 Hz as well as the acceleration obtained from the sensor. From Figure 20 and Figure 21, one cycle of walking was shown, but one foot was grounded outside of the force plate until time 0.13[s], so the measurement of the

force plates indicated by the solid blue line did not measure the composite floor reaction force until time 0.13[s]. Table 5 showed the correlation coefficients between the measured values of the force plates and the floor reaction force derived using the average motion mode function, the floor reaction force derived using the motion mode function of one subject, and the floor reaction force derived using nine measured parts, respectively. Each correlation coefficient was derived using the data after 0.13[s].

TABLE V. CORRELATION COEFFICIENTS BETWEEN THE FLOOR REACTION FORCE MEASURED BY THE FORCE PLATE AND THE FLOOR REACTION FORCE DERIVED USING THE AVERAGE MOTION MODE FUNCTION, THE FLOOR REACTION FORCE DERIVED USING THE MOTION MODE FUNCTION OF ONE SUBJECT, AND THE FLOOR REACTION FORCE DERIVED USING NINE MEASUREMED PARTS IN THE VERTICAL DIRECTION AND THE WALKING DIRECTION.

Function	Vertical direction	Walking direction
Average motion mode function	0.934	0.887
Motion mode function of one subject	0.924	0.952
Nine measured parts	0.952	0.933

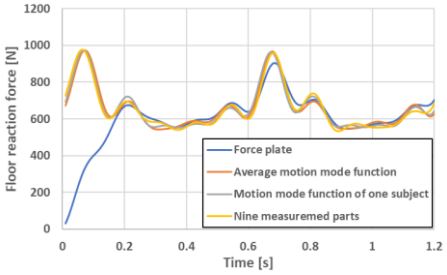


Figure 20. Comparison of the floor reaction force derived from the force plate measurements and the average motion mode function, the floor reaction force derived from the motion mode function of one subject, and the floor reaction force derived from nine measuremed parts in the vertical direction. The time up to 0.13 seconds is not included in the comparison because it is the point where the composite floor reaction force cannot be measured because one foot is grounded outside the force plate.

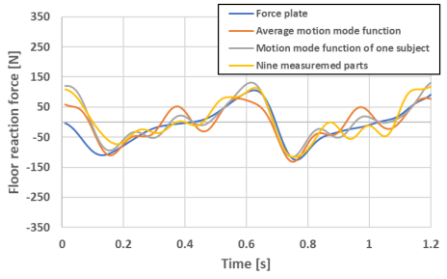


Figure 21. Comparison of the floor reaction force derived from the force plate measurements and the average motion mode function, the floor reaction force derived from the motion mode function of one subject, and the floor reaction force derived from nine measuremed parts in the walking direction. The time up to 0.13 seconds is not included in the comparison because it is the point where the composite floor reaction force cannot be measured because one foot is grounded outside the force plate.

Table 5 showed that in the vertical direction, the correlation coefficients of the method using the average motion mode function were almost equal to those of the method using the motion mode function of one subject and the method using nine measured parts, indicating that the method using the average motion mode function also ensured accuracy. Also, from Figure 20, the bimodal feature [14] that appears in the vertical floor reaction forces of one foot on each side in healthy walking appears as a trimodal feature in the composite vertical floor reaction forces presented. This result can be attributed to the fact that the acceleration of each unmeasured part was estimated accurately.

Table 5 showed that in the direction of travel, the correlation coefficients of the method using the average motion mode function and the method using nine measured parts were slightly lower than those of the method using the motion mode function of one subject. The method using the averaged motion mode function, shown by the solid orange line in Figure 21, included high-frequency waves, resulting in error from the target value measured by the force plates. This factor is thought to have been caused by a decrease in the estimation accuracy of the upper and lower limbs in the acceleration estimation of the unmeasured parts. However, the walking direction, which did not show good accuracy in the acceleration estimation, showed a much higher correlation coefficient and improved accuracy in the floor reaction force than in the acceleration. This can be attributed to the fact that the accuracy of acceleration estimation for the large mass part was high, which reduced the influence of the small mass part whose acceleration estimation accuracy was low. In this paper, the acceleration of the unmeasured part was estimated using a common order. However, since the magnitude of the higher-order components differs depending on the part, there is a limit to the accuracy of the estimation using a common order, but it is thought that the amplitude and phase are aligned by using a common order to the higher-order components.

VI. CONCLUSION AND FUTURE WORK

In this paper, in order to reduce the number of sensors used for estimating the floor reaction force, we proposed a method to estimate the acceleration of the unmeasured part from the measured part using the average motion mode function derived from the trials of multiple subjects, and to estimate the composite floor reaction force in two directions from these results. The average motion mode function can be derived by dividing the unmeasured part by the measured part for each order, and is a function that shows the relationship between the motion of the measured part and the unmeasured part. In this paper, we improved the versatility of the system by using the average motion mode function, which is the average of the motion mode functions of five subjects for each order.

In this paper, we first showed the correlation coefficients between the measured values of MC and the acceleration estimated using the average motion mode function and the acceleration estimated from the motion mode function of one subject, respectively, and indicated the usefulness of the

method for estimating the acceleration of the unmeasured part shown in Section Ⅳ.A using the average motion mode function. Next, the correlation coefficients were used to examine which part of the body is suitable to be selected as the measured part when estimating the acceleration of the unmeasured part. In this paper, the measured and unmeasured parts were selected by focusing on the parts with a large mass, which is important in estimating the floor reaction force. The optimal part where the acceleration of the unmeasured part was estimated with high accuracy differed depending on the subject and direction. However, when the measured part was the upper trunk, the acceleration of the upper trunk, average thigh, and lower trunk, where the mass of the whole body is particularly large, was estimated accurately. To reduce the number of sensors, we thought it appropriate to select the upper trunk as the measured part. In the future, we will investigate whether it is possible to estimate the floor reaction force more accurately by changing the measured part according to the estimated direction. Finally, the floor reaction forces in the vertical and walking directions were derived by estimating the accelerations of the remaining 8 parts using the average motion mode function with the upper trunk as the measured part. The correlation coefficient between the floor reaction force measured by the force plates and the floor reaction force derived using the average motion mode function was 0.934 in the vertical direction and 0.887 in the walking direction. The walking direction, which did not show good accuracy in the acceleration estimation, showed a much higher correlation coefficient and improved accuracy in the floor reaction force. In addition, the bimodal characteristics that appear in the vertical floor reaction forces of one foot on each side in healthy walking were trimodal in the composite vertical floor reaction forces shown in this study, which well captured the characteristics of healthy walking. Therefore, the usefulness of the proposed method for estimating the floor reaction force using the average motion mode function was confirmed.

In the future, we will aim to further improve the versatility of the system using the average motion mode function, and propose a system that can accurately estimate the floor reaction force even for unknown subjects with different paces.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number JP18K11106.

REFERENCES

[1] S. Hontama, K. Shibata, Y. Inoue, and H. Satoh," Estimation of Body Part Acceleration While Walking Using Frequency Analysis: Estimating head acceleration from movement of upper trunk," CENTRIC 2020.

[2] V. L. Chester, E. N. Biden, and M. Tingley, "Gait Analysis," Biomed Instrum Techno, vol. 39, pp. 64–74, January 2005

[3] W. R. Johnson, A. Mian, C. J. Donnelly, D. Lloyd, and J. Alderson, "Predicting athlete ground reaction forces and moments from motion capture," Medical & Biological Engineering & Computing, vol. 56, pp. 1781–1792, 2018.

[4] A. Rajagopal, C. L. Dembia, M. S. DeMers, and D. D. Delp, "Full-Body Musculoskeletal Model for Muscle-Driven Simulation of Human Gait," IEEE Transactions on Biomedical Engineering, vol. 63, pp. 2068–2079, October 2016.

[5] L. Wang, Y. Sun, Q. Li, and T. Liu, "Estimation of Step Length and Gait Asymmetry Using Wearable Inertial Sensors," IEEE Sensors Journal, vol. 18, pp. 3844–3851, May 2018.

[6] S. Zihajehzadeh and E. J. Park, "Regression Model-Based Walking Speed Estimation Using Wrist-Worn Inertial Sensor," PLoS ONE, October 2016. Available from: https://doi.org/10.1371/journal.pone.0165211

[7] T. Liu, Y. Inoue, K. Shibata, and K. Shiojima, "A Mobile Force Plate and Three-Dimensional Motion Analysis System for Three-Dimensional Gait Assessment," IEEE Sensors Journal, vol. 12, pp. 1461–1467, May 2012.

[8] A. Isshiki, Y. Inoue, K. Shibata, and M. Sonobe, "Estimation of Floor Reaction Force During Walking Using Physical Inertial Force by Wireless Motion Sensor," HCI Int'l, vol. 714, pp. 249-254, May 2017, DOI: 10.1007/978-3-319-58753-0_37, 2017, pp.22-33, ISSN:1348-711

[9] M. Ae, H. Tang, and T. Yokoi, "Estimation of Inertia properties of the Body Segments in Japanese Athletes," Soc. Biomechanisms Jpn., vol. 11, 1992, pp. 22-33 (in Japanese).

[10] A. Isshiki, Y. Inoue, K. Shibata, M. Sonobe, and A. Hashiguchi, "Estimation of Floor Reaction Forces during Walking Using a Small Number of Wearable Inertial Sensors," The 47th Student Graduation Research Presentation Lecture, No.312, 2017 (in Japanese)

[11] S. Hontama, Y. Inoue, and K. Shibata," Characteristics of walking motion by using Frequency analysis: Transfer function for Upper body," The Japan Society of Mechanical Engineers Chugoku-Shikoku Bracch , The 50th Student Graduation Research Presentation Lecture, No.06b3, 2020 (in Japanese)

[12] T. Yokoi, K. Shibukawa, and A. Michiyoshi, " Body segment parameters of Japanese children," Physical Education Research., 31 (1), 53−66, 1986

[13] R. K. Jensen, "Estimation of the biomechanical properties of three body types using a photogrammetric method," Journal of Biomechanics, vol. 11, pp. 349–358, 1978.

[14] M. Masatoshi and K. Fukui, " Research on walking of healthy people using force plate," Rehabilitation Medicine, vol. 24, pp. 93–101, 1987. Available from: https://doi.org/10.2490/jjrm1963.24.93

Towards a Component Reference Implementations Frame for Achieving Multi-disciplinary Coherent Conceptual and Chorological Contextualisation in Prehistory and Prehistoric Archaeology

Claus-Peter Rückemann
Westfälische Wilhelms-Universität Münster (WWU), Germany;
Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), Germany;
Leibniz Universität Hannover, Germany
Email: ruckema@uni-muenster.de

***Abstract*—This paper presents contributions towards a new component reference implementations frame, which can be used with conceptual knowledge models. The extended research allows a general chorological and coherent conceptual knowledge contextualisation across disciplines. The frame will enable to coherently address multi-disciplinary knowledge by employing conceptual knowledge frameworks and component reference implementations. The presented achievements and scenarios concentrate on new information science approaches for multi-disciplinary contexts in prehistory and archaeology, targeting the use of inherent aspects and creation of new insight, strategies, and perspectives. The new advanced approach can be applied to any type of knowledge, e.g., factual, conceptual, procedural, metacognitive, and structural. The integration of chorological knowledge includes integration of spatial and geospatial knowledge and features. Goal of this research is a new component reference implementations frame for multi-disciplinary coherent conceptual and chorological contextualisation. For the case scenarios, we are focussing on coherent knowledge in contexts with prehistory and archaeology disciplines, natural sciences, and advanced geoscientific information systems. The solution also integrates knowledge from satellite data and soil diversity and respective properties. The knowledge approach allows a multi-directional utilisation of coherent conceptual knowledge. Future research concentrates on further continued development of components, analysis, and application in multi-disciplinary scenarios of prehistory and archaeology.**

Keywords–Reference Component Implementations Frame; Prehistory; Natural Sciences; Chorology; Contextualisation; Coherent Conceptual Knowledge.

I. INTRODUCTION

This paper presents the new component reference implementations frame, which can be used with conceptual knowledge reference implementations. The extended research allows a general chorological and coherent conceptual knowledge contextualisation across disciplines. The frame enables to coherently address multi-disciplinary knowledge by employing conceptual knowledge frameworks and component reference implementations. This paper is an extended and updated presentation of the research based on the publication and presentation at the GEOProcessing 2021 conference in Nice, France [1].

These days, mostly all Geographic Information Systems and even more advanced and more complex Geoscientific Information Systems (GIS) are –by themselves– not yet taking multi-disciplinary coherent knowledge and contexts into consideration. From scientific point of view, it is a questionable approach to think of and practice a distinct discipline while considering any other required scientific discipline being auxiliary, which, for further simplification, may even be reduced to 'data delivery', 'technical', and 'procedural' contributions.

Many measurements and and contexts in prehistorical archaeology and natural sciences cannot be 'sensed' directly and require further endeavours. Multi-disciplinary scenarios often require to consider a wide range of contexts with disciplines put to their level. It is the coherent knowledge of contexts, which is most relevant for new insight. Therefore, contextualisation should not be done without considering multi-disciplinary coherency and expert views from different disciplines put on a par with respective further scientific collaboration and support.

Goal of this research is a new component reference implementations frame for multi-disciplinary coherent conceptual and chorological contextualisation, based For the case scenarios, we are focussing on coherent knowledge in contexts with prehistory and archaeology disciplines, natural sciences, and advanced geoscientific information systems.

The new advanced approach enables a systematical, coherent contextualisation and can be applied with knowledge complements, e.g., factual, conceptual, procedural, metacognitive, and structural. The approach is knowledge-centric, in a way "knowledge-driven" but explicitly not "development-procedure-driven" or "software-driven".

The rest of this paper is organised as follows. Section II presents the method of component integration. Section III illustrates the implementation and realisation scenario for prehistory's and archaeology's conceptual and chorological knowledge reference implementation contexts. Section IV discusses the coherent conceptual knowledge integration scenario case study. Section V delivers the new component reference implementation frame with major building blocks. Sections VI and VII discuss the state of the achieved results and summarise lessons learned, conclusions, and future work.

II. COMPONENT INTEGRATION AND METHOD

From knowledge complements' point of view, chorology is for place what chronology is for time. The integration of chorological knowledge includes integration of spatial and geospatial knowledge and features. Here, we are focussing on scenarios of coherent multi-disciplinary knowledge in context with prehistory and archaeology, natural sciences, and advanced geoscientific information system components. The solution also integrates knowledge from satellite data and soil diversity with respective properties.

Contexts in prehistory are peculiar in a way that there are no direct historical sources and respectively no literary reference and documentation. Contextualisation is therefore a main intrinsic task in prehistory and protohistory.

An approach has to conform with information science fundaments and universal knowledge and has to enable an integration of the required components from methodologies to realisations for knowledge representations of realia and abstract contexts [2] while many facets of knowledge, including prehistory, need to be continuously acquired and reviewed [3]. There is no published approach known, which can be reasonably compared with the implemented and presented method. Therefore, there was a strong need for an advanced methodology to contextualise knowledge, e.g., from practically available knowledge resources. This paper presents the methodological fundaments for a chorological and coherent multi-disciplinary contextualisation. The potential for finding and integrating multi-disciplinary inherent aspects and creation of new insight, strategies, and perspectives by development of components employing coherent conceptual knowledge has been a major motivation.

Commonly used tools are not aware of features for contextualisation from multi-disciplinary components. Therefore, advanced individual workflows need proper preparation of components and workflow procedures. Preparation requires methods for deployment of respective knowledge characteristics and properties. Further, any workflow should be created being aware of the individuality of these characteristics.

Many aspects of knowledge, including meaning, can be described using knowledge complements supporting a modern definition of knowledge and subsequent component instrumentation [4] [5], e.g., considering factual, conceptual, procedural, metacognitive, and structural knowledge. Especially, conceptual knowledge can relate to any of factual, conceptual, procedural, and structural knowledge. To a comparable extent, metacognitive knowledge can relate to any of factual, conceptual, procedural, and structural knowledge.

Knowledge complements are a means of understanding and targeting new insight, e.g., enabling advanced contextualisation, integration, analysis, synthesis, innovation, prospection, and documentation. The approach can be summarised based on the methodological fundaments.

- Selection and development of a coherent, multi-disciplinary reference implementation. Knowledge complements are addressing reference implementations.
- Multi-disciplinary knowledge resources and integrated components are realised with knowledge-centric focus.
- Contextualisation employing knowledge complements.
- Analysis, synthesis, documentation can employ reference implementations for new insight and development.

An approach to the multi-disciplinary nature of this research requires significant developments of coherently integrationable, fundamental context components especially

- multi-disciplinary contexts of prehistory and archaeology and respective resources,
- chorological contexts, e.g.,
- homogeneously consistent high resolution Digital Elevation Model (DEM) for land and sea bottom, and
- natural sciences Knowledge Resources (KR), e.g., soil classification resources, standardised soil reference systems, and parameters.

A more detailed, comprehensive discussion and examples regarding the fundaments are available with the research on methodology, contextualisation, and conceptual knowledge. Relevant pre-existing and ongoing component developments addressing knowledge with multi-disciplinary KR have been summarised [6] and discussed.

Further, major practical groups of component implementations and developments required and addressed for integration with this research are:

1) Conceptual knowledge frameworks.
2) Conceptual knowledge base.
3) Integration of scientific reference frameworks.
4) Formalisation.
5) Methodologies and workflows integration.
6) Prehistory Knowledge Resources.
7) Natural Sciences Knowledge Resources.
8) Inherent representation groups.
9) Scientific context parametrisation.
10) Structures and symbolic representation.

The following sections will present the implementation and realisation and integration scenarios and the new resulting component reference implementation frame.

The implementation and realisation scenario is based on the coherent conceptual knowledge reference implementations. The integration scenario employs these reference implementations and delivers the realisation of integration and symbolic representation and facilities for further analysis. General chorological and coherent conceptual aspects are in focus of realisations. The resulting component reference implementation frame delivers the practical integration frame of components for implementation and realisation and consistently includes available coherent conceptual knowledge reference implementations for this and future multi-disciplinary scenarios.

III. Implementation and realisation scenario

A means of choice in order to achieve overall efficient realisations even for complex scenarios, integrating arbitrary knowledge, is to use the principles of Superordinate Knowledge. The core assembly elements of Superordinate Knowledge are methodology, implementation, and realisation [7].

In the following example solutions, scenario targets are contexts of prehistoric cemeteries and burials at the North Sea coast, in North-Rhine Westphalia, Lower Saxony, and The Netherlands. Integration targets are natural sciences and speleological contexts, caves and cave systems in North-Rhine Westphalia, Lower Saxony, and The Netherlands, soil diversity, and overall integration with chorological, symbolical, spatial context representations, e.g., place, spatial planning, auxiliary subdivisions for boundaries and spatial forms and administrative units.

A. Coherent conceptual knowledge reference implementation

We can select relevant references from the implemented prehistory-protohistory and archaeology Conceptual Knowledge Reference Implementation (CKRI) [8] (E.0.4.4). The methodology allows to address any other references on a coherent information science knowledge base, e.g., geoscientific knowledge from natural sciences KR components. Further, the reference implementation enables to address chorology on the coherent knowledge base. Universally consistent conceptual knowledge is based on UDC references for demonstration, spanning the main tables [9] shown in Table I.

TABLE I. COHERENT CONCEPTUAL KNOWLEDGE DEPLOYED FOR CONTEXTUALISATION, SELECTED UDC CODE REFERENCES (EXCERPT).

Code / Sign Ref.	*Verbal Description (EN)*
UDC:0	Science and Knowledge. Organization. Computer Science. Information. Documentation. Librarianship. Institutions. Publications
UDC:004	*Computer science and technology. Computing.*
UDC:1	Philosophy. Psychology
UDC:2	*Religion. Theology*
UDC:3	Social Sciences
UDC:5	Mathematics. Natural Sciences
UDC:52	*Astronomy. Astrophysics. Space research. Geodesy*
UDC:53	Physics
UDC:539	Physical nature of matter
UDC:54	Chemistry. Crystallography. Mineralogy
UDC:55	Earth Sciences. Geological sciences
UDC:550.3	Geophysics
UDC:551	General geology. Meteorology. Climatology. Historical geology. Stratigraphy. Palaeogeography
UDC:551.44	*Speleology. Caves. Fissures. Underground waters*
UDC:551.46	Physical oceanography. Submarine topography. Ocean floor
UDC:551.7	Historical geology. Stratigraphy
UDC:551.8	Palaeogeography
UDC:56	Palaeontology
UDC:6	Applied Sciences. Medicine, Technology
UDC:63	Agriculture and related sciences and techniques. Forestry. Farming. Wildlife exploitation
UDC:631.4	*Soil science. Pedology. Soil research*
UDC:7	The Arts. Entertainment. Sport
UDC:8	Linguistics. Literature
UDC:9	Geography. Biography. History
UDC:902	Archaeology
UDC:903	*Prehistory. Prehistoric remains, artefacts, antiquities*
UDC:904	Cultural remains of historical times
UDC (1/9)	Common auxiliaries of place
UDC:(1)	Place and space in general. Localization. Orientation
UDC:(2)	Physiographic designation
UDC:(20)	Ecosphere
UDC:(21)	Surface of the Earth in general. Land areas in particular. Natural zones and regions
UDC:(23)	*Above sea level. Surface relief. Above ground generally. Mountains*
UDC:(24)	*Below sea level. Underground. Subterranean*
UDC:(25)	Natural flat ground (at, above or below sea level). The ground in its natural condition, cultivated or inhabited
UDC:(26)	Oceans, seas and interconnections
UDC:(28)	Inland waters
UDC:(3/9)	Individual places of the ancient and modern world
UDC:(3)	Places of the ancient and mediaeval world
UDC:(4/9)	Countries and places of the modern world
UDC:(4)	*Europe*
UDC:"..."	Common auxiliaries of time.
UDC:"6"	Geological, archaeological and cultural time divisions
UDC:"62"	*Cenozoic (Cainozoic). Neozoic (70 MYBP - present)*

For this research, major references from both main and auxiliary tables are highlighted in italics with bluish colour.

B. Multi-disciplinary views: Prehistory and archaeology

Table II shows an excerpt of UDC:903.. . .:2 ritual/burial object and subgroup examples, and conceptual view groups [10] (PAKA, [11] [12]), referenced according the prehistory-protohistory and archaeology CKRI (E.0.4.4).

TABLE II. CKRI: PREHISTORY AND PROTOHISTORY RITUAL/BURIAL OBJECT AND SUBGROUP EXAMPLES, AND VIEW GROUPS [10] (EXCERPT).

Major Object Group	*Selected Objects*	*Conceptual View Group*
Ritual places, burials	yes	UDC:903. . .:2
Cemetery	–	UDC:903. . .:2
Barrow	–	UDC:903. . .:2
round	–	UDC:903. . .:2
long	–	UDC:903. . .:2
Cist	–	UDC:903. . .:2
Dolmen	–	UDC:903. . .:2
Tomb	–	UDC:903. . .:2
chamber	–	UDC:903. . .:2
court	–	UDC:903. . .:2
portal	–	UDC:903. . .:2
rock cut	–	UDC:903. . .:2
wedge	–	UDC:903. . .:2
Pithos burial	–	UDC:903. . .:2
Cave	–	UDC:903. . .:2
Body finding	–	UDC:903. . .:2
Urn	–	UDC:903. . .:2
. . .	–	UDC:903. . .:2

Regarding the conceptual properties, these example objects and objects groups represent knowledge facets rather than simple hierarchical references. Contextualisation based on conceptual knowledge can this way be employed in combinatory fashion and can address peculiarities by flexible means, e.g., by knowledge complements, symbolic representation, and conceptual knowledge views of chorology and chronology.

For this illustrative object scenario, the excerpt does not show individual micro-groups and individual differences. Besides different distributions and different origins, object context can be referred, e.g., artificial origin and natural origins as well as relevant object properties, materials, and soil contexts can be considered and systematically contextualised.

C. Multi-disciplinary views: Soil diversity

A soil diversity CKRI development edition (E.0.6.2) was created based on WRB standard soil type reference groups and soil type specifications [13], [14], and a suitable UDC:631.4. . . base soil reference system for prehistory and archaeology, which has been compiled along with this research. The results are available in Table III.

For this research, the reference system is based on standard soil references and UDC, both enabling a systematic and coherent approach. Soil diversity groups are relevant for prehistorical and archaeological objects and contexts. Contextualised soil diversity groups are referenced in a consistent, standardised way. From this base compilation, a properties based reference system can be created for further contextualisation, parametrisation, and processing with the ongoing research on soil diversity for prehistory and archaeology. Associated information, e.g., on soil

- drainage,
- wetness,
- pH status,
- base saturation,
- chloride,
- subsoil organic material, and
- stiffness

can be found as reference in the World Reference Base (WRB) for soil resources [13], [14] from the Food and Agriculture Organisation (FAO), United Nations.

TABLE III. CKRI: SOIL DIVERSITY, BASED ON WRB STANDARD AND CONCEPTUAL REFERENCE SYSTEM (UDC:631.4...), (E.0.6.2).

Soil Type Reference Group	*Soil Type Specification Name in WRB 2006 / WRB 1998*
Acrisol	Haplic / Ferric, Gleyic, Haplic, Humic, Plinthic
Alisol	Plinthic
Albeluvisol	Haplic / Endoeutric, Gleyic, Haplic, Histic, Stagnic, Umbric
Andosol	Aluandic / Dystric, Humic, Umbric, Mollic, Vitric
Anthrosol	Anthrosol, Plaggic
Arenosol	Albic, Haplic, Protic
Calcisol	Aridic
Chernozem	Calcic, Haplic, Gleyic, Haplic, Luvic
Cambisol	Haplic / Calcaric, Haplic / Chromic, Haplic / Dystric, Haplic / Eutric, Gleyic, Haplic, Mollic, Vertic
Fluvisol	Haplic / Calcaric, Haplic / Dystric, Haplic / Eutric, Gleyic, Haplic, Histic, Mollic, Salic, Thionic
Gleysol	Haplic / Calcaric, Haplic / Dystric, Haplic / Eutric, Haplic / Haplic, Histic, Humic, Mollic, Thionic
Gypsisol	Haplic / Aridic
Histosol	Histosol, Hemic / Dystric, Hemic / Eutric, – / Fibric, – / Gelic, – / Sapric
Kastanozem	Calcic, Haplic, Luvic
Leptosol	Haplic / Calcaric, Haplic / Dystric, Haplic / Eutric, Haplic / Haplic, Haplic / Humic, Rendzic, Lithic
Luvisol	Albic, Haplic / Arenic, Calcic, Haplic / Chromic, Haplic / Dystric, Haplic / Ferric, Gleyic, Haplic, Vertic
Phaeozem	– / Albic, Haplic / Calcaric, Gleyic, Haplic, Luvic, Haplic / Sodic
Planosol	Haplic / Dystric, Haplic / Eutric, Haplic
Podzol	Haplic / Carbic, Haplic / Entic, Gleyic, Haplic, Leptic, Placic, Haplic / Rustic, Umbric
Regosol	Haplic / Calcaric, Haplic / Dystric, Haplic / Eutric, Haplic
Solonchak	Gleyic, Haplic, Haplic / Takyric, Mollic
Solonetz	Gleyic, Haplic, Mollic
Umbrisol	Arenic, Gleyic
Vertisol	Haplic / Chromic, Haplic, Haplic / Pellic

In this context, the conceptual references are referring to the respective categories, e.g., UDC:631.4...:903+"4...".

D. Multi-disciplinary integration facets

Table IV shows the reference facets of a respective multi-disciplinary target contextualisation.

TABLE IV. REFERENCE FACETS OF A MULTI-DISCIPLINARY TARGET CONTEXTUALISATION, BASED ON CKRI (EXCERPT).

Code / Sign Ref.	*Verbal Description (EN)*
	Geography. Biography. History
UDC:903...	Prehistory, prehistoric remains, artefacts, antiquities
...:2	referring to religion and rituals
...,"62..."	from Holocene
...,(4...DENW)	... in North-Rhine Westphalia, Germany
...,(4...DENI)	... in Lower Saxony, Germany
...,(4...NL)	... in The Netherlands
	Earth sciences, geological sciences
UDC:551.44	Speleology, caves, fissures, underground waters
	Applied sciences, agriculture in general
UDC:631.4	Soil research data
	Geodesy. Photogrammetry
UDC:52...,(23)	Remote sensing data, above sea level
UDC:52...,(24)	Remote sensing data, below sea level
	Contextualisation Place
UDC:(4)	Europe

Reference facets of a multi-disciplinary target contextualisation are based on CKRI, implemented and realised using UDC code references. The contextualisation uses coherent conceptual knowledge and refers to the chorological references for consequent knowledge integration and symbolic representation.

IV. COHERENT KNOWLEDGE INTEGRATION SCENARIO

Figure 1 shows a generated, resulting coherent conceptual knowledge integration sketch for the realisation based on the KR. The sketch considers the major conceptual references for illustration. Detailed research can further detail on prehistoric object groups, characteristics, and properties, topographic properties, soil properties, and many more. Therefore, the conceptual sketch view can result in levels of arbitrary numbers of different integrations of complements and associated properties as resulting from the KR, which are discussed in the following. The result integrates required KR components based on coherent conceptual knowledge and systematical chorological knowledge for multi-disciplinary contexts, e.g., arbitrary group representations, classification based representations, and geospatial representations.

Knowledge objects and contexts are provided by The Prehistory and Archaeology Knowledge Archive (PAKA) [11] [12]. The multi-disciplinary coherent contextualisation employs the base of a new soil system reference development with soil types (UDC:631.4...) of WRB standard, reference contexts, especially for UDC:903...:2,551.7+"628"..., prehistorical, protohistorical time spans and artefacts related to religion and rituals, geology, especially stratigraphy and paleogeography, quaternary, especially late glacial and Holocene.

The integrated natural sciences KR further provide information on caves in the respective region. Contextualisation is enabled by the Conceptual Knowledge Reference Implementation (CKRI), including multi-disciplinary contexts of natural sciences and humanities [8]. The conceptual knowledge base is The Universal Decimal Classification (UDC) [10].

In this illustration plain Digital Chart of the World (DCW) data are used [15]. The coastline database is the Global Self-consistent Hierarchical High-resolution Geography (GSHHG) [16], [17], which was mainly compiled from the World Vector Shorelines (WVS) [18], the CIA World Data Bank II (WDBII) [19], and the Atlas of the Cryosphere (AC).

An equal area projection (Eckert IV) is advised due to the type of discipline knowledge representation. The compilation uses the World Geodetic System (WGS). The symbolic representation of the contextualisation is done via LX Professional Scientific Content-Context-Suite (LX PSCC Suite) deploying the Generic Mapping Tools (GMT) [20] for visualisation.

Concrete details of knowledge complements and components have been provided [6]. All basic technical aspects can be created on that base for individual application scenarios. As illustrated, the solution is explicitly not a database concept and the goal is explicitly not just to link different multi-disciplinary concepts. The solution allows to create individual conceptual knowledge based algorithms and to integrate with new and available spatial and temporal processing algorithms. Basic components and functions are given in the references.

The presented integration approach for chorological and coherent contextualisation provides a solid base for multi-disciplinary contexts in prehistory and archaeology. The implemented system of components, continuously in development,

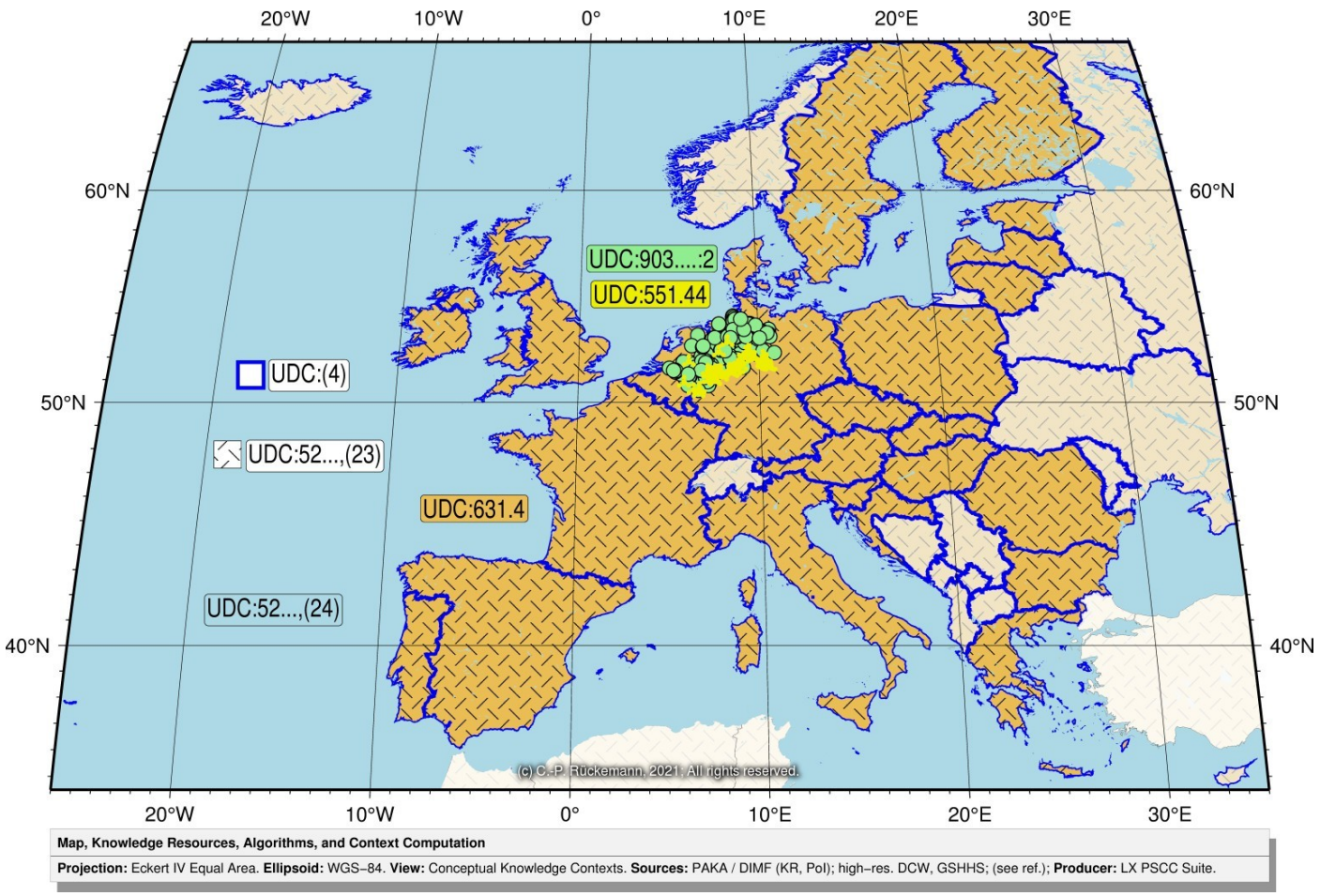


Figure 1. Resulting coherent conceptual knowledge integration sketch diagram showing knowledge resources for a prehistoric, natural sciences, and spatial contextualisation for excerpts of prehistoric cemeteries' and caves' distributions, remote sensing data, and soil properties with respective knowledge references.

integrates relevant and beneficial methodologies and properties, especially

- coherent conceptual knowledge views,
- multi-disciplinary contextualisation,
- application approaches for multi-disciplinary contexts in prehistory and archaeology,
- allows systematical chorological consideration of knowledge, e.g., arbitrary group representations, classification based representations, geospatial representations,
- further development and valorisation of resources,
- integration of multi-disciplinary resources,
- choice for homogeneity of components,
- deployment of systematical procedures,
- effective and efficient integration and analysis,
- automation of workflows and procedures, and
- provides multi-lingual conceptual knowledge support.

The methodological approach also allows the multi-directional utilisation: Conceptual knowledge and resulting integration, e.g., symbolic representation, on the one hand and integration results delivering references to conceptual knowledge and new integrated knowledge contexts on the other hand. As demonstrated with the integration for this research, besides coherency, general flexibility, robustness, and scalability, criteria for components employed with implementation and realisation should be evaluated carefully for being able to consider solid information science fundaments and knowledge centricity, beyond plain technical and proprietary features.

Resulting from the methodology, the realisation integrates a wide range of relevant selection criteria, e.g., in this scenario:

- Conceptual selection (esp. prehistory, cemetery; natural sciences, caves).
- Spatial, mathematical selection.
- Regional spatial selection.
- Topographic selection (esp. above sea level).
- Contextualised selection (esp. with availability of sufficient natural sciences, soil, and other context data).

Many solutions resulting from the above scenario were implemented and realised on the fundament framework of flexibly coupled components. The following sections deliver the component reference implementations frame created resulting from the long-term research, achieving the framework for the required implementations and realisations and for the multi-disciplinary coherent knowledge integration. The component reference implementations frame provides a systematical approach for a large variety of coherent conceptual, consistent approaches for multi-disciplinary contextualisation for many scenarios beyond the examples for prehistory and prehistorical archaeology, natural sciences, and humanities.

V. Component reference implementations frame

The next sections give a compact overview of a major Component Reference Implementations (CRI) development stage edition (E.0.3.5) integrated with this research on knowledge complements and frameworks for coherent knowledge integration in prehistory, natural sciences, and humanities. A more detailed, comprehensive discussion and examples regarding the fundaments are available with the research on methodology, contextualisation, and conceptual knowledge. Relevant pre-existing and ongoing component developments addressing knowledge with multi-disciplinary Knowledge Resources (KR) have been summarised [6].

The following component reference implementations provide a resulting, comprehensive set of complements and components required for an advanced practical long-term case scenario, integrating knowledge on prehistory and prehistorical archaeology, natural sciences, remote sensing, and soil science.

A. CRI: Conceptual knowledge frameworks

Conceptual knowledge frameworks were created as reference implementations and are further continuously in development and used in practice with ongoing long-term research and applied for KR [6], e.g.:

- *Prehistory-protohistory and archaeology Conceptual Knowledge Reference Implementation (CKRI)*, including multi-disciplinary contexts of natural sciences and humanities and any facets (E.0.4.4) [8] [21].
- *Mathematical and computational conceptual knowledge framework* [22].
- *Environmental information systems conceptual knowledge framework* [23].

Additional CKRI can be created by and for disciplines and scenarios and coherently integrated with CKRI already in use.

B. CRI: Conceptual knowledge base

Conceptual knowledge base is The *Universal Decimal Classification (UDC)* [10], a general plan for knowledge classification, providing an analytico-synthetic and faceted classification, designed for subject description and indexing of content of information resources irrespective of the carrier, form, format, and language. UDC is the world's foremost document indexing language in the form of a multi-lingual classification scheme covering all fields of knowledge and constitutes a sophisticated indexing and retrieval tool. UDC-based references for demonstration are taken from the multilingual UDC summary [10] released by the UDC Consortium, Creative Commons license [24].

C. CRI: Integration of scientific reference frameworks

Relevant scientific practices, frameworks, and standards from disciplines and contexts. Natural sciences, geosciences, and soil science are continuously delivering updated state of the art research and insight, including geodiversity and standardisation [25] [26]. Associated information, e.g., on soil drainage, wetness, pH status, base saturation, chloride, subsoil organic material, and stiffness can be found as reference in the *World Reference Base (WRB) for soil resources* [13], [14] from the Food and Agriculture Organisation (FAO), United Nations.

D. CRI: Formalisation

All integration components, for all disciplines, require an explicit and continuous formalisation [27] process in order to conform with the information science principles according to the practices in the disciplines [28]. This includes knowledge objects and entities as well as procedural components (e.g., *C* [29], *Fortran* [30], *Perl* [31], *Shell wrapper*, *Julia* [32], [33], [34]), computation model support, e.g., *parallelisation standards, OpenMP* [35], [36], Reg Exp patterns, e.g., *Perl Compatible Regular Expressions (PCRE)* [37], further standard tools, e.g., *Structured Query Language (SQL)*, *Tool Command Language (TCL)* [38], Extract Transform Load (ETL), Extract Load Transform (ELT), and hybrid solutions. Addressing aspects of discipline related parole [39].

E. CRI: Methodologies and workflows integration

Methodologies for creating and utilising methods include model processing, remote sensing, spatial mapping, high information densities, and visualisation. Respective contextualisation of (prehistoric) scenarios should each be done under specific (prehistoric) conditions, especially supported by state-of-the-art methods, especially, consistent sources of standard algorithms [40], multi-dimensional criteria, spatial operations, interpolation geodesic computation [41], triangulation [42], gradient computation [43], and projection [44]. Workflow integration includes the overall spectrum of problem solving, e.g., mathematical algorithms, mathematical processes, filter processes, phonetic and linguistic context support [45]. Visualisation, *Generic Mapping Tools (GMT)* [20].

F. CRI: Prehistory Knowledge Resources

Common sources of information in many disciplines are often not yet aware of universal knowledge concepts and multilingual approaches. Common sources mostly lack sufficient coherency, consistency, and structure and more often they show to be fragmented and heterogeneous. In order to overcome basic shortcomings of public 'data collections' the objects, entities, and respective conceptual knowledge references' excerpts and examples are taken from *The Prehistory and Archaeology Knowledge Archive (PAKA)*, in continuous development for more than three decades [11] and is released by DIMF [12].

G. CRI: Natural Sciences Knowledge Resources

Several coherent systems of major natural sciences' context object groups from *KR realisations* have been implemented [6], [10], [46].

H. CRI: Inherent representation groups

The methodology can consider a wide range of representation groups for major disciplines and context object groups regarding their inherent representation and common utilisation, e.g., *points*, *polygons*, *lines*, *Digital Elevation Model (DEM), Digital Terrain Model (DTM), and Digital Surface Model (DSM) representations* sources, e.g., from *satellites*, *drones* (raster data, RAdio Detection And Ranging (RADAR), Synthetic Aperture Radar (SAR), Light Detection And Ranging (LiDAR), etc.), *positioning/navigation* (common satellite systems / satellite navigation systems, e.g., Galileo, Europe; Global Positioning System (GPS), USA; GLObalnaja

NAwigazionnaja Sputnikowaja Sistema (GLONASS), Russia; Quasi-Zenith Satellite System (QZSS), Japan; Indian Regional Navigation Satellite System (IRNSS) / Navigation Indian Constellation (NAVIC), India), *z-value representations*, *distance representations*, *area representations*, *raster*, *vector*, *binary*, and *non-binary data*. Essential base context sources should provide *worldwide homogeneous and consistent data* [26] allowing extrapolation and interpolation in various dimensions, e.g., from the School of Ocean and Earth Science and Technology (SOEST), National Aeronautics and Space Administration (NASA), Goddard National Space Science Data Center (NSSDC), National Oceanographic and Atmospheric Administration (NOAA), Central Intelligence Agency (CIA) resources, European Community (EC) resources, and national and federal organisations and initiatives for further integration and future solutions.

I. CRI: Scientific context parametrisation

Scientific *context parametrisation of prehistoric targets* can use the overall insights, e.g., from geoscientific disciplines [47] [48]. A relevant example is contextualisation with palaeolandscapes [49]. In case of prehistory, parametrisation depends on the prehistorical context, e.g., the geoscientific parametrisation and geoscientific contextualisation depend of the respective selected prehistorical object groups and associated properties. The highly inter-dependent complexity can make the scientific parametrisation an extremely advanced long-term challenge.

J. CRI: Structures and symbolic representation

The deployment of long-term universal structure and data standards is essential. Relevant examples of sustainable implementations are *NetCDF* [50] based standards, including advanced features, hybrid structure integration, and parallel computing support (*PnetCDF*) and generic multi-dimensional table data, standard xyz files, universal source and text based structure and code representations, e.g., American Standard Code for Information Interchange (ASCII) and LaTeX [51].

Structure is an organisation of interrelated entities in a material or non-material object or system [46]. Structure is essential in logic as it carries unique information. Structure means features and facilities. There are merely higher and lower facility levels of how structures can be addressed, which result from structure levels. Structure can, for example, be addressed by logic, names, references, address labels, pointers, fuzzy methods, phonetic methods. 'Non-structures' can, for example, be addressed by locality, source, context, logic, attributes, size, quantity. Structure is and especially reflects knowledge (especially factual, conceptual, procedural, metacognitive, and structural complements), context, experience, persistence, reusability, sustainability, value, and formalisation, including abstraction and reduction. Structure systematics, meaning, levels of structures, and means of addressing were discussed in detail [46]. We should be aware that lower structure levels can only be addressed on higher formalisation levels, independent of the fact that structure may either be not available or not recognised. Substantial deficits of lower level structured knowledge representations cannot be compensated by (procedural) tools. Therefore, addressing structures on cognostic levels is preferable to isolated procedural means and can be utilised for symbolic representations. Symbolic representations of prehistoric context information include graphs, e.g., diagrams using visualisation techniques, for logical, quantitative, schematic, and semi-schematic characteristics. Concrete examples are relationships of entity representations, variables, topological and spatial properties, and combined representations of abstract and realia properties. Structures and standards, in integration with formalisation processes, knowledge system, and components should foster seamless long-term development and sustainable realisation. Nevertheless, the complements, which enable flexible automation capabilites are up to vast parts depending on the context of how realia are viewed and in consequence how they should be described and managed, e.g., by formalisation, standards, consistency, systematics, methodological procedures, structure, and object groups.

VI. Discussion

The presented practical component reference frame implementations are based on a number of component groups, which are in continuous long-term development. The methodology, methods, case scenarios and the fundaments of the component reference implementations and conceptual knowledge frameworks were lately publicly presented and discussed at the Informational Modeling - Theory and Practice - International Conference, Sofia, Bulgaria [8], at the International Conference on Mathematics of Informational Modeling, Varna, Bulgaria [52], both Bulgarian Academy of Sciences, at the Delegates' Summit, Symposium on Advanced Computation and Information in Natural and Applied Sciences, Rhodes, Greece [53], and at the Machine Learning for Industry Forum hosted by the High-Performance Computing Innovation Center and Data Science Institute at the Lawrence Livermore National Laboratory, USA [54]. Practical applications and concrete developments of conceptual knowledge reference implementations and a component reference implementations frame were discussed regarding ongoing and future research initiatives in prehistory and prehistorical archaeology, natural sciences, and integrative industry applications. Conceptual knowledge solutions are recognised a fundament of future industrial learning, collaborative, and multi-disciplinary information science. The component reference implementations frame resulting from this research has shown to provide a flexible framework for the required implementations and realisations including the multidisciplinary coherent knowledge integration. Especially, the respective conceptual knowledge reference edition implementation stages [55] (E.0.4.4 and future) to integrate seamlessly with the coupled components.

VII. Conclusion

The new component reference implementations frame enables efficient implementations and realisations of multidisciplinary coherent conceptual and chorological context integration and contextualisation in prehistory and prehistorical archaeology. The frame allows coherent contextualisation in prehistorical archaeology and natural sciences even where contexts cannot be 'sensed' directly. The components allow a seamless integration of Conceptual Knowledge Reference Implementations (CKRI) for coherent conceptual and chorological context integration and contextualisation in prehistory and prehistorical archaeology.

The practical solution for implementation and realisation of the new knowledge-based methodology showed to enable

coherent conceptual knowledge for contextualisation in prehistory, archaeology, and natural sciences. The more, the approach enables to integrate multi-disciplinary contexts by a consistent system (editions) and supports multi-lingualism. Multi-disciplinary scenarios can be considered in multi-fold ways, e.g., knowledge can be documented, analysed, integrated, and selected deploying conceptual knowledge. The methodological approach also allows a multi-directional utilisation. Any KR result can be considered starting point, intermediate result, and final result, depending on a defined task and workflow.

The general methodology provides flexibility for solid information science based methods and enables a wide range of benefits for scenarios and implementations, e.g., coherent multi-disciplinary and multi-lingual documentation and systematical knowledge based geo-spatial processing, aware of inherent knowledge spanning arbitrary disciplines. With that, geospatial scenarios are special cases of chorological contextualisation. Further recommendations from practical experiences with knowledge complements and component integration are:

- Create (specialised) coherent CKRI.
- Add consistent and coherent conceptual knowledge to objects and entities of your resources and make workflows deploy conceptual knowledge.
- Choose multi-disciplinary resources with homogeneous properties, e.g., resolution and coverage.
- Use long-term standards.
- Practice scientific state-of-the-art parametrisation of respective knowledge, data, and algorithms.
- Create context dependent, suitable symbolic representations and individual methods.
- Proceed the knowledge-centric integration reasoned and levelheaded.

Future research concentrates on continued development of components and reference implementations for creation of models for analysis and development of multi-disciplinary knowledge and context, taking new findings and context integration of application scenarios in prehistory and archaeology into account and further considering chorological and coherent conceptual knowledge contextualisation, especially regarding dedicated prehistory and prehistorical archaeology research cases.

Acknowledgements

This ongoing research is supported by scientific organisations and individuals. We are grateful to the "Knowledge in Motion" (KiM) long-term project, Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), for partially funding this research, implementation, case studies, and publication under grants D2018F1P04962, D2018F3P04932, D2018F6P04938, D2019F1P04998, and D2020F1P05228 and to its senior scientific members and members of the permanent commission of the science council, especially to Dr. Friedrich Hülsmann, Gottfried Wilhelm Leibniz Bibliothek (GWLB) Hannover, to Dipl.-Biol. Birgit Gersbeck-Schierholz, Leibniz Universität Hannover, to Dipl.-Ing. Martin Hofmeister, Hannover, and to Olaf Lau, Hannover, Germany, for fruitful discussion, inspiration, and practical multi-disciplinary case studies. We are grateful to Dipl.-Geogr. Burkhard Hentzschel and Dipl.-Ing. Eckhard Dunkhorst, Minden, Germany, for prolific discussion and exchange of practical spatial and drone case scenarios. We are grateful to Dipl.-Ing. Hans-Günther Müller, Göttingen, Germany, for providing specialised, manufactured high end computation and storage solutions. We are grateful to The Science and High Performance Supercomputing Centre (SHPSC) for long-term support. / DIMF-PIID-DF98_007; URL: https://scienceparagon.de/cpr .

References

[1] C.-P. Rückemann, "Methodology of Chorological and Coherent Conceptual Knowledge Contextualisation: Approaches for Multi-disciplinary Contexts in Prehistory and Archaeology," in Proceedings of The Thirteenth International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2021), July 18 – 22, 2021, Nice, France. XPS Press, Wilmington, Delaware, USA, 2021, ISSN: 2308-393X, ISBN-13: 978-1-61208-871-6, URL: http://www.thinkmind.org/download.php?articleid=geoprocessing_2021_1_60_30045 [accessed: 2021-11-28].

[2] C.-P. Rückemann, "From Knowledge and Meaning Towards Knowledge Pattern Matching: Processing and Developing Knowledge Objects Targeting Geoscientific Context and Georeferencing," in Proc. of The Twelfth Internat. Conf. on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2020), November 21–25, 2020, Valencia, Spain. XPS Press, Wilmington, Delaware, USA, 2020, pp. 36–41, ISSN: 2308-393X, ISBN-13: 978-1-61208-762-7.

[3] R. Gleser, Zu den erkenntnistheoretischen Grundlagen der Prähistorischen Archäologie. Leiden, 2021, 2021, (title in English: On the Epistemological Fundaments of Prehistorical Archaeology), in: M. Renger, S.-M. Rothermund, S. Schreiber, and A. Veling (Eds.), Theorie, Archäologie, Reflexion. Kontroversen und Ansätze im deutschsprachigen Diskurs, (in print).

[4] C.-P. Rückemann, F. Hülsmann, B. Gersbeck-Schierholz, P. Skurowski, and M. Staniszewski, Knowledge and Computing. Post-Summit Results, Delegates' Summit: Best Practice and Definitions of Knowledge and Computing, Sept. 23, 2015, The Fifth Symp. on Adv. Comp. and Inf. in Natural and Applied Sciences (SACINAS), The 13th Int. Conf. of Num. Analysis and Appl. Math. (ICNAAM), Sept. 23–29, 2015, Rhodes, Greece, 2015, pp. 1–7, DOI: 10.15488/3409.

[5] L. W. Anderson and D. R. Krathwohl, Eds., A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. Allyn & Bacon, Boston, MA (Pearson Education Group), USA, 2001, ISBN: 978-0801319037.

[6] C.-P. Rückemann, "Prehistory's and Natural Sciences' Multi-disciplinary Contexts: Contextualisation and Context Integration Based on Universal Conceptual Knowledge," in Proceedings of The Eleventh International Conference on Advanced Communications and Computation (INFOCOMP 2020), May 30 – June 3, 2021, Valencia, Spain. XPS Press, Wilmington, Delaware, USA, 2021, ISSN: 2308-3484, ISBN: 978-1-61208-865-5.

[7] C.-P. Rückemann, "Principles of Superordinate Knowledge: Separation of Methodology, Implementation, and Realisation," in The Eighth Symp. on Adv. Comp. and Inf. in Natural and Applied Sciences (SACINAS), Proceedings of The 16th Int. Conf. of Num. Analysis and Appl. Math. (ICNAAM), Sept. 13–18, 2018, Rhodes, Greece, AIP CP, Vol. 2116. AIP Press, American Institute of Physics, Melville, New York, USA, 2019, ISBN: 978-0-7354-1854-7, DOI: 10.1063/1.5114033.

[8] C.-P. Rückemann, "The Information Science Paragon: Allow Knowledge to Prevail, from Prehistory to Future – Approaches to Universality, Consistency, and Long-term Sustainability," The International Journal "Information Models and Analyses" (IJ IMA), vol. 9, no. 3, 2020, pp. 203–226, Markov, K. (ed.), ISSN: 1314-6416 (print), Submitted accepted article: November 18, 2020, Publication date: August 17, 2021, URL: http://www.foibg.com/ijima/vol09/ijima09-03-p01.pdf [accessed: 2021-11-28].

[9] "UDC Summary Linked Data, Main Tables," 2020, Universal Decimal Classification (UDC), UDC Consortium, URL: https://udcdata.info/078887 [accessed: 2021-11-28].

[10] "Multilingual Universal Decimal Classification Summary," 2012, UDC Consortium, 2012, Web resource, v. 1.1. The Hague: UDC Consortium (UDCC Publication No. 088), URL: http://www.udcc.org/udcsummary/php/index.php [accessed: 2021-11-28].

[11] C.-P. Rückemann, "Information Science and Inter-disciplinary Long-term Strategies – Key to Insight, Consistency, and Sustainability: Conceptual Knowledge Reference Methodology Spanning Prehistory, Archaeology, Natural Sciences, and Humanities," International Tutorial, DataSys Congress 2020, Sept. 27 – Oct. 1, 2020, Lisbon, Portugal, 2020, URL: http://www.iaria.org/conferences2020/ProgramINFOCOMP20.html [accessed: 2021-11-28].

[12] "The Prehistory and Archaeology Knowledge Archive (PAKA) license," 2021, (release 2021), Unabhängiges Deutsches Institut für Multidisziplinäre Forschung (DIMF): All rights reserved. Rights retain to the contributing creators.

[13] "World reference base for soil resources," 1998, World Soil Resources Report No. 84, ISSS-ISRIC-FAO, Rome.

[14] "World reference base for soil resources," 2006, World Soil Resources Report No. 103, ISSS-ISRIC-FAO, Rome.

[15] P. Wessel, "The Digital Chart of the World for GMT 5 or later," 2018, URL: ftp://ftp.soest.hawaii.edu/gmt [accessed: 2021-11-28], URL: http://www.soest.hawaii.edu/pwessel/dcw/ [accessed: 2021-11-28].

[16] P. Wessel, "Global Self-consistent Hierarchical High-resolution Geography," 2016, URL: ftp://ftp.soest.hawaii.edu/gmt [accessed: 2021-11-28], URL: http://www.soest.hawaii.edu/pwessel/gshhg/ [accessed: 2021-11-28].

[17] P. Wessel and W. H. F. Smith, "A global, self-consistent, hierarchical, high-resolution shoreline database," J. Geophys. Res., vol. 101, no. B4, 1996, pp. 8741–8743.

[18] E. A. Soluri and V. A. Woodson, "World Vector Shoreline," Int. Hydrograph. Rev., vol. LXVII, no. 1, 1990, pp. 27–35.

[19] A. J. Gorny, "World Data Bank II General User Guide," 1977.

[20] P. Wessel, W. H. F. Smith, R. Scharroo, J. Luis, and F. Wobbe, "The Generic Mapping Tools (GMT)," 2020, URL: http://www.generic-mapping-tools.org/ [accessed: 2021-11-28], URL: http://gmt.soest.hawaii.edu/ [accessed: 2021-11-28].

[21] C.-P. Rückemann, "Keynote Lecture: The Information Science Paragon: Allow Knowledge to Prevail, from Prehistory to Future – Approaches to Universality, Consistency, and Long-term Sustainability," Informational Modeling – Theory and Practice, International Conference Webinar, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, October 20–21, 2020, Sofia, Bulgaria, [Lecture], URL: http://math.bas.bg, 2020.

[22] C.-P. Rückemann, "Superordinate Knowledge Based Comprehensive Subset of Conceptual Knowledge for Practical Mathematical-Computational Scenarios," in The Ninth Symp. on Adv. Comp. and Inf. in Natural and Applied Sciences (SACINAS), Proceedings of The 17th Int. Conf. of Num. Analysis and Appl. Math. (ICNAAM), Sept. 23–28, 2019, Rhodes, Greece, AIP CP, Volume 2293 A. AIP Press, American Institute of Physics, Melville, New York, USA, Nov. 2020, pp. 040 002–1 – 040 002–4, ISBN-13: 978-0-7354-4025-8, DOI: 10.1063/5.0026478.

[23] C.-P. Rückemann, Sustainable Knowledge and Resources Management for Environmental Information and Computation. Business Expert Press, Manhattan, New York, USA, Mar. 2018, Ch. 3, pp. 45–88, in: Huong Ha (ed.), Climate Change Management: Special Topics in the Context of Asia, ISBN: 978-1-94784-327-1, in: Robert Sroufe (ed.), Business Expert Press Environmental and Social Sustainability for Business Advantage Collection, ISSN: 2327-333X (collection, print).

[24] "Creative Commons Attribution Share Alike 3.0 license," 2012, URL: http://creativecommons.org/licenses/by-sa/3.0/ [accessed: 2021-11-28], (first release 2009, subsequent update 2012).

[25] B. M. Das, Advanced Soil Mechanics; Fifth Edition, 5th ed. CRC Press, Taylor & Francis Group, Boca Raton, London, New York, 2019, ISBN: 978-0-8153-7913-3.

[26] "Land Processed Distributed Active Archive Center," 2021, LP DAAC, URL: https://lpdaac.usgs.gov/ [accessed: 2021-11-28].

[27] C.-P. Rückemann, R. Pavani, B. Gersbeck-Schierholz, A. Tsitsipas, L. Schubert, F. Hülsmann, O. Lau, and M. Hofmeister, Best Practice and Definitions of Formalisation and Formalism. Post-Summit Results, Delegates' Summit: The Ninth Symp. on Adv. Comp. and Inf. in Natural and Applied Sciences (SACINAS), The 17th Int. Conf. of Num. Analysis and Appl. Math. (ICNAAM), Sept. 23–28, 2019, Rhodes, Greece, 2019, DOI: 10.15488/5241.

[28] C.-P. Rückemann, "Keynote Lecture: Information Science and Structure: Beware - Ex Falso Quod Libet," The Tenth Symposium on Advanced Computation and Information in Natural and Applied Sciences (SACINAS), The 18th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM), September 17–23, 2020, Rhodes, Greece, September 17, 2020, [Lecture], URL: http://icnaam.org/sites/default/files/ICNAAM%202020%20Preliminary%20Program%20Virtual%20ver15.pdf [accessed: 2021-11-28], 2020.

[29] "GCC, the GNU Compiler Collection," 2021, URL: http://gcc.gnu.org/ [accessed: 2021-11-28].

[30] "GNU Fortran," 2021, URL: https://gcc.gnu.org/fortran/ [accessed: 2021-11-28].

[31] "The Perl Programming Language," 2021, URL: https://www.perl.org/ [accessed: 2021-11-28].

[32] I. Balbaert, Julia 1.0 Programming: Dynamic and high-performance programming to build fast scientific applications, 2nd ed. Packt Publishing, 2018, ISBN: 978-1-78899-909-0.

[33] M. Sherrington, Mastering Julia - Tackle the Contemporary Challenges of Programming and Data Science with Julia. Packt Publishing, 2015, ISBN: 978-1-78355-331-0.

[34] A. Joshi, Julia for Data Science. Packt Publishing, 2016, ISBN: 978-1-78528-969-9.

[35] L. Dagum and R. Menon, "OpenMP: an industry standard API for shared-memory programming," Computational Science & Engineering, (IEEE), vol. 5, no. 1, 1998, pp. 46–55.

[36] OpenMP Architecture Review Board, "OpenMP API 5.1 Specification," Nov. 2020, URL: https://www.openmp.org/wp-content/uploads/OpenMP-API-Specification-5-1.pdf [accessed: 2021-11-28].

[37] "Perl Compatible Regular Expressions (PCRE)," 2021, URL: https://www.pcre.org/ [accessed: 2021-11-28].

[38] "Tcl Developer Xchange," 2021, URL: https://www.tcl.tk/ [accessed: 2021-11-28].

[39] F. de Saussure, Cours de linguistique générale, 1916, (title in English: Course in General Linguistics), C. Bally and A. Sechehaye (eds.).

[40] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, Numerical Recipes in Fortran 90, The Art of Parallel Scientific Computing. Cambridge University Press, 1996.

[41] T. Vincenty, "Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations," Surv. Rev., vol. XXII, no. 176, 1975, pp. 88–93.

[42] R. J. Renka, "Algorithm 772: STRIPACK: Delaunay Triangulation and Voronoi Diagram on the Surface of a Sphere," AMC Trans. Math. Software, vol. 23, no. 3, 1997, pp. 416–434.

[43] B. K. P. Horn, "Hill-Shading and the Reflectance Map," Proceedings of the IEEE, vol. 69, no. 1, 1981, pp. 14–47.

[44] A. J. Kent and P. Vujakovic, Eds., The Routledge Handbook of Mapping and Cartography. Taylor & Francis Group, London and New York, 2018, ISBN: 978-1-138-83102-5.

[45] C.-P. Rückemann, "Archaeological and Geoscientific Objects used with Integrated Systems and Scientific Supercomputing Resources," Int. Jour. on Adv. in Systems and Measurements, vol. 6, no. 1&2, 2013, pp. 200–213, ISSN: 1942-261x, URL: http://www.thinkmind.org/download.php?articleid=sysmea_v6_n12_2013_15 [accessed: 2021-11-28].

[46] C.-P. Rückemann, "The Impact of Information Science Accompanied Structural Information on Computation of Knowledge Pattern Matching and Processing: A Prehistory, Archaeology, Natural Sciences, and Humanities Conceptual Integration Perspective," in Proc. of The Tenth Int, Conf. on Adv. Comm. and Comp. (INFOCOMP 2020), Sept. 27 – Oct. 1, 2020, Lisbon, Portugal, 2020, ISSN: 2308-3484, ISBN: 978-1-61208-807-5, URL: http://www.thinkmind.org/index.php?view=article&articleid=infocomp_2020_1_10_60015 [accessed: 2021-11-28].

[47] R. Brinkmann, Abriß der Geologie, Bd. 2, Historische Geologie, neubearbeitet von Karl Krömmelbein, 12th/13th ed. Ferdinand Enke Verlag,

Stuttgart, 1986, ISBN: 3-432-80603-5, (title in English: Epitome of Geology, Vol. 2, Historical Geology).

[48] P. Woldstedt and K. Duphorn, Norddeutschland und angrenzende Gebiete im Eiszeitalter, 3rd ed., 1974, (title in English: Northern Germany and Adjacent Regions in the Ice Age).

[49] G. N. Bailey, J. Harff, and D. Sakellariou, Eds., Under the Sea: Archaeology and Palaeolandscapes of the Continental Shelf, ser. Coastal Research Library. Springer International Publishing AG 2017, 2017, vol. 20, ISSN: 2211-0577, ISBN: 978-3-319-53158-8, DOI: 10.1007/978-3-319-53160-1.

[50] "Network Common Data Form (NetCDF)," 2021, DOI: 10.5065/D6H70CW6, URL: http://www.unidata.ucar.edu/software/netcdf/ [accessed: 2021-11-28].

[51] "Comprehensive TeX Archive Network (CTAN)," 2021, (TEX/LATEX), URL: http://www.ctan.org [accessed: 2021-11-28].

[52] C.-P. Rückemann, "Keynote Lecture: The Multi-disciplinary Case of Prehistorical Insight: Information Science at the Edge of Structured Data Comprehension," The XI-th International Conference - Mathematics of Informational Modeling, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, July 1–2, 2021, Varna, Bulgaria, URL: http://math.bas.bg [accessed: 2021-11-28], 2021.

[53] C.-P. Rückemann, R. Pavani, Z. Kovacheva, B. Gersbeck-Schierholz, F. Hülsmann, and I. Naydenova, Best Practice and Definitions – Concepts of Cognostic Addressing Structured and Non-structured Data. Post-Summit Results, Delegates' Summit: Best Practice and Definitions – Concepts of Cognostic Addressing Structured and Non-structured Data, September 20, 2021, The Eleventh Symposium on Advanced Computation and Information in Natural and Applied Sciences (SACINAS), The 19th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM), September 20–26, 2021, Rhodes, Greece, 2021, delegates and other contributors: Claus-Peter Rückemann, Westfälische Wilhelms-Universität Münster (WWU), Germany / Knowledge in Motion, Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), Germany / Leibniz Universität Hannover, Germany; Raffaella Pavani, Department of Mathematics, Politecnico di Milano, Italy; Zlatinka Kovacheva, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences / University of Mining and Geology, Sofia, Bulgaria; Birgit Gersbeck-Schierholz, Knowledge in Motion, Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), Germany; Friedrich Hülsmann, Knowledge in Motion, Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), Germany; Ina Naydenova, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria; URL: http://www.scienceparagon.de/cpr/x/publ/2021/delegatessummit2021/rueckemann_icnaam2021_summit_summary.pdf, (to appear).

[54] C.-P. Rückemann, "Coherent Knowledge Solutions From Prehistory to Future – Towards Coherent Multi-disciplinary Knowledge Reference Implementation Blueprints for Industrial Learning: Insight from Consistent Coherent Conceptual Integration of Prehistory, Archaeology, Natural Sciences, and Humanities," ML4I – Machine Learning for Industry Forum 2021; hosted by the High-Performance Computing Innovation Center (HPCIC) and Data Science Institute (DSI), at the Lawrence Livermore National Laboratory (LLNL), August 10–12, 2021, Livermore, California, U.S.A., (Invited Speech), URL: http://www.llnl.gov [accessed: 2021-11-28], 2021.

[55] C.-P. Rückemann, "Towards Conceptual Knowledge Reference Implementations for Context Integration and Contextualisation of Prehistory's and Natural Sciences' Multi-disciplinary Contexts," International Journal on Advances in Systems and Measurements, vol. 14, no. 1&2, 2021, (to be published).

Towards Conceptual Knowledge Reference Implementations for Context Integration and Contextualisation of Prehistory's and Natural Sciences' Multi-disciplinary Contexts

Claus-Peter Rückemann
Westfälische Wilhelms-Universität Münster (WWU), Germany;
Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), Germany;
Leibniz Universität Hannover, Germany
Email: ruckema@uni-muenster.de

***Abstract*—This paper delivers contributions towards new conceptual knowledge reference implementations and the results of the requirements study on a new context integration system for conceptual contextualisation of prehistory's and natural sciences' universal multi-disciplinary contexts. The conceptual knowledge reference implementations focus on general multi-disciplinary coherent conceptual knowledge integration, for case studies here especially on prehistory-protohistory and archaeology. The conceptual knowledge reference implementations can be used in flexible ways, e.g., creating and employing a frame of component reference implementations. The methodology provides a coherent conceptual knowledge integration, fostering consecutive coherent analysis based on information science fundaments. The methodological approach enables an inclusion of new insight and newly created knowledge, e.g., via deployment of knowledge resources and structures. The programmatic approach and new conceptual knowledge reference implementation span multi-disciplinary knowledge in a coherent, consistent, and multilingual way. The methodology to consistently integrate knowledge context from prehistory and archaeology disciplines with knowledge in natural sciences and humanities is accompanied by ongoing multi-disciplinary case studies implementing the required methods. The focus of this research is on knowledge-based methodologies and deployment of information science methods, especially, universal conceptual knowledge, for the goal of creating and developing conceptual knowledge reference implementations towards coherent and general multi-disciplinary contextualisation and context integration, targeting further future creation of new insight, strategies, and perspectives.**

Keywords–Conceptual Knowledge Reference Implementations; Prehistory; Natural Sciences; Humanities; Contextualisation; Conceptual Knowledge.

I. INTRODUCTION

Multi-disciplinary context integration and contextualisation are relevant for most scientific disciplines. This paper presents new conceptual knowledge reference implementations and the results of the requirements study on a new context integration system for conceptual contextualisation of prehistory's and natural sciences' universal multi-disciplinary contexts. The conceptual knowledge reference implementations focus on multi-disciplinary coherent conceptual knowledge integration, e.g., on prehistory-protohistory and archaeology natural sciences, and humanities. The paper delivers practical development stages of future Conceptual Knowledge Reference Implementations (CKRI) and the results of previous and ongoing research initiatives, which integrate based on information science fundaments for a coherent conceptual integration, enabling consecutive coherent analysis. The conceptual knowledge reference implementations can be used in a flexible way, e.g., employing a frame of component reference implementations.

This paper is an extended presentation of the research based on the publication and presentation at the INFOCOMP 2021 conference in Valencia, Spain [1]. When dealing with context, the signification of the terms 'complex' and 'complicated' is often mixed up. Complexity is a concomitant phenomenon of context, which, when not well comprehended, appears complicated. Both, complexity and context, are linked by manyfold inter-dependencies and often experienced together. Context is witness-context in many cases, in prehistory and natural sciences. Understanding context and gathering complexity in a coherent, consistent, and methodological way are therefore important fundaments, which can lead to consequent systematical instrumentalisation, consecutive coherent analysis, and aspiring new insight.

The new conceptual knowledge reference implementations are employed in practical and presented regarding their methodological and systematical fundaments and components for implementing a multi-disciplinary integration of prehistory and its context. The paper summarises the results of immanent milestones and, based on these, proposes the next complementary methodological and practical resources' developments. Further, details on complements and results of specific application scenarios will be discussed in separate extended papers.

The rest of this paper is organised as follows. Section II introduces into the information science fundaments. Section III gives the essential background of motivation resulting from different disciplinary views. Section IV introduces to disciplinary background and requirements. Section V presents the methodological fundaments and components. Section VI presents a short overview of component implementations for integration. Section VII delivers the results and state of major coherent conceptual knowledge reference implementations. Section VIII presents a practical example of an integration case scenario. Sections IX and X discuss the lessons learned and summarise conclusions and future work.

II. FUNDAMENTS

The fundaments of terminology and understanding the essence of knowledge are layed out by Aristotle, being a central part of 'Ethics' [2]. Information sciences can very much benefit from Aristotle's fundaments and a knowledge-centric approach [3] but for building holistic and sustainable solutions, supporting a modern definition of knowledge and subsequent component instrumentation [4], they need to go beyond the available technology-based approaches and hypothesis [5] as analysed in Platon's Phaidon. Aspects of meaning can be described using knowledge complements, e.g., considering factual, conceptual, procedural, metacognitive [3], and structural

knowledge. Especially, conceptual knowledge can relate to any of factual, conceptual, and procedural knowledge.

To a comparable extent, metacognitive knowledge can relate to any of factual, conceptual, and procedural knowledge. Knowledge complements are a means of understanding, e.g., enabling advanced contextualisation, documentation, prospection, integration, and analysis. From an information science point of view, the classical fundaments of episteme, techne, and doxa are intrinsically tied complements. However, knowledge complements, when consequently applied, do not make the creation and development of resources instantaneously easier. They do not make problem solving algorithms simpler. Knowledge complements do not make scientific contexts obsolete, they do neither make qualified expertise unneeded nor do they lead to faster education or cheaper gain of research results and insight.

The focus of this research is on knowledge-based methodologies and deployment of information science methods, especially universal conceptual knowledge, for the goal of creating a component framework of reference implementations for coherent and general multi-disciplinary contextualisation targeting the creation of new insight, strategies, and perspectives. This research is part of several extensive long-term strategies and concentrating on contextualisation and context integration for prehistory, protohistory, archaeology and their associated contexts, especially natural sciences and humanities. Contexts in prehistory are special in a way that there are no direct historical sources and respectively no literary reference and documentation. Contextualisation is therefore a main intrinsic task in prehistory and protohistory. From the knowledge point of view, also when looking on methodological conditions, prehistory shares many characteristics and factual conditions with natural sciences, e.g., geology and soil science. A coherent conceptual knowledge approach can enable to establish ties and building bridges between contributing knowledge, including future methodologies and contributions from disciplines.

III. Motivation

Complexity is carrying information. Therefore, from information science point of view, we should take care not to loose complexity whenever dealing with information. The complexity of appearing context is commonly even increased when applying methods from multiple disciplines. So far, there are no other comparably holistic and systematical approaches and implementations on conceptual contextualisation known and published besides the presented approach. Contrary, during the last decades, it has become common practice to tackle challenges regarding knowledge and related content solely with procedural approaches, contrary to the fact that creation processes, handling, and management may allow more effective and efficient measures in context of analysis, long-term development of resources, computation, and processing, Common ways of implementing procedural approaches as plain technical solutions are often neither effective nor efficient. In addition, such approaches often lack long-term adaptability and scalability.

How can we create a suitable, practical system of coherent knowledge? Such a system has to conform with information science fundaments and universal knowledge and has to enable an integration of the required components from methodologies to realisations for knowledge representations of realia and abstract contexts [6]. Many facets of knowledge, including prehistory, need to be continuously acquired and reviewed [7]. Knowledge itself is part of cognitive processes and requires an understanding of epistemological fundaments, depending on participated disciplines and views [8], [9].

We should therefore create a system of balanced fundaments of sustainable, complementary solutions based on information science and contextualisation-aware methodologies and complements [10], which allows the application of coherent conceptual knowledge in theory and practice. The conceptual knowledge approach should provide facilities expressing instances of mental concepts and the state of research of their perception. The creation of object types may be influenced by criteria, e.g., by education, experience, and social context. The approach should enable further development of practical disciplinary terminology assignment, e.g., adaption and synchronisation of terminology.

IV. Consideration of disciplinary background

Prehistoric context, even for chorologically, chronologically, and thematically restricted object groups [11], [12], [13] comprises of a wide and highly variable spectrum of knowledge, applied approaches, and formalisation, including abstraction [14] and documentation [15]. Almost all knowledge is further referring to complex contexts of many associated disciplines and views.

In complex scenarios, like multi-disciplinary prehistoric realia founded contexts, we should utilise as much complexity and structure with knowledge complements as possible in order to achieve a high level of integration of factual, conceptual, and structural but also of procedural and metacognitive knowledge. In practice, these approaches are often not followed. In many cases, simple convenience of workflow tasks might suggest to integrate ‘data’ available as is. Even standards and implementations may not be optimal, they can result from the fact that information and data are often determined by technological means. The result may be a limitation regarding the fundamental coherence of knowledge and it may limit the applicability and use of methods and algorithms.

Basic deficits of simplified approaches and many commonly used frameworks (e.g., context-unaware approaches for maps/earth services) make these approaches undesirable for a general and coherent scientific and methodological realisation.

Further, the application of not well satisfying approaches and methodological deficits, especially in multi-disciplinary context, are often fragmented, heterogeneous, and lacking required coherence and precision [16] or require unnecessary estimations and approximations [17].

Further, in addition to such contrary practice there are multifold cases, which should direct to more feasible approaches, e.g., in situations

- when terminology does –in any case– not reflect the context of respective findings,
- when relocated objects require contextualisation and descriptive conceptual knowledge,
- where indications of resources are available without respective artefacts,
- isolated findings of various levels exist,
- when objects with presently isolated contexts require coherent chronology.

Examples of associated, guiding questions are: How can existing and emerging knowledge from prehistory and other disciplines be methodologically integrated? Which multi-disciplinary contexts and approaches can be considered on a coherent, consistent information science base? What are context areas of special characteristics, e.g., where are possible regions of interest and further research? Which fundaments and component implementations should be in focus of contextualisation?

A basic approach for prehistoric contextualisation should be characterised by modular components and premises, namely,

- a coherent, multi-disciplinary methodology, spanning disciplines and fields,
- an overall coherent and consistent knowledge base,
- principle concepts for knowledge description,
- implementing the state of the art in information science and knowledge,
- considering long-term time ranges for continuous developments,
- enabling wide context integration,
- enabling representation of different views,
- enabling representation of different actual perceptions,
- allowing to complement terminology where required,
- and integration of standards and frameworks.

The premise of coherency of the knowledge base is important in a way that solutions should not be restricted to procedural components and interfaces, which intrinsically require additional multi-level formalisation. The coherent approach can provide required descriptive complements to otherwise prescriptive terminologies. The integration should be aware of cognitive visualisation aspects. The contextualisation should further enable to continuously integrate results of past and ongoing research of prehistoric on-site context surveys.

A means of choice in order to achieve overall efficient realisations even for complex scenarios, integrating arbitrary knowledge, is to use the principles of Superordinate Knowledge. The core assembly elements of Superordinate Knowledge are methodology, implementation, and realisation [18]. The presented implementations strictly follow the fundamental methodological algorithm base of the the Conceptual Knowledge Pattern Matching (CKPM) methodology [6], providing and accessing knowledge object patterns based on the Superordinate Knowledge Methodology, which allows systematical use and thorough processing. Respective results from a methodology targeting structures, including implementation, and knowledge-aware application of the methodology were layed out and are available with practical examples [19].

V. Methodological fundaments and components

The methodological and systematical fundaments for contextualisation of prehistory, protohistory, and contexts require modularity and flexibility with structure levels and multi-dimensional knowledge context, especially regarding

- prehistoric object groups,
- prehistoric objects,
- inter-object group context and references,
- chorological and chronological context,
- context correlation for soil context,
- material context, and
- toponymic context,

with further natural and environmental context, regarding methods and extendability, valorisation, analysis, and potential for new insight. At these conditions and based on the previous research and project practice, basic fundaments are:

- Universal, coherent, and consistent conceptual knowledge system.
- Integration of scientific reference frameworks from disciplines and contexts.
- Formalisation for complements, coherence, consistency.
- Methodologies, general problem solution, workflow integration. Implementation and deployment of methods and algorithms.
- Prehistory and protohistory knowledge resources and complements.
- Natural sciences knowledge context resources and complements.
- Inherent representation groups of context resources.
- Scientific context parametrisation.
- Universal structures and data standards.
- Facilities for analysis.
- Spatial mapping.
- Symbolic representation of context information.
- Facilities for automation.
- Long-term development and sustainability.

Besides obvious reasons, e.g., spatial ranges, serious dependencies are made up by conditions of required mathematical algorithms and the context of available data. These dependencies cannot be overcome in many cases as, e.g., it is not possible to get direct data from the original context of a prehistorical site. Targeting contextualisation, the conceptual implementation should integrate knowledge for natural conditions and processes, soil-affine and respective soil-related, e.g., agricultural or geoforensic, contexts. The implementation should consider different systems of chorologies and chronologies, e.g., prehistorical and geological time frames, palaeolithic to neolithic in coexistence with Pleistocene to Holocene and other conceptual and absolute chronologies. The achieved results of respective developments and implementations of the components will be discussed in the following sections.

VI. Methodological component implementations

Focus is on required methodological, conceptual, non-procedural, non-interactive, and non-technical components. Practical components for systematical and methodological implementations are defined and developed according to the analyses in already realised projects and case studies of practical scenarios as cited here and described in the references.

These items can only give a minimal necessary background for CKRI development editions regarding the component groups of an implementation frame. Overall results on a fully compatible respective component reference implementation frame are to be published including details on component implementations and realisations [20]. The next sections briefly summarise components used for addressing knowledge with multi-disciplinary Knowledge Resources (KR).

A. Conceptual knowledge frameworks

Examples of frameworks developed and used in practice with ongoing long-term research and applied for multi-disciplinary KR are stages of the future prehistory-protohistory and archaeology Conceptual Knowledge Reference Implementation (CKRI), including multi-disciplinary contexts of natural sciences and humanities (E.0.4.4) [10], an environmental information systems conceptual knowledge framework [21], and a mathematical and computational conceptual knowledge framework [22].

The methodology allows to address any other references on a coherent information science knowledge base, e.g., geoscientific knowledge from natural sciences KR components. Further, the reference implementation enables to address chorology on the coherent knowledge base.

Table I shows a general overview containing an excerpt of major references of coherent conceptual knowledge, relevant for the case scenario. Universally consistent conceptual knowledge is based on UDC references [23] for demonstration cases of references implementations and examples.

B. Conceptual knowledge base

The Universal Decimal Classification (UDC) [24] is a general plan for knowledge classification. UDC is also the world's foremost document indexing language in the form of a multi-lingual classification scheme covering all fields of knowledge and constitutes a sophisticated indexing and retrieval tool. UDC-based references in this publication are taken from the multi-lingual UDC summary [24] released by the UDC Consortium under a Creative Commons license [25].

C. Integration of scientific reference frameworks

The integration includes relevant scientific practices, frameworks, and standards from disciplines and contexts, e.g., natural sciences. Geosciences and soil science are continuously delivering updated insight on state of the art research, including the geodiversity and standardisation [26] as required for contextualisation. Essential base context sources should provide worldwide homogeneous and consistent data [27] allowing extrapolation and interpolation in various dimensions.

D. Formalisation

All integration components, for all disciplines, require an explicit and continuous formalisation [28] process in order to conform with the information science principles according to the practices in the disciplines. This includes knowledge objects and entities as well as procedural components and addressing aspects of discipline related parole [29]. Components can deploy Perl Compatible Regular Expressions (PCRE) [30] syntax for specifying common string patterns.

E. Methodologies and workflows integration

Methodologies for creating and utilising methods include model processing, remote sensing, spatial mapping, high information densities, and visualisation. The respective contextualisation of prehistoric scenarios should each be done under individual prehistoric conditions, supported by state-of-the-art methods, especially, consistent sources of standard algorithms [31], multi-dimensional criteria, spatial operations, interpolation geodesic computation [32], triangulation [33], gradient computation [34], and projection [35]. Workflow integration also includes the overall spectrum of problem solving, e.g., mathematical algorithms, mathematical processes, filter processes, but also phonetic and linguistic context support [36].

F. Prehistory Knowledge Resources

Common sources of information in many disciplines are often not yet aware of universal knowledge concepts and multi-lingual approaches, not sufficiently coherent, consistent, and structured and more often they show to be fragmented and heterogeneous. In order to be independent of these basic shortcomings, all of the objects, entities, and respective conceptual knowledge references' excerpts and examples are taken from The Prehistory and Archaeology Knowledge Archive (PAKA). PAKA has been in continuous development for more than three decades [37] and is released by DIMF [38].

G. Natural Sciences Knowledge Resources

Sources of information regarding natural sciences and contexts are provided by Natural Sciences Knowledge Resources, an implemented system of major natural sciences' context object groups from KR realisations [24], [19].

H. Inherent representation groups

Most relevant data in participated disciplines is addressed in defined ways regarding structure and form. According to their inherent representation and common utilisation, major discipline and context object groups can be referred by conceptual knowledge.

I. Scientific context parametrisation

Scientific context parametrisation of prehistoric targets can use the overall insights, e.g., from geoscientific disciplines [39], [40]. A relevant example is contextualisation with palaeolandscapes [41]. In case of prehistory, parametrisation depends on the prehistorical context, e.g., the geoscientific parametrisation and geoscientific contextualisation depend of the respective selected prehistorical object groups and associated properties.

J. Structures and symbolic representation standards

The deployment of long-term universal structure and data standards is essential. Relevant examples of sustainable implementations are NetCDF [42] based standards, including advanced features, hybrid structure integration, and parallel computing support (PnetCDF).

VII. Coherent Conceptual Knowledge Reference Implementations (CKRI)

Relevant knowledge can be referenced, e.g., from the implemented development stage prehistory-protohistory and archaeology CKRI (E.0.4.4) [10]. The methodology allows to address any other references on a coherent conceptual knowledge base, e.g., geoscientific knowledge from natural sciences KR components (Table I). The following sections present excerpts of the development stages of the long-term

TABLE I. CONCEPTUAL KNOWLEDGE REFERENCE IMPLEMENTATION: COHERENT CONCEPTUAL KNOWLEDGE DEPLOYED FOR CREATION OF CONCEPTUAL KNOWLEDGE REFERENCE IMPLEMENTATIONS, CONTEXT INTEGRATION, AND CONTEXTUALISATION, SELECTED UDC CODE REFERENCES (EXCERPT, INCLUDING PREHISTORY-PROTOHISTORY AND ARCHAEOLOGY CKRI DEVELOPMENT STAGE E.0.4.4).

Code / Sign Ref.	*Verbal Description (EN)*
UDC:0	Science and Knowledge. Organization. Computer Science. Information. Documentation. Librarianship. Institutions. Publications
UDC:004	*Computer science and technology. Computing.*
UDC:1	Philosophy. Psychology
UDC:2	*Religion. Theology*
UDC:3	Social Sciences
UDC:5	Mathematics. Natural Sciences
UDC:52	*Astronomy. Astrophysics. Space research. Geodesy*
UDC:528	Geodesy. Surveying. Photogrammetry. Remote sensing. Cartography
UDC:528.2	Figure of the Earth. Earth measurement. Mathematical geodesy. Physical geodesy. Astronomical geodesy
UDC:528.3	Geodetic surveying
UDC:528.4	Field surveying. Land surveying. Cadastral survey. Topography. Engineering survey. Special fields of surveying
UDC:528.5	Geodetic instruments and equipment
UDC:528.7	Photogrammetry: aerial, terrestrial
UDC:528.8	Remote sensing
UDC:528.9	Cartography. Mapping (textual documents)
UDC:53	Physics
UDC:539	Physical nature of matter
UDC:54	Chemistry. Crystallography. Mineralogy
UDC:55	Earth Sciences. Geological sciences
UDC:550.3	Geophysics
UDC:550.7	Geobiology. Geological actions of organisms
UDC:550.8	Applied geology and geophysics. Geological prospecting and exploration. Interpretation of results
UDC:551	General geology. Meteorology. Climatology. Historical geology. Stratigraphy. Palaeogeography
UDC:551.44	*Speleology. Caves. Fissures. Underground waters*
UDC:551.46	Physical oceanography. Submarine topography. Ocean floor
UDC:551.7	Historical geology. Stratigraphy
UDC:551.8	Palaeogeography
UDC:56	Palaeontology
UDC:6	Applied Sciences. Medicine, Technology
UDC:63	Agriculture and related sciences and techniques. Forestry. Farming. Wildlife exploitation
UDC:631	Agriculture in general
UDC:631.4	*Soil science. Pedology. Soil research*
UDC:7	The Arts. Entertainment. Sport
UDC:8	Linguistics. Literature
UDC:9	Geography. Biography. History
UDC:902	*Archaeology*
UDC:903	*Prehistory. Prehistoric remains, artefacts, antiquities*
UDC:904	Cultural remains of historical times
UDC (1/9)	Common auxiliaries of place
UDC:(1)	Place and space in general. Localization. Orientation
UDC:(2)	Physiographic designation
UDC:(20)	Ecosphere
UDC:(21)	Surface of the Earth in general. Land areas in particular. Natural zones and regions
UDC:(23)	*Above sea level. Surface relief. Above ground generally. Mountains*
UDC:(24)	*Below sea level. Underground. Subterranean*
UDC:(25)	Natural flat ground (at, above or below sea level). The ground in its natural condition, cultivated or inhabited
UDC:(26)	Oceans, seas and interconnections
UDC:(28)	Inland waters
UDC:(3/9)	Individual places of the ancient and modern world
UDC:(3)	Places of the ancient and mediaeval world
UDC:(4/9)	Countries and places of the modern world
UDC:(4)	*Europe*
UDC:"..."	Common auxiliaries of time.
UDC:"6"	Geological, archaeological and cultural time divisions
UDC:"62"	*Cenozoic (Cainozoic). Neozoic (70 MYBP - present)*

results of CKRI realisations, which are used in this context with ongoing research:

- Prehistory-protohistory and archaeology contexts.
- Natural Sciences.
- Soil diversity.
- Inherent representation groups.

A. CKRI: Prehistory-protohistory and archaeology contexts

The Prehistory-protohistory and Archaeology CKRI (E.0.4.4) can, in future, address all facets, referencing multi-disciplinary prehistory-protohistory and archaeology contexts.

Table II shows examples of prehistory and protohistory ritual/burial objects and subgroups, and conceptual view groups, an excerpt of UDC:903. . . :2 groups [24] for prehistory and protohistory.

TABLE II. CKRI: PREHISTORY-PROTOHISTORY AND ARCHAEOLOGY, VIEW GROUPS [24] (EXCERPT, E.0.4.4).

Major Object Reference Group	*Conceptual View Group*
Earthworks	UDC:[903,902].. . .
Settlements	UDC:[903,902].. . .
Fortifications	UDC:[903,902].. . .
Architectures	UDC:[903,902].. . .
Structures and arrangements	UDC:[903,902].. . .
Timber	UDC:[903,902].. . .
Stone	UDC:[903,902].. . .
. . .	UDC:[903,902].. . .
Rock art	UDC:[903,902].. . .
Sculptured objects	UDC:[903,902].. . .
Resources (usage, mining, etc.)	UDC:[903,902].. . .
Ritual places, burials	UDC:[903,902].. . .
Cemetery	UDC:[903,902].. . .
Barrow	UDC:[903,902].. . .
round	UDC:[903,902].. . .
long	UDC:[903,902].. . .
Cist	UDC:[903,902].. . .
Dolmen	UDC:[903,902].. . .
Tomb	UDC:[903,902].. . .
chamber	UDC:[903,902].. . .
court	UDC:[903,902].. . .
portal	UDC:[903,902].. . .
rock cut	UDC:[903,902].. . .
wedge	UDC:[903,902].. . .
Pithos burial	UDC:[903,902].. . .
Cave	UDC:[903,902].. . .
Body finding	UDC:[903,902].. . .
Urn	UDC:[903,902].. . .
. . .	UDC:[903,902].. . .
Relics, organic and non-organic	UDC:[903,902].. . .
Organic	UDC:[903,902].. . .
Metal	UDC:[903,902].. . .
. . .	UDC:[903,902].. . .
Artefacts, organic and non-organic	UDC:[903,902].. . .
Organic	UDC:[903,902].. . .
Metal	UDC:[903,902].. . .
. . .	UDC:[903,902].. . .
. . .	UDC:[903,902].. . .

The conceptual view group is prehistory and archaeology, including prehistoric remains, artefacts, and antiquities. Context integration and contextualisation include multi-disciplinary contexts, e.g., natural sciences and humanities.

The reference groups and CKRI (E.0.4.4) can be referenced in any knowledge resources, e.g., the shown groups are practical examples from the PAKA [37], [38]. Objects are selected from a major object group, including subgroups.

B. CKRI: Natural Sciences

The Natural Sciences CKRI development stage (E.0.2.6) can address all facets, referencing multi-disciplinary natural sciences' contexts. Table III shows a plain representation excerpt of an implemented system of major natural sciences' context object groups from respective KR realisations employing a system of context object groups and conceptual views groups [24], [19].

TABLE III. CONCEPTUAL KNOWLEDGE REFERENCE IMPLEMENTATION: NATURAL SCIENCES / NATURE CKRI, VIEWS GROUPS [24] (EXCERPT, E.0.2.6).

Major Object Reference Group	*Conceptual View Group*
Landmarks	UDC:55+539. . .
Height	UDC:55+539. . .
Depth	UDC:55+539. . .
. . .	UDC:55+539. . .
Caves	UDC:55+539. . .
Natural resources	UDC:55+539. . .
Rock outcrops	UDC:55+539. . .
Well springs	UDC:55+539. . .
Soil features	UDC:55+539. . .
Volcanological features	UDC:55+539. . .
Impact features	UDC:55+539. . .
. . .	UDC:55+539. . .

The conceptual view group is earth sciences and geological sciences, physical nature of matter, including geophysics, historical geology, and palaeogeography, soil science and research.

The reference groups and CKRI can be referenced in any KR realisations e.g., the shown groups are practical examples from the Natural Sciences Knowledge Resources, which provide sources of information regarding natural sciences and contexts.

C. CKRI: Soil diversity

The Soil Diversity CKRI development stage (E.0.6.2) provides an implemented and realised soil type conceptual reference system compilation (UDC:631.4. . . base), shown in Table IV. The Soil Diversity CKRI (E.0.6.2) can be seamlessly integrated with respective prehistory and archaeology contexts.

The soil type reference system is based on soil diversity standards and universal conceptual knowledge. The reference system deploys implemented and realised World Reference Base (WRB) standard soil type reference groups and soil type specifications for soil resources [43], [44].

Associated information, e.g., on soil drainage, wetness, pH status, base saturation, chloride, subsoil organic material, and stiffness can be found as reference in the World Reference

Base (WRB) for soil resources [43], [44] from the Food and Agriculture Organisation (FAO), United Nations.

For this research, the created reference system is based on standard soil references and UDC, both enabling a systematic and coherent approach. In this context, the conceptual references are referring to the respective categories, e.g., UDC:631.4. . .:903+"4. . .". The conceptual view group is soil science, pedology, and soil research.

Soil diversity groups are relevant for prehistorical and archaeological objects and contexts. Contextualised soil diversity groups are referenced in a consistent, standardised way.

Properties based reference systems can be created from this base compilation, e.g., a properties based reference system for further contextualisation, parametrisation, and processing with the ongoing research on soil diversity for prehistory and archaeology.

D. CKRI: Inherent representation groups

The Inherent Representation CKRI (Table V) development stage (E.0.2.2) refers to conceptual knowledge of inherent representations objects, e.g., objects used for implementations and realisations of structure and form in disciplines, processing, symbolic representation, and scientific analysis.

TABLE V. CONCEPTUAL KNOWLEDGE REFERENCE IMPLEMENTATION: INHERENT REPRESENTATION CKRI, VIEW GROUPS [24] (EXCERPT, E.0.2.2).

Major Object Reference Group	*Conceptual View Group*
Points, (Points of Interest, PoI)	UDC:52+004. . .
Polygons	UDC:52+004. . .
Lines	UDC:52+004. . .
Digital Elevation Model (DEM) repr.	UDC:52+004. . .
z-value representations	UDC:52+004. . .
Distance representations	UDC:52+004. . .
Area representations	UDC:52+004. . .
Raster	UDC:52+004. . .
Vector	UDC:52+004. . .
Binary	UDC:52+004. . .
Non-binary	UDC:52+004. . .
. . .	UDC:52+004. . .

The excerpt shows a plain representation of major discipline and context object groups regarding their inherent representation and common utilisation using a system of context object groups and conceptual view groups [24]. The conceptual view group is astronomy, astrophysics, space research, and geodesy, computer science and technology, computing, and data processing, including earth measurement, field surveying, photogrammetry, remote sensing, data processing, interpretation, mapping, data representation, data handling, and computer languages. The reference groups and CKRI can be referenced in any KR realisations e.g., the shown groups are practical examples used with the case scenarios.

VIII. INTEGRATION CASE EXAMPLE

The integration may be illustrated by a compact case example of contributions, which employs the presented

- coherent conceptual knowledge reference implementations, with realised KR components and
- systematical chorological knowledge for multi-disciplinary contexts, e.g., arbitrary group representations, classification based representations, and geospatial representations, with further components.

A. Contributions from implementations and realisations

A means of choice in order to achieve overall efficient realisations even for complex scenarios, integrating arbitrary knowledge, is to use the principles of Superordinate Knowledge. The core assembly elements of Superordinate Knowledge are methodology, implementation, and realisation [18]. In the following example solution [1], [45], scenario targets with required contributions are named in the following listing.

- *Contexts of prehistoric cemeteries and burials*
- *at the North Sea coast, in North-Rhine Westphalia, Lower Saxony, and The Netherlands.*
- *Integration targets are natural sciences and*
- *speleological contexts, caves and cave systems*
- *in North-Rhine Westphalia, Lower Saxony, and The Netherlands,*
- *soil diversity, and overall integration with*
- *chorological, symbolical, spatial context representations, e.g., place, spatial planning,*
- *auxiliary subdivisions for boundaries and*
- *spatial forms,*
- *administrative units, and*
- *resources suitable for computing,*

for all of which knowledge needs to be created, continuously developed, and re-used [1], [45] A prominent prehistoric valorisation example using auxiliaries is the swimming reindeer [46], included in detail in [10]. Therefore, this integration case, considers the major conceptual references. Specific research can address further detail on any contexts and properties, e.g.,

- prehistoric object groups,
- object characteristics, and
- object properties,
- topographic properties,
- soil properties, and many more.

B. Multi-disciplinary coherent knowledge integration

An excerpt, deploying CKRI based reference facets of coherent knowledge, of an implemented and realised multi-disciplinary target contextualisation for a case scenario is shown in Table VI. The case scenario integrates facets for references from a number of respective reference implementations the prehistory-protohistory and archaeology CKRI (E.0.4.4), the natural sciences CKRI (E.0.2.6), the soil diversity CKRI (E.0.6.2), the inherent representation CKRI (E.0.2.2), and auxiliary tables.

TABLE IV. CONCEPTUAL KNOWLEDGE REFERENCE IMPLEMENTATION: SOIL DIVERSITY CKRI, VIEW GROUPS (UDC:631.4...), IMPLEMENTED AND REALISED WRB STANDARD SOIL TYPE REFERENCE GROUPS AND SOIL TYPE SPECIFICATIONS (CKRI EXCERPT, E.0.6.2).

Soil Type Reference Group	*Soil Type Specification Name in WRB 2006 / WRB 1998*	*Conceptual View Group*
Acrisol	Haplic / Ferric, Gleyic, Haplic, Humic, Plinthic	UDC:631.4...
Alisol	Plinthic	UDC:631.4...
Albeluvisol	Haplic / Endoeutric, Gleyic, Haplic, Histic, Stagnic, Umbric	UDC:631.4...
Andosol	Aluandic / Dystric, Humic, Umbric, Mollic, Vitric	UDC:631.4...
Anthrosol	Anthrosol, Plaggic	UDC:631.4...
Arenosol	Albic, Haplic, Protic	UDC:631.4...
Calcisol	Aridic	UDC:631.4...
Chernozem	Calcic, Haplic, Gleyic, Haplic, Luvic	UDC:631.4...
Cambisol	Haplic / Calcaric, Haplic / Chromic, Haplic / Dystric, Haplic / Eutric, Gleyic, Haplic, Mollic, Vertic	UDC:631.4...
Fluvisol	Haplic / Calcaric, Haplic / Dystric, Haplic / Eutric, Gleyic, Haplic, Histic, Mollic, Salic, Thionic	UDC:631.4...
Gleysol	Haplic / Calcaric, Haplic / Dystric, Haplic / Eutric, Haplic / Haplic, Histic, Humic, Mollic, Thionic	UDC:631.4...
Gypsisol	Haplic / Aridic	UDC:631.4...
Histosol	Histosol, Hemic / Dystric, Hemic / Eutric, – / Fibric, – / Gelic, – / Sapric	UDC:631.4...
Kastanozem	Calcic, Haplic, Luvic	UDC:631.4...
Leptosol	Haplic / Calcaric, Haplic / Dystric, Haplic / Eutric, Haplic / Haplic, Haplic / Humic, Rendzic, Lithic	UDC:631.4...
Luvisol	Albic, Haplic / Arenic, Calcic, Haplic / Chromic, Haplic / Dystric, Haplic / Ferric, Gleyic, Haplic, Vertic	UDC:631.4...
Phaeozem	– / Albic, Haplic / Calcaric, Gleyic, Haplic, Luvic, Haplic / Sodic	UDC:631.4...
Planosol	Haplic / Dystric, Haplic / Eutric, Haplic	UDC:631.4...
Podzol	Haplic / Carbic, Haplic / Entic, Gleyic, Haplic, Leptic, Placic, Haplic / Rustic, Umbric	UDC:631.4...
Regosol	Haplic / Calcaric, Haplic / Dystric, Haplic / Eutric, Haplic	UDC:631.4...
Solonchak	Gleyic, Haplic, Haplic / Takyric, Mollic	UDC:631.4...
Solonetz	Gleyic, Haplic, Mollic	UDC:631.4...
Umbrisol	Arenic, Gleyic	UDC:631.4...
Vertisol	Haplic / Chromic, Haplic, Haplic / Pellic	UDC:631.4...

TABLE VI. CKRI BASED REFERENCE FACETS OF MULTI-DISCIPLINARY TARGET CONTEXTUALISATION (EXCERPT).

Code / Sign Ref.	*Verbal Description (EN)*
	Geography. Biography. History
UDC:903....	Prehistory, prehistoric remains, artefacts, antiquities
...:2	referring to religion and rituals
...,"62..."	from Holocene
...,(4...DENW)	... in North-Rhine Westphalia, Germany
...,(4...DENI)	... in Lower Saxony, Germany
...,(4...NL)	... in The Netherlands
	Earth sciences, geological sciences
UDC:551.44	Speleology, caves, fissures, underground waters
	Applied sciences, agriculture in general
UDC:631.4	Soil research data
	Geodesy. Photogrammetry
UDC:52...,(23)	Remote sensing data, above sea level
UDC:52...,(24)	Remote sensing data, below sea level
	Contextualisation Place
UDC:(4)	Europe
	Organization. Computer Science
UDC:004...	Computer science and technology. Computing.

Respective sketch maps for various practical examples, considering major conceptual references, have been recently published [45], [47]. Conceptual sketch views can result in levels of arbitrary numbers of different integrations of complements and associated properties as resulting from the integrated KR. The individual possibilities of views and contextualisation are literally unlimited.

C. Knowledge Resources and components

The knowledge integration for prehistoric, natural sciences, and spatial contextualisation for excerpts of prehistoric cemeteries' and caves' distributions, remote sensing data, and soil properties with respective knowledge references [1], [45] is achieved by following major reference implementation components as given. The knowledge integration enables universal conceptual knowledge for all components, a wide range of knowledge contextualisation and high flexibility for implementation and realisation. Especially, the methodology allows flexible component realisations for integration of contexts, enabling management of KR, symbolic representation, and projection. In this integration case, the coherent multi-disciplinary contextualisation employs a base of the following contexts:

- Knowledge objects and contexts are provided by The Prehistory and Archaeology Knowledge Archive (PAKA) [37] [38].
- New soil system reference development / (UDC:631.4...), WRB standard,
- reference contexts are defined, especially for UDC:903....:2,551.7+"628"...,
- prehistorical, protohistorical time & artefacts related to religion and rituals,
- geology, especially stratigraphy and
- palaeogeography, quaternary, especially late glacial and Holocene.
- The integrated natural sciences KR further provide information on caves in the respective region.
- The reference implementations allow for flexible multi-disciplinary scientific, chorological contextualisation,

e.g., Digital Elevation Model (DEM) data from satellites and drones.

For contextualisation and symbolic representation, plain Digital Chart of the World (DCW) data [48] can be used. Coastline database can be the Global Self-consistent Hierarchical High-resolution Geography (GSHHG) [49] [50], which was mainly compiled from the World Vector Shorelines (WVS) [51], the CIA World Data Bank II (WDBII) [52], and the Atlas of the Cryosphere (AC).

An equal area projection (Eckert IV) is advised due to the type of discipline knowledge representation. The compilation can use the World Geodetic System (WGS).

The symbolic representation of the contextualisation can be done via LX Professional Scientific Content-Context-Suite (LX PSCC Suite) deploying the Generic Mapping Tools (GMT) [53] for visualisation.

IX. Discussion

Many measurements and and contexts in prehistorical archaeology and natural sciences cannot be 'sensed' directly and require further endeavours. [20]. Multi-disciplinary scenarios often require to consider a wide range of contexts with disciplines put to their level. It is the coherent knowledge of contexts, which is most relevant for new insight. Therefore, contextualisation should not be done without considering multi-disciplinary coherency and expert views from different disciplines put on a par with respective further scientific collaboration and support.

The range of application scenarios of the presented coherent conceptual knowledge reference implementations is far beyond any examples, which might be given. Nevertheless, the presented practical conceptual knowledge reference implementations and conceptual knowledge base are in continuous long-term development.

The status of integration potentials and the outlook on concrete targets were reviewed based on the lessons learned from the methodological component implementations. The component related processes are challenging and not trivial, especially formalisation and parametrisation. This is the more true for the integration processes. The resulting bases, CKRI and CRI, are the start of a continuous long-term integration on contextualisation for prehistory and multi-disciplinary contexts. All the presented components were created, developed, and evaluated with the referred practical project results and case studies. The conceptual knowledge reference implementations, especially the prehistory CKRI and components showed that they are best choice addressing required properties and features for the tasks. The presented components' set of reference implementations and components also allows further development, targeting the integration for coherent contextualisation including required standards from information science, conceptual knowledge, prehistory and archaeology, natural sciences and geosciences, soil science, satellite and spatial data, and processing algorithms, for the purpose of contextualisation and further utilisation and prospection in prehistory and context.

It should be explicitely noticed, that the integrated methods, resources, and workflows have to support features beyond methodological compatibility, suitability, modularity, and flexibility on the task, e.g., with development, storage, transfer, and utilisation. Especially, the presented conceptual knowledge system enables to respect the rights of participated parties and conform with and adhere to intellectual properties, privacy, and licensing of resources and components, e.g., with intermediate, and resulting structures, formats, and procedural components.

In consequence, practical integration can refer to involved resources and components from all disciplines, prehistory, geosciences, soil science, remote sensing, application of reference implementations and standards, creation of knowledge, procedural realisations, e.g., algorithms and model processing, and results, e.g., symbolic representation of prehistoric context.

The conceptual knowledge reference implementations and methodologies were lately publicly presented and discussed at the Informational Modeling - Theory and Practice - International Conference, Sofia, Bulgaria [10], at the International Conference on Mathematics of Informational Modeling, Varna, Bulgaria [54], both Bulgarian Academy of Sciences, at the Delegates' Summit, Symposium on Advanced Computation and Information in Natural and Applied Sciences, Rhodes, Greece [55], and at the Machine Learning for Industry Forum hosted by the High-Performance Computing Innovation Center and Data Science Institute at the Lawrence Livermore National Laboratory, USA [56].

Practical applications and concrete developments of conceptual knowledge reference implementations and a component reference implementations frame were discussed regarding ongoing and future research initiatives in prehistory and prehistorical archaeology, natural sciences, and integrative industry applications. Conceptual knowledge solutions are recognised a fundament of future industrial learning, collaborative, and multi-disciplinary information science.

The component reference implementations frame resulting from this research has shown to provide a flexible framework for the required implementations and realisations including the multi-disciplinary coherent knowledge integration. Especially, the respective coherent conceptual knowledge reference implementations integrate seamlessly with coupled components addressed by provided Component Reference Implementations (CRI) frames [20]. The integration allows a flexible contextualisation for many peculiarities and situations, e.g., for multi-disciplinary coherent conceptual and chorological contexts.

X. Conclusion

This research successfully achieved to deliver new, widely deployable, coherent Conceptual Knowledge Reference Implementations (CKRI) from the results of the long-term requirements study on a new context integration system for conceptual contextualisation of prehistory's and natural sciences´ universal multi-disciplinary contexts.

Coherent conceptual knowledge reference implementations integrate seamlessly with Knowledge Resources and coupled components, e.g., via Component Reference Implementations (CRI) frames. The integration allows a flexible contextualisation for many peculiarities and situations, e.g., for multi-disciplinary coherent conceptual and chorological contexts. Employed resources are in continuous development. In addition, the conceptual views groups are a unique, flexible, and extendable approach of addressing multi-lingual verbal descriptions with a systematic approach and standardised implementation framework for multi-disciplinary and multi-dimensional scenarios of coherent conceptual knowledge, beyond plain representation.

This research contributed on knowledge-based methodologies and deployment of information science methods, especially universal conceptual knowledge, for contextualisation and context integration of multi-disciplinary contexts of prehistory and natural sciences, which can enable coherent future analysis. The results of long-term research projects in different disciplines leading to this publication contributed to the achieved goal to create reference implementations for coherent and general multi-disciplinary contextualisation, which represents more than its component parts. The integration enables to deal with knowledge complements, e.g., factual, structural, and formalised like time periods but also with metacognitive like experience, meaning, and symbolism. The presented results are nevertheless the start of a consecutive long-term integration project and continuing projects in participated disciplines. The presented methodological approach allows to systematically overcome conceptual fragmentation and to foster on a multi-level coherency for multi-disciplinary knowledge. The multilingual conceptual reference implementation allows to address problems of various language dependent fragmentation, e.g., to resolve national and local terminology fragmentation. This is increasingly relevant for coherency of inter-disciplinary knowledge in contextualisation. The new CKRI and facilities for a component reference implementations frame enable a coherent conceptual integration of prehistory and context disciplines and can foster the consideration and visibility of inherent aspects. CKRI and facilities for a component reference implementations frame allow to address, correlate, and integrate contexts of prehistory-protohistory and archaeology and context of natural sciences and related disciplines, e.g., diversity of soil and relevant properties.

Methodology and implementation allow a wide range or multi-disciplinary contexts and approaches for prehistoric context research for arbitrary regions on interest based on context knowledge, which can globally kept homogeneous and consistent as allowed by publicly available state-of-the-art resources. Examples are geoscientific and mathematical parametrisation and model computations for prehistoric scenarios. The developed reference implementations and components have been in continuous further development to address the continuous development of multi-disciplinary knowledge resources and new methodological implementations.

Overall, in result, contextualisation fosters careful and diligent scientific analysis and interpretation, besides global applicability of the methodology and implementations.

Future research targets on long-term development of a coherent consistent conceptual knowledge framework focussing on model creation and analysis of prehistory and archaeology case scenarios, taking new findings, contexts, and context integration into account. Foci are on new advanced, practical editions of a prehistory-protohistory and archaeology CKRI, a natural sciences CKRI, a soil diversity CKRI, and an inherent representation groups CKRI. That includes context-aware surveys on prehistoric object groups, multi-disciplinary contextualisation of prehistory-geodiversity scenarios, modular integration, analysis, and symbolic representation models for prehistory and context disciplines, e.g., concentrating on Central European supra-regional and on micro-regional studies in Northern Germany (North-Rhine Westphalia, Lower Saxony) and The Netherlands coast areas. The integration and priorities with information science research depend on the state-of-the-art results and development in contributing disciplines.

ACKNOWLEDGEMENTS

This ongoing research is supported by scientific organisations and individuals. We are grateful to the "Knowledge in Motion" (KiM) long-term project, Unabhängiges Deutsches Institut für Multi-disziplinäre Forschung (DIMF), for partially funding this research, implementation, case studies, and publication under grants D2018F1P04962, D2018F3P04932, D2018F6P04938, D2019F1P04998, and D2020F1P05228 and to its senior scientific members and members of the permanent commission of the science council, especially to Dr. Friedrich Hülsmann, Gottfried Wilhelm Leibniz Bibliothek (GWLB) Hannover, to Dipl.-Biol. Birgit Gersbeck-Schierholz, Leibniz Universität Hannover, to Dipl.-Ing. Martin Hofmeister, Hannover, and to Olaf Lau, Hannover, Germany, for fruitful discussion, inspiration, and practical multi-disciplinary case studies. We are grateful to Dipl.-Geogr. Burkhard Hentzschel and Dipl.-Ing. Eckhard Dunkhorst, Minden, Germany, for prolific discussion and exchange of practical spatial and drone case scenarios. We are grateful to Dipl.-Ing. Hans-Günther Müller, Göttingen, Germany, for providing specialised, manufactured high end computation and storage solutions. We are grateful to The Science and High Performance Supercomputing Centre (SHPSC) for long-term support. / DIMF-PIID-DF98_007; URL: https://scienceparagon.de/cpr .

REFERENCES

[1] C.-P. Rückemann, "Prehistory's and Natural Sciences' Multi-disciplinary Contexts: Contextualisation and Context Integration Based on Universal Conceptual Knowledge," in Proceedings of The Eleventh International Conference on Advanced Communications and Computation (INFOCOMP 2021), May 30 – June 3, 2021, Valencia, Spain. XPS Press, Wilmington, Delaware, USA, 2021, ISSN: 2308-3484, ISBN: 978-1-61208-865-5, URL: http://www.thinkmind.org/download.php?articleid=infocomp_2021_1_20_60011 [accessed: 2021-11-28].

[2] Aristotle, The Ethics of Aristotle, 2005, Project Gutenberg, eBook, EBook-No.: 8438, Rel. Date: Jul., 2005, Digit. Vers. of the Orig. Publ., Produced by Ted Garvin, David Widger, and the DP Team, Edition 10, URL: http://www.gutenberg.org/ebooks/8438 [accessed: 2021-11-28].

[3] L. W. Anderson and D. R. Krathwohl, Eds., A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. Allyn & Bacon, Boston, MA (Pearson Education Group), USA, 2001, ISBN: 978-0801319037.

[4] C.-P. Rückemann, F. Hülsmann, B. Gersbeck-Schierholz, P. Skurowski, and M. Staniszewski, Knowledge and Computing. Post-Summit Results, Delegates' Summit: Best Practice and Definitions of Knowledge and Computing, Sept. 23, 2015, The Fifth Symp. on Adv. Comp. and Inf. in Natural and Applied Sciences (SACINAS), The 13th Int. Conf. of Num. Analysis and Appl. Math. (ICNAAM), Sept. 23–29, 2015, Rhodes, Greece, 2015, pp. 1–7, DOI: 10.15488/3409.

[5] Plato, Phaedo, 2008, (Written 360 B.C.E.), Translated by Benjamin Jowett, Provided by The Internet Classics Archive, URL: http://classics.mit.edu/Plato/phaedo.html [accessed: 2021-11-28].

[6] C.-P. Rückemann, "From Knowledge and Meaning Towards Knowledge Pattern Matching: Processing and Developing Knowledge Objects Targeting Geoscientific Context and Georeferencing," in Proc. of The Twelfth Internat. Conf. on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2020), November 21–25, 2020, Valencia, Spain. XPS Press, Wilmington, Delaware, USA, 2020, pp. 36–41, ISSN: 2308-393X, ISBN-13: 978-1-61208-762-7.

[7] R. Gleser, "Klassifikation, Verbreitung und chemische Zusammensetzung kupferzeitlicher Metallartefakte and Rhein, Mosel und Saar," in Proceedings, Beiträge des international Symposiums zur Archäologie in der Großregion, 14.–17. April 2016, Europäische Akademie Otzenhausen, 2017, pp. 163–189, (title in English: Classification, Distribution, and Chemical Composition of Copper-age Metal Artefacts at the Rhine, Moselle, and Saar), Archäologentage Otzenhausen, Band 3, Archäologie in der Großregion, Koch, M. (ed.), ISBN: 578-3-941509-14-6.

[8] R. Gleser, Rekonstruktion der Vergangenheit: Zur methodischen Eigenart prähistorischen Erkennens. mentis, Paderborn, 2018, pp. 199–237, (title in English: Reconstruction of the Past: On the Methodical Peculiarity of Prehistorical Cognition), in: A.-S. Naujoks, J. Stelling, and O. R. Scholz (eds.), Von der Quelle zur Theorie. Vom Verhältnis zwischen Objektivität und Subjektivität in den historischen Wissenschaften, ISBN: 978-3-95743-136-3.

[9] R. Gleser, Zu den erkenntnistheoretischen Grundlagen der Prähistorischen Archäologie. Leiden, 2021, 2021, (title in English: On the Epistemological Fundaments of Prehistorical Archaeology), in: M. Renger, S.-M. Rothermund, S. Schreiber, and A. Veling (Eds.), Theorie, Archäologie, Reflexion. Kontroversen und Ansätze im deutschsprachigen Diskurs, (in print).

[10] C.-P. Rückemann, "The Information Science Paragon: Allow Knowledge to Prevail, from Prehistory to Future – Approaches to Universality, Consistency, and Long-term Sustainability," The International Journal "Information Models and Analyses" (IJ IMA), vol. 9, no. 3, 2020, pp. 203–226, Markov, K. (ed.), ISSN: 1314-6416 (print), Submitted accepted article: November 18, 2020, Publication date: August 17, 2021, URL: http://www.foibg.com/ijima/vol09/ijima09-03-p01.pdf [accessed: 2021-11-28].

[11] H.-J. Häßler, Ur- und Frühgeschichte in Niedersachsen. Nikol, Hamburg, 2002, ISBN: 3-933203-61-9.

[12] H. Kröber, Natur und Landschaft in Niedersachsen, Die Naturdenkmal-Typen. Schlütersche Verlagsgesellschaft mbH & Co. KG, 2001, ISBN: 3-87706-616-X.

[13] R. Wiechers-Weidner, Großsteingräber in Westfalen. Landschaftsverband Westfalen-Lippe, Landesbildstelle Westfalen, 1985, (title in English: Megalithic Tombs in Westphalia).

[14] A. Cochrane and A. M. Jones, Eds., Visualising the Neolithic: Abstraction, Figuration, Performance, Representation. Oxbow Books, Oxford and Oakville, 2012, ISBN: 978-1-84218-477-7, Neolithic Studies Group Seminar Papers 13.

[15] P. Skoglund, J. Ling, and U. Bertilsson, Eds., Picturing the Bronze Age. Published in the United Kingdom by Oxbow Books, Oxford & Philadelphia, 2015, Swedish Rock Art Series: Volume 3, ISBN: 978-1-78297-879-4.

[16] B. Fritsch, M. Furholt, M. Hinz, L. Lorenz, H. Nelson, G. Schafferer, S. Schiesberg, and K.-G. Sjögren, "Dichtezentren und lokale Gruppierungen – Eine Karte zu den Großsteingräbern Mittel- und Nordeuropas," jungsteinSITE, Oct. 2010, ISSN: 1868-3088, (title in English: Centres of Density and Local Groups – A Map of Megalithic Tombs in Central and Northern Europe).

[17] C. Nimura, Prehistoric Rock Art in Scandinavia: Agency and Environmental Change. Published in the United Kingdom by Oxbow Books, Oxford & Philadelphia, 2016, Swedish Rock Art Series: Volume 4, ISBN: 978-1-78570-119-1.

[18] C.-P. Rückemann, "Principles of Superordinate Knowledge: Separation of Methodology, Implementation, and Realisation," in The Eighth Symp. on Adv. Comp. and Inf. in Natural and Applied Sciences (SACINAS), Proceedings of The 16th Int. Conf. of Num. Analysis and Appl. Math. (ICNAAM), Sept. 13–18, 2018, Rhodes, Greece, AIP CP, Vol. 2116. AIP Press, American Institute of Physics, Melville, New York, USA, 2019, ISBN: 978-0-7354-1854-7, DOI: 10.1063/1.5114033.

[19] C.-P. Rückemann, "The Impact of Information Science Accompanied Structural Information on Computation of Knowledge Pattern Matching and Processing: A Prehistory, Archaeology, Natural Sciences, and Humanities Conceptual Integration Perspective," in Proc. of The Tenth Int, Conf. on Adv. Comm. and Comp. (INFOCOMP 2020), Sept. 27 – Oct. 1, 2020, Lisbon, Portugal, 2020, ISSN: 2308-3484, ISBN: 978-1-61208-807-5, URL: http://www.thinkmind.org/index.php?view=article&articleid=infocomp_2020_1_10_60015 [accessed: 2021-11-28].

[20] C.-P. Rückemann, "Towards a Component Reference Implementations Frame for Achieving Multi-disciplinary Coherent Conceptual and Chorological Contextualisation in Prehistory and Prehistoric Archaeology," International Journal on Advances in Systems and Measurements, vol. 14, no. 1&2, 2021, (to be published).

[21] C.-P. Rückemann, Sustainable Knowledge and Resources Management for Environmental Information and Computation. Business Expert Press, Manhattan, New York, USA, Mar. 2018, Ch. 3, pp. 45–88, in: Huong Ha (ed.), Climate Change Management: Special Topics in the Context of Asia, ISBN: 978-1-94784-327-1, in: Robert Sroufe (ed.), Business Expert Press Environmental and Social Sustainability for Business Advantage Collection, ISSN: 2327-333X (collection, print).

[22] C.-P. Rückemann, "Superordinate Knowledge Based Comprehensive Subset of Conceptual Knowledge for Practical Mathematical-Computational Scenarios," in The Ninth Symp. on Adv. Comp. and Inf. in Natural and Applied Sciences (SACINAS), Proceedings of The 17th Int. Conf. of Num. Analysis and Appl. Math. (ICNAAM), Sept. 23–28, 2019, Rhodes, Greece, AIP CP, Volume 2293 A. AIP Press, American Institute of Physics, Melville, New York, USA, Nov. 2020, pp. 040 002–1 – 040 002–4, ISBN-13: 978-0-7354-4025-8, DOI: 10.1063/5.0026478.

[23] "UDC Summary Linked Data, Main Tables," 2020, Universal Decimal Classification (UDC), UDC Consortium, URL: https://udcdata.info/078887 [accessed: 2021-11-28].

[24] "Multilingual Universal Decimal Classification Summary," 2012, UDC Consortium, 2012, Web resource, v. 1.1. The Hague: UDC Consortium (UDCC Publication No. 088), URL: http://www.udcc.org/udcsummary/php/index.php [accessed: 2021-11-28].

[25] "Creative Commons Attribution Share Alike 3.0 license," 2012, URL: http://creativecommons.org/licenses/by-sa/3.0/ [accessed: 2021-11-28], (first release 2009, subsequent update 2012).

[26] B. M. Das, Advanced Soil Mechanics; Fifth Edition, 5th ed. CRC Press, Taylor & Francis Group, Boca Raton, London, New York, 2019, ISBN: 978-0-8153-7913-3.

[27] "Land Processed Distributed Active Archive Center," 2021, LP DAAC, URL: https://lpdaac.usgs.gov/ [accessed: 2021-11-28].

[28] C.-P. Rückemann, R. Pavani, B. Gersbeck-Schierholz, A. Tsitsipas, L. Schubert, F. Hülsmann, O. Lau, and M. Hofmeister, Best Practice and Definitions of Formalisation and Formalism. Post-Summit Results, Delegates' Summit: The Ninth Symp. on Adv. Comp. and Inf. in Natural and Applied Sciences (SACINAS), The 17th Int. Conf. of Num. Analysis and Appl. Math. (ICNAAM), Sept. 23–28, 2019, Rhodes, Greece, 2019, pp. 1–16, DOI: 10.15488/5241.

[29] F. de Saussure, Cours de linguistique générale, 1916, (title in English: Course in General Linguistics), C. Bally and A. Sechehaye (eds.).

[30] "Perl Compatible Regular Expressions (PCRE)," 2021, URL: https://www.pcre.org/ [accessed: 2021-11-28].

[31] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, Numerical Recipes in Fortran 90, The Art of Parallel Scientific Computing. Cambridge University Press, 1996.

[32] T. Vincenty, "Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations," Surv. Rev., vol. XXII, no. 176, 1975, pp. 88–93.

[33] R. J. Renka, "Algorithm 772: STRIPACK: Delaunay Triangulation and Voronoi Diagram on the Surface of a Sphere," AMC Trans. Math. Software, vol. 23, no. 3, 1997, pp. 416–434.

[34] B. K. P. Horn, "Hill-Shading and the Reflectance Map," Proceedings of the IEEE, vol. 69, no. 1, 1981, pp. 14–47.

[35] A. J. Kent and P. Vujakovic, Eds., The Routledge Handbook of Mapping and Cartography. Taylor & Francis Group, London and New York, 2018, ISBN: 978-1-138-83102-5.

[36] C.-P. Rückemann, "Archaeological and Geoscientific Objects used with Integrated Systems and Scientific Supercomputing Resources," Int. Jour. on Adv. in Systems and Measurements, vol. 6, no. 1&2, 2013, pp. 200–213, ISSN: 1942-261x, URL: http://www.thinkmind.org/download.php?articleid=sysmea_v6_n12_2013_15 [accessed: 2021-11-28].

[37] C.-P. Rückemann, "Information Science and Inter-disciplinary Long-term Strategies – Key to Insight, Consistency, and Sustainability: Conceptual Knowledge Reference Methodology Spanning Prehistory, Archaeology, Natural Sciences, and Humanities," International Tutorial, DataSys Congress 2020, Sept. 27 – Oct. 1, 2020, Lisbon, Portugal, 2020, URL: http://www.iaria.org/conferences2020/ProgramINFOCOMP20.html [accessed: 2021-11-28].

[38] "The Prehistory and Archaeology Knowledge Archive (PAKA) license," 2021, (release 2021), Unabhängiges Deutsches Institut für Multidisziplinäre Forschung (DIMF): All rights reserved. Rights retain to the contributing creators.

[39] R. Brinkmann, Abriß der Geologie, Bd. 2, Historische Geologie, neubearbeitet von Karl Krömmelbein, 12th/13th ed. Ferdinand Enke Verlag, Stuttgart, 1986, ISBN: 3-432-80603-5.

[40] P. Woldstedt and K. Duphorn, Norddeutschland und angrenzende Gebiete im Eiszeitalter, 3rd ed., 1974.

[41] G. N. Bailey, J. Harff, and D. Sakellariou, Eds., Under the Sea: Archaeology and Palaeolandscapes of the Continental Shelf, ser. Coastal Research Library. Springer International Publishing AG 2017, 2017, vol. 20, ISSN: 2211-0577, ISBN: 978-3-319-53158-8, DOI: 10.1007/978-3-319-53160-1.

[42] "Network Common Data Form (NetCDF)," 2021, DOI: 10.5065/D6H70CW6, URL: http://www.unidata.ucar.edu/software/netcdf/ [accessed: 2021-11-28].

[43] "World reference base for soil resources," 1998, World Soil Resources Report No. 84, ISSS-ISRIC-FAO, Rome.

[44] "World reference base for soil resources," 2006, World Soil Resources Report No. 103, ISSS-ISRIC-FAO, Rome.

[45] C.-P. Rückemann, "Methodology of Chorological and Coherent Conceptual Knowledge Contextualisation: Approaches for Multi-disciplinary Contexts in Prehistory and Archaeology," in Proceedings of The Thirteenth International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2021), July 18 – 22, 2021, Nice, France. XPS Press, Wilmington, Delaware, USA, 2021, ISSN: 2308-393X, ISBN-13: 978-1-61208-871-6, URL: http://www.thinkmind.org/download.php?articleid=geoprocessing_2021_1_60_30045 [accessed: 2021-11-28].

[46] J. Cook, The Swimming Reindeer, ser. British Museum Objects in Focus. The British Museum Press, 2015, ISBN: 978-0-7141-2821-4, (first published 2010).

[47] C.-P. Rückemann, "Approaches to Coherent Conceptual Knowledge Integration for Prehistory, Archaeology, Natural Sciences, and Humanities: Information Science Based Computation of Structural Knowledge and Spatial Context Information," International Journal on Advances in Software, vol. 14, no. 1&2, 2021, (to be published).

[48] P. Wessel, "The Digital Chart of the World for GMT 5 or later," 2018, URL: ftp://ftp.soest.hawaii.edu/gmt [accessed: 2021-11-28], URL: http://www.soest.hawaii.edu/pwessel/dcw/ [accessed: 2021-11-28].

[49] P. Wessel, "Global Self-consistent Hierarchical High-resolution Geography," 2016, URL: ftp://ftp.soest.hawaii.edu/gmt [accessed: 2021-11-28], URL: http://www.soest.hawaii.edu/pwessel/gshhg/ [accessed: 2021-11-28].

[50] P. Wessel and W. H. F. Smith, "A global, self-consistent, hierarchical, high-resolution shoreline database," J. Geophys. Res., vol. 101, no. B4, 1996, pp. 8741–8743.

[51] E. A. Soluri and V. A. Woodson, "World Vector Shoreline," Int. Hydrograph. Rev., vol. LXVII, no. 1, 1990, pp. 27–35.

[52] A. J. Gorny, "World Data Bank II General User Guide," 1977.

[53] P. Wessel, W. H. F. Smith, R. Scharroo, J. Luis, and F. Wobbe, "The Generic Mapping Tools (GMT)," 2020, URL: http://www.generic-mapping-tools.org/ [accessed: 2021-11-28], URL: http://gmt.soest.hawaii.edu/ [accessed: 2021-11-28].

[54] C.-P. Rückemann, "Keynote Lecture: The Multi-disciplinary Case of Prehistorical Insight: Information Science at the Edge of Structured Data Comprehension," The XI-th International Conference - Mathematics of Informational Modeling, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, July 1–2, 2021, Varna, Bulgaria, URL: http://math.bas.bg [accessed: 2021-11-28], 2021.

[55] C.-P. Rückemann, R. Pavani, Z. Kovacheva, B. Gersbeck-Schierholz, F. Hülsmann, and I. Naydenova, Best Practice and Definitions – Concepts of Cognostic Addressing Structured and Non-structured Data. Post-Summit Results, Delegates' Summit: Best Practice and Definitions – Concepts of Cognostic Addressing Structured and Non-structured Data, The Eleventh Symposium on Advanced Computation and Information in Natural and Applied Sciences (SACINAS), The 19th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM), September 20–26, 2021, Rhodes, Greece, 2021, (to appear).

[56] C.-P. Rückemann, "Coherent Knowledge Solutions From Prehistory to Future – Towards Coherent Multi-disciplinary Knowledge Reference Implementation Blueprints for Industrial Learning: Insight from Consistent Coherent Conceptual Integration of Prehistory, Archaeology, Natural Sciences, and Humanities," ML4I – Machine Learning for Industry Forum 2021; hosted by the High-Performance Computing Innovation Center (HPCIC) and Data Science Institute (DSI), at the Lawrence Livermore National Laboratory (LLNL), August 10–12, 2021, Livermore, California, U.S.A., (Invited Speech), URL: http://www.llnl.gov [accessed: 2021-11-28], 2021.

Clinopyroxene/Melt Partitioning: Models for Higher Upper Mantle Pressures Applied to Sodium and Potassium

Julia Marleen Schmidt
Institute of Geological Sciences)
Freie Universität Berlin
Berlin, Germany
email: julia.schmidt@fu-berlin.de

Lena Noack
Institute of Geological Sciences
Freie Universität Berlin
Berlin, Germany
email: lena.noack@fu-berlin.de

***Abstract*—Mineral/melt partition coefficients describe the redistribution of trace elements during melting processes. They are highly influenced by changing pressure, temperature, and composition. We find that for sodium, partition coefficients rise in the order of two magnitudes from 0-15 GPa along the peridotite solidus temperature. Still, because of lacking high-pressure experimental data and the limited pressure range of most partitioning models, mantle and crust evolution models normally rely on constant partition coefficients. This is especially unfavorable for models which want to explore the thermal evolution of a terrestrial planet by tracing the heat producing elements. Based on a thermodynamic model, we parametrise sodium partitioning equations which are applicable from 0-12 GPa and 1410-2350 K. With the help of the sodium partitioning model, we calculate clinopyroxene/melt partition coefficients for the heat producing element potassium. Knowing both, the sodium and potassium partitioning behaviour, will help us to improve elemental redistribution calculations and therefore also thermal evolution models.**

Keywords*— *high-pressure partition coefficients; partial melting; thermodynamic modelling; trace element redistribution; radiogenic heat sources

I. Introduction

Inside the Earth and other terrestrial planets, partial melts usually control the redistribution of trace elements from the mantle to the crust. If melt is buoyant due to a lower density than the surrounding rock, it moves upwards and transports the previously incorporated incompatible trace elements towards the surface. This leads to enriched reservoirs near the surface and depleted upper mantle areas. Using thermodynamic models, the incorporation of an incompatible element can be predicted as has been outlined in [1] based on sodium as a reference for other trace elements. Expanding the redistribution concept to a larger number of elements (as is shown here exemplarily for the 1+ charge element potassium) leads to the creation of trace element patterns, which are unique geochemical markers of the events that led to the exact appearance and compositions of a rock of interest [2][3][4].

During melting, partition coefficients describe if an element prefers to be incorporated into a mineral or, if it is incompatible, into melt. Partitioning data for peridotite including the minerals olivine, orthopyroxene, and clinopyroxene show that in mantle rocks, clinopyroxene is the mineral taking in most incompatible trace elements [5][6]. This makes clinopyroxene the mineral with the largest influence on trace element redistribution.

Since partition coefficients of sodium between clinopyroxene and silicate melt are highly affected by changing pressure and temperature, partition coefficient models should consider these parameters. Because of analytical limitations, there are only few measured partition coefficients available for high-pressure conditions. Additionally, partition coefficient models are often only applicable for rather low pressures. This is the reason why so far, partial melt simulations in some mantle evolution models either neglected partition coefficients and chose pre-defined elemental abundances in mantle melts [7], or settled for constant experimentally derived and estimated partition coefficients [8][9][10][11].

In 1937, Goldschmidt postulated that both the matching size and charge of the mineral's lattice site and element of interest determine an element's ability to partition into a mineral [12]. This leads to the assumption that not only it is possible to determine the P-T sensitive partition coefficients experimentally, but also by means of numerical modelling. Based on a model of Brice (1975) [13], Blundy et al. (1995) [14] developed a quantitative model and determined a parametrised fit function for clinopyroxene/melt partition coefficients for sodium in the range of 0-4 GPa and 1000-1800°C. In Schmidt and Noack (2021) [1], this temperature and pressure range was already extended to higher pressure conditions, leading to a general parametrisation over a broad temperature range and pressures from 4-12 GPa. Here, we derive new parametrisations from surface to upper mantle conditions focusing on specific melt partitioning scenarios related to both formation and freezing of melt.

As is shown in Figure 1, partition coefficients of trace elements have a parabolic relationship in an Onuma diagram. The parabolas' curves change in broadness and shift along the x-axis depending on the bulk modulus E and the ideal lattice site radius r_0, respectively. Because of this relationship, it is

possible to calculate other trace elements' partition coefficients based on the reference coefficient D_{Na}.

The aim of this study is to implement a thermodynamic model for $D_{Na}^{cpx/melt}$ for a large range of pressures suitable for mantle melt simulations. Furthermore, we parametrise a scaling law from the mentioned model that is applicable for upper mantle pressures. This will open the possibility to include partition coefficients depending on pressure and temperature into mantle evolution models and to acquire more realistic model results [15].

After proving that both the thermodynamic models and the scaling laws are applicable, we derive clinopyroxene/melt partition coefficients for potassium from them. As one of the most abundant radiogenic elements, potassium is one of the largest contributors of heat inside the Earth [16][17]. Being able to calculate pressure and temperature dependent partition coefficients of potassium in clinopyroxene/melt provides us therefore an opportunity to make thermal evolution and mantle melting models more realistic.

The rest of the paper is organized as follows. In section II, we describe the methods used for our parametrisation, the thermodynamic model, and potassium partitioning in the appropriate subsections. Section III is dedicated to the results for the thermodynamic model, the parametrisations, and the application to potassium partitioning respectively. Section IV discusses the results and compares them to experimental data. Section V concludes the paper, summarizes the findings, and points out what can be done in future works.

II. Methods

A. Thermodynamic Model

To obtain the most realistic trace element partitioning results for mantle melting processes, it is useful to determine a strain-compensated partition coefficient D_0 [18]. Strain-compensation means that the redistributed element has the same charge and size as the regular element on the lattice site of the crystal [13]. Therefore, it is assumed that the radius of the element of interest r_i equals the lattice site radius r_0. This way the partition coefficient D_i can be calculated with the following equation:

$$D_i = D_0 \exp\left(\frac{-4\pi E_{M2} N_A (\frac{r_0}{2}(r_i - r_0)^2 + \frac{1}{3}(r_i - r_0)^3)}{RT}\right), \quad (1)$$

where E_{M2} is the bulk modulus of the M2 lattice site, N_A the Avogadro's number, R the gas constant and T the temperature. Here, the lattice site of interest is the crystallographic M2 site of clinopyroxene. Sodium is used as the strain-compensated partition coefficient D_0 because it is the 1+ charge element that has a radius r_i which is closest to the M2 lattice site radius r_0. Therefore, we assume $D_0 = D_{Na}$.

To calculate the strain-compensated partition coefficient D_0, we took a thermodynamic approach described by Blundy et al. (1995) [14]. For this, we used the melting curve of jadeite and linked it to the activity of jadeite's components in the melt. To calculate the partition coefficients of sodium between clinopyroxene and silicate melt, jadeite was the clinopyroxene of choice because of both, its high concentration in sodium and its ability to mix nearly ideally with diopside [19] and enstatite.

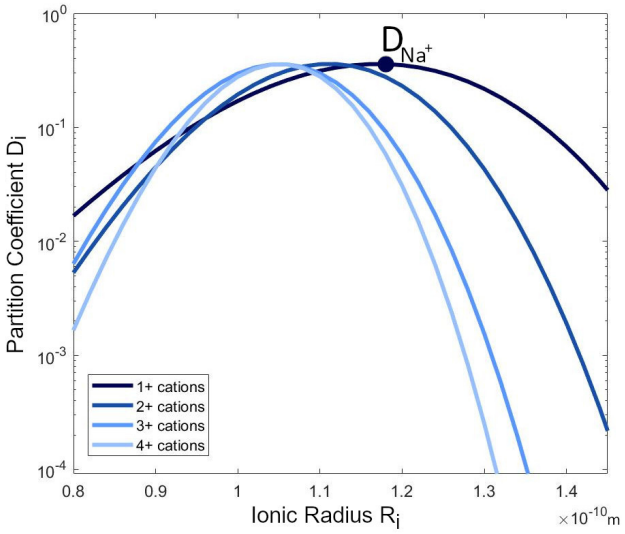


Figure 1. Partition coefficients of 1+ to 4+ charged cations in the clinopyroxene M2 lattice site. Modelled after Wood et al. (1997) [21]. Here, D_{Na} equals D_0.

Generally, partition coefficients are given as the weight fraction ratio

$$D_i = \frac{X_i^{crystal}}{X_i^{melt}}, \quad (2)$$

with X_i being the weight percentage of the given component. However, since we obtain partition coefficients with the help of thermodynamic properties, we calculate molar partition coefficients. Based on experiments on plagioclase/fluid partitioning [20], it is broadly assumed that also for other minerals molar partition coefficients nearly equal weight percentage coefficients [21]. Thus, we can make use of Flood's equation for exchange equilibria in molten salts [22]:

$$RT\ lnK_f = \sum N_i \Delta G^0, \quad (3)$$

where K_f is the molar equilibrium constant, N_i the concentration of the component and ΔG^0 the Gibbs' free energy of change. Because we use a thermodynamic approach for our calculations, N_i can be neglected and is set to one. The molar partitioning of jadeite D^{jd} can be expressed by using K_f in the expression

$$D_{jd} = \frac{1}{K_f} = \frac{n_{jd}^{cpx}}{n_{jd}^{melt}}, \quad (4)$$

where n_{jd}^{cpx} and n^{melt} are the mole of the solid clinopyroxene and molten components, respectively. Here, we assume an ideal case where the activity equals the mole fraction of the elements of interest. Thus, we assume $n_{jd}^{cpx} = X_{Na}^{cpx}$ and

$n_{jd}^{melt} = X_{Na}^{melt}$. Taking (4) and rearranging (3) leads to the following equation:

$$D_{Na}^{cpx/melt} = \exp\left(\frac{\Delta G_{f(P,T)}}{RT}\right). \quad (5)$$

Besides (5), one more equation is needed to determine D_{Na} thermodynamically. Equation (6) relates the entropy of the melting reaction to its heat capacity and melting point T_f of jadeite at the pressure P_f=0.0001 GPa. By including the enthalpy, we determine the Gibb's free energy change of reaction with the following equation:

$$\begin{aligned}\Delta G_{f(P,T)} =& \Delta H_{f(P_f,T_f)} + \int_{T_f}^{T} \Delta C_p dT \\ & - T\left(\Delta S_{f(P_f,T_f)} + \int_{T_f}^{T} \frac{\Delta C_p}{T} dT\right) \\ & + \int_{P_f}^{P} \Delta V_{f(T)} dP, \end{aligned} \quad (6)$$

where $\Delta H_{(f(P_f,T_f)}$ is the entropy, $\Delta S_{f(P_f,T_f)}$ the enthalpy and ΔC_p the difference between the local P,T values and P_f,T_f. $\Delta V_{f(T)}$ is the molar volume of fusion at pressure P_f extrapolated to the temperature of interest [14]. We calculate D_{Na} by inserting (6) into (5).

To be able to solve (6), we use the Maier-Kelly power function [23] to calculate the heat capacity C_p (as stated in Table I). With the given parameters we can solve the first two integrals in (6) analytically, by evaluating the heat capacity of the solid at temperature T and the melt heat capacity at T_f. To solve the third part of the integral in (6) we use a Riemann integral, where we approximate the intervals on subintervals defined over a pressure vector of n=1000 steps from P=0.0001 to 15 GPa and assume that the volumes of melt and solid are constant in each subinterval.

$$\begin{aligned}\int_{P_f}^{P} \Delta V_{f(T)} dP =& \sum_{i=1}^{n-1} \int_{P_i}^{P_{i+1}} \Delta V_{f(T),i} dP \\ \approx& \sum_{i=1}^{n-1} \Delta V_{f(T),i} \int_{P_i}^{P_{i+1}} dP \\ =& \sum_{i=1}^{n-1} \Delta V_{f(T),i} \left(P_{i+1} - P_i\right) \end{aligned} \quad (7)$$

where $\Delta V_{f(T),i}$ is the volume of fusion, i.e., the volume difference between melt and solids. The calculation of V_{solid} and V_{melt} is done separately using an isothermal Birch-Murnaghan equation of state:

$$P = 1.5K_T\left(x^{\frac{7}{3}} - x^{\frac{5}{3}}\right)\left(1 + 0.75(K' - 4)\left(x^{\frac{2}{3}} - 1\right)\right), \quad (8)$$

where $x = V^0/V$. K_T is the bulk modulus for pyroxene and K' its derivative. The volumes further depend on temperature via

$$V(T) = V \cdot \exp\left(\int_{T_0}^{T} \alpha^0(T) dT\right). \quad (9)$$

Here, we assume that the thermal expansion coefficient α^0 is constant for melts and solids (see Table I), hence

$$\begin{aligned}\Delta V_{f(T),i} =& \\ & V_{melt}(P_i) \cdot \exp\left(a_{melt}^0 (T - T_f)\right) \\ & - V_{solid}(P_i) \cdot \exp\left(a_{solid}^0 (T - T_0)\right). \end{aligned} \quad (10)$$

T_0 is set to 298 K. The partition coefficient D_{Na} is then derived for varied pressures and temperatures by inserting ΔG into (5).

TABLE I
THERMODYNAMIC INPUT DATA FOR LIQUID AND SOLID JADEITE AND COMPARISON OF MODEL PARAMETERS.

Parameters	Pyroxene	Melt	Units
V^0	60.4	79.9	kJ GPa^{-1}
α^0	2.81·10^{-5}	6.28·10^{-5}	K^{-1}
K_T	141.2	14.8	GPa
$\partial K/\partial T$	-0.025	-0.0015	GPa K^{-1}
K'	4.5	4.5	
$\Delta H_{f(0.1,T_f)}$	61.1±1.3		kJ mol^{-1}
$\Delta S_{f(0.1,T_f)}$	64.8±0.6		J mol^{-1}
T_f	943±22		K
C_p^a			
1	0.30113	0.28995[b]	kJ mol^{-1} K^{-1}
2	1.0143· 10^{-5}		kJ mol^{-1} K^{-1}
3	-2239.3		kJ mol^{-1} K^{-1}
4	-2.0551		kJ mol^{-1} K^{-1}

Parameters as in Blundy et al. (1995) [14] at 0.0001 GPa and 298 K (unless stated otherwise).
[a] $C_p = C_{p1} + C_{p2} \cdot T + C_{p3} \cdot T^{-2} + C_{p4} \cdot T^{-0.5}$ (T in kelvins).
[b] At temperatures >1200 K.

B. Scaling Law Development

The thermodynamic model indicates that partition coefficients are not only dependent on pressure, but temperature and therefore melt fraction as well. Thus, to develop a mantle melt fit function for the reference coefficient D_{Na}, we calculated partition coefficients between 0 and 15 GPa along the peridotite solidus temperature, peridotite liquidus temperature, and corresponding melt fraction F temperatures. The respective partition coefficients were determined with the thermodynamic approach described above. Melt fraction F is included because as is suggested by the batch melting equation [24], a rising melt fraction F leads to a decrease in the total amount of incompatible trace elements in the melt. Therefore, we arranged the melt fractions in $F = 0.2$ increments while neglecting the extraction of the melt, by using

$$F(P,T) = \frac{T - T(P)_{sol}}{T(P)_{liq} - T(P)_{sol}}. \quad (11)$$

The solidus T_{sol} and liquidus temperatures T_{liq} equations which we used [25] are third-order fits to experimental data

[26][27]:

$$T_{sol,ini} = 1409K + 134.2\frac{K}{GPa}\cdot P - 6.581\frac{K}{GPa^2}\cdot P^2 + 0.1054\frac{K}{GPa^3}\cdot P^3, \quad (12)$$

$$T_{liq,ini} = 2035K + 57.46\frac{K}{GPa}\cdot P - 3.487\frac{K}{GPa^2}\cdot P^2 + 0.0769\frac{K}{GPa^3}\cdot P^3. \quad (13)$$

For pressures from 0-15 GPa, the solidus temperatures range from 1409-2297 K, while the liquidus temperatures go from 2035-2372 K. The pressure step size for each temperature profile is 0.1 GPa, which corresponds to 151 D_{Na}-P-T data points for each melt temperature profile. With the least square function, these data points were fitted to a function with parameters a, b, c, d, e, and f (14) in Python 3. Because the function has the same form as the fit function of Blundy et al. (1995) [14], their fit parameters are compared in Table II.

$$\mathrm{D}_{Na}(\mathrm{T[K],P[GPa]}) = \exp\left(\frac{a + b\cdot P - c\cdot P^2}{T} - d + e\cdot P - f\cdot P^2\right). \quad (14)$$

By including the melt fraction P-T profiles, we make sure that the resulting model function will satisfy a broader range of P-T conditions and is not only valid for modelling partial melting in the mantle (where temperatures are close to the solidus), but also for crystallisation of melt (where temperatures are close to the liquidus). In contrast to our model, partition coefficients for crystallising liquid were often determined by taking only the liquidus temperatures into account [28][29][30].

Depending on the application (e.g magma ocean crystallisation or partial melt models), it may be advantageous to consider a narrower temperature/melt degree range, parametrised with a higher accuracy. For these parametrisations, we again used equations which calculate the respective temperatures from pressures in the range of 0-15 GPa. The stepsize of the pressure is again 0.1 GPa. For the parametrisation, we varied the temperature by $\pm$50 K along the melting curve to optimise the accuracy if there are smaller temperature diversions from the solidus (12) and liquidus (13) equations.

For both the thermodynamic model and the fit function development, replication data is accessible in the TRR170-DB [31].

C. Application to Other Trace Elements

To be able to model clinopyroxene/melt partition coefficients for anhydrous melts on the basis of D_{Na}, we mainly based our calculations on previous studies [21][32]. In these calculations, the melt composition is added to determine the bulk modulus E and lattice site radius r_0. The calculations are applicable to the M2 lattice site of clinopyroxene and ions with a 1+ charge. In the literature model [21], the method iterates the temperature over a fixed pressure until the clinopyroxene saturation in the melt is reached. For our model, however, we did not take the clinopyroxene saturation into account in favor of illustrating the partition coefficient without temperature adjustments for a fixed pressure.

As a first step in the model, the melt composition is given in weight% oxides (S_iO_2, T_iO_2, Al_2O_3, $Cr2O_3$, FeO, MnO, MgO, CaO, Na_2O, K_2O). As in [21], we consider iron only in the form of FeO. After adding temperature (in K) and pressure (in GPa) to the code, we converted the melt composition to moles and calculated the mol amounts of each cation to a total of 6 oxygen atoms in the melt. The resulting mol amounts are then used to determine the amount of the mineral components [21]:

$$Na = NaAlSi_2O_6\ (JD)$$

$$K = KAlSi_2O_6\ (KT)$$

$$Ti = CaTiAl_2O_6\ (CT)$$

$$(Al - 2Ti - Na - K) = CaAl_2SiO_6\ (CATS)$$

$$Ca - CaAl_2SiO_6 - CT = Ca(Mg, Fe, Mn)Si_2O_6\ (DI)$$

$$\tfrac{1}{3}[Mg + Fe + Mn - Ca(Mg, Fe, Mn)Si_2O_6] = (Mg, Fe, Mn)_3Si_{1.5}O_6\ (OL)$$

$$\tfrac{1}{3}[Si - 2K - 2Na - 2Ca(Mg, Fe, Mn)Si_2O_6 - CaAl_2SiO_6 - 1.5(Mg, Fe, Mn)_3Si_{1.5}O_6] = Si_3O_6\ (QZ).$$

Next, we calculated the model fraction of the jadeite component $NaAlSi_2O_6$ (X_{jd}^{cpx}) in the crystal with the following equation [14][21]:

$$\mathrm{X}_{jd}^{cpx}(T[K], P[GPa]) = JD\cdot\exp\left(\frac{10367 + 2100P - 165P^2}{T}\right) - 10.27 + 0.358P - 0.0184P^2. \quad (15)$$

To calculate the Ti content of the clinopyroxene M1 site X_{Ti}^{cpx}, we use the approximate relationship between CT in the liquid and X_{jd}^{cpx}:

$$X_{Ti}^{cpx} = CT\cdot(0.374 + 1.5X_{jd}^{cpx}). \quad (16)$$

We calculate the Cr content in the crystal M1 site X_{Cr}^{cpx} similarly:

$$X_{Cr}^{cpx} = 5Cr. \tag{17}$$

The activity of the $CaAl_2SoO_6$ component in clinopyroxne γ_{CaTs}^{cpx} is calculated from the product Ca for 6 oxygens in the liquid and Al per 6 oxygens:

$$\gamma_{CATS}^{cpx} = (CT + CATS + DI) \cdot (CATS + JD + KT) \cdot exp\left(\frac{76469 - 62.078T + 12430P - 870P^2}{8.314T}\right). \tag{18}$$

To be sure that pyroxene is not unrealistically aluminous, we apply an upper limit for the activity γ_{CaTs}^{cpx} of 0.4. The activity of the diopside component $CaMgSi_2O_6$ (DI) of the liquid γ_{Di}^{cpx} is calculated with the help of a parametrised parabolic pressure dependence of volume:

$$\gamma_{Di}^{cpx} = DI \cdot Mg\#_{melt} \cdot exp\left(\frac{132650 - 82.152T + 13860P - 1130P^2}{8.314T}\right). \tag{19}$$

With the activities calculated above, we calculate the Ca content of the M2 lattice site X_{Ca}^{M2}:

$$X_{Ca}^{M2} = \frac{\gamma_{CaTs}^{cpx} Mg\#_{melt} + \gamma_{Di}^{cpx}}{Mg\#_{melt}(1 - X_{Cr}^{cpx} - X_{Ti}^{cpx})}. \tag{20}$$

With the help of the activity of the CATS component γ_{CaTs}^{cpx} and the Ca content of the M2 lattice site X_{Ca}^{M2}, we are now able to calculate the lattice site radius r_0 [14]:

$$\mathrm{r_0} = 10^{-10}. \tag{21}$$

$$\left(\left(\frac{(0.974 + 0.067X_{Ca}^{M2})\ 0.051\gamma_{CaTs}^{cpx}}{X_{Ca}^{M2}}\right) + 0.12\right)[m].$$

In turn, the bulk modulus E is calculated as follows [14]:

$$E = (318.6 + 6.9P - 0.036T) \cdot \frac{1}{3}\ [GPa]. \tag{22}$$

For 1+ charge ions D_i^{1+}, (1) is slightly adjusted:

$$D^{1+} = D_{Na} \exp\left(\frac{-4\pi E_{M2} N_A (\frac{r_0}{2}(r_0^2 - r_i^2) + \frac{1}{3}(r_i^3 - r_0^3))}{RT}\right). \tag{23}$$

A detailed description of the adaptations to calculate partition coefficients for higher charges is also available in the literature [21][32]. For the model developed in this study, we used the programming language MATLAB.

III. Results

A. Thermodynamic Model Along Melting Temperature Profiles

In a pressure range between 0 and 15 GPa and a temperature range between 1400 and 2400 K, the thermodynamic model produces increasing partition coefficients from low P/high T (0 GPa/2400 K) to high P/low T (10-12 GPa/1400 K) conditions (Figure 2). The increase is in the order of four magnitudes and coincides with experimental observations [33][34]. At higher pressure above 12 GPa and low temperatures, the model starts to invert the trend and the partition coefficients decrease with increasing pressure. However, for our fit function development we have used the P-T space between the solidus and liquidus, which is not affected by this inversion. Along the melting temperature profiles, increasing pressure always leads to increasing partition coefficients, while increasing temperatures cause the partition coefficients to decrease (Figure 2).

In Table II, we compare the thermodynamic model results with experimental literature data. It is notable that most of the experimental data fit well to the model results. On average, the coefficients calculated by the thermodynamic model deviate from the experimental data by 26%. The best fitting value deviates only 2.9% from the experiments at 0.0001 GPa and 1526 K, while the worst fit deviates by 46.9% at 0.0001 GPa and 1524 K. The implications of this variance will be discussed in section IV.

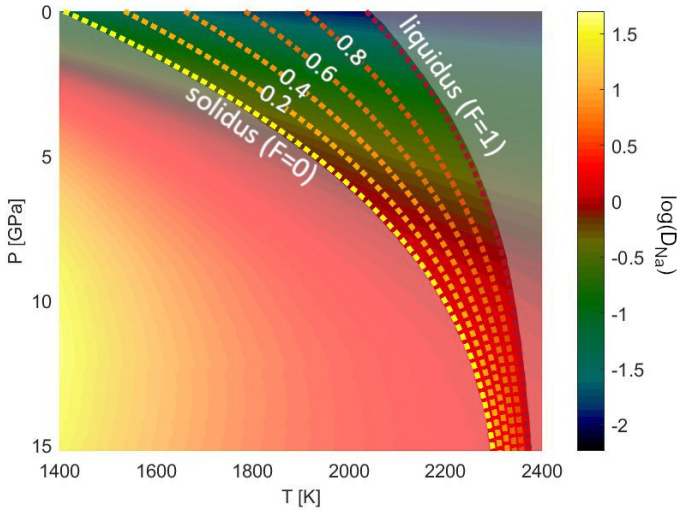


Figure 2. Clinopyroxene/melt partition coefficients for sodium in a P-T space, calculated thermodynamically with the methods of Blundy et al (1995) [14]. Solidus (12), liquidus (13) and intermediate melt fraction outlines are calculated from de Smet (1999) [25].

B. Parametrisation of Thermodynamic Model Results

As described in section II-B, we developed a scaled equation along the mantle peridotite solidus, liquidus, and corresponding melt fraction temperatures in between. For this, we used the least square function to fit the thermodynamic

TABLE II
COMPARISON OF EXPERIMENTAL DATA FOR CLINOPYROXENE/MELT PARTITION COEFFICIENTS WITH THE THERMODYNAMIC DATA

P [GPa]	T [K]	D_{Na} (experimental data)	Ref.	therm. model (this study)	Fit values by Blundy et al. (1995) [14]	Fit[a] (this study)
0.0001	1370	0.050(4)	[14]	0.0696	0.0670	0.0507[b]
0.0001	1449	0.042(4)	[14]	0.0496	0.0444	0.0465[b]
0.0001	1524	0.070(19)	[14]	0.0372	0.0312	0.0432[b]
0.0001	1526	0.040(4)	[14]	0.0369	0.0309	0.0431[b]
0.0001	1526	0.038(5)	[14]	0.0369	0.0309	0.0431[b]
0.0001	1598	0.046(5)	[14]	0.0288	0.0228	0.0404[b]
1	1663	0.075(7)	[14]	0.0891	0.0794	0.0991[b]
2	1773	0.113(8)	[14]	0.1708	0.1680	0.1827[b]
2	1843	0.144(13)	[14]	0.1363	0.1247	0.1587[b]
3	1938	0.237(30)	[14]	0.2241	0.2170	0.2465[b]
6	2038	0.52(12)	[14]	0.7789	0.6507*	0.7774
1.2	1588	0.225(5)	[35]	0.1497	0.1494	0.1366[b]
1.2	1458	0.221(5)	[35]	0.2671	0.3039	0.1807[b]
1.6	1643	0.283(4)	[35]	0.1866	0.1927	0.1750[b]

Parameter	Fit [14] 0-4 GPa	Fit[a] (this study) 4-12 GPa
a	10367	2183
b	2100	2517
c	165	-157
d	-10.27	-4.575
e	0.358	-0.5149
f	-0.0184	0.0475

D_{Na} is the experimentally determined weight partition coefficient.
[a] Parameters as in the scaling law fit for various melt fractions (24).
[b] Note that these are extrapolated values (beyond valid P-T range).
Parameters to be inserted into (14).

model results presented in Figure 2 to the following equation:

$$\mathrm{D}_{Na}(\mathrm{T[K]},\mathrm{P[GPa]}) = \exp\left(\frac{2183 + 2517P - 157P^2}{T} - 4.575 - 0.5149P + 0.0475P^2\right). \quad (24)$$

With the new resulting scaling law (24), it is now possible to calculate partition coefficients for sodium at varied temperature and pressure conditions.

Figure 3 illustrates how the rising D_{Na} values correlate with rising pressure along the solidus temperature. In contrast to the scaling law of Blundy et al. (1995) [14], our resulting scaling law produces steadily rising values up to 13.2 GPa at the respective peridotite solidus temperatures. After this point, the calculated values are starting to decrease. In contrast to our study, the model values of Blundy et al. (1995) [14] start to decline after approximately 5 GPa along the solidus temperature. However, Blundy et al. (1995) [14] themselves state that their scaling law is not to be used over a pressure of 4 GPa.

In Figure 4, the decrease of partition coefficients from the solidus towards the liquidus becomes visible. This indicates that not only pressure, but also temperature and therefore melt

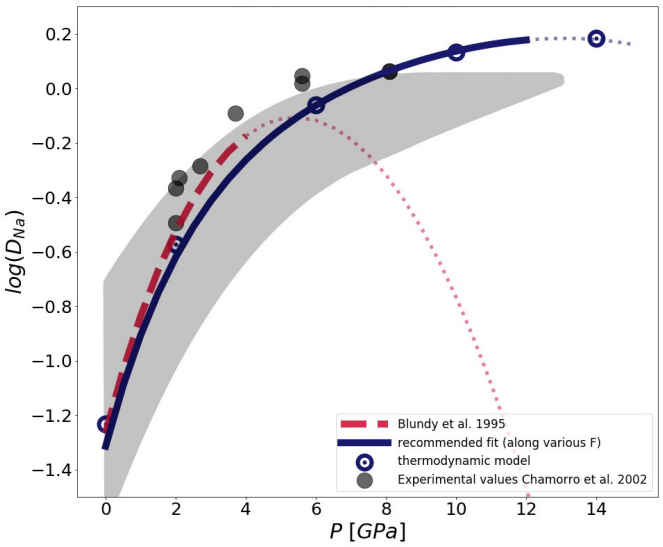


Figure 3. D_{Na} fits along the solidus from Blundy et al. (1995) [14] (red, dashed line) and this study (blue, solid line) compared to experimental data [34] (shaded grey area and grey dots) for comparison. Thin dotted lines are extrapolations of the fit functions beyond the range of validity.

fractions have an impact on the redistribution. It should be noted that at low pressures and high melt fractions, our fit seems to divert more from the thermodynamic model than the previous fit [14]. Consequences of this diversion on the

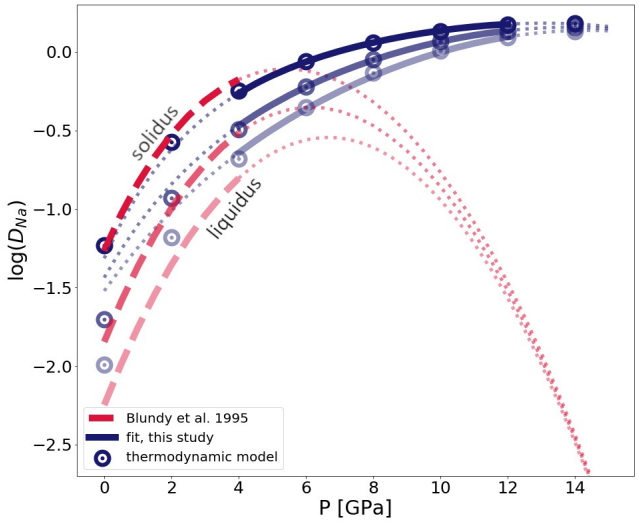


Figure 4. Partition coefficients depending on melt fractions. The line in between liquidus and solidus indicates a melt fraction F=0.5. Thin dotted lines are extrapolations of the fit functions beyond the range of validity.

applicability of our scaling law (24) are explored in section IV.

C. Application to Potassium

Potassium is a prominent radiogenic heat source in the Earth's mantle [16][17] and therefore an important trace element in mantle evolution models. Because potassium is, just like sodium, a 1+ charged element, D_{Na} can be used as an almost strain-free "reference" coefficient.

In Figure 5, we plotted the resulting potassium partition coefficients D_K over solidus P-T conditions with (1) the D_{Na} scaling law which can be found in the literature [14], (2) D_{Na} values modeled from the thermodynamic model, and (3) our D_{Na} scaling law fitted to higher pressures (24). The diagram shows that the D_K calculated with our scaling law (24) only diverts slightly from the D_K values which were calculated with the help of the thermodynamic model D_{Na} values: Below 4 GPa, the variation between the three models is negligible. The largest difference occurs at 12 GPa with a difference of 16.8%.

The scaling law from [14] also shows a good fit to the thermodynamic model in the range of its applicability (up to 4 GPa), but it diverts slightly more from the thermodynamic model than our own scaling law, which was fitted along these thermodynamic model values.

D. Solidus and Liquidus Scaling Laws

To avoid the wide spread between thermodynamic model and scaling law at low pressures (Figure 4), we created two more scaling laws for solidus and liquidus conditions, respectively. As described in section II-B, we added and substracted 50 K along (12) and (13) to ensure no large discrepancies with the thermodynamic model for small temperature perturbations. The results are presented in Figure 6 and the two new scaling laws along solidus and liquidus are, respectively:

$$\mathrm{D}_{Na}^{sol}(\mathrm{T[K],P[GPa]}) = \exp\left(\frac{2508 + 2333P - 138.5P^2}{T} - 4.514 - 0.4791P + 0.0415P^2\right), \quad (25)$$

$$\mathrm{D}_{Na}^{liq}(\mathrm{T[K],P[GPa]}) = \exp\left(\frac{-13728 + 4827P - 237P^2}{T} + 2.496 - 1.540P + 0.0829P^2\right). \quad (26)$$

As is shown in Figure 6, both scaling laws fit well to the thermodynamic model along their respective P-T conditions. However, the scaling laws seem to slightly overestimate the partition coefficients for D_{Na} at low pressures below approximately 4.5 GPa, while they underestimate the partition coefficients for higher pressures. This results in a small tilting of the scaling laws in respect to the thermodynamic model. As a consequence, the largest difference between solidus scaling law (25) and thermodynamic model is 33.5% at 15 GPa. For the liquidus scaling law (26), the largest difference is as well at 15 GPa with 29.8%.

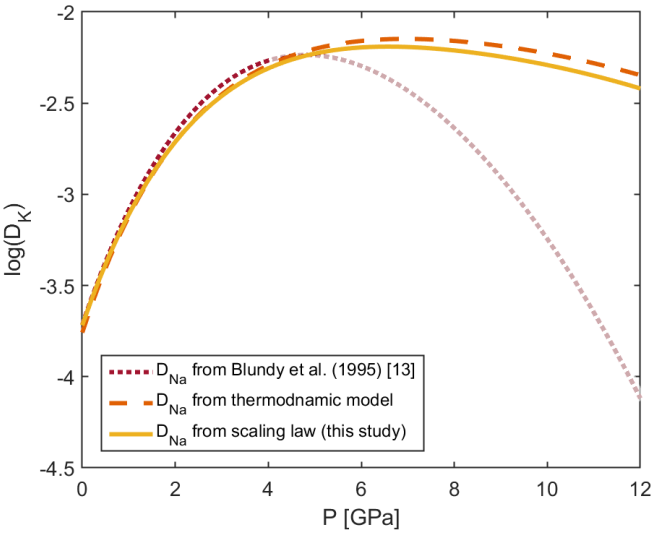


Figure 5. Potassium partition coefficients over solidus P-T conditions. Calculations are based on D_{Na} of either this study (24), thermodynamic model, and a model of Blundy et al. (1995) [14].

TABLE III
MELT STARTING COMPOSITIONS OF EXPERIMENTAL LITERATURE DATA.

	Hauri (1994) [36]	Hauri (1994) [36]	Hart and Dunn (1993) [37]	Klemme (2002) [38]	Salters and Stracke (2004) [39]	Salters and Stracke (2004) [39]
P [GPa]	2.5	1.7	3	3	3	2
T [K]	1703.15	1678.5	1653.15	1673.15	not specified[2]	not specified[2]
Si_O_2	46.7	46.7	47.67	56.7	46.18	57.53
Ti_O_2	0.5	0.5	2.67	6.3	1.45	0.59
Al_2O_3	18.0	18.0	16.06	19.3	25.29	16.28
$Cr2O_3$	0.0	0.0	0.0	0.0	0	0
FeO	8.03	8.03	10.88	0.0[1]	9.72	4.1
MnO	0.15	0.15	0.23	0.0[1]	0[1]	0[1]
MgO	9.52	9.52	5.91	4.2	15.49	5.27
CaO	12	12	9.5	6.8	7	6.99
Na_2O	2.05	2.05	3.92	3.3	2.71	6.87
K_2O	0.07	0.07	1.34	3.3	2.16	2.37
D_K	0.0067	0.0081	0.0072	0.007	0.0001	0.001

Melt composition of the experimental literature values stated in Figure 7 in wt%. For our example calculations, an alkaline basaltic melt composition was used [37].
[1]In order to calculate the partition coefficients with the method described in section II-C, values were adjusted to 0.0001.
[2] The authors did not specify the temperature. For Figure 7, the temperature was set to the corresponding solidus temperature.

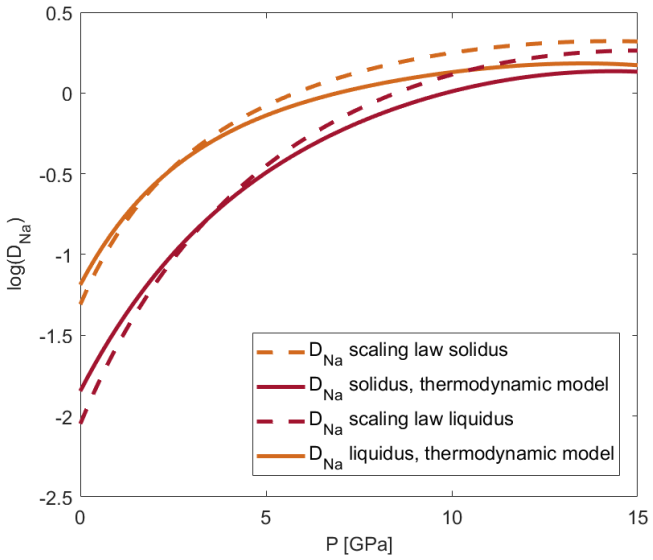


Figure 6. Scaling laws for parametrised solidus and liquidus [25].

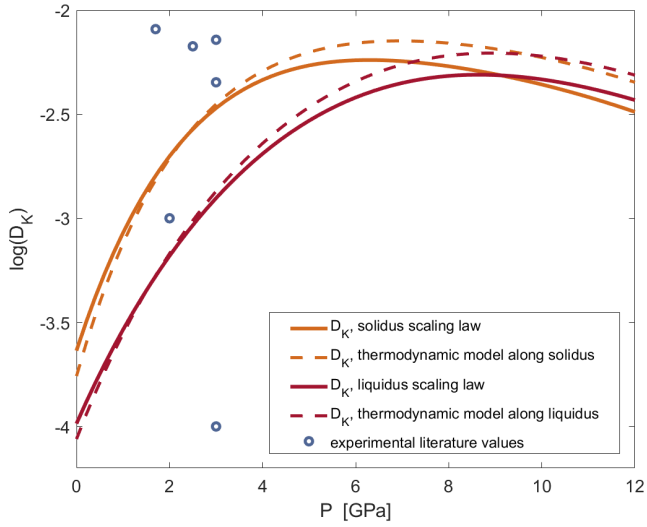


Figure 7. Scaling laws for parametrised solidus and liquidus [25]. Corresponding temperature and melt composition of the experimental literature values can be found in Table III.

In addition to the described D_K calculations in section III-C, we calculated D_K as well with the help of the solidus and liquidus scaling laws (Figure 7). As for D_{Na}, a variation can be observed between calculated D_{Na} from the thermodynamic model and from the scaling law for higher pressures. For the D_K calculations, we have used a literature alkali basaltic melt composition [37].

As a comparison, we plotted experimental literature values for the corresponding pressures against the calculated partition coefficients. The apparent wide spread of the data is discussed below in section IV-C.

The melt compositions of the experimental data as used in our calculations [36][37][38][39] can be found in Table III.

IV. DISCUSSION

A. Sodium Partitioning from low to higher Pressure

Partition coefficients are highly controlled by pressure, temperature, and composition. Since it is not possible for models to include all compositional interactions between a mineral and melt, comparing them to experimental data is sometimes difficult. This is especially true for higher pressures, where partition coefficient data is, to this date, often lacking. However, the existing experimental data for lower pressures already indicate if we can expect realistic results from our calculations. As is shown in Table II, our thermodynamic model results fit well to the experimental results [14][35]. Best matches were achieved for experimental

data at 0.0001 GPa/1526 K, 2 GPa/1843 K and 3 GPa/1938 K [14], where the model data deviates from the experimental values by only 2.9% and for the latter two by 5.4%, respectively. Interestingly, with 46.9%, the second largest diversion from the experimental data is at 0.0001 GPa/1524 K, which is at the same pressure and only 2 K below a very well fitting value (Table II). However, the experimentally derived partition coefficients vary more than what could be solely explained by changing P-T conditions. Deviating experimental and thermodynamic model results can happen because of several effects. For example, analytical error margins have to be considered. Furthermore, the composition of melt as well as solid clinopyroxene can be much more diverse in experiments than it is for thermodynamic model calculations, which does not take these changes into account.

It should be noted that the fit parameters in (14) differ from the parametrisation in Blundy et al. (1995) [14], which is due to the larger parameter space applied in our model. For a comparison of the two scaling laws for various melt fractions, we refer to Figure 3 and Figure 4. A direct comparison of the resulting parameters for our scaling law (24) and the previous study [14] is given in Table II.

Figure 3 illustrates that the partition coefficients calculated with our scaling law (24) over solidus temperatures is close to the experimental sodium partitioning data. This data is indicated by a shaded grey area and, for D_{Na} in the $Ab_{80}Di_{20}$ system, by darker grey dots [34]. A direct comparison with appropriate temperatures in the range of the experimental data [34] shows that here, the thermodynamic model is within an error of on average 32%.

As is shown in Figure 3 and Figure 4, our and the previous fit [14] do not completely agree with each other. Blundy et al. (1995) [14] parametrised their thermodynamically calculated results over a peridotite solidus of McKenzie and Bickle (1988) [40], while we took the peridotite solidus equation of de Smet (1999) [25]. This could have produced a small shift in the D_{Na}-P-T field and, therefore, in the resulting fit function. Also, a different parametrisation method could have been used by the previous study [14], which might have produced slightly different fits. Finally, the previous study [14] concentrated more on low pressure partitioning, whereas we have tried to parametrise the function up to the 410 km discontinuity. The inclusion of higher pressures could also be the reason why at pressures below 4 GPa (especially at higher melt fractions) our scaling law (24) fits less well to the thermodynamic model values than the older model [14].

Between 4 and 12 GPa, our scaling law (24) fits well to the thermodynamic data over all melt fraction values from solidus to liquidus. However, at high melt fractions below 4 GPa, larger diversions from the original fit function [14] appear (Figure 4). The reason for this lies in the nature of the least square parametrisation used in Python 3, which allows the scaling law to divert from the thermodynamic model at low pressures for the expense of being applicable to higher pressures and to a wide range of melt fractions. If we would only parametrise the model over the solidus, the shift would disappear at the expense of the model being applicable to any other P-T conditions. Thus, the included P-T conditions for the varying melt fractions ensure that the model is useful over a wide range of pressures and temperatures, but limits the applicability of the scaling law for higher melt fractions to a range between 4 and 12 GPa.

In the upper mantle, buoyant melt can occur up to the depth where it becomes gravitationally stable and a density inversion of melt and solid surrounding material takes place (the so-called density crossover). In other words, melts formed at higher pressures may not be able to rise to the surface [41][42]. However, this is only true for upper mantle melts and melts rising upwards in lower regions of the Earth can not be ruled out [43]. Inside Earth, the density crossover exists at approximately 11-12 GPa at 2000°C [44], and in Mars between 7 GPa [44] and 7.5 GPa [45]. To include pressure and temperature dependent partition coefficients into a mantle evolution model, it would often suffice to be able to calculate them up to the density crossover. Overlapping with the density crossover, in the Earth's 410 km mantle discontinuity (i.e., pressures of approximately 12-15 GPa), phase changes occur [46] and pyroxene slowly starts to dissolve into a pyrope-rich garnet to form majorite [47]. This and the density crossover indicate that a parametrise up to approximately 12-15 GPa is sufficient.

As the experimental data [33][34] suggests, the partition coefficients of sodium in clinopyroxene increase with temperature and pressure before they remain constant. At solidus P-T conditions, our fit function curve slowly starts to flatten and starts to fall at approximately 13 GPa (Figure 3). This coincides with the thermodynamic model, where (along the solidus) the partition coefficients start to decrease at 12.5 GPa. Combining these findings with the occurence of a density crossover and transition zone at approximately 12 GPa, we suggest to not use our scaling law (24) above this pressure.

As is discussed above, the scaling law (24) works well for melting P-T conditions between the peridotite solidus and liquidus between 4 and 12 GPa. Thus, it can be considered as a useful expansion of the previous scaling law for D_{Na} [14]. Because of the broad P-T range, the model should not only be useful for mantle melting, but also for models which crystallise melt, as is the case in a magma ocean. However, as recent studies have suggested, the solidus and liquidus temperature may change heavily depending on if material is melting or crystallising [48]. If this is the case, our model could lie outside of the P-T range between solidus and liquidus for crystallising liquids and would have to be extrapolated. Therefore, one has to be careful with the usage of the fits.

B. Solidus and Liquidus Scaling Law

To avoid the usage of two scaling laws to cover the pressure range of 0-12 GPa, we developed separate scaling laws which are only useful along the solidus and liquidus P-T conditions.

It is shown in Figure 6 that, in contrast to the scaling law for various melt fractions (24), especially the liquidus scaling law (26) fits much better to the thermodynamic model than the more general approach (24) and works just as well as the scaling law of the previous study for low pressures [14]. As a consequence, along the liquidus it is possible to use the respective scaling law (26) over the pressure range 0-12 GPa.

Along the solidus, the general scaling law over various melt fractions (24) fits well to the thermodynamic model (Figure 3). In fact, it seems to fit slightly better to the thermodynamic model than the solidus scaling law (25). As a conclusion, it might not have a big impact if either the solidus scaling law or the more general scaling law are used from 0-12 GPa along solidus temperatures.

C. Potassium Partitioning

To calculate clinopyroxene/melt partition coefficients for potassium, we have used D_{Na} as a nearly strain-free "reference" coefficient as described in section II. In Figure 5, the general D_{Na} scaling law (24) leads to a very good fit of D_K when comparing the thermodynamic D_{Na} approach with the scaling law. In contrast to that, Figure 7 shows D_K values calculated with the solidus and liquidus scaling laws and compares them to experimental data. The experimental data exhibits a very wide spread for close pressure ranges. This spread can occur because (1) the experimental values were measured for different temperatures, (2) the composition of the melt differs, (3) components in the experimental starting composition were neglected in our calculations but lead to additional chemical reactions and influence the partition coefficients.

D. Application to other Terrestrial Planets

Just as the thermodynamic model, phase transitions and density crossover behaviour all depend on pressure. Thus, even for planets with a radius or mass different from Earth, the fit should be applicable if the mantle composition is comparable to Earth. Additionally, the mineralogy of the terrestrial planet has to be taken into account. If there is no or if there are only very minor portions of clinopyroxene in the planets upper mantle, our partition coefficient calculations for clinopyroxene cannot be used.

V. Conclusion and Future Work

With the new high-pressure scaling laws (24)-(26), it is now possible to include partition coefficient models depending on pressure and temperature into mantle convection models for the entire pressure range over which upper mantle melts are buoyant. The newly developed fit functions can be used to calculate clinopyroxene/melt redistribution behaviour of sodium starting from 0 or 4 GPa up to the mantle transition zone of the Earth. This is in contrast to the scaling law by Blundy et al. (1995) [14], which can be applied only from 0 to 4 GPa. We apply our new scaling laws as basis for calculating the partition coefficients for 1+ charge element D_K with (23). Possible approaches to model clinopyroxene/melt partition coefficients for the charges 2+ to 4+ are described in Wood and Blundy (2014) [32] and are based on adjusted calculations for the mineral's lattice site radius r_0 and the bulk modulus E.

Compared to the existing experimental data, our scaling law (24) allows for a good approximation of clinopyroxene partition coefficients of trace elements between solid and melt. This enables us to do self-consistent calculations of local partition coefficients for variable pressures and temperatures. Because we parametrise our scaling law (24) over a wide range of P-T conditions and melt fractions between the peridotite solidus and liquidus, our model can be applied for any $D_{Na}^{cpx/melt}$ calculation between 1850-2360 K and 4-12 GPa. However, if only the solidus or liquidus conditions are needed, the respective scaling laws (25) and (26) might be more useful, as they can be used over the whole range of 0-12 GPa.
The new, higher-pressure-tolerant scaling laws enable us to model the redistribution of trace elements in terrestrial planets in much more detail. Our partition coefficient calculations for clinopyroxene should be applicable as long as clinopyroxene is present in the planet's upper mantle in sufficient abundance.

Future works could not only include the application of our scaling laws in numerical simulations, but also further investigations on partitioning behaviour in mantle material as well. For instance, adding an orthopyroxene/melt trace element partitioning model to a mantle evolution model would provide an even more detailed tool to study on the trace element redistribution from mantle to crust if used alongside the clinopyroxene/melt partitioning model. Furthermore, it is known that water lowers partition coefficients significantly [49]. Consequently, our models would greatly benefit from additional adaptations for water-saturated melts.

For the thermodynamic model, fit function development, and the partitioning of the 1+ charge elements in clinopyroxene/melt, the source code is available in the TRR170-DB [31].

Acknowledgment

We would like to thank Jon Blundy and Timm John for fruitful discussions on partition coefficient modelling. This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 263649064 – TRR 170. This is TRR 170 Publication No. 147.

References

[1] J. M. Schmidt and L. Noack "Parametrising a Model of Clinopyroxene/Melt Partition Coefficients for Sodium to Higher Upper Mantle Pressures", GEOProcessing conference 2021, July 2021.

[2] H. Neumann, J. Mead, and C. Vitaliano "Trace element variation during fractional crystallization as calculated from the distribution law," Geochimica et Cosmochimica Acta, vol. 6 (2-3), pp. 90-99, February 1954.

[3] P. W. Gast "Trace element fractionation and the origin of tholeiitic and alka-line magma types," Geochimica

et Cosmochimica Acta, vol. 32 (10), pp. 1057-1086, October 1968.

[4] B. B. Jensen "Patterns of trace element partitioning," Geochimica et Cosmochimica Acta, vol. 37 (10), pp. 2227-2242, October 1973.

[5] K. T. Johnson, H. J. Dick, and N. Shimizu "Melting in the oceanic upper mantle: an ion microprobe study of diopsides in abyssal peridotites," Journal of Geophysical Research: Solid Earth, 95 (B3), 2661-2678, March 1990.

[6] T. Skulski, W. Minarik, and E. B. Watson, E. B. "High-pressure experimental trace-element partitioning between clinopyroxene and basaltic melts," Chemical Geology, vol. 117 (1-4), pp. 127-147, November 1994.

[7] U. Walzer, and R. Hendel "Mantle convection and evolution with growing continents" Journal of Geophysical Research: Solid Earth, vol. 113 (B9), September 2008.

[8] A. Morschhauser, M. Grott, and D. Breuer "Crustal recycling, mantle dehydration, and the thermal evolution of Mars," Icarus, vol. 212 (2), pp. 541-558, April 2011.

[9] T. Ruedas, P. J. Tackley, and S. C. Solomon "Thermal and compositional evolution of the martian mantle: Effects of phase transitions and melting," Physics of the Earth and Planetary Interiors, vol. 216 , pp. 32-58, March 2013.

[10] A.-C. Plesa, and D. Breuer "Partial melting in one-plate planets: Implications for thermo-chemical and atmospheric evolution," Planetary and Space Science, vol-405 98 , pp. 50-65, August 2014.

[11] R. E. Jones, P. E. van Keken, E. H. Hauri, J. M. Tucker, J. Vervoort, and C. J. Ballentine "Origins of the terrestrial Hf-Nd mantle array: evidence from a combined geodynamical-geochemical approach," Earth and Planetary Science Letters, vol. 518 , pp. 26-39, July 2019.

[12] V. M. Goldschmidt "The principles of distribution of chemical elements in minerals and rocks," Journal of the Chemical Society (Resumed), pp. 655-673, March 1937.

[13] J. Brice "Some thermodynamic aspects of the growth of strained crystals," Journal of Crystal Growth, vol. 28 (2), pp. 249-253, March 1975.

[14] J. Blundy, T. Falloon, B. Wood, and J. Dalton "Sodium partitioning between clinopyroxene and silicate melts," Journal of Geophysical Research: Solid Earth, vol. 340 100 (B8), pp. 15501-15515, August 1995.

[15] J. M. Schmidt and L. Noack "Modelling clinopyroxene/melt partition coefficients for higher upper mantle pressures," EGU General Assembly 2021, online, EGU21-12478, https://doi.org/10.5194/egusphere-egu21-12478, April 2021.

[16] R. Arevalo Jr, W. F. McDonough, and M. Luong "The K/U ratio of the silicate Earth: Insights into mantle composition, structure and thermal evolution," Earth and Planetary Science Letters, vol. 278(3-4), pp. 361-369, February 2009.

[17] A. Gando, D. A. Dwyer, R. D. McKeown, and C. Zhang "Partial radiogenic heat model for Earth revealed by geoneutrino measurements," Nature geoscience, vol. 4(9), pp. 647-651, July 2011.

[18] P. McDade, J. D. Blundy, and B. J. Wood "Trace element partitioning on the Tinaquillo Lherzolite solidus at 1.5 GPa," Physics of the Earth and Planetary Interiors, vol. 139 (1-2), pp. 129-147, September 2003.

[19] B. J. Wood, and J. Nicholls "The thermodynamic properties of reciprocal solid solutions," Contributions to Mineralogy and Petrology, vol. 66 (4), pp. 389-400, June 1978.

[20] J. Blundy, and B. Wood "Crystal-chemical controls on the partitioning of Sr and Ba between plagioclase feldspar, silicate melts, and hydrothermal solutions," Geochimica et Cosmochimica Acta, vol. 55 (1), pp. 193-209, January 1991.

[21] B. J. Wood, and J. D. Blundy "A predictive model for rare earth element partitioning between clinopyroxene and anhydrous silicate melt," Contributions to Mineralogy and Petrology, vol. 129 (2-3), pp. 166-181, October 1997.

[22] K. Grjotheim, C. Krohn, and J. Toguri "Thermodynamic evaluation of activities in molten mixtures of reciprocal salt systems," Transactions of the Faraday Society, vol. 57 , pp. 1949-1957, March 1961.

[23] C. G. Maier, and K. Kelley "An equation for the representation of high-temperature heat content data," Journal of the American chemical society, vol. 54 (8), pp. 3243-3246, August 1932.

[24] J. Hertogen, and R. Gijbels "Calculation of trace element fractionation during partial melting," Geochimica et Cosmochimica Acta, vol. 40 (3), pp. 313-322, March 1976.

[25] J. De Smet, A. Van Den Berg, and N. Vlaar, N. "The evolution of continental roots in numerical thermo-chemical mantle convection models including differentiation by partial melting," In: Developments in geotectonics. Elsevier, vol. 24, pp. 153-170, 1999.

[26] T. Gasparik "Phase relations in the transition zone," Journal of Geophysical Research: Solid Earth, vol. 95 (B10), pp. 15751-15769, September 1990.

[27] E. Takahashi "Speculations on the Archean mantle: missing link between komatiite and depleted garnet peridotite," Journal of Geophysical Research: Solid Earth, vol. 95 (B10), pp. 15941-15954, September 1990.

[28] D. C. Presnall, Y.-H. Weng, C. S. Milholland, and M. J. Walter "Liquidus phase relations in the system MgO-MgSiO3 at pressures up to 25 GPa - constraints on crystallization of a molten Hadean mantle," Physics of the Earth and Planetary Interiors, vol. 107 (1-3), pp. 83-95, April 1998.

[29] M. Walter, E. Nakamura, R. Tronnes, R., and D. Frost, D. "Experimental constraints on crystallization differentiation in a deep magma ocean," Geochimica et Cosmochimica Acta, vol. 68 (20), pp. 4267-4284, October 2004.

[30] D. M. Shaw "Trace element fractionation during anatexis," Geochimica et Cosmochimica Acta, vol. 34 (2),

pp. 237-243, February 1970.

[31] J. M. Schmidt, and L. Noack "Replication Data for: Clinopyroxene/Melt Partitioning: Models for Higher Upper Mantle Pressures Applied to Sodium and Potassium," TRR170-DB, (V01), 2021, doi.org/10.35003/GIAZCQ.

[32] B. J. Wood, and J. D. Blundy, J. D. (2014) "Trace element partitioning: The influences of ionic radius, cation charge, pressure, and temperature," In H. D. Holland and K. K. Turekian (Eds.), Treatise on geochemistry (second edition) (Second Edition ed., p. 421 - 448). Oxford: Elsevier, 2014.

[33] W. Wang, and E. Takahashi "Subsolidus and melting experiments of a K-rich basaltic composition to 27 GPa: Implication for the behavior of potassium in the mantle," American Mineralogist , vol. 84 (3), pp. 357-361, November 1999.

[34] E. Chamorro, R. Brooker, J. A. Wartho, B. Wood, S. Kelley, and J. Blundy "Ar and K partitioning between clinopyroxene and silicate melt to 8 GPa," Geochimica et Cosmochimica Acta, vol. 66 (3), pp. 507-519, February 2002.

[35] G. A. Gaetani, A. J. Kent, T. L. Grove, I. D. Hutcheon, and E. M. Stolper "Mineral/melt partitioning of trace elements during hydrous peridotite partial melting," Contributions to Mineralogy and Petrology, vol. 145 (4), pp. 391-405, May 2003.

[36] E. H.Hauri, T. P. Wagner, and T. L. Grove "Experimental and natural partitioning of Th, U, Pb and other trace elements between garnet, clinopyroxene and basaltic melts," Chemical Geology, vol. 117(1-4), pp. 149-166, 1994.

[37] S. R. Hart and T. Dunn "Experimental cpx/melt partitioning of 24 trace elements," Contributions to Mineralogy and Petrology, vol. 113(1), pp. 1-8, 1993.

[38] S. Klemme, J. D. Blundy, B. J. Wood "Experimental constraints on major and trace element partitioning during partial melting of eclogite," Geochimica et Cosmochimica Acta, vol 66(17), pp. 3109-3123, 2002.

[39] V. J. Salters, A. Stracke "Composition of the depleted mantle," Geochemistry, Geophysics, Geosystems, vol. 5(5), 2004

[40] D. Mckenzie, and M. Bickle "The volume and composition of melt generated by extension of the lithosphere," Journal of petrology, vol. 29 (3), pp. 625-679, June 1988.

[41] E. Stolper, D. Walker, B. H. Hager, and J. F. Hays "Melt segregation from partially molten source regions: the importance of melt density and source region size," Journal of Geophysical Research: Solid Earth, vol. 86 (B7), pp. 6261-6271, July 1981.

[42] Z. Jing, and S.-i. Karato "The density of volatile bearing melts in the Earth's deep mantle: The role of chemical composition," Chemical Geology, vol. 262 (1-2), pp. 100-107, May 2009.

[43] M. J. Beuchert and H. Schmeling "A melting model for the lowermost mantle using Clapeyron slopes derived from experimental data: Consequences for the thickness of ultralow velocity zones (ULVZs)," Geochemistry, Geophysics, Geosystems, vol. 14 (1), pp. 197-208, January 2013.

[44] E. Ohtani, Y. Nagata, A. Suzuki, and T. Kato "Melting relations of peridotite and the density crossover in planetary mantles," Chemical Geology, vol. 402 120 (3-4), pp. 207-221, March 1995.

[45] L. T. Elkins-Tanton, E. Parmentier, and P. Hess "Magma ocean fractional crystallization and cumulate overturn in terrestrial planets: Implications for Mars," Meteoritics & Planetary Science, vol. 38 (12), pp. 1753-1771, January 2003.

[46] A. E. Ringwood, A. E. "Phase transformations and their bearing on the constitution and dynamics of the mantle," Geochimica et Cosmochimica Acta, vol. 55 (8), pp. 2083-2110, August 1991.

[47] M. Akaogi "Phase transitions of minerals in the transition zone and upper part of the lower mantle," Special Papers-Geological Society Of America, vol. 421 , p. 1, 2007.

[48] C.-E. Boukare, Y. Ricard, and G. Fiquet "Thermodynamics of the MgO-FeO-SiO2 system up to 140 GPa: Application to the crystallization of Earth's magma ocean," Journal of Geophysical Research: Solid Earth, vol. 120 (9), pp. 6085-6101, August 2015.

[49] J. Blundy and B. Wood "Partitioning of trace elements between crystals and melts," Earth and Planetary Science Letters, vol. 210(3-4), pp. 383-397, January 2003.

Dynamics of Momentary Reserves in Different Grid Topologies under Contingency: Observations from Numerical Experiments

Kosisochukwu Pal Nnoli
Dept. of Computer Science and Electrical Engineering
Jacobs University, 28759 Bremen
Bremen, Germany
email: k.nnoli@jacobs-university.de

Stefan Kettemann
Dept. of Physics and Earth Science and Dept. of Computer Science
Jacobs University, 28759 Bremen
Bremen, Germany
Division of Advanced Materials Science, Pohang 790-784
South Korea
email: s.kettemann@jacobs-university.de

***Abstract*—This paper presents the studies and investigations on the dynamics and mechanism of momentary reserves contribution in electrical power systems under contingency. Momentary reserve through the machine's inertia serves the purpose of primary frequency control and prevents voltage collapse in the case of reactive power reserves. A simulation was performed on a realistic Nigerian 330 kV transmission network in PowerFactory software to study and investigate the mechanism of these reserve functions on the network buses as an inertia active power control method. Moreover, we investigated the influence of geodesic increment of momentary reserve on the decay of disturbances and in turn estimated the value of the post contingent frequency at the network buses. The results indicated that the momentary reserve by inertia alone reduces the frequency deviation from its nominal value, delays the transmission of disturbances and enhances the damping of oscillations by reducing the final frequency settling time at the buses under contingency. These results were further extrapolated by the numerical experiments carried out on a synthetic Square 330 kV power network to account for the effects of meshedness as a distinguishing factor in grid topologies. These numerical experiments also suggest the optimal placement of the momentary reserves in the grid in order to improve system stability against power outage disturbances.**

***Index Terms*—momentary reserves; system perturbation; inertia control, oscillations damping; grid meshedness.**

I. Introduction

The active and reactive power reserves of synchronous generators or Battery Energy Storage Systems (BESS) are the keys to a successful system control in power systems. For each power generator in the grid, power reserve represents the total amount of power remaining after the supply of system loads and losses. The definition and importance of many power reserves have been discussed in [1] and would be reiterated here. This definition of power reserves does not extend to exceeding the power capability curve of the generators. These reserves are particularly referred to as spinning reserves in synchronous machines more than in other kinds of power generators, like the wind and solar generators using BESS [2]. These power reserves can be used for both or either primary frequency control, secondary control and tertiary control [3] [4] [5]. Again, not all of these remaining power reserves from generators are assigned for primary control function alone, we classify the ones momentarily made available through the generators' droop functions specifically for few seconds primary frequency and active power control as momentary reserves of the generators. A lot about system stability and reliability rest on the grid's momentary reserves.

The control and response of generators to network disturbances or contingencies depend heavily on the kind and magnitude of the disturbance and amount of momentary reserves available for use in the primary frequency control or in voltage security. Since primary control requires fast control action to be taken within few seconds of contingency, momentary reserves and their placement play important roles in the dynamics of the disturbance from the point of contingency (PoC) or fault location to the rest of the electrical power network. In this paper, we investigate how the dynamics of a disturbance is influenced by momentary reserves first at the contingent node, its neighbour nodes and other nodes located far away from the fault location. We would see whether a disturbance is damped as it travels across a grid and whether it could be contained on fewer nodes (i.e., localized) based on the function of the grid's momentary reserves. These investigations are carried out in the DigSILENT PowerFactory software [6], using the Nigerian 330 kV grid as a case study. In order to account for a distinctive grid topology characteristic, we would also consider a synthetic Square 330 kV power network as another case study because of its high degree of meshedness.

In this paper, we would start with a brief literature review of reserve functions in Section II. We would then proceed with the description, modeling and simulation of our test transmission networks in Section III. In Section IV, we would investigate the dynamics of momentary reserve on the test networks with a concise summary of the results in Section V and conclude with our findings and recommendation in Section VI.

II. Literature Review

For many years, synchronous machine reserves have been the major source of power compensation in complex power grids prior to the popularity and grid integration of major successful storage devices like the BESS and hydrogen storage. These power compensation reserves, now from both synchronous machines and other storage sources can be called on during grid contingencies, especially in power outage events due to loss of generators or major transmission lines to ensure sufficient system frequency oscillation damping, control and stability. These power reserves can be categorized as momentary, primary, secondary and tertiary based on the grids' request time, duration and magnitude of need of them as shown in Figure 1.

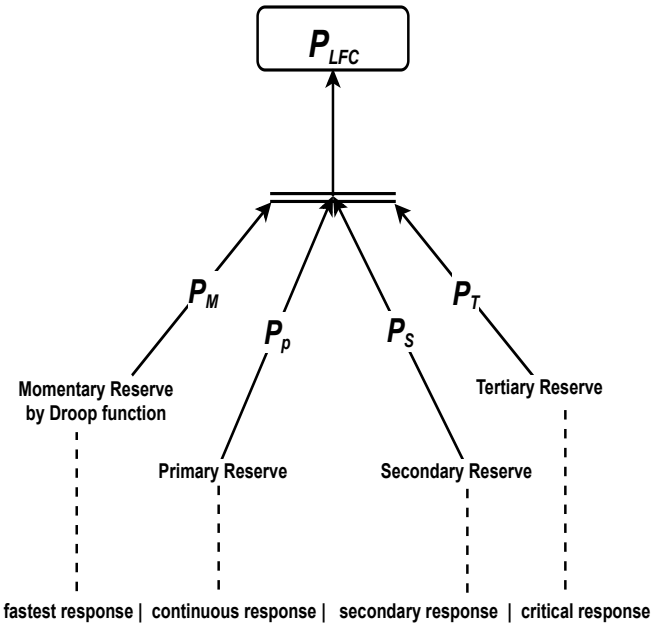


Figure 1: Categorization of Power Reserves based on Grid's Compensation needs in request time, duration and magnitude. P_{LFC} indicates power reserves injected at low frequency contingencies.

According to [3] [4] [7], the time, duration and magnitude factors characterizing the need for grid's power compensation are determined by forecast and sudden imbalances between the power generation, consumption, losses and minutes to hourly contingencies. While the primary reserves are expected to be fully activated within 30 seconds of local request time, the secondary reserves are expected to be automatically and fully activated between 10 to 15 minutes of request time in order to restore power balance and to bring the frequency back to its nominal value. The tertiary reserves account for additional dispatches possible through market negotiations to bring the grid to stability in case of prolonged contingencies, which could not be handled by finite continuous response of secondary reserves. While the secondary and tertiary reserves are controlled and implemented by system operators and where multiple network authorities may be involved [7] [8], the momentary reserves, which are constantly available by droop call function within seconds (i.e., fastest response) followed by continuous response (i.e., primary reserves) made available through agreements between the grid operators and power generating firms for quarterly compensation or contingency purposes are our focus in this numerical experiments. Many authors have carried out extensive research on the effects and importance of primary, secondary and tertiary control reserves [9]–[11]. Machowski, Bialek and Bumby in [4] have confirmed the importance of primary frequency control to the overall frequency stability of the network, especially for inter-area modes of oscillation. A new nonlinear droop control function was studied for wind turbine in [12] where a new dynamic droop coefficient is connected as the negative feedback of the original droop control coefficient in order to bring the product of the original droop control coefficient and nominal frequency within a controllable range. This was done in real time according to the target frequency to improve the primary frequency control. The adaptive droop control action was simulated to show its frequency control rate but there was no concrete understanding of the mechanism of this action on the network bus rather than reducing the frequency dip at post-contingency. This means that little has been numerically investigated to practically understand the mechanics and mechanisms of contribution of momentary control reserves in the stability of power grids. We predict that these first few seconds of primary control response determines the system's need for continuous (i.e., full activation of primary control reserves) and secondary reserves and is therefore essential to system recovery from contingencies. As a result, this paper is directed towards studying the numerical mechanics of dynamics and effective distribution of control response function of momentary reserves at a fault event in realistic model of Nigerian network and a Square grid in order to account for varying network topologies.

III. Modeling and Simulation of the Nigerian Transmission Network Model

To understand our case study system and its parameter interactions, we would describe the components that make up the network. The Nigerian 330 kV transmission grid consist of $N_S = 71$ substations/nodes, $N_L = 81$ over-head transmission lines (made from an alloy of aluminium and steel) with an average length of 92 km, each with a limiting current of 1320 A. The grid is comprised of 107 less-decommissioned units of generators, accounting for the present 29 power stations. The apparent power capacity of the Nigerian grid is about $13,208MW$ as of 2020 [13]. There are other lower voltage networks including the 132 kV and 33 kV sub-transmission networks. For household utilities, there are 11 kV and 0.415 kV 3-phase distribution networks. The Nigerian network operates at $\nu_0 = 50$ Hz frequency and can be described as a grid where most of the nodes are connected to one another in a ring form [14], as seen in the diagram shown in Figure 2.

To control the voltage outputs of the generators through their excitation control, the simplified excitation Automatic Voltage Regulator (AVR) model is used [15]. Other controllers

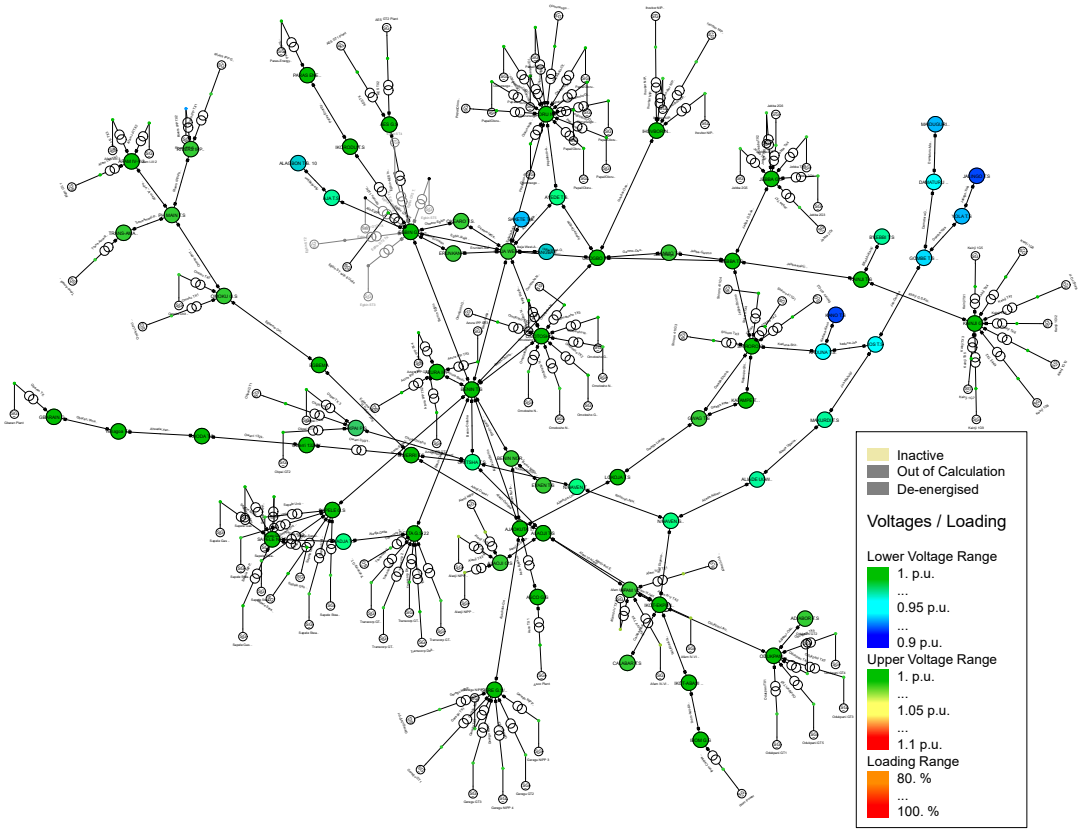


Figure 2: The Nigerian 330 kV Electrical Power Grid.

include power system stabilizers (PSS2A-model) tasked with enhancing the damping of the entire power system's oscillations through excitation control.

The input signal to the PSS2A controller is the derivative of generator's rotor speed and its output is injected to the AVR through the excitation system. This injection works to terminate the phase-lag between the voltage reference and the windings' torque of the generator [16]. For the synchronous machine model, the choice of the model is influenced by the IEEE guide in [17]. The speed governor in the model is the Turbine Governor (TGOV) model used to maintain the frequency operational limits according to swing equation [18]. Here, the swing equation describes the torque balance between the mechanical torque T_t in N.m. of each synchronous machine's turbine and the electromagnetic torque T_e in N.m. as governed by the differential equation given as [3] [4] [11],

$$J_i \frac{d\omega_i^{\Theta}}{dt} + D_{r_i}\omega_i^{\Theta} = T_t - T_e - D_{r_i}\omega_0, \tag{1}$$

where $J_i = \frac{2H_i}{\omega_0^2} S_i$ is the combined moment of inertia of the generator and turbine in kg.m^2 with H_i the generator inertia constant in seconds and S_i the generator apparent power in Volt-Amperes (VA). D_{r_i} represents the rotational loss due to generator rotor windings for each i^{th} generator in N.m.s and i denotes the index of power generators in the grid. Here, t is time in seconds and ω_i^{Θ} is the angular velocity of the rotor in electrical rad/s with ω_0 as its rated synchronous value in electrical rad/s. If we assume that a change in the rotor's angular velocity $(\omega_i^{\Theta} - \omega_0)$ is a derivative of its angular position δ in electrical radians with respect to its rotating setpoint, δ_0 at t $= 0$ given as

$$\omega_i^{\Theta} - \omega_0 = \frac{d\delta_i}{dt}, \tag{2}$$

then, with respect to time, the derivative of ω_i^{Θ} would give

$$\frac{d\omega_i^{\Theta}}{dt} = \frac{d}{dt}\left(\frac{d\delta_i}{dt}\right) + \frac{d\omega_0}{dt}, \tag{3}$$

where ω_0 is the constant rated synchronous value whose derivative with respect to time gives zero (i.e., $\frac{d\omega_0}{dt} = 0$), (3) becomes,

$$\frac{d\omega_i^{\Theta}}{dt} = \frac{d^2\delta_i}{dt^2}. \tag{4}$$

In practice, ω_0 is related to the grid frequency (v_o) by $2\pi v_o$, where v_o is 50 Hz in the Nigerian power grid. If we represent the net mechanical shaft torque at grid frequency to be $T_m = T_t - D_{r_i}\omega_0$, substituting (4) into (1), we then have,

$$J_i \frac{d^2\delta_i}{dt^2} + D_{r_i}\left(\frac{d\delta_i}{dt}\right) = T_m - T_e. \tag{5}$$

Here, we assumed that the network perturbation effected on the rotors from the fault location is small. Multiplying both

sides of (5) by the rated speed (ω_0) in order to ensure that we maintain a synchronous 50 Hz revolution throughout the system, balancing the power, we have

$$J_i\omega_0\frac{d^2\delta_i}{dt^2}+D_{r_i}\omega_0\left(\frac{d\delta_i}{dt}\right)=T_m\omega_0-T_e\omega_0. \quad (6)$$

As power $P=T\omega$, the right side of (6) can now be written as

$$J_i\omega_0\frac{d^2\delta_i}{dt^2}+D_{r_i}\omega_0\left(\frac{d\delta_i}{dt}\right)=P_m-P_e, \quad (7)$$

where P_m is the turbine's mechanical power and P_e is the generator's air-gap electrical power. If we represent the rotor angular momentum at rated speed with M_i (i.e., $M_i=J_i\omega_0=\frac{2H_i}{\omega_0}S_i$) and also represent the damping coefficient at rated synchronous speed with D_i (i.e., $D_i=D_{r_i}\omega_0$), the swing equation can then be re-written in many forms as,

$$M_i\frac{d^2\delta_i}{dt^2}+D_i\frac{d\delta_i}{dt}=P_{m,i}-P_{e,i}, \quad (8)$$

and also as [19] [20],

$$\frac{2H_i}{\omega_o}S_i\frac{d^2\delta_i}{dt^2}+D_i\frac{d\delta_i}{dt}=P_i+\sum_{j=1}^{N_S}W_{ij}\sin(\delta_j-\delta_i), \quad (9)$$

where P_i is the power in the grid's i^{th} node, N_S is the number of nodes/buses and $W_{ij}=V_iV_jB_{ij}$ is the power capacity in Watt of the transmission lines and it is dependent on the network voltage with $\sin(\delta_j-\delta_i)$ modeling the dependence of their phase differences, which informs the direction of powerflow and the transmission of disturbances in the case of contingencies. We performed load-flow calculations using the Newton-Raphson method and electromechanical simulations in DigSILENT PowerFactory software, as documented in [21]. Here, we report results applying these simulations to study the effect of momentary reserves on system dynamics and its contribution to the overall system stability.

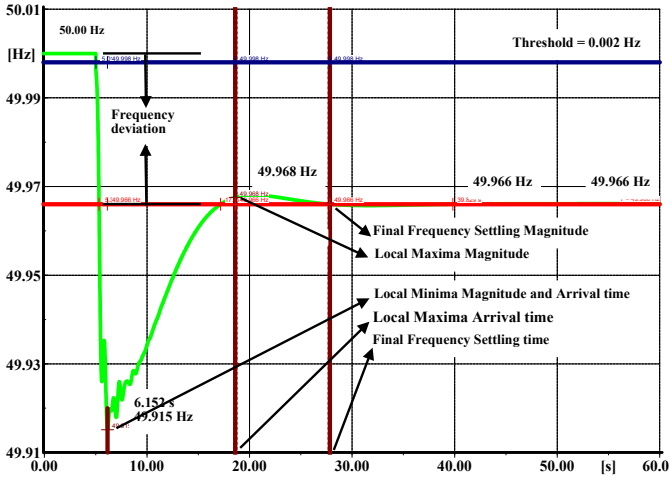


Figure 3: Green Line: Frequency as a function of time at a node where a Synchronous Machine (SM) event occurred at $t=5$s.

IV. Dynamics of Momentary Reserve

For a power system to be in a balanced state, the total power generated (P) has to be equal to the sum of the total load and the transmission losses (P^{x}). This means that for the system to always remain balanced, any change in the load demand and/ losses would require a balancing change in the generated power P. Therefore, the condition $\Delta P >> \Delta P^{\mathrm{x}}$ should always be fulfilled to ensure the security and integrity of power grids. These dynamic changes result in a permanently changing frequency according to the speed-droop characteristics of power turbines, given by [4]

$$\frac{\Delta\nu_i}{\nu_0}=-\sigma_i\frac{\Delta P_i}{P_i^{\mathrm{x}}}, \quad (10)$$

where ν_0 is the nominal frequency in Hertz and σ_i is the local droop of the generation characteristics at the i^{th} generator node, $\sigma_i=P_i^{\mathrm{x}}/P_i$. Thus, any change in frequency resulting from a change in load demand or losses requests a generator response. In this way, if the node is a synchronous generator node, its inertia is expected to slow down the transmission of disturbances to other nodes connected to it, while attempting to damp the frequency oscillations with its momentary reserve according to its droop σ_i function. However, its ability to do so depends on the amount of momentary reserve that the generator can supply, according to its dead zone speed-droop characteristics setting (i.e., possibility of ΔP_i and in the absence of no generator overload in the case of sudden load injection). Since it is not feasible to inject power equally at every node, we aim to understand the dynamics of these reserves to see how an available momentary reserve at one node influences the frequency dynamics at other nodes as function of their distance from the fault location. A better understanding of this influence in the realistic Nigerian grid model would be advantageous for its optimal placement and could improve primary frequency control in real power grids.

To show the influence of geographical distribution of inertia with available momentary reserve, we chose the Nigerian transmission grid as the case study grid. As reported in Section III, we modelled this network in PowerFactory software and calculated the load flow of the network. We considered the voltage dependency of grid loads, and the active power control according to the grid inertia. The effect of provision of momentary reserves by system inertia on the frequency stability of the electrical power networks is studied by varying the aggregated inertia constant H_{agg} of the entire grid, as defined by [22] [23],

$$H_{\mathrm{agg}}=\frac{\sum_{i=1}^{N_S}H_iS_i}{S_n}, \quad (11)$$

where S_i and H_i are the rated apparent power and inertia constant of the i^{th} bus, respectively, $S_n=\sum_{i=1}^{N_S}S_i$ is the total rated power in the grid. Note that H_i on non-generating buses are set to zero. The H_{agg} was varied in the entire grid by multiplying the S_i with a common factor. Note that the Nigerian transmission network does not comprise tie-line connections and power transfer [13] [24]. The grid frequency

is operated uniformly at 50 Hz across the entire network.

Here, we choose 11 buses for the investigation. The fault location bus 24, two buses at the same geodesic distance (unweighted), $r = 2$ from fault location with no inertia (i.e., buses 8 and 10) and three buses with inertia (i.e., buses 22, 55, and 57). We again choose two buses at the same geodesic distance (unweighted), $r = 7$ with no inertia (i.e., buses 7 and 30) and three buses with inertia (i.e., buses 3, 28, and 69). The reason is that we want to investigate the effect of momentary reserve both in the vicinity of the reserve (i.e., near to the fault location) and at far distances from the injection node. We also want to understand how momentary reserve tentatively contributes to damping of disturbances as they propagate along the network. A reference to the network buses numbers are in [23].

Keeping all system parameters, generations, losses and loads constant under undisturbed operations, a synchronous machine outage event (disturbance/contingency) is induced at bus 24 at exactly $t = 5$s of the 90s transient electromechanical stability simulation time frame with a 200ms switching, enabling the observations on the buses in PowerFactory power simulation software. The change in frequency propagated across the grid is related to the change of power in (10). The Figure 3 shows the frequency dynamics at a network bus. Here, the red line marks the frequency magnitude at final settling time. The black arrows point to the local maxima or minima of the frequency magnitude and arrival times. From this nodal points described in Figure 3, we observe at each study case node, the frequency's time of arrival (ToA), defined as the time the frequency deviation first reaches a small threshhold of $\delta\nu = \pm 0.002$ Hz, as defined in more detail in [23]. Furthermore we record the time of the first maximum (maxima_t) of the transient and its magnitude ($\text{maxima}_{\text{mag}}$), the time of the first minimum (minima_t) and its magnitude ($\text{minima}_{\text{mag}}$). We also observe the final frequency's settling time (FS_t), its magnitude (FS_{mag}) and frequency deviation (Dev_{mag}) from the nominal 50Hz value, for each of these nodes. For case studies I-V in Tables I-V, we show the observations in milliseconds for the case study nodes, with static network power flow of the Nigerian transmission grid.

TABLE I: N-NODAL OBSERVATIONS WITH LARGE DISTURBANCE AND NO RESERVE AT FAULT LOCATION GIVEN THAT $H_{\text{agg}} = 2$s

Bus	r	ToA (s)	minima_t (s)	$\text{minima}_{\text{mag}}$ (Hz)	FS_t (s)	FS_{mag} (Hz)	Dev_{mag} (Hz)
24	0	5.012	6.522	49.596	27.200	49.829	0.171
8	2	5.012	6.522	49.596	27.201	49.829	0.171
10	2	5.012	6.522	49.596	27.200	49.829	0.171
22	2	5.013	6.282	49.596	27.195	49.829	0.171
55	2	5.013	6.532	49.596	27.201	49.829	0.171
57	2	5.012	6.532	49.597	27.196	49.829	0.171
7	7	5.013	6.372	49.598	27.205	49.829	0.171
30	7	5.013	6.322	49.600	27.206	49.829	0.171
3	7	5.013	6.422	49.590	27.185	49.829	0.171
28	7	5.013	6.942	49.596	27.184	49.829	0.171
69	7	5.013	6.352	49.592	27.186	49.829	0.171

TABLE II: N-NODAL OBSERVATIONS WITH LARGE DISTURBANCE AND LARGE RESERVE AT FAULT LOCATION GIVEN THAT $H_{\text{agg}} = 2$s

Bus	r	ToA (s)	minima_t (s)	$\text{minima}_{\text{mag}}$ (Hz)	FS_t (s)	FS_{mag} (Hz)	Dev_{mag} (Hz)
24	0	5.014	6.132	49.920	25.700	49.967	0.033
8	2	5.014	6.112	49.920	25.691	49.967	0.033
10	2	5.014	6.102	49.920	25.690	49.967	0.033
22	2	5.015	6.282	49.919	25.687	49.967	0.033
55	2	5.014	6.082	49.920	25.691	49.967	0.033
57	2	5.015	6.192	49.919	25.688	49.967	0.033
7	7	5.016	6.292	49.920	25.696	49.967	0.033
30	7	5.016	6.372	49.920	25.698	49.967	0.033
3	7	5.017	6.252	49.918	25.679	49.967	0.033
28	7	5.017	6.142	49.916	25.678	49.967	0.033
69	7	5.017	6.182	49.918	25.680	49.967	0.033

TABLE III: N-NODAL OBSERVATIONS WITH LARGE DISTURBANCE AND LARGE RESERVE AT FAULT LOCATION GIVEN THAT $H_{\text{agg}} = 6$s

Bus	r	ToA (s)	minima_t (s)	$\text{minima}_{\text{mag}}$ (Hz)	FS_t (s)	FS_{mag} (Hz)	Dev_{mag} (Hz)
24	0	5.015	9.312	49.928	29.304	49.967	0.033
8	2	5.015	9.262	49.928	29.304	49.967	0.033
10	2	5.015	9.272	49.928	29.304	49.967	0.033
22	2	5.017	8.352	49.928	29.302	49.967	0.033
55	2	5.015	9.442	49.928	29.305	49.967	0.033
57	2	5.017	9.142	49.928	29.302	49.967	0.033
7	7	5.019	9.182	49.928	29.308	49.967	0.033
30	7	5.020	8.932	49.928	29.309	49.967	0.033
3	7	5.028	8.672	49.927	29.296	49.967	0.033
28	7	5.074	8.602	49.927	29.295	49.967	0.033
69	7	5.029	8.652	49.927	29.296	49.967	0.033

TABLE IV: N-NODAL OBSERVATIONS WITH LARGE DISTURBANCE AND LARGE RESERVE AT FAULT LOCATION AND AN INCREASED RESERVE AT BUS 22 GIVEN THAT $H_{\text{agg}} = 2$s

Bus	r	ToA (s)	minima_t (s)	$\text{minima}_{\text{mag}}$ (Hz)	FS_t (s)	FS_{mag} (Hz)	Dev_{mag} (Hz)
24	0	5.014	6.612	49.920	27.645	49.967	0.033
8	2	5.014	6.122	49.921	27.647	49.967	0.033
10	2	5.014	6.612	49.920	27.645	49.967	0.033
22	2	5.015	6.352	49.919	27.641	49.967	0.033
55	2	5.014	6.582	49.920	27.647	49.967	0.033
57	2	5.015	6.152	49.921	27.642	49.967	0.033
7	7	5.016	6.342	49.921	27.653	49.967	0.033
30	7	5.016	6.452	49.921	27.655	49.967	0.033
3	7	5.017	6.282	49.919	27.630	49.967	0.033
28	7	5.017	6.142	49.917	27.629	49.967	0.033
69	7	5.017	6.252	49.919	27.631	49.967	0.033

TABLE V: N-NODAL OBSERVATIONS WITH LARGE DISTURBANCE AND LARGE RESERVE AT FAULT LOCATION AND WITH A NEWLY INSTALLED RESERVE AT BUS 7 GIVEN THAT $H_{\text{agg}} = 2$s

Bus	r	ToA (s)	minima_t (s)	$\text{minima}_{\text{mag}}$ (Hz)	FS_t (s)	FS_{mag} (Hz)	Dev_{mag} (Hz)
24	0	5.014	6.662	49.920	39.118	49.967	0.033
8	2	5.014	6.632	49.920	39.116	49.967	0.033
10	2	5.014	6.662	49.920	39.118	49.967	0.033
22	2	5.015	6.332	49.918	39.127	49.967	0.033
55	2	5.014	6.632	49.920	39.115	49.967	0.033
57	2	5.015	6.202	49.919	39.124	49.967	0.033
7	7	5.016	6.602	49.920	39.102	49.967	0.033
30	7	5.016	6.492	49.920	39.097	49.967	0.033
3	7	5.017	6.292	49.919	39.149	49.967	0.033
28	7	5.017	6.152	49.917	39.151	49.967	0.033
69	7	5.017	6.272	49.919	39.148	49.967	0.033

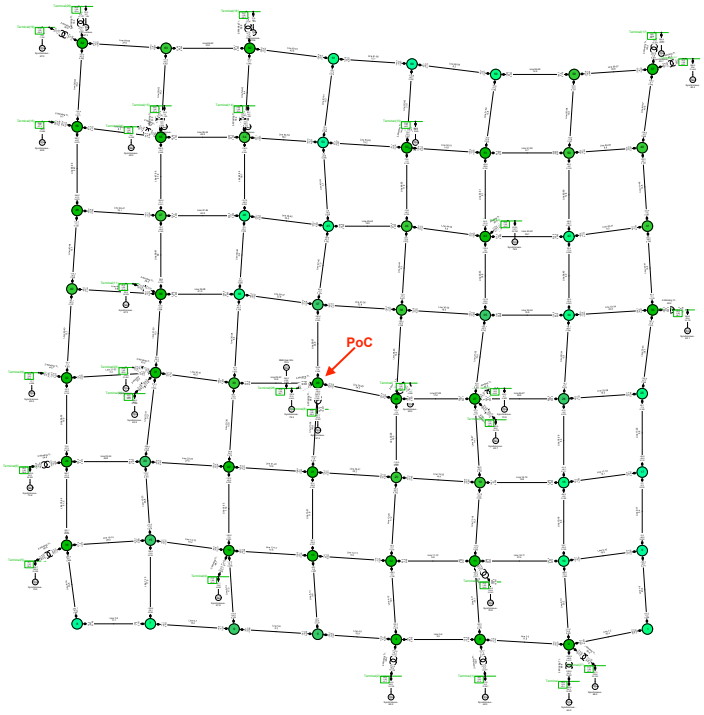


Figure 4: The Square 330 kV Power Grid. The Red pointer is the Point of Contingency (PoC, i.e., fault location) [23].

In Table I and with network H_{agg} at 2s, a typical behaviour of high renewable energy source injected grids, we observe a high frequency deviation (i.e., 0.171 Hz) induced by the large disturbance (of 1320MW power magnitude outage) at bus 24 fault location. We also observe that the disturbance arrived at the fault location first and at the same time as most of its nearest neighbours at $r = 2$, but arrived a little later at the distant buses 7, 30, 28 and 69. The frequency dip described by the minima$_{mag}$ shows the lowest magnitude of frequency deviation before the actions of turbines' governors.

At $H_{agg} = 2$s and in comparison with Table I, Table II shows the observations when the fault location is injected with inertia and large reserve. We observe that there are delays in the frequency ToA at the buses, a reduction in minima$_t$, which corresponds to increased frequency dips (i.e., minima$_{mag}$), shorter final frequency settling times with smaller magnitudes and hence, a decrease in the frequency deviation from the nominal value (0.033 Hz).

In Table III, we kept the network parameters constant and only increased the grid inertia, $H_{agg} = 6$s using (11) without changing the active power injections at any network node. We did not observe any further decrement in the frequency deviation or any increment in the frequency final settling magnitude, rather we observe a delay in its time of arrival at the buses with a corresponding increase in its dip and final settling time. This suggests that increasing the H_{agg} without a corresponding increase in the reserve does not improve the frequency dynamics during contingencies, rather it increases the arrival time of the disturbance while reducing the frequency dip across the network (i.e., improved minima$_{mag}$). This further delay in ToA also increases the frequency final settling time.

Table IV shows our nodal observations on the case study buses when we injected more power reserve prior to contingency at another bus with $r = 2$, from the fault location and keeping all other system operation parameters constant from observations in Table II. Here, the generator's power and reserve at the fault location are higher in magnitude than the injected ones. We observe that the frequency final settling magnitude did not increase, the ToA at the buses remained the same but there is an observable reduction in the frequency dip at some nodes.

To investigate the effect of the same reserve at a bus geometrically farther away from the fault location but with a higher degree of connectivity, which defines the number of edges connected to it, we removed the reserve at bus 22 with a node degree of 2 and installed it at another bus with $r = 7$ but with a node degree of 5 and keeping all other system operation parameters constant from our observations in Table II. The result of this new installation in Table V compared with Table IV shows no observable change in the frequency ToA at the buses and in the frequency final settling magnitude. Rather, we observe a delay in frequency dip time (i.e., minima$_t$) with a further delay in the final frequency settling time at the buses.

To further understand and verify if the above findings on the influence of momentary reserves would be general in different power network topological settings with special emphasis on

network's degree of meshedness, which is calculated by the coefficient of meshedness given as [25],

$$\phi_M = \frac{N_{\mathrm{E}} - N_B + 1}{2N_B - 5}, \tag{12}$$

where N_{E} and N_B are respectively the total number of transmission edges/lines and buses, we transformed the Nigerian 330 kV transmission network (N) into a Square power grid (i.e., S or S-grid) as shown in Figure 4. Note: ϕ_M has a maximum value of 1 for a completely meshed power grid. The ϕ_M for Nigerian 330 kV grid is 0.11 while that of Square power grid is calculated to be 0.40. See Ref: [23] for the details of the transformation.

In the Square grid shown in Figure 4, we choose for the purpose of verification, the point of contingency (i.e., bus 29), buses 13 and 20 with no inertia and bus 31 with inertia, all at the same geodesic distance of r= 2 from fault location. Also, at a geodesic distance of r= 5 from fault location, we choose buses 42 and 48 with no inertia and bus 55 with inertia. We choose r= 5 in the Square power grid because of the inertia criterion of selection and the limited number of buses with r= 7 from the fault location. Keeping all system parameters constant and inducing a synchronous machine event at bus 29 exactly at t =5s in a 90s transient electromechanical simulation time frame, we measure the change in frequency transmitted across the grid by this disturbance by placing a small threshold at $\delta\nu = \pm 0.002$ Hz exactly as in the case of the Nigerian transmission network studies. Again, we show the dynamic observations in milliseconds for the case study nodes, with static Square network power flow, recording the ToA, minima_t, maxima_t, (FS_t), $\mathrm{FS}_{\mathrm{mag}}$ and $\mathrm{Dev}_{\mathrm{mag}}$. For case studies in Tables VI-X similar to the ones studied in Tables I-V.

In Table VI and with H_{agg} at 2s in comparison with Table I, we again observe a high frequency deviation (i.e., 0.187 Hz) effected by the same large disturbance magnitude but induced at bus 29 point of contingency (PoC). Here, we observe the disturbance arrived at the PoC first but the arrival at the neighbouring buses with r= 2 are exponentially delayed. As geodesic distance, r is increasing from the PoC, the frequency dip described by the minima_t is reducing while the $\mathrm{minima}_{\mathrm{mag}}$ of the buses are increasing as expected.

At H_{agg} = 2s and in comparison with Table VI, Table VII shows the observations when the PoC is injected with inertia and a large reserve. We observe that the delays in the frequency ToA at the buses improved. There are reductions in the minima_t with correspondingly improved frequency dips (i.e., $\mathrm{minima}_{\mathrm{mag}}$) similar to Table II and with reductions in the final frequency settling time (FS_t) and magnitude ($\mathrm{FS}_{\mathrm{mag}}$). These lead to a decrease in the frequency deviation, $\mathrm{Dev}_{\mathrm{mag}}$ value (0.035 Hz) from the nominal.

In Table VIII, we increased only the grid inertia, H_{agg} = 6s by using (11) without changing the active power injections at any network bus and keeping all other network parameter constant as in Table VII. Again, we did not observe any change in the frequency deviation or any increment in the frequency final settling magnitude, rather we observe a delay in the frequency's ToA at the buses increasing with geodesic distance (r) and with a corresponding increase in the frequency dip (minima_t) and final settling times. This again suggests that increasing the H_{agg} alone without a corresponding increase in the reserve provides a delay in propagation of disturbances while reducing the frequency dip across the network (i.e., improved $\mathrm{minima}_{\mathrm{mag}}$). However, it does not improve the $\mathrm{Dev}_{\mathrm{mag}}$ during contingencies. This further delay in ToA resulting from increasing the H_{agg} also increases the frequency final settling times, similar to our observations in Table III above.

Table IX shows our nodal observations on the Square grid case study buses when we injected more power reserve prior to contingency at bus 31 with $r = 2$ from the PoC and keeping all other system operation parameters constant from the observations in Table VII. Here, the power capacity of the generator and reserve at the PoC are higher in magnitude than the injected one.

TABLE VI: S-NODAL OBSERVATIONS WITH LARGE DISTURBANCE AND NO RESERVE AT FAULT LOCATION GIVEN THAT $H_{\mathrm{agg}} = 2$s

Bus	r	ToA (s)	minima_t (s)	$\mathrm{minima}_{\mathrm{mag}}$ (Hz)	FS_t (s)	$\mathrm{FS}_{\mathrm{mag}}$ (Hz)	$\mathrm{Dev}_{\mathrm{mag}}$ (Hz)
29	0	5.001	6.243	49.562	27.500	49.813	0.187
13	2	5.002	6.302	49.566	27.496	49.813	0.187
20	2	5.002	6.372	49.564	27.496	49.813	0.187
31	2	5.003	6.202	49.565	27.500	49.813	0.187
42	5	5.003	6.512	49.562	27.503	49.813	0.187
48	5	5.004	6.142	49.569	27.506	49.813	0.187
55	5	5.005	6.132	49.569	27.505	49.813	0.187

TABLE VII: S-NODAL OBSERVATIONS WITH LARGE DISTURBANCE AND LARGE RESERVE AT FAULT LOCATION GIVEN THAT $H_{\mathrm{agg}} = 2$s

Bus	r	ToA (s)	minima_t (s)	$\mathrm{minima}_{\mathrm{mag}}$ (Hz)	FS_t (s)	$\mathrm{FS}_{\mathrm{mag}}$ (Hz)	$\mathrm{Dev}_{\mathrm{mag}}$ (Hz)
29	0	5.007	6.072	49.915	27.275	49.965	0.035
13	2	5.015	6.242	49.916	27.273	49.965	0.035
20	2	5.013	6.452	49.916	27.274	49.965	0.035
31	2	5.020	6.202	49.915	27.278	49.965	0.035
42	5	5.023	6.532	49.916	27.282	49.965	0.035
48	5	5.027	6.142	49.917	27.285	49.965	0.035
55	5	5.030	6.142	49.917	27.284	49.965	0.035

TABLE VIII: S-NODAL OBSERVATIONS WITH LARGE DISTURBANCE AND LARGE RESERVE AT FAULT LOCATION GIVEN THAT $H_{\mathrm{agg}} = 6$s

Bus	r	ToA (s)	minima_t (s)	$\mathrm{minima}_{\mathrm{mag}}$ (Hz)	FS_t (s)	$\mathrm{FS}_{\mathrm{mag}}$ (Hz)	$\mathrm{Dev}_{\mathrm{mag}}$ (Hz)
29	0	5.010	9.272	49.925	30.357	49.965	0.035
13	2	5.020	8.732	49.925	30.356	49.965	0.035
20	2	5.019	8.682	49.925	30.356	49.965	0.035
31	2	5.028	9.032	49.925	30.359	49.965	0.035
42	5	5.031	8.652	49.925	30.361	49.965	0.035
48	5	5.042	9.312	49.925	30.363	49.965	0.035
55	5	5.055	9.412	49.925	30.363	49.965	0.035

TABLE IX: S-NODAL OBSERVATIONS WITH LARGE DISTURBANCE AND LARGE RESERVE AT FAULT LOCATION AND AN INCREASED RESERVE AT BUS 31 GIVEN THAT $H_{agg} = 2$s

Bus	r	ToA (s)	minima$_t$ (s)	minima$_{mag}$ (Hz)	FS$_t$ (s)	FS$_{mag}$ (Hz)	Dev$_{mag}$ (Hz)
29	0	5.007	6.452	49.917	27.618	49.966	0.034
13	2	5.015	6.382	49.919	27.615	49.966	0.034
20	2	5.014	6.422	49.918	27.616	49.966	0.034
31	2	5.023	6.212	49.915	27.700	49.966	0.034
42	5	5.023	6.582	49.917	27.625	49.966	0.034
48	5	5.028	6.182	49.918	27.630	49.966	0.034
55	5	5.031	6.082	49.919	27.628	49.966	0.034

TABLE X: S-NODAL OBSERVATIONS WITH LARGE DISTURBANCE AND LARGE RESERVE AT FAULT LOCATION AND AN INCREASED RESERVE AT BUS 55 GIVEN THAT $H_{agg} = 2$s

Bus	r	ToA (s)	minima$_t$ (s)	minima$_{mag}$ (Hz)	FS$_t$ (s)	FS$_{mag}$ (Hz)	Dev$_{mag}$ (Hz)
29	0	5.007	6.462	49.917	27.562	49.966	0.034
13	2	5.015	6.412	49.919	27.561	49.966	0.034
20	2	5.014	6.442	49.918	27.562	49.966	0.034
31	2	5.021	6.222	49.915	27.566	49.966	0.034
42	5	5.023	6.532	49.917	27.570	49.966	0.034
48	5	5.028	6.162	49.918	27.574	49.966	0.034
55	5	5.032	6.142	49.918	27.572	49.966	0.034

We observe that the final frequency settling magnitude increased by 1 mHz, the ToAs at the buses relatively increased but there is an observable delay in the minima$_t$ with improvements in the magnitude of the frequency dips as the geodesic distance increases.

Furthermore, to investigate the effect of the same reserve at a bus geometrically farther with r= 5 from the PoC but with the same degree of connectivity, we keep all other system operation parameters constant from our observations in Table VII. The result of this new reserve placement in Table X compared with Table IX shows little observable changes only in the frequency ToAs at buses 31 and 55, which happens to be the locations of inertia. We also observe no change in the final frequency settling magnitude. Also, we observe that the final frequency settling time at the buses improved but with no observable change in its magnitude (i.e., FS$_{mag}$) similar to our observations in the case study 5 in Table V.

From the above numerical experiments, we can note that the frequency ToA at any network bus is a function of system parameters, in particular inertia while the frequency dip time (i.e., Minima$_t$) and its final settling magnitude are respectively functions of turbine's governor time constant and reserves. we can then estimate the deviation in frequency and the resultant frequency magnitude from the mechanics of momentary reserve dynamics if we consider important times of arrival at various points on any network bus after contingency. If a bus with inertia and momentary reserve undergoes contingency specifically of power outage event type and in a low H_{agg} (i.e., high renewable energy source penetration), we can estimate the final frequency settling time (FS$_t$) at the network buses to be,

$$\mathrm{FS}_t = (T_{\mathrm{oA}} M_t) - T_{\mathrm{oA}} + \phi_M, \qquad (13)$$

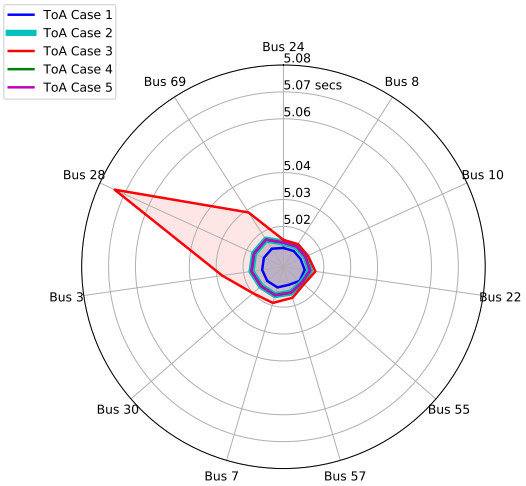


Figure 5: Visualization of Times of Arrival (ToA) in the Nigerian grid study cases I-V as tabulated in Tables I-V.

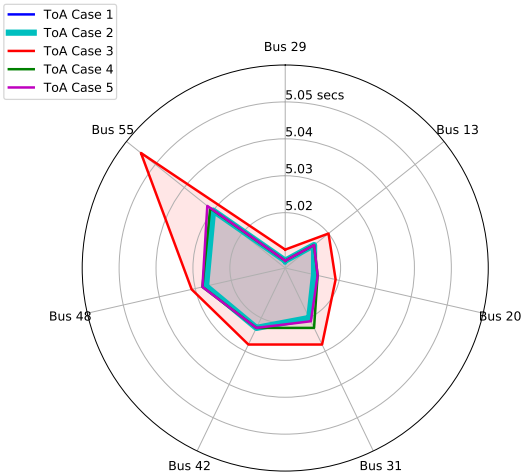


Figure 6: Visualization of Times of Arrival (ToA) in the Square grid study cases VI-X as tabulated in Tables VI-X.

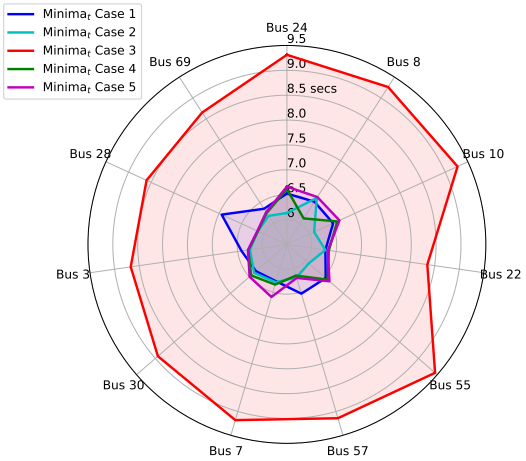


Figure 7: Visualization of frequency dip time (Minima$_t$) in the Nigerian grid study cases I-V as tabulated in Tables I-V.

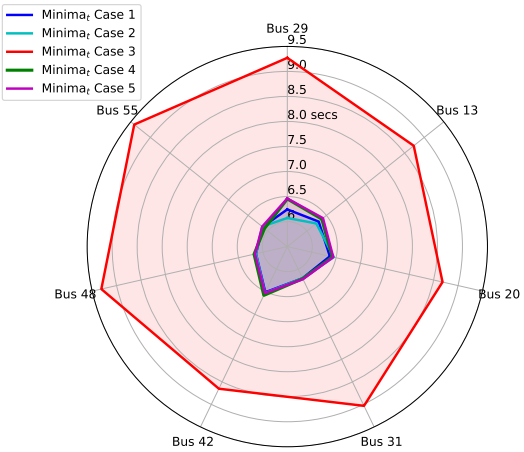


Figure 8: Visualization of frequency dip time (Minima$_t$) in Square grid study cases VI-X as tabulated in Tables VI-X.

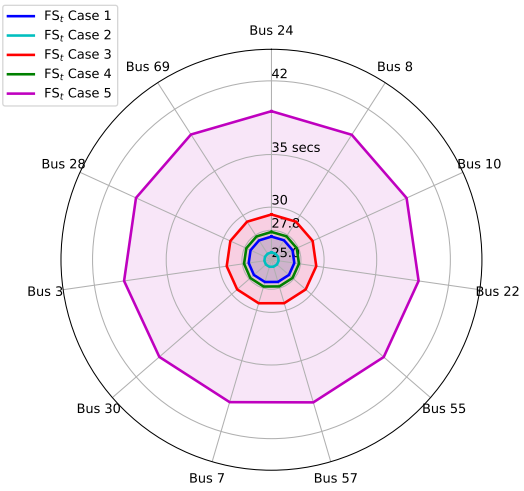


Figure 9: Visualization of final frequency settling time (FS$_t$) in the Nigerian grid study cases I-V as tabulated in Tables I-V.

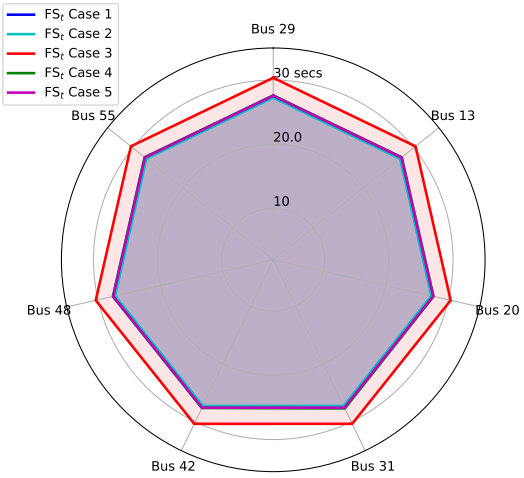


Figure 10: Visualization of final frequency settling time (FS$_t$) in the Square grid study cases VI-X as tabulated in Tables VI-X.

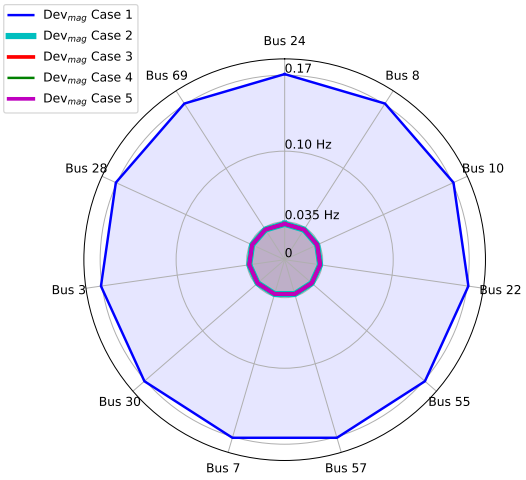


Figure 11: Visualization of final frequency deviation magnitude (Dev$_{\text{mag}}$) in the Nigerian grid study cases I-V as tabulated in Tables I-V.

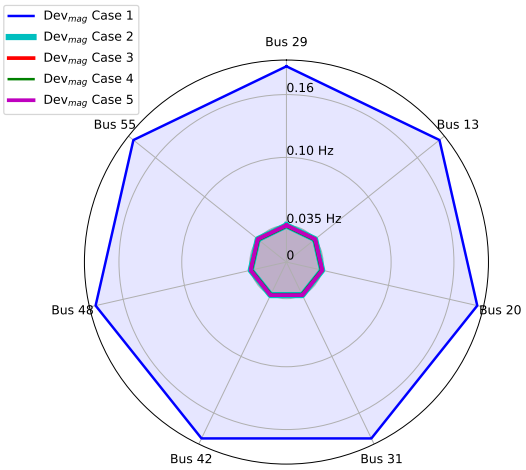


Figure 12: Visualization of final frequency deviation magnitude (Dev$_{\text{mag}}$) in the Square grid study cases VI-X as tabulated in Tables VI-X

where T_{oA} is the frequency time of arrival (ToA), ϕ_M is the grid's coefficient of meshedness as defined in Section IV and M_t is an abbreviation of Minima$_t$, which is a function of speed-droop characteristics of turbine governor defined in (10). The value of FS$_t$ in (13) is true provided that H_{agg} is low. With more momentary reserve distribution at contingent neighbouring buses in the low inertia grid, FS$_t$ can be estimated and given as,

$$\text{FS}_t = (T_{\text{oA}} M_t) + T_{\text{oA}} + \phi_M. \tag{14}$$

However, if a network has a high H_{agg} (e.g. $H_{\text{agg}} = 6$s) with no corresponding increment in the magnitude of momentary reserves, we can estimate FS$_t$ at the network buses to be,

$$\text{FS}_t = (T_{\text{oA}} M_t) - T_{\text{oA}} - M_t - \phi_M. \tag{15}$$

Irrespective of the conditions for the value of FS$_t$ in (13) - (15), the resultant final frequency settling magnitude (with unit in

Hertz) as a result of $1/\mathrm{FS}_t$ deviation can now be estimated as,

$$\mathrm{FS}_{\mathrm{mag}} = \nu_1 - \frac{1}{\mathrm{FS}_t}, \tag{16}$$

where ν_1 and $\mathrm{FS}_{\mathrm{mag}}$ are the pre-contingent and post-contingent network bus frequencies, respectively. The value of ν_1 at normal operation is ν_0 (i.e., 50 or 60 Hz).

V. Results Summary

To properly summarize the characteristics of the case study grids, we combine the 5 study cases in spider graphs to enhance the visualizations, differences and similarities between these grids with different topologies and degree of meshedness (ϕ_M). We reiterate that an increase in the nodal momentary reserve generally delays the travel and arrival of disturbances in a power grid at contingencies as shown in Figures 5 and 7. These observations are found to be similar to the ones in Figures 6 and 8, respectively. In particular, momentary reserves improve the frequency magnitudes at the local minima points and reduces the final frequency settling time in Nigerian network as shown in Figures 7 and 9. These findings are also valid for the Square grid as respectively shown in Figures 8 and 10 but in addition, we observe that its final frequency arrival times are more clustered. This is maybe due to its high degree of meshedness.

Moreover, we found out that the optimal placement of momentary reserve is at the point of contingency as it contributes more in the damping of disturbance across the network more than at any other place as visualized in Figures 11 and 12. Since we may not always be able to predict a fault location, the optimal solution would be to place momentary reserve at all buses where resources and costs allow. In this way, the power system could quickly recover most contingencies within few seconds after their occurrences. Again, we observed that injecting momentary reserve at a bus with high connectivity does not improve the frequency dip time (i.e., minima_t) and final settling time at the buses if the bus is geometrically farther away from the fault location. Hence, the farther away the reserve is from the fault location, the more time it would take for the frequency to stabilize at the buses.

Also, we conclude that increasing the grid inertia without a corresponding increase in the magnitude of the reserve could only delay the travel and arrival of disturbances in electrical network but does not reduce the frequency deviation from the nominal value as seen in case study 3, irrespective of the grid's ϕ_M. Moreover, we established an estimation for the final frequency settling time and magnitude at network buses after contingency, especially of power outage event type. These results would be important to the Transmission System Operators (TSOs) when injecting virtual inertia in the energy transition to renewable schemes.

VI. conclusion

In this paper, we have studied the dynamics of power system reserves in a realistic model of Nigerian power grid as well as synthetic Square power network. By realistic numerical experiments, we explored the mechanisms of momentary reserve contributions to the damping of system oscillations at contingencies and thereby work to restore the grid frequency to its nominal value. We have shown that the optimal placement of momentary reserve in a highly penetrated renewable energy source grid would be at the fault location (i.e., PoC), particularly in the case of generator outage events at plant stations irrespective of the network's degree of meshedness. Since this situation could not always be predicted, we suggest placement of reserves at all nodes where resources permit as this would improve the overall final frequency settling time, frequency dip, and reduce overall frequency deviation from the nominal value, thereby contributing to primary frequency control and reducing the amount of secondary control power needed.

Acknowledgment

We gratefully acknowledge the support of Bundesministerium für Bildung und Forschung (BMBF) CoNDyNet-2, FK. 03EK3055D.

References

[1] K.P. Nnoli and S. Kettemann, "Dynamics of momentary reserves under contingency: Observations from numerical experiments," *ENERGY 2021, The Eleventh International Conference on Smart Grids, Green Communications and IT Energy-Aware Technologies*, 2021.

[2] K. Shi, H. Ye, W. Song, and G. Zhou, "Virtual inertia control strategy in microgrid based on virtual synchronous generator technology," *IEEE Access*, vol. 6, pp. 27 949–27 957, 2018.

[3] P. Kundur, *Power System Stability and Control*. McGraw-Hill, New York, 1994.

[4] J. Machowski, J. Bialek, and J. R. Bumby, *Power System Dynamics: Stability and Control*. Wiley, 2008.

[5] P. Anderson and A. Fouad, "Power system control and stability," *Iowa State University Press, Ames, Iowa*, vol. 1, 1977.

[6] DigSILENT GmbH, *PowerFactory version 2020 Software Manual*, Heinrich-Hertz-Straße 9, 72810, Gomaringen, Germany, 2020.

[7] M. Rajagopalan and M. Rautkivi, "System level value of power system reserve optimization in Singapore and South Korea." Power-Gen Asia, 2014.

[8] Y. Zhang, J. Tan, I. Krad, R. Yang, V. Gevorgian, and E. Ela, *Investigating power system primary and secondary reserve interaction under high wind power penetration*. National Renewable Energy Laboratory (NREL), 2016.

[9] UCTE, "Load-frequency control and performance," *UCTE Operation Handbook Policy 1*, 2004.

[10] P. Kundur, J. Paserba, V. Ajjarapu, G. Andersson, A. Bose, C. Canizares, N. Hatziargyriou, D. Hill, A. Stankovic, C. Taylor, T. V. Cutsem, and V. Vittal, "Definition and classification of power system stability," *IEEE Transactions on Power Systems*, vol. Vol. 19, No. 2, (2004).

[11] A. A. Sallam and O. P. Malik, *Power system stability: Modelling, analysis and control*. IET Power and Energy Series 76, 2015.

[12] G. Tan, C. Xu, F. Wu, C. Qi, D. Wang, P. Yang, and Y. Feng, "Research on primary frequency regulation of wind turbine based on new nonlinear droop control," in *2020 4th International Conference on HVDC (HVDC)*, 2020, pp. 170–174.

[13] K. P. Nnoli, "Implementation of a dynamic network model of the Nigerian transmission grid for investigations on power system stability," *https://doi.org/10.31224/osf.io/r82zn*, 2019.

[14] National Control Center Osogbo, "Transmission data update of the transmission company of Nigeria," *Field Research at National Transmission Company of Nigeria in March*, 2019.

[15] ENTSO-E, *Documentation on controller test in test grid configurations*. Entso-E, 2013.

[16] G. Rogers, "Demystifying power system oscillations," *IEEE Computer Applications in Power*, vol. 9, no. 3, pp. 30–35, 1996.

[17] IEEE Power Engineering Society, *IEEE guide for synchronous generator modelling practice and applications in power system stability analysis*. IEEE Press, 2002.

[18] IEEE Power and Energy Society, *Dynamic models for turbine-governor in power system studies*. IEEE Press, 2013.

[19] A. R. Bergen and D. J. Hill, "A structure preserving model for power system stability analysis," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-100, no. 1, pp. 25–35, 1981.

[20] D. Manik, M. Rohden, H. Ronellenfitsch, X. Zhang, S. Hallerberg, D. Witthaut, and M. Timme, "Network susceptibilities: Theory and applications," *Physical Review E 95, 012319*, 2017.

[21] K.P. Nnoli, S. Kettemann, "Supplementary material in DataPort," *[IEEE DataPort] https://dx.doi.org/10.21227/pjpt-nk47*, 2021.

[22] A. Ulbig, T. S. Borsche, and G. Anderson, "Impact of low rotational inertia on power system stability and control," vol. 14, no. 3, pp. 7290–7297, 2014.

[23] K.P. Nnoli and S. Kettemann, "Spreading of disturbances in realistic models of transmission grids: Dependence on topology, inertia and heterogeneity," *[engrarxiv] https://dx.doi.org/10.31224/osf.io/c8awt*, 2021.

[24] Nigerian Electricity Regulatory Commission (NERC), *The grid code for the Nigerian electricity transmission system*. NERC Press, 2014.

[25] J. Buhl, J. Gautrais, N. Reeves, R. V. Solé, S. Valverde, P. Kuntz, and G. Theraulaz, "Topological patterns in street networks of self-organized urban settlements," *The European Physical Journal B*, vol. 49, no. 4, pp. 513–522, 2006.

How Resonance Works for Development and Propagation of Memes

Muneo Kitajima
Nagaoka University of Technology
Niigata, Japan
https://orcid.org/0000-0002-0310-2796

Makoto Toyota
T-Method
Chiba, Japan
pubmtoyota@mac.com

Jérôme Dinet
Université de Lorraine, CNRS, INRIA, Loria
Nancy, France
jerome.dinet@univ-lorraine.fr

***Abstract*—Meme is a type of behavior that is passed from one member of a group to another, not through the genes but by other means. This paper claims that the concept of resonance should play an important role in the propagation of memes among group members and the development of meme in individual members. Dinet et al. pointed out that the concept of resonance originally issued from physics that has been successfully applied to cognitive processes for behavior selection; the Model Human Processor with Real-Time Constraints (MHP/RT) model proposed by Kitajima and Toyota is a unique model that incorporates a resonance mechanism to connect perceptual-cognitive-motor (PCM) processes that work synchronously, and memory processes that work asynchronously with the environment. This paper elaborates the resonance mechanism implemented in MHP/RT and its associated multi-dimensional memory frames from three perspectives: how resonance works in a single-action selection process that happens in a hierarchically organized structure of behavior; what types of memes can exist in multi-dimensional memory frames; how PCM processes and memory develop from birth to the end of adolescence in terms of the detailed workings of resonance. This paper concludes with the implications of the mechanistic understanding of the role of resonance, the effect of resonance malfunction and the effect of the existence of memory that does not resonate with the real environment.**

***Index Terms*—Resonance; Meme; MHP/RT; Development.**

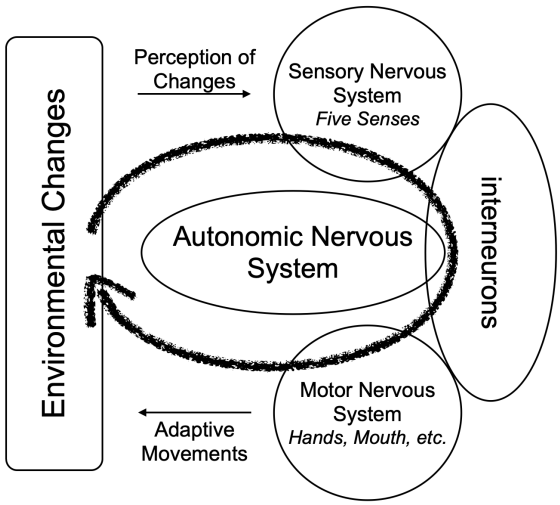


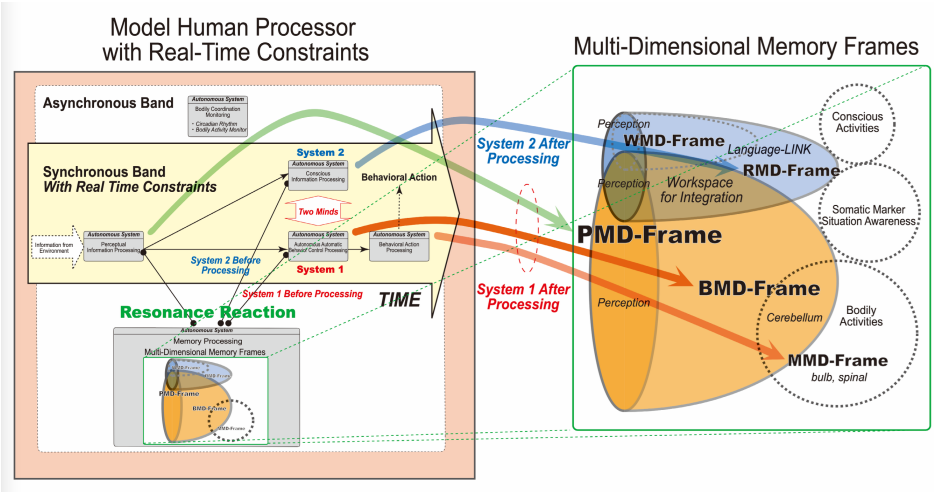
Fig. 1. Continuous cyclic loop of perception and movement (adapted from [8, Figure 1]).

I. Introduction

This paper is based on the previous work originally presented in [1]. It extends the described concepts in Section III by providing new Sections III-D, III-E, and III-F.

At the 0-th order approximation, a person interacts with the environment by running an endless stream of perceptions involving the external and internal environments through the five senses, i.e., taste, sight, touch, smell, and sound, via sensory neurons as parallel processing, and acting in response to the external environment using body parts, e.g., limbs, eye balls, and so on, via motor neurons using serial processing (see Figure 1). As s/he perceives the results of movement of his/her body parts as well as the changes in the external environment with the passage of time, the next cycle of perceptual-motor processes occur. Interneurons between the sensory neurons and motor neurons convert the input patterns to the output patterns – this constitutes a Perceptual-Cognitive-Motor (PCM) process. Starting from this basic cycle (see Figure 1), Kitajima and Toyota [2][3] constructed a comprehensive theory of action selection and memory, Model-Human Processor with Real-time Constraints (MHP/RT), that provides a basis for constructing any models for an acting person (see Figure 2). MHP/RT is an extension of Model Human Processor proposed by Card, Moran, and Newell [4] that can simulate routine goal-directed behaviors. The processes involved in action selection is a dynamic interaction that evolves in the irreversible time dimension. The purpose of MHP/RT is to explain the following three facts: 1) the fundamental processing mechanism of the brain is Parallel Distributed Processing (PDP) [5], which is referred to as Organic PDP (O-PDP) system in the development of MHP/RT; 2) human behavior emerges as a result of competition of the dual processes of System 1, fast *unconscious* processes for intuitive reaction with feedforward control that connect perception with motor movements, and System 2, slow *conscious* processes for deliberate reasoning with feedback control; this is called Two Minds [6]; 3) human behavior is organized under happiness goals [7].

MHP/RT consists of two parts. The first part comprises cyclic PCM processes (see Figure 2, left), in which PDP for these processes is implemented in hierarchically organized bands with characteristic times for operations by associating relative times (not absolute) to the PCM processes that carry



WMD (Word MD)-frame is the memory structure for language. It is constructed on a very simple one-dimensional array.
RMD (Relation MD)-frame is the memory structure associated with the conscious information processing. It combines a set of BMD-frames into a manipulable unit.
BMD (Behavior MD)-frame is the memory structure associated with the autonomous automatic behavior control processing. It combines a set of MMD-frames into a manipulable unit.
PMD (Perceptual MD)-frame constitutes perceptual memory as a relational matrix structure. It incrementally grows as it creates memory from the input information and matches it against the past memory in parallel.
MMD (Motion MD)-frame constitutes behavioral memory as a matrix structure. It gathers a variety of perceptual information as well to connect muscles with nerves using spinals as a reflection point. In accordance with one's physical growth, it widens the range of activities the behavioral action processing can cover autonomously.

Fig. 2. MHP/RT and the distributed memory system implemented as multi-dimensional memory frames (modified from [3, Figure 3]).

TABLE I
NEWELL'S TIME SCALE OF HUMAN ACTION. (ADAPTED FROM NEWELL [9, PAGE 122, FIG. 3-3]).

Scale (sec)	Time Units	System	World (Theory)
10^7	months		
10^6	weeks		SOCIAL
10^5	days		BAND
10^4	hours	Task	
10^3	10min	Task	RATIONAL
10^2	minutes	Task	BAND
10^1	10sec	Unit Task	
10^0	1sec	Operations	COGNITIVE
10^{-1}	100ms	Deliberate Act	BAND
10^{-2}	10ms	Neural Circuit	
10^{-3}	1ms	Neuron	BIOLOGICAL
10^{-4}	100μsec	Organelle	BAND

out a series of events *in synchronous with* changes in the external environment. Table I shows the bands, i.e., biological, cognitive, rational, and social bands, as defined by Newell [9]. It should be noted that there is a gap between two adjacent bands; these two bands are non-linearly connected and therefore it is inappropriate to understand the phenomena that happen across these bands by constructing a linear model. The phenomena occur by connecting what happens in a band to what happens in its adjacent band non-linearly. *A mechanism is required to connect the phenomena; MHP/RT suggests that this connection is provided by the resonance mechanism.*

The bottom left and middle portions of Figure 2 show the autonomous memory system consisting of multi-dimensional memory frames of perception, motion, behavior, relation, and word. These memory frames store information associated with the corresponding autonomous processes defined in the PCM processes. The middle portion of Figure 2 shows the five memory frames and their relationship with the PCM processes. The right portion of Figure 2 provides brief explanations of the respective memory frames. The important feature of the memory system is that it works *asynchronously* with the external environment. MHP/RT assumes that the *synchronous* PCM processes, including the perceptual system, System 1, System 2, and the motor system, and the asynchronous memory system communicate with each other through a resonance mechanism. The concept of resonance has been borrowed from physics to describe the linkage between the asynchronous memory system and the synchronous PCM processes. As Dinet et al. [10] suggested, apprehension of psychological phenomena using concepts issued from physics is useful because the majority of the interactions, including psychological interactions, between humans and the environment (social or physical environment) can be derived from physical processes.

The PCM process cycle in a human being continues from his or her birth until death in the ecological system that consists of the self and the environment. The linkage between the memory system and the PCM processes through resonance supports the development of the PCM processes. At the same time, the PCM processes are accompanied by changes in the connections of neurons that constitute the multi-dimensional memory frames. The purpose of this paper is to understand the development process from the viewpoint of the memory-PCM linkage through resonance.

This paper is organized as follows. Section II focuses on the role of resonance in the use of memory and its relationship with the changes in memory when a human being interacts with the environment. Section III extends the snap-shot view of resonance explored in the previous section in the time

dimension to understand the nature of external objects that generate sustainable resonance. We define these sustainable objects as "memes" and this section describes how memes propagate over time and space. Section IV describes the development of the PCM processes and the multi-dimensional memory frames from birth to the end of adolescence. Section V concludes the paper by summarizing the contents and pointing out implications derived from the considerations of the role of resonance described in this paper.

II. Memorizing the Results of Resonance

The network of sensory neurons, motor neurons, and interneurons shown in Figure 1 develop as a human being interacts with the external environment in time. This section starts by describing a general feature of the development process in Section II-A, followed by the descriptions of what is memorized in Section II-B, and how the resonance mechanism relates to memory usage and memory formation in Section II-C.

A. Forming an O-PDP System

Life emerges in a preexisting structure, which is the global environment that surrounds the human being who is born, and exists as a system that operates under the mechanism of PDP. Life has evolved through adaptation to the global environments, which are characterized by the following preexisting structures:

- **Atmosphere:** fluctuates at the meteorological scales showing chaotic behavior,
- **Sea:** has tidal currents and the surface layer is affected by the atmosphere showing chaotic behavior,
- **Periodic circular structures:** stable periods of rotation and revolution around the sun with a close-circular orbit as a planet of the solar system,
- **Gravity:** upper vs. lower directions, defined by the earth's gravity, and
- **Energy:** future vs. past directions, defined by the direction of energy flow as the dissipative structure of the earth.

Life is formed under these preexisting structures and the direction of the evolution of life is determined by the pressure caused by these structures. Under the dissipative structural space, i.e., the earth, the fundamental structural pressure is a prerequisite to life, which is born and active in it, and prescribes the direction of life evolution. Life is formed as an adaptive body with the functional and structural features that work most efficiently in the environment having the features best conceived by the four concepts of Goals, Operators, Methods, and Selection rules (GOMS) [4].

In the global environment before life occurred, there existed multiple-layered structures equipped with means of communicating information in a multi-dimensional space, including such physical and chemical entities as light, sound, heat, ion, etc. Under these preexisting conditions, life emerged. Life interacts with the multi-dimensional environment by using perception and motion as interfaces to it. Assuming the size of dimension of perception and that of motor to be M and N, respectively, the function of interface is represented as mappings in the $M \otimes N$ space. Specific numbers of M and N reflect the preexisting conditions of the environment.

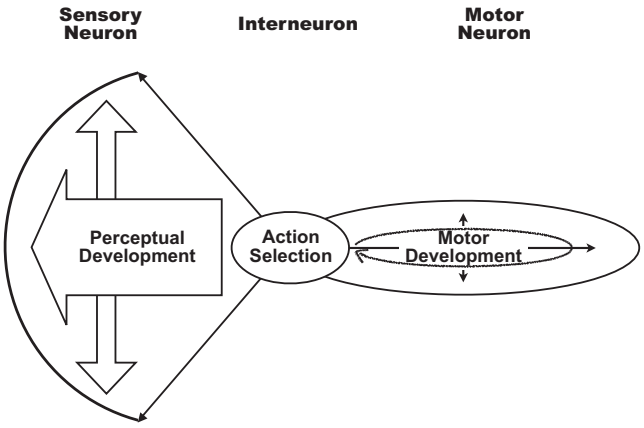


Fig. 3. Development of the sensory nervous system and the motor nervous system, and interneurons connecting them with action selection process.(adapted from [11, Figure 1]).

B. Interneurons to memorize effective $M \otimes N$ mappings

Table I shows that human actions are hierarchically organized in four bands. In general, in-band closed processes are executed in feedforward. Processes carried out in the upper bands provide feedback to the processes carried out in the lower bands. Life activities perform mappings of M-dimensional perceptual input (Biological band) to N-dimensional motor output (Biological band) by the system of interneurons that belongs to the Cognitive band. Activities at an upper band emerge, each of which is associated with a part of the entire sequence of activities, i.e., a sub-sequence, performed at a lower band. The action sequence belonging to the lower band is segmented into sub-sequences, each of which corresponds to one of activities in the upper band; due to the fluctuations in processing times of the upper-band activities, relativization of processing times and functionalization of subsequences occur. It is possible to establish an association between the processing in the upper band activity (Cognitive band) with an effective sequence of $M \otimes N$ mappings in the lower band (Biological band) to form a hard-wired circuit of interneurons that is characteristic of this processing, which is the process of memory formation and development. Once the memory is formed, life activities perform mappings of M-dimensional perceptual input to N-dimensional motor output by using the memory stored in the system of interneurons. In addition, trajectories of the life activities fluctuate, which provides opportunities to the O-PDP system to develop by adaptation.

The genealogy of DNA of vertebrates suggests that Perception, Interneuron, and Motion as the basis for development. Figure 3 shows that PIM is the basis of the formation of the concept of the body based on neural circuits and the whole

is formed as PIM develops. Perception (P) captures various kinds of environmental changes via the sensors with different properties (*M*-dimension). Motor movement (M) is carried out continuously and cyclically from one's birth with gradual development via the accuracy and strength (*N*-dimension). Interneurons (I) memorize effective interlocked relationships between P and M to form neural circuits naturally in the form of feedforward (System 1) and more complex feedback to increase the effectiveness of reactions (System 2). PIM develops by expanding behavioral-ecological bandwidth. Behavioral ecological categories of vertebrates such as acquiring food, raising children, and so on, are almost identical and within a limited range. Everyday human life is performed cyclically in a behavioral ecological band, which is expanded by performing new actions to realize a goal within the limited behavioral ecological categories and adding these to the existing band.

Actions are realized as an *ad hoc* adaptive combination of various elements of the O-PDP system each time. Quasi-stable relationships between elements are generated by local relativization of the time relationship between the participating elements and added to the existing band. Since each process of performing actions is autonomous, it is possible for multiple processes to be initiated in a certain environment, resulting in *coincidental* parallel activities. Even if the timing of the execution of two or more parallel processes changes, it is possible for all of the processes to be completed within a certain time range with or without any relationships in the results. If the overall results are good, behavior selection in the future might be changed by this memory associated with rewards. Arbitrarily activated processes through conscious System 2 thinking, which may or may not be directly related to the current on-going external situation, can become a part of these expanded memories. Conscious thinking facilitates the expansion of memory.

C. Conscious/Unconscious Processes Before/After an Event

MHP/RT maps $M \otimes N$ with the help of memory (Figure 2, left). It assumes the multi-dimensional memory frames and the resonance mechanism for incorporating memory, i.e., interneurons, in the mapping process. Perception with the M-dimensional perceptual information resonates with the contents in the PMD-frame. The activation in the PMD-frame spreads to the other memory frames. The activated portions of memory frames resonate with the System 1 and System 2 processes and are included in these processes. In the left portion of Figure 2, the resonance processes are shown graphically by •——•. It is suggested that the mirror-neuron system is the physical/neurological support for the resonance in human behavior [10] that can be regarded as the process of $M \otimes N$ mapping. The mirror system is thought to directly match visual input from an observed action with a stored motor program for the same behavior (e.g. [12][13][14]). If that motor program is then executed, the result is imitation.

MHP/RT assumes that actions are carried out by feedforward processes (System 1) or feedback processes (System 2) with the help of the resonance mechanism for utilizing existing memory. The processing principle of MHP/RT is that it should work in one of four different modes when one looks at it from *a particular event* that occurred at the absolute time T [15]. These modes are called Four-Process in the MHP/RT architecture. In Figure 2, these four modes are indicated where appropriate to show how they are supported by the resonance mechanism to utilize the contents stored in memory (System 1 and System 2 Before modes), and how they affect in forming memory (System 1 and System 2 After modes) as described below.

1) Rise of Resonance: Two modes work before the event. In the time range of $T-\beta \leq t < T-\beta'$, MHP/RT uses memory for *consciously* preparing for what would happen in the future (*System 2 Before Mode*), and in the time range of $T-\beta' \leq t < T$, it *unconsciously* coordinates motor activities to the interacting environment (*System 1 Before Mode*), where $\beta' \sim 500$msec and β ranges from a few seconds to hours, and even to months. In these two modes, the part of memory activated through resonance in response to perceptual processing could resonate with System 1 processing and System 2 processing (see Figure 2, left).

2) Convergence of Resonance: The other two modes work after the event. In the time range of $T < t \leq T+\alpha'$, MHP/RT *unconsciously* tunes the connections between sensory inputs and motor outputs for better performance for the same event in the future (*System 1 After Mode*), and in the time range of $T+\alpha' < t \leq T+\alpha$, it *consciously* recognizes what has happened and then modifies memory concerning the event (*System 2 After Mode*), where $\alpha' \sim 500$msec and α ranges from a few seconds to minutes, and even to hours. In these two modes, the results of action selection for the event at T would be reflected in the network connections of the respective multi-dimensional memory frames (see Figure 2, middle).

III. Meme Propagation by Means of Resonance

Memory is represented as a network of neurons. It is initially activated through an M-dimensional multimodal sensory input through a resonance mechanism, a large part of which originates from the other human beings or artifacts they have created; then the activation spreads in the connected memory regions. The information generated by human beings propagates among them by means of imitative behavior in which resonance connects memory and the PCM processes to carry out imitation. This raises the question of what causes resonance to be maintained in humans. This section introduces the Structured Meme Theory [16] which provides an answer to this question and provides the micro mechanism of resonance at the level of PCM processes, which is called weak synchronization [17], to see how the process of resonance could occur.

A. What Can Resonate

The O-PDP system that defines a human being interacts with the environment, which consists of not only inanimate objects but also other O-PDP systems (people who interact with him/her). Resonance is the mechanism by which action is taken at appropriate times. Resonance increases when the

organism is in System 1 and 2 Before Modes of the MHP/RT, but converges when in the System 1 and 2 After Modes which effectively update the internal neural networks for future resonance. This section describes what resonates in an O-PDP system. The resonance mechanism, which is implemented by means of mirror neurons [10] in the human brain, is incorporated in the MHP/RT as the mechanism for making available the part of memory activated by perception in System 1 and 2 (see Figure 2, left).

The cyclic loop of perception and movement shown in Figure 1 carries out continuous updates of network connections of constituent neurons during the convergence of resonance period in the MHP/RT. Human behavior is structured in four bands (biological, cognitive, rational, and social bands) [9], and memory is constructed as five distinctive multi-dimensional memory frames (see Figure 2, middle). This suggests that the information is coming from an environment where the informant could either be other humans or inanimate objects. This information will generate resonance in the recipient of the information in a specific way that reflects the states of the development of the multi-dimensional memory frames. For a set of given external stimuli, a human being generates resonance by using the current memory for carrying out the next action and updates it. On the one hand, the PCM processes utilize the resonance mechanism to create the subsequent actions. On the other hand, the external entities, that make resonance generate internally, are created ultimately by humans. Human action associated with a band of the Newell's time scale could be associated with the part of memory that participates via the resonance mechanism.

From the above argument, it is plausible to assume that a human could create an entity that makes the others generate resonance for performing the subsequent actions and vice versa. This assumption represents what was meant by the term "meme" that Richard Dawkins [18] coined in the 1970s, which was conceptual and was not defined clearly. The Structured Meme Theory (SMT) proposed by Toyota et al. [16] defines memes as entities that represent the information associated with the objects that the brain can recognize.

B. Propagation of Meme

Figure 4 depicts how memes propagate in the reality field. The process of propagation is facilitated by symbolization. A symbolized meme enables human beings to think at abstract levels. Suppose that Object-O_1, Object-O_2, or Object-O_3 appear in the environment. Each of these resonates with memory by using patterns stored in the PMD-frame and causes activation of RMD-, BMD-, and MMD-frames. The part of memory that is activated, Pattern-P_a (see Figure 4, bottom), will be used in action selection through resonance. Any entities in the environment that match Pattern-P_a are treated in the same way in the action selection process as Object-$O[P_a]$. Pattern-P_a will be given a unique name Symbol-S_a and it will be stored in WMD-frame. In the future, when Symbol-S_a appears in the environment in someone's utterances, it might resonate with Pattern-P_a. Symbol-S_a could be associated with Object-O_1, Object-O_2, or Object-O_3. Therefore Symbol-S_a effectively functions as its abstraction, AbstractSymbol-AS_a, when it is used in communication. It is reasonable to assume that the higher the resonance level of Symbol-S_a becomes, the longer AbstractSymbol-AS_a is maintained and inherited as an effective communication medium.

C. Structured Meme

The recent consensus is that the range of information inherited by the genes is limited to physical functions and infantile behavior. Human beings need to acquire basic behavioral skills and communication skills through the experience of acting in the environment. As described in Section III-A, memes that can cause resonance constitute an important part of experience. This section explains that memes are structured in such a way that the memory is structured in multi-dimensional memory frames.

Figure 5 expands Pattern-P_a, that appears at the bottom of Figure 4, and the processes shown in Figure 5 as "Mapping Patterns on Brain" and "Mapping Patterns on WMD-frame" in order to show which portions of the neural networks would participate while a human being utilizes the WMD-frame in communication. As shown by Figure 4, symbols in the WMD-frame are gradually incorporated into the environment in the form of a thesaurus, i.e., lists of words in groups of synonyms and related concepts, languages used for person-to-person communication (individual language), which might include not only direct but also metaphorical uses, and languages used in cultural contexts (cultural language), in which appropriate understanding of common sense that has been established in the specific community, is essential for successful communication. Thesauruses, individual languages, and cultural languages increase their complexity in this order in terms of the patterns they are linked with the objects in the environment. Thesauruses are associated with the objects in the environment that are encoded in the neural networks in the initial development stage from the birth to 3 years. Individual languages are associated with not only the objects in the environment but also the symbols that have already been incorporated in the environment. The same is true for cultural languages.

The process of "Mapping patterns on symbols in WMD-frame" shown in Figure 4 can be subdivided into three processes based on the degree of complexity of mapping. The patterns that were mapped on the thesauruses, individual languages, and cultural languages are shown as Action-Level Meme, Behavior-Level Meme, and Culture-Level Meme (see Figure 5), respectively, that were introduced in the Structured Meme Theory proposed by Toyota et al. [16] . The mechanism by which the three levels of memes and genes inherit information is analogous to an information system. Genes serve as firmware that mimics behavior-level activities. Action-level memes serve as the operating system that defines general patterns of spatial-temporal behavioral functions. Behavior-level memes represent middleware that extends the general patterns to concrete patterns. Culture-level memes act as

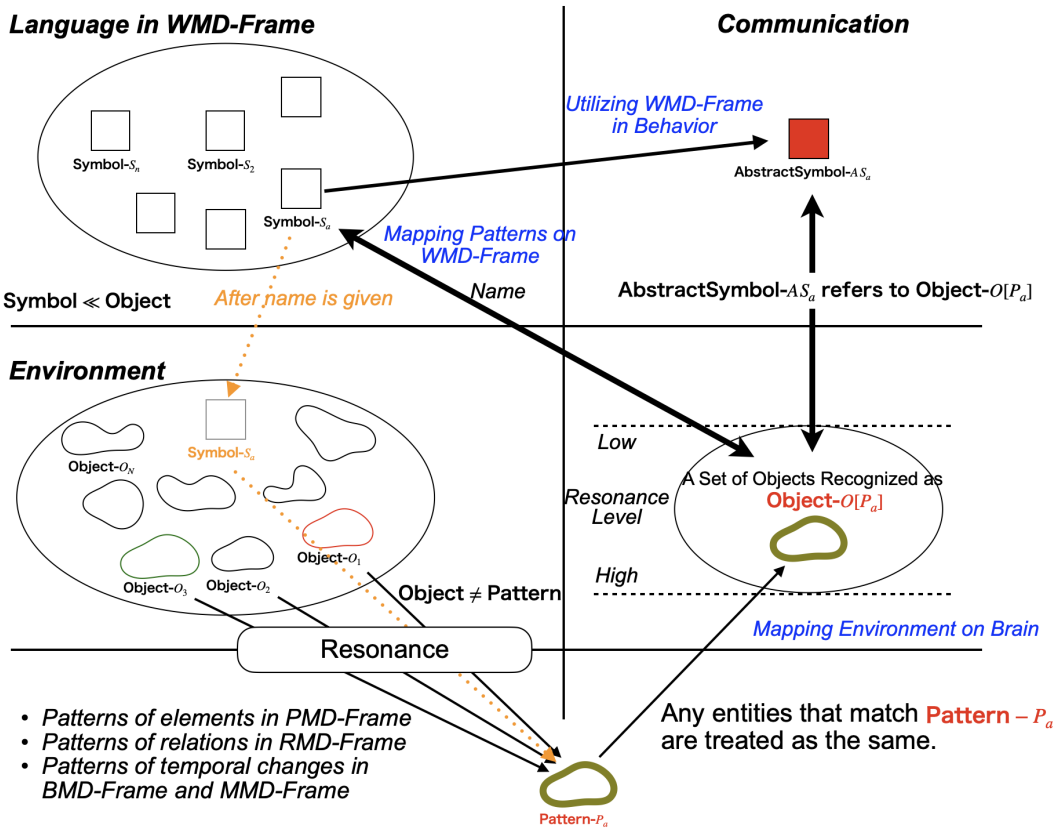


Fig. 4. Propagation of Meme (adapted from [1])

application tools that extend the concrete patterns to the ones that work in a number of groups of people.

The relationships between the three levels of memes and multi-dimensional memory frames are as follows:

- Action-level memes represent bodily actions stored in the MMD-frame;
- Behavior-level memes represent behaviors in the environment stored in the BMD-frame; and
- Culture-level memes represent culture stored in the RMD-frame and the WMD-frame.

D. Process and Structure, and Resonance

Figure 4 shows that the appearances of Object-O_1, Object-O_2, Object-O_3, or Symbol-S_a in the environment causes resonance with memory by using *patterns* stored in the PMD-frame. This resonance process is shown by the leftmost •——• in Figure 2. Figure 6 schematically illustrates the process of chain-firing in the PMD-frame that follows the resonance with the objects. The accessible memory region, depicted in the red oval in the figure, corresponds to the region where activation levels are kept above certain threshold values necessary for maintaining chain-firing using primarily perceptual stimuli. Chain-firing originates from continuous input from the current environment.

This initial resonance process follows activation of the patterns of relations stored in the RMD-frame and the patterns of temporal changes stored in the BMD- and MMD-frames. The part of memory that is activated, Pattern-P_a (see Figure 4, bottom), will be used in action selection by means of System 1 and System 2 through the subsequent resonance processes, which correspond to the rightmost and the middle •——•'s in the left portion of Figure 2, respectively. These resonance processes occur in the "Rise of Resonance" phase introduced in Section II-C, where the current network structure of memory is used. The memory is updated in the "Convergence of Resonance" phase by reflecting the outcome of the execution of the selected action. The cyclic process of resonating with memory using *patterns*, performing action by applying *patterns*, and updating *patterns* in memory continues endlessly from one's birth to death.

A pattern refers to the snapshot of the state of activation of the current memory network. However, the activated network involves not only memories which are directly relevant to the current situation but also sequences of actions that have lead to success in the past. The latter is very useful in determining which action to select at the moment. In other words, past successful experiences are integrated in the memory as part of the process of resonance, i.e., the Convergence of Resonance.

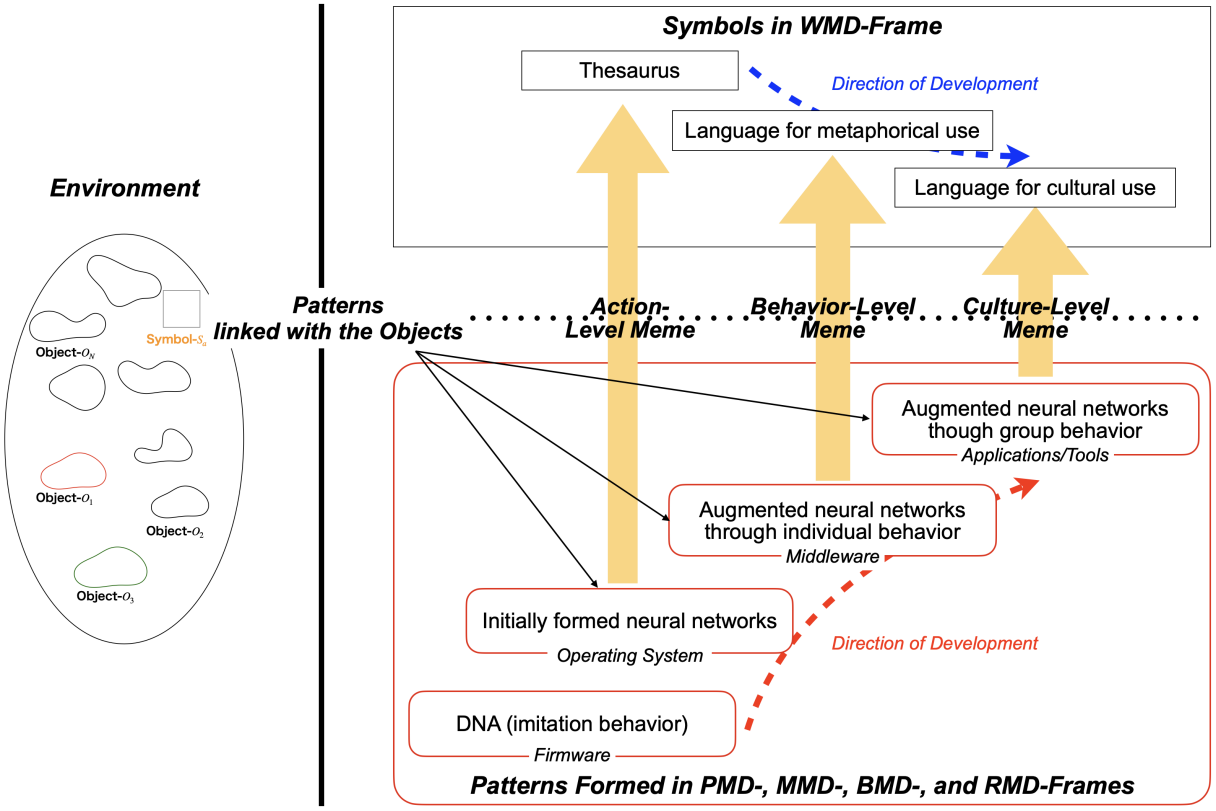


Fig. 5. Structure of meme (adapted from [1])

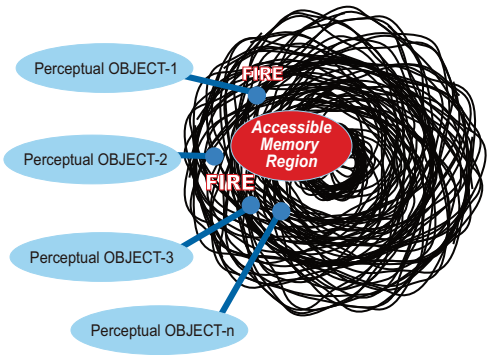


Fig. 6. Chain-firing triggered by perceptual stimuli in a cross-networked memory structure. (adapted from [19, Figure 2]).

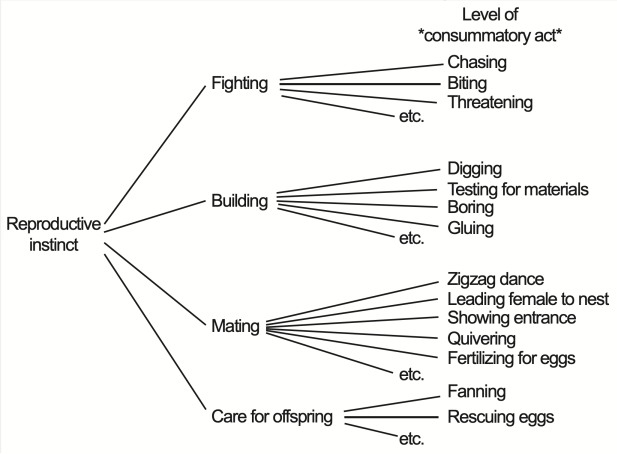


Fig. 7. Fixed hierarchy in the behaviors accompanied with instinct of procreation (redrawn Figure 8.10 in [20]).

The pattern stored in memory develops as one accumulates experiences. In the ever-changing environment in which the degree of variations is restricted by Mother Nature, the patterns that *can* occur are restricted, which results in a structure of patterns. In other words, a certain structure emerges which should restrict the ranges of possible patters which can occur, which circularly restricts the direction of development of the structure. At the primitive level, Figure 7 shows the fixed hierarchy in behaviors accompanied with instinct of procreation through the observation of behavior of Nemipterus virgatus [20]. Since the evolution of vertebrata is no more than a history of increasing complexity, the results of any ecological analyses of animal behavior would tend to lead to implicit structures, which is free from he complications of society and cultural diversity.

Exploring the phenomenon wherein a living creature acts in

the ever-changing environment through reciprocal relationship between processes and structures, and that it establishes a quasi-stable universe through self-organization is worth considering for a deeper investigation of the role of resonance. Given a certain stable system, a question may arise: why is the system stable? A system can be viewed as the processes that are carried out in the system stably, or the structure that makes the processes stable [21]. Newell's time scales of human actions as shown in Table I define the *structure* of human actions that are carried out by conducting the *processes* shown in MHP/RT (Fig. 2). Chain-firing defines the *processes* which take place in the memory defined by multi-dimensional memory frames (the middle portions of Figure 2).

Patterns encode a time series of events at the levels of perceived elements in the PMD-frame, relations in the RMD-frame, and temporal changes in the BMD- and MMD-frames. The encodings of the current situation at these levels include not only a time series of encodings up to the present, S_t $(t = -n, -(n-1), \cdots, 0)$, but also other encodings of future time series, $S_{t'}$ $(t' = 1, \cdots)$. A pattern can resonate with the current situation if its past encodings match well with how the current situation was achieved. In other words, the time series of the current situation matches the pattern in memory, and synchronization is established between the external and internal worlds in memory. The degree of synchronization will be kept at a good level if the external world develops as the memorized pattern; however, it will become weak if the degree of discrepancy between them becomes large. The patterns that have a good match both in the past and in the future can result in the development of action-level memes, behavior-level memes, and cultural level memes. On the other hand, or those that cannot establish good matches, the corresponding part of memory will lose its strength and the associated memes will become extinct.

E. Resonance Implemented as Weak Synchronization

Patterns encode the time series of an event. Resonance activates patterns. Living creatures need to take appropriate actions in a timely manner as the external environment develops. In other words, synchronization occurs between a living creature and the external environment. The term "synchronization" places emphasis on the dimension of time that proceeds independently in relation to the living creature and the external environment. On the other hand, the term "resonance" emphasizes the phenomenon that happens between the two entities.

A state of resonance may be described from the view point of synchronization, and it is referred to as *weak synchronization* [17]. Its mechanism, shown by Figure 8, is described as follows. Any event that occurs at T can be viewed by the person as an event that has happened in the time range of $[T-\beta, T+\alpha]$, which corresponds to the beginning of "Rise of Resonance" and the end of "Convergence of Resonance" introduced in Section II-C. Regarding the event that occurs at T, it is important to note that the consciousness which is related to the event comes to play in System 2 After Mode at $T+\alpha'$. This implies that consciousness lags by a certain amount of time, α', in the real world. Therefore, a person is consciously blind during the period of $[T-\beta', T+\alpha']$. However, this person would consciously integrate the blind period into conscious activities during the time range of $[T-\beta, T+\alpha]$. Therefore, a consciously retrievable experience has to be considered as a memory structure related to the event at T in the extended time range of $[T-\beta, T+\alpha]$, with a consciously inaccessible but integrated memory region corresponding to the consciously blind time range of $[T-\beta', T+\alpha']$.

The external event at T is mapped internally as a consciously retrievable event. Specifically, a person's activity related to an event has to be considered from the four processing modes, which ranges between a relatively long time before and after the actual time the event happens. Therefore, "synchronization" has to be considered as an alternative to the phenomena as follows.

A person's activities during the time range of $[T-\beta, T+\alpha]$ are linked with a specific recognizable event at time T. Furthermore, the activities are carried out using any of the four processing modes. When this is satisfied, the event is considered synchronized with a person's activities, which is referred to as weak synchronization.

The schematic presented in Figure 8 shows how MHP/RT works under weak synchronization during the time range of $[T-\beta, T+\alpha]$. During $[T-\beta, T-\beta')$, activities at upper bands, social, rational, and upper cognitive bands are carried out consciously, The conscious activities are taken over to unconsciously performed activities, which are carried out during the period of $[T-\beta', T)$, the time when the event happens. This is followed by unconscious tuning activities during the period of $(T, T+\alpha']$. Finally, these activities are taken over to the activities carried out during the period of $(T+\alpha', T+\alpha]$ for conscious reflection. Now, we must explain the mechanism for establishing chains between activities.

Firstly, we think about what are combined. An O-PDP system is composed of autonomous elements in a band-structure. The bands are: Biological, Cognitive, Rational, and Social Bands as suggested by Newell [9]. Activities carried out in the respective bands are combined with each other to form a function, then coherently integrated for the event at T. Now, how are these activities combined? The combined entity indirectly reflects the circularity of the existing environment and fluctuations inherent to the environment, which is called autopoiesis [22]. Circularity connotes the relativization of time: no distinction between "after" and "before," and these fluctuations indicate that the resulting system is a complex system. The event appears recursively in the PCM processes, and is likely to extend the relevant time range, i.e., β and α would become larger. The larger β becomes, the farther the person can foresee. The larger α becomes, the more the person can elaborate on the event. These should make the person live smarter by widening the range of resonance.

Now we consider what results from weak synchronization. Various combinations of chains between parallel processes can occur, causing path proliferation within the network. When the

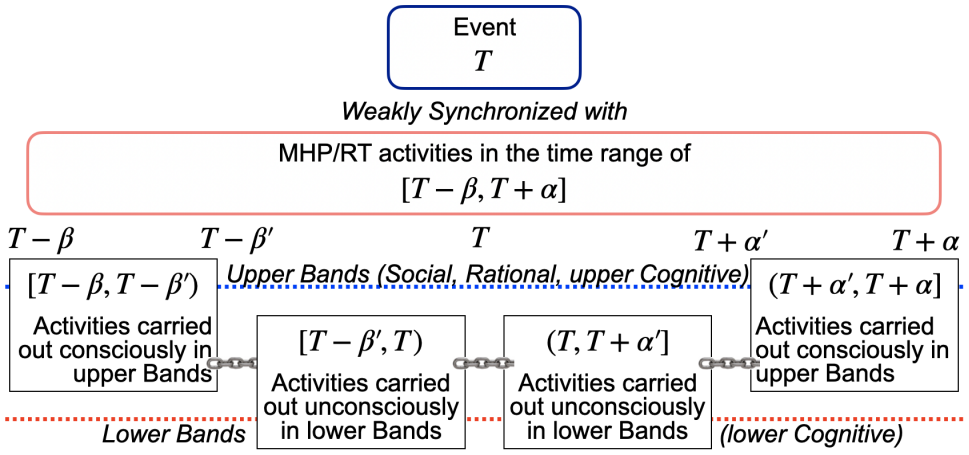


Fig. 8. A schematic illustration of the mechanism of weak synchronization during the period of $[T-\beta, T+\alpha]$.

recall rate of specific paths becomes higher, proliferation along these paths is suppressed to centralize the activations on the subsequent paths. This makes it possible for the O-PDP system to survive in the cyclic and fluctuating environment. Each element of the O-PDP system plays a certain role in achieving the overall goal of the whole O-PDP system. However, its role is not determined from the beginning and it exists only as the result of each element's own efforts to survive since its activation.

F. Resonance in a Circular Environment

At the microscopic level of the Newell's time scale of human action (see Table I), i.e., at the Cognitive Band, resonance can be understood as the phenomenon that occurs between the external environment and the PMD-frame, and between the multi-dimensional memory frames and Two Minds, which is implemented by means of weak synchronization as described in Section III-E. However, the PCM process for performing weak synchronization is circular even at this level because memories stored under System 2 After Mode during the "Convergence of Resonance" may be used in the future under System 2 Before Mode, and this happens recursively. This recursive process accumulates past experiences in the BMD- and RMD- frames. Since the characteristic times become longer as the more the number of experiences accumulate, the appropriate time scale for conceiving the phenomena shifts from Cognitive Band to the upper band of Rational and Social bands.

Figure 9 illustrates the results of the process of accumulation of experiences which cause the extension of networks in the BMD- and RMD-frames. This figure shifts attention from the microscopic processes of weak synchronization to the resultant networks in BMD- and RMD-frames that should be formed after a number of recursions of resonances, and that should support human actions at the Rational and Social bands. Input of perceptual stimuli as presented in the middle-left of Figure 9 awakens a person who receives it, and causes resonance with the patterns in the PMD-frame, followed by activation of relevant patterns in the BMD-frame. The activation of memory leads to System 2's conscious activities and System 1's unconscious activities, before leading to physical activities under the MMD-frame. The networks under the BMD-frame are extended by reflecting on the physical activities. Even during sleep, contents of the networks in the BMD-frame are continuously updated by autonomous activities. Under the BMD-frame, a circular neural network formed as a result of physical activities that are continuously performed from the time the BMD-frame was originally generated to the time when the organism dies and stops its activities.

The top half of Figure 9 shows the contents of the RMD-frame. A variety of relations are stored in the RMD-frame. Relations pertain to those between the person and his or her concerns at the time of performing physical activities. As shown in the figure, the targets of the person's concerns could be individual, i.e., self, his or her family, any communities to which they belong, or any enterprises they work for. If physical activities are characterized as *individual* activities as shown in the top-left of Figure 9, its related memory in the RMD-frame is continuously formed as a neural circuit in relation with the activated PMD-frame from the time that the conscious activity is initiated. Since conscious activities which are executed intermittently are accumulated, it becomes discontinuous. However, since conscious activity is executed periodically in the external environment, this results in a circular network that reflects it. These formed networks should reside at the Rational or Social bands and serve as patterns which resonate in the future when the person encounters similar situations. If the physical activities are characterized as *group* activities as shown in the top-right of Figure 9, their related memories in the RMD-frame are continuously formed as neural circuits in relation with the activated PMD-frame from the time when the activity was initiated. Since this is formed by an accumulation of each group activity, i.e., family, community, and enterprise, it becomes discontinuous.

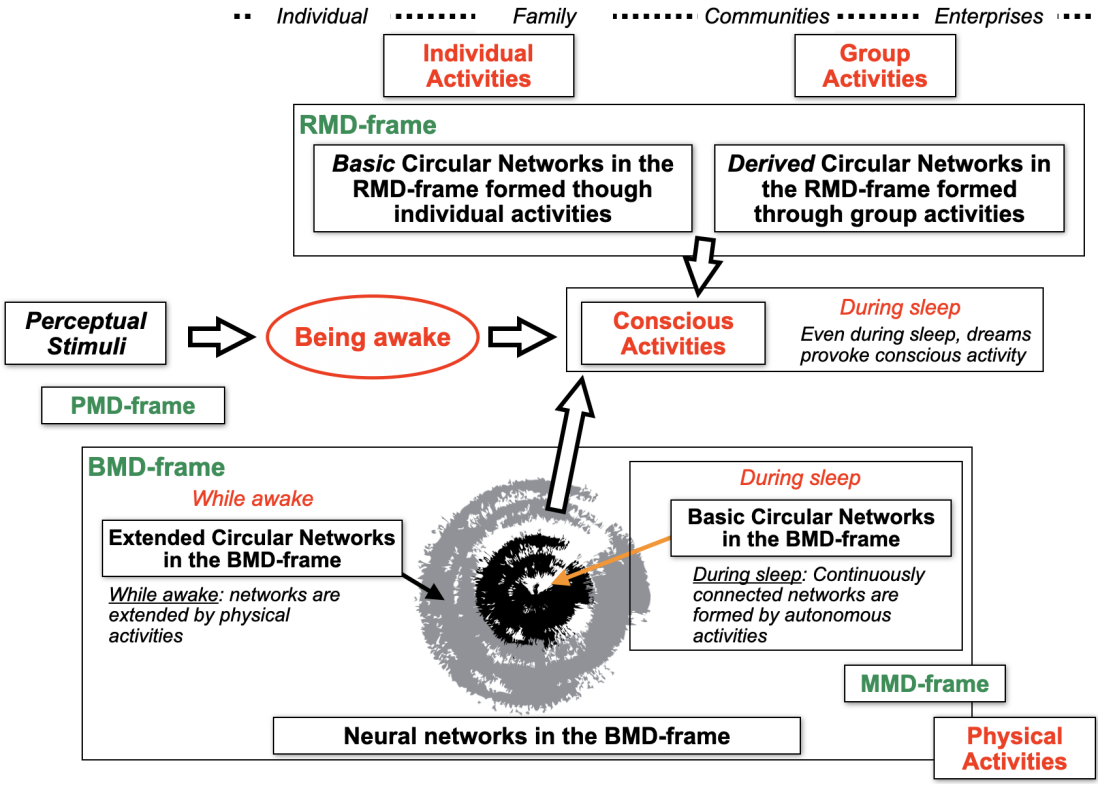


Fig. 9. Development of the BMD- and RMD-frames in a circular environment.

However, since these group activities are executed periodically, it becomes a circular network that reflects it.

It is important to note that the different portions of the RMD-frame resonate with the same perceptual stimuli that should resonate with the same part of the PMD-frame. However, this might lead to different physical activities. The mappings of a single perceptual stimulus in relation to physical activities could multiply via the social-role dependent activation of the RMD-frame. This has deep implications for development of a person, i.e., *a person behaves differently depending on the environment of the group in which he or she participates*.

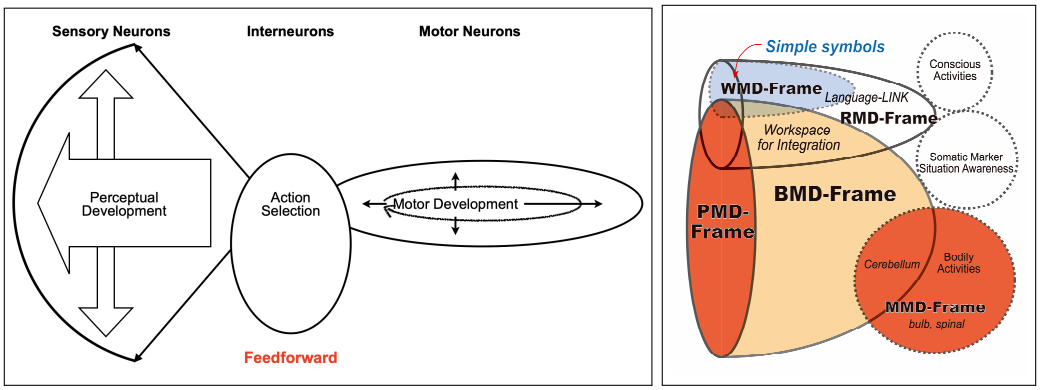
IV. Development of PIM and Multi-dimensional Memory Frames

In Section II, we described the mechanism by which PIM develops with the Four-Process by utilizing the resonance mechanism. Section III described how the contents of multi-dimensional memory frames, which could resonate with the PCM processes running synchronously with the environment, are maintained among human beings through communication. This section describes the changes in the contents of PIM and multi-dimensional memory frames over time as human beings grow up from birth to the end of adolescent. This explains human development in terms of PIM and multi-dimensional memory frames (Figure 5, bottom).

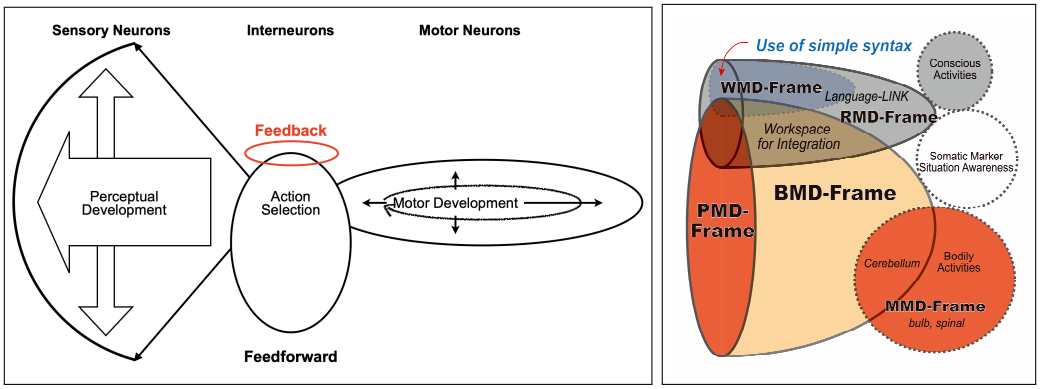
This section elaborates the four stages in PIM development as described in [8]. Table II shows the age range, rate of autonomous development of perception and motion, and the method of connecting sensory and motor neurons via interneurons and the resultant control method. Figure 10 depicts schematically the development of PIM (the sensory nervous system and the motor nervous system characterized by the functioning of interneurons connecting them (Figure 3)) and the development of multi-dimensional memory frames that shows the structured inter-connections of interneurons along with each stage (Figure 2, center). The multi-dimensional memory frames that develop at a particular stage are shown by adding shades.

A. Stage 1

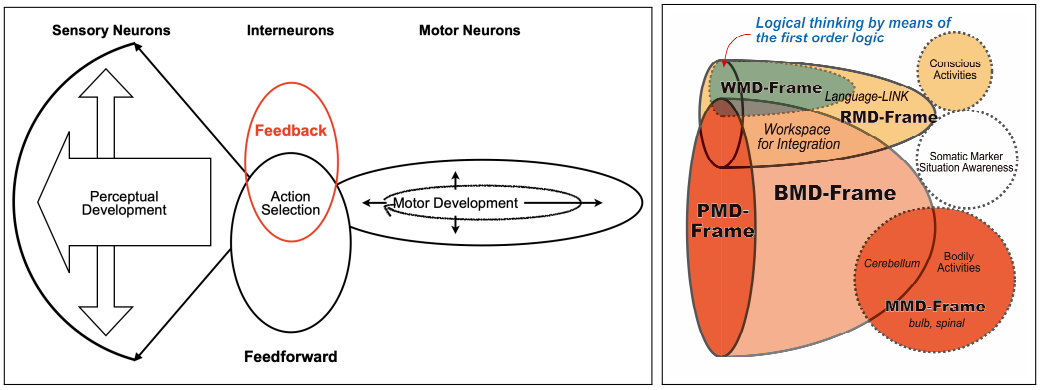
At $0 \sim 6$ years of age, feedforward loops are the dominant control mechanism and these establish fundamental relationships between MMD- and BMD-frames by means of "uplinking." In the first half of this period, at $0 \sim 3$ years of age, sensory and motor neurons experience rapid autonomous development. Interneurons connect them by subordinate-intervention, resulting in feedforward control (Figure 10-1A). During this period, human beings establish inter-connections between PMD- and MMD-frames, and PMD-, MMD-, and BMD-frames. The former is related to action-level memes and the latter, to behavior-level memes. These make integrated movements of bodily actions possible on the basis of the



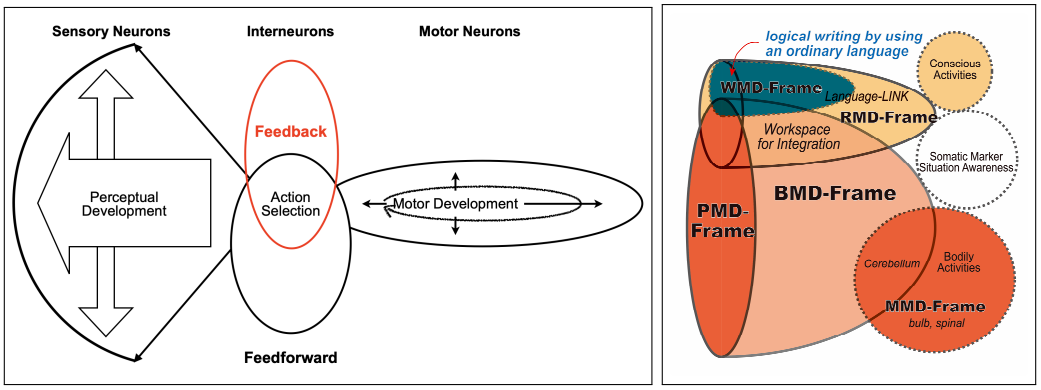
Stage 1 1A 1B



Stage 2 2A 2B



Stage 3 3A 3B



Stage 4 4A 4B

Fig. 10. Development of PIM and multi-dimensional memory frames (adapted from [1])

TABLE II
PIM DEVELOPMENT PROCESS (ADAPTED FROM [1])

		Speed of Autonomous Development		Method for Establishing Connection	Resulting Control Mechanism
Stage	Age	Perception	Motion	Interneurons	
1	$0\sim 3$	Rapid	Rapid	subordinate	Feedforward control
2	$4\sim 7$	Fast	Fast	memory-mediated	Limited feedback control
3	$7\sim 12$	Stable	Stable	memory-mediated-proactive	Wide feedback control
4	$13\sim 18$	Settled	Settled	memory-mediated-autonomous	Extended feedback control

relationships between the input from the perceptual system in the PMD-frame and the output expressed as reflexive movements in the MMD- and BMD-frames, for example, simple utterances (Figure 10-1B).

B. Stage 2

In the latter half of this period, at $4 \sim 6$ years of age, there is rapid autonomous development of sensory and motor neurons. Interneurons connect these by memory-mediated interventions, resulting in feedback controls (Figure 10-2A). During this period, human beings acquire the skill of behaving in relation with other persons and the methods for conversing with others such as the ability to explain using simple syntax. Acquisition of utterances and combined motor movements become possible by using the RMD-frame that connects simple syntax of symbols in the WMD-frame by linking PMD-, BMD-, and MMD-frames. At this stage, the conscious processes of System 2 Before and System 2 After that make feedback controls possible are initiated (Figure 10-2B).

C. Stage 3

Later, at $7 \sim 12$ years of age, parallel distributed activities are almost complete. There is stable autonomous development of the sensory and motor neurons. Interneurons connect them by memory-mediated proactive interventions, resulting in wider feedback controls (Figure 10-3A). During this period, human beings acquire the skill of first-order logical thinking by using letters or symbols and that of cooperation with other persons. These activities facilitate the development of interconnections among the multi-dimensional memory frames, resulting in very complex networks. Figure 10-3B) shows the development of interconnections in the RMD- and WMD-frames. The key is the presence of symbols that intervene between the various input and output connections; these symbols are related to culture-level memes.

Structural development precedes procedural development. The former refers to the development of connections between PMD- and MMD-frames with the support of the BMD-frame. The BMD-frame stores repetitive sequences of motor actions in the MMD-frame as single units; therefore, the action sequences in the BMD-frame can be carried out unconsciously. The latter refers to the development of connections between WMD- and RMD-frames. The connections start from spoken informal language to more abstract structural representations of formal language. The WMD-frame stores sequences of symbols as language dissociated from the real world; the RMD-frame stores a set of BMD-frames into a manipulable unit. The connections between the contents in the WMD-frame and those in the RMD-frame give reality to the contents in the WMD-frame. Structural development and procedural development interact with each other in the RMD-frame occasionally and proceed in parallel. Procedural development in the RMD-frame is critical to extend the range of behavior. Once elements in the RMD-frame are compiled in the System 2 After mode, the compiled rule can be used to initiate a behavior with longer action steps.

D. Stage 4

Finally, at age $13 \sim 18$ years, parallel distributed activities are complete. There is settled autonomous development of sensory and motor neurons. Interneurons connect these using memory-mediated autonomous intervention, resulting in even wider feedback control (Figure 10-4A). During this period, feedback loops are established; these are used to form language processing circuits in a single-layer, WMD-frame. During this period, the ability of logical writing and the generation of grammatically correct sentences by using ordinary language affects significantly the development process. The WMD-frame can evolve autonomously and extensively; part of the WMD-frame is associated with the BMD-frame and ultimately connected with the real world but the rest of the WMD-frame is dissociated from the real world because there is no connection between the WMD-frame and RMD-frame (Figure 10-4B).

V. CONCLUSION

Resonance connects remote systems enabling them to work jointly. Human beings act in the environment by using two systems: a cyclic PCM system that works in synchronization with the environment and an autonomous memory system that operates asynchronously with the environment. Based on MHP/RT, this paper describes the hypothesis that the PCM system and the memory system interact with each other via a resonance mechanism. First, this paper shows that a single action selection carried out by the PCM system was conceived as a mapping of the input from the environment represented

in M-dimensional information in the PMD-frame on the output to the environment represented in N-dimensional motor actions in the MMD-frame, connected with the BMD- and RMD-frames, and discusses the role played by resonance in the PCM processes defined by the MHP/RT and updates of the memory represented by the multi-dimensional memory frames. Second, in situations where the environment comprises objects created by humans, i.e., artifacts such as symbols, manners, language, and culture, this paper discusses how artifacts are maintained over time and space with the help of resonance. A necessary condition for the survival of artifacts is that they resonate with human beings. This paper introduces action-level, behavior-level, and culture-level memes to represent resonating entities stored in memory. Third, we also described the human development process from birth to the end of adolescence in terms of the changes in PIM and memory, taking into account the changes in memes that human beings develop in memory as they grow older.

The argument this paper has profound implications for cases where there is malfunction of resonance, and where memory that fails to resonate as reality develops.

An example of the first case is Autism Spectrum Disorder (ASD). Is ASD a resonance problem? Imitation is a potentially crucial aspect of social cognitive development [23] because imitation is an efficient tool for two main adaptive functions: learning and communication. Imitation-based communication is possible through the use of the two aspects of imitation: imitating and being imitated [24]. These two aspects give rise to two roles that partners can exchange by taking turns while they synchronize matched activities. Neuroimaging studies of interactive imitation have shown that such communicative systems involve a coordination of bottom-up and top-down processes. In accordance with some authors (e.g. [25]) and based on the MHP/RT model, we hypothesize that perceptive-motor resonance plays a more central role in imitation in infancy than does a rational evaluation of the observed action. Thus, because individuals with ASD have difficulties with imitation, the nature of which remains unclear, it is plausible to postulate that one of the main problems of ASD is related to difficulties with the mechanisms underlying resonance. Several studies have shown that very young children with ASD (M Age $<$ 48 months; for a synthesis: [26]) used imitation less often when copying the actions of others, spent less time looking at others' faces and more time looking at the actions of others; attentional, social and executive factors underlie different aspects of imitation difficulties in this population. In other words, ASD is characterized by a deficiency in the development of imitation capacity, and this deficiency implies a deficit in mapping neural codings for actions between sensory and motor modalities, rather than in motivation or executive function [27]. Thus, ASD could be characterized by abnormal development of these mappings and that of resonance mechanisms.

Another important implication of our hypothesis is related to the autonomous development of the WMD-frame at the later stage of development. This paper showed that part of the contents in the WMD-frame do not necessarily have reality, which is established by linking the WMD-frame with the chain of RMD-, BMD-, and MMD-frames. Systematic and logical knowledge as a group norm could become part of the WMD-frame through education. It could resonate with the symbols that refer to the knowledge but would not resonate with *real* objects in the environment. The larger the non-resonant part of memory, the less the person will resonate with reality in the environment.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number 20H04290. The author would like to thank Editage (www.editage.com) for English language editing.

REFERENCES

[1] M. Kitajima, M. Toyota, and J. Dinet, "The Role of Resonance in the Development and Propagation of Memes," in *COGNITIVE 2021 : The Thirteenth International Conference on Advanced Cognitive Technologies and Applications*, 2021, pp. 28–36.

[2] M. Kitajima, *Memory and Action Selection in Human-Machine Interaction*. Wiley-ISTE, 2016.

[3] M. Kitajima and M. Toyota, "Decision-making and action selection in Two Minds: An analysis based on Model Human Processor with Realtime Constraints (MHP/RT)," *Biologically Inspired Cognitive Architectures*, vol. 5, pp. 82–93, 2013.

[4] S. K. Card, T. P. Moran, and A. Newell, *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.

[5] J. L. McClelland and D. E. Rumelhart, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition : Psychological and Biological Models*. The MIT Press, 6 1986.

[6] D. Kahneman, "A perspective on judgment and choice," *American Psychologist*, vol. 58, no. 9, pp. 697–720, 2003.

[7] D. Morris, *The nature of happiness*. London: Little Books Ltd., 2006.

[8] M. Kitajima and M. Toyota, "Hierarchical Structure of Human Action Selections – An Update of Newell's Time Scale of Human Action," in *Procedia Computer Science, BICA 2014. 5th Annual International Conference on Biologically Inspired Cognitive Architectures*, vol. 41, 2014, pp. 321–325.

[9] A. Newell, *Unified Theories of Cognition (The William James Lectures, 1987)*. Cambridge, MA: Harvard University Press, 1990.

[10] J. Dinet and M. Kitajima, "The Concept of Resonance: From Physics to Cognitive Psychology," in *COGNITIVE 2020 : The Twelfth International Conference on Advanced Cognitive Technologies and Applications*, 2020, pp. 62–67.

[11] M. Kitajima and M. Toyota, "Guidelines for designing artifacts for the dual-process," in *Procedia Computer Science, BICA 2015. 6th Annual International Conference on Biologically Inspired Cognitive Architectures*, vol. 71, 2015, pp. 62–67.

[12] J. Decety, T. Chaminade, J. Grèzes, and A. Meltzoff, "A pet exploration of the neural mechanisms involved in reciprocal imitation," *NeuroImage*, vol. 15, no. 1, pp. 265 – 272, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811901909383[retrieved:March,2021]

[13] M. Iacoboni and M. Dapretto, "The mirror neuron system and the consequences of its dysfunction," *Nature Reviews Neuroscience*, vol. 7, no. 12, pp. 942–951, 2006. [Online]. Available: https://doi.org/10.1038/nrn2024[retrieved:March,2021]

[14] M. Iacoboni, R. P. Woods, M. Brass, H. Bekkering, J. C. Mazziotta, and G. Rizzolatti, "Cortical mechanisms of human imitation," *Science*, vol. 286, no. 5449, pp. 2526–2528, 1999. [Online]. Available: https://doi.org/10.1126/science.286.5449.252[retrieved:March,2021]

[15] M. Kitajima and M. Toyota, "Four Processing Modes of *in situ* Human Behavior," in *Biologically Inspired Cognitive Architectures 2011 - Proceedings of the Second Annual Meeting of the BICA Society*, A. V. Samsonovich and K. R. Jóhannsdóttir, Eds. Amsterdam, The Netherlands: IOS Press, 2011, pp. 194–199.

[16] M. Toyota, M. Kitajima, and H. Shimada, "Structured Meme Theory: How is informational inheritance maintained?" in *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, B. C. Love, K. McRae, and V. M. Sloutsky, Eds. Austin, TX: Cognitive Science Society, 2008, p. 2288.

[17] M. Kitajima, J. Dinet, and M. Toyota, "Multimodal Interactions Viewed as Dual Process on Multi-Dimensional Memory Frames under Weak Synchronization," in *COGNITIVE 2019 : The Eleventh International Conference on Advanced Cognitive Technologies and Applications*, 2019, pp. 44–51.

[18] R. Dawkins, *The Selfish Gene*. New York City: Oxford University Press, 1976.

[19] M. Kitajima and M. Toyota, "Topological Considerations of Memory Structure," in *Procedia Computer Science, BICA 2014. 5th Annual International Conference on Biologically Inspired Cognitive Architectures*, vol. 41, 2014, pp. 45–50.

[20] L. W. Swanson, *Brain Architecture*. Oxford University Press, 2011.

[21] E. Jantsch, *The Self-Organizing Universe: Scientific and Human Implications of the Emerging Paradigm of Evolution (Systems Science and World Order Library. Innovations in Systems Science)*. Pergamon, 3 1980.

[22] H. Maturana and F. Varela, *Autopoiesis and Cognition (Boston Studies in the Philosophy and History of Science)*, softcover reprint of the original 1st ed. 1980 ed. D. Reidel Publishing Company, 8 1991.

[23] H. J. Stewart, R. D. McIntosh, and J. H. Williams, "A specific deficit of imitation in autism spectrum disorder," *Autism Res*, vol. 6, no. 6, pp. 522–530, Dec 2013.

[24] J. Nadel, "Perception-action coupling and imitation in autism spectrum disorder," *Dev Med Child Neurol*, vol. 57 Suppl 2, pp. 55–58, Apr 2015.

[25] M. Paulus, S. Hunnius, M. Vissers, and H. Bekkering, "Imitation in infancy: Rational or motor resonance?" *Child Development*, vol. 82, no. 4, pp. 1047–1057, 2011. [Online]. Available: http://www.jstor.org/stable/41289825[retrieved:March,2021]

[26] G. Vivanti, D. Trembath, and C. Dissanayake, "Mechanisms of imitation impairment in autism spectrum disorder," *J Abnorm Child Psychol*, vol. 42, no. 8, pp. 1395–1405, Nov 2014.

[27] J. H. Williams, A. Whiten, and T. Singh, "A systematic review of action imitation in autistic spectrum disorder," *J Autism Dev Disord*, vol. 34, no. 3, pp. 285–299, Jun 2004.

Asia Techno Farm Initiatives for growing Future Farmers of Asia

Nobutaka Ito
Faculty of Engineering
Khon Kaen University
Khon Kaen 40002 Thailand
e-mail: nobuito@kku.ac,th

***Abstract*—Humans are now facing the four global issues (Global Tetralemma) related to Population, Food, Energy and Environment. The key point is to find out the cause of these problems and how to solve them. Most of the problems are caused by human activities prioritizing economic promotion. It can be, however, noted that Asia has a high potentiality to play an important role in solving most of the problems in question. In this paper, the hopeful technologies applicable to Asian agrifuture are introduced mainly from the viewpoint of the sustainable development of future Asian society building, by establishing and maintaining a harmonic balance between ecology and economy leading to stable regional peace keeping.**

Keywords- global tetralemma; high-tech agricultural mechanization; Asia Food Project; Future Farmers of Asia growing program; Asian agrifuture;

I. INTRODUCTION

Earth is the only planet on which humans can live a normal life. The global tetralemma refers to the four global issues consisting of population, food, energy and environment. Figure 1 shows this ecological concept. When the population increases, the food production must be increased. To increase the food production, more energy is needed. Energy consumption increases CO_2 production, which jeopardizes the environment. This creates a situation in which we, humans, are facing global problems related to food, energy and environment [1].

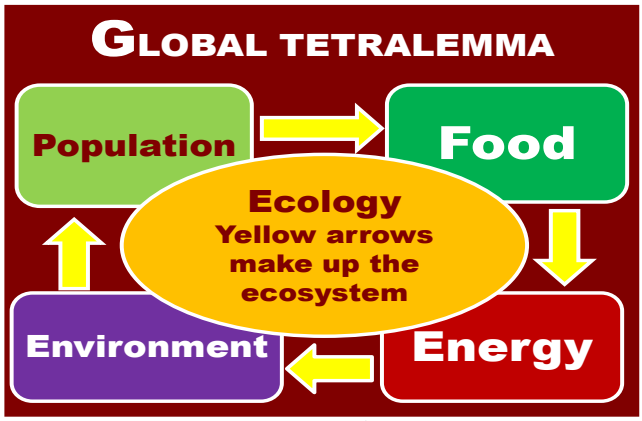


Figure 1. Schematic concept of global tetralemma

Figure 2 shows how economy breaks ecology and produces all the other problems. How to tackle and find the solutions for those problems is one of the main objectives of this paper.

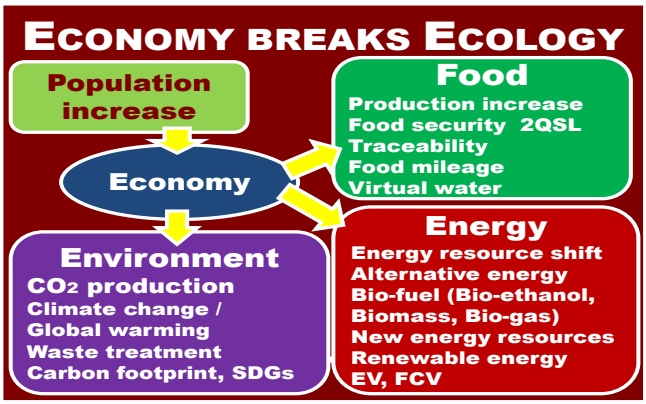


Figure 2. Economy breaks ecology and contributes to other problems

Currently, the world population is 7.8 billion in 2021 and increasing by 80 million per year. Considering this figure carefully, about 140 million people are born and 60 million are dying each year. The difference of 80 million is the population increase per year. As the population grows, more food production is needed. It is said that the population engaged with food production is around 20% of the world population. This simply means that one farmer produces food for 5 persons, including that farmer.

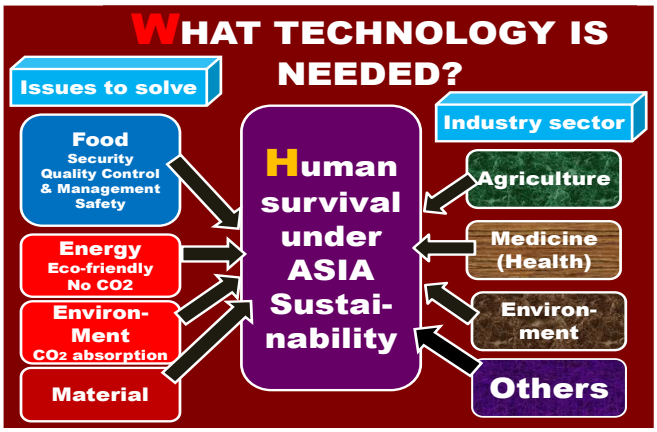


Figure 3. What technology is needed?

Figure 3 shows the relationship between the issues we are facing and the industry sector. We are already facing two emerging global issues related to energy and the environment that we need to solve immediately, however, the final and optimum solution has not been found yet, in spite of many efforts being done by various scientists, researchers and engineers. This is really a competition of the production cost and price of petroleum with the other energy resources in which most of them are eco-friendly or environment friendly ones. Recently, petroleum price has been drastically coming down almost 1/3 compared to ten years ago, due to the new energy resource discovery of shale oil / gas, in which the production cost is very much lower and more competitive. Methane hydrate is another possibility. Japan successfully developed the technology to take it out of the seabed in 2014, however, it is still not cost competitive. It will take a little bit more time to make it for the commercial base production. It may be guessed that the decrease in price of the petroleum may slow down the speed of technology development due to enough production to supply to the market. On the other hand, the mitigation control of CO_2 gas may be delayed, and the global warming issue cannot be improved if this situation is continued. Which one should be chosen, energy or environment, is one of the good examples to think about the real meaning of sustainable development or low carbon society building from the viewpoint of how the harmonic balance can be well maintained under this condition in promoting economy without jeopardizing the environment. The economy always makes us hesitate easily in making a choice. We, humans, already know that the fossil energy should not be chosen due to the cause of the global warming, however, the lower prices of oil entice us.

The author delivered a lecture entitled “Asian Agriculture Growth Strategy” in Prime Minister ABE's doctrine "ABEnomics” as one of the invited keynote speakers at the Agricultural Mechanization Session, of the JSAMFE (Japanese Society of Agricultural Machinery & Food Engineers) annual meeting held at Ryukyu University, Okinawa, Japan, May 17, 2014 [2]. Figure 4 is a schematic diagram showing ASEAN Economic Community should pick up the industrial sector for promoting the economy and how it should be done. It is already known that the ASEAN Economic Community was officially established in December 2015, however, it looks a little bit difficult to know what industrial sector should be set as the main framework of the body to meet and achieve the final goal of the community. In the author's understanding, the goals should be 1) To make ASIA the world's food pantry, enough to supply safe food on demand and 2) To stabilize the economy and maintain the regional peace based on agriculture. The following shows the concept outline of ASEAN Economic Community proposed by the author and how it should be promoted step by step towards the final goal achievement [3].

Most Asian countries are more or less still relying on the economy of the agricultural production. ASEAN member countries are also in similar situation, except some of them. It is already well known that, as mentioned above, the human population is still rapidly increasing at the rate of 80 million per year. It can obviously be guessed that sooner or later, food production must be increased to feed.

Figure 4. Asian Agriculture Growth Strategy

Otherwise, the food issue will become a serious problem later. Asia is fortunately producing a huge amount of bio-resources, especially food resources, however the quality is not well controlled and managed. As far as food production is concerned, the priority should be to secure the safety. It should be noted that there are two kinds of countries in Asia: one kind is resource-oriented countries, and the other kind is technology-oriented countries, called ASEAN, plus three like China, Japan and South Korea. To secure the quality, especially food safety, higher technology is needed and applied. The author's proposal is based on the collaboration and mutual competition among these two types of countries. One needs the resources and the other needs the technologies. Both need them mutually and interactively. The best way is to collaborate and compete to promote a stable economy in Asia. In addition, the following two things should be done on the process of final goal achievement. They are: 1) Technology transfer and 2) Human Resources Development. Both should be desirably promoted in parallel, however, due to some inconvenience of laws and regulations, it is hard to achieve in practice. Which one should be started first? That should be the human resources development, especially focusing on universities. The extension must be followed through the process of technology development and its validation; therefore, it takes time for final technology transfer to extension. Even for the human resources development, it takes time until the effective result could be obtained, therefore the academic mobility program should be started earlier and promoted in the regional area. It is required that highly educated human resources will be absolutely needed by the mature global society. Two stages of goal achievement are shown towards the final one. The first step is to make Asia one of the world food pantries in huge amount of food resources production.

Secondly, the most reliable Asian food brands should be created enough to guarantee both quality and safety. Following the process based on this proposed concept, the Asian economy can be developed and promoted stably, and regional peace keeping can also be maintained. Both resource-oriented and technology-oriented countries can get mutual prosperity to survive together contributing a lot in safe food production and supply to all over the world (Figure 4). The author already tried to propose the project under the title of Asia Food Project consisting of two main parts. They are Asia Techno Farm and FFA growing program, where FFA means Future Farmer of Asia [4][5][6]. These two programs are combined together and the whole framework is explained later in detail as whole involved in Asia Food Project. This paper therefore discusses the following two contents: 1) introduction and application of the various technologies seemed to be required for future Asian agriculture achievement and 2) Asia Food Project proposal consisting of two programs, named Asia Techno Farm and FFA (Future Farmer of Asia) growing program, which is one of the human resources development program.

II. SMART AGRICULTURE

The Ministry of Agriculture, Forestry and Fisheries of The Government of Japan defines smart agriculture as "New agriculture that enables hyper labor saving and high-quality products production by utilizing cutting-edge technologies such as robotic technology and ICT (Information & Communication Technology)". According to the Ministry, the materialization of smart agriculture can achieve hyper labor saving and large-scale farming production by automatic control of agricultural machinery, high-yield, high-quality products production that makes full use of sensing technology and large data, and heavy labor by using robot technology. It can be expected to have more merits such as CO_2 mitigation and labor saving, simplification of agricultural operation by combining know-how with data and assisting operating function and providing important and necessary information to consumers by providing final products information (traceability) [7].

A. Precision Agriculture

Precision Agriculture has variable rate control function for reducing loss and saving materials and energy. The concept of this farming system is similar to the TOYOTA car manufacturing system named “Kanban (or Kaizen)" system mainly consisting of the three conditions listed below.

1) Car manufacturing company provide the required information on which part, how many sets amount, and by when parts must be prepared should be informed to the related parts supply manufacturing company in advance based on the production schedule every day.
2) Parts supply company supplies the ordered parts just enough amounts required for one day manufacturing plan.
3) The parts supply company is strictly requested to bring them by the time when appointed without any delay

In actual farming, three conditions of “What”, “How many or How much” and “When” must be decided from time to time knowing the data provided by GIS is specifically matched with site by site [8][9].

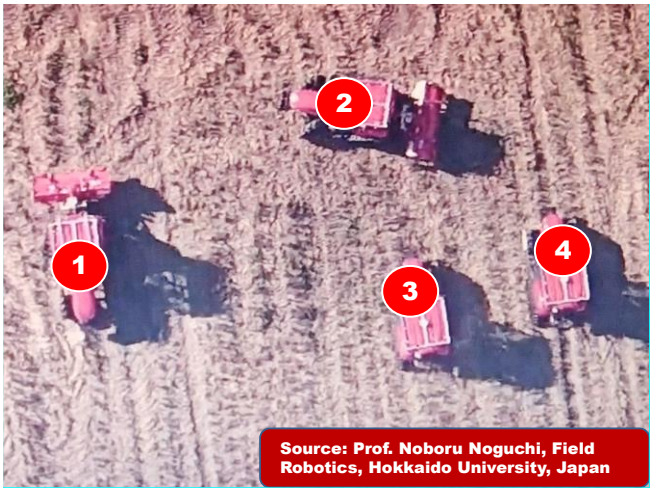


Figure 5. Farm tractor group guidance driving control

Figure 5 shows a group guidance control consisting of four vehicles (tractors in this case). Tractor No. 1 is leading the other three. At the corner or headland of the farm field, the leading vehicle (tractor) starts to turn left. The other three are waiting for the leading one until it to completes the turning and gets in the new path keeping suitable interval and space. In this case the four tractors were used for demonstration running test, however the total number of the vehicles to be controlled is not the problem. Even the bigger group consisting of more number of the vehicles can be controlled in the same way.

Figure 6. Truck group guidance driving control

Figure 6 shows the group guidance driving test consisting of three trucks in this case demonstrated under

the leadership of the Ministry of Land, Infrastructure, Transport and Tourism [10]. The merits are as follows: 1) The benefits are: 1) Aging drivers and a shortage of long-distance truck drivers due to heavy labor within a set limited time, 2) Air resistance can be reduced and fuel consumption can be reduced since the distance between vehicles can be kept short during platooning driving.

Communication and information technology has become more familiar and easier due to the dramatic spread of its technology due to the relaxation of GPS for civilian use, which was initially developed for military purposes. Smart phone is a typical example, and we are now in an era where we can hardly feel the distance of the national sense of the world. The era of IoT (Internet of Things) that connects various information devices to the Internet is really coming.

Until now, unmanned driving technology for automobiles has been the goal of car companies, and the development deadline was initially set to the time of the Olympic Games 2020, but it has been significantly delayed due to the COVID-19 disaster. In the field of agricultural machinery and mechanization, unmanned operation of mobile vehicles such as tractors and combines has already been achieved at the laboratory level. Of course, agricultural machinery has an extremely slow running speed compared to automobiles, traffic is light on agricultural land, and the conditions are relatively good and better, which made it possible to develop technology ahead of automobiles, but in addition to being small scale in agriculture. Since it is a declining industry, it has been considered difficult to be accepted and spread out even if it is commercialized in terms of cost.

B. Robotics

There are two kinds of robots. One is something like an industrial robot set and used to complete the work for the post-harvest products, such as selection, weighing, grading, sorting, packaging etc. in a specially prepared building or facility. The other is a mobile vehicle uch as tractor, combine and transplanter doing the original operation while moving, such as tillage, fertilizer application etc. The location of the robotic machine is autonomously guided by the GPS signal provided from the satellite. The optimal operation can be done under the variable rate control based on the final decision derived from the data collected and provided by GIS continuously from time to time while moving. A laser scanner is mounted in front of the vehicle (tractor, in this case) to detect the obstacle and it functions to stop the machine immediately. The other direct contact type sensor is also mounted for a double check in safety. Three way motions of the vehicle, namely pitching, yawing, and rolling, reduces accuracy. It is also possible to control the automatic guidance operation of a group of vehicles consisting of multiple vehicles while maintaining a master-slave relationship [11]. Figure 7 shows a comparison between agricultural and industrial robots function by some major items. It can be seen from this that there is a fundamental difference in the handling of artificially designed and manipulated agricultural products from the ones grown in the natural environment between two kinds of robots.

Difference between Industrial and Agricultural Robots

No.	Item	Agricultural robot	Industrial robot
1	Robot motion	Move to work Search, Find, Identify, Off road	Stay and wait for the work
2	Objective work	Non standardized Size, Color, Shape, Maturity Hardness, Location	Standardized Designated set position
3	Operation	Autonomous	Program based
4	Function	Learning	Teaching
5	Structure	More complicated	Comparatively simple

Figure 7. Difference between agricultural & industrial robots

The biggest difference between agricultural robots and industrial robots is that industrial robots handle the artificially designed standardized (normalized) products, while agricultural robots handle non-standard products. Non-standard products are different in color, shape, size, weight, ripeness, fruit hardness, etc. for each solid. The harvesting robot makes a decision as to whether or not to harvest by collating this information with preset standard code. Agricultural robot needs to know where the fruit is hanging in what part of position of the fruit tree. Robot has to check each fruit to see if it should be harvested and only fruits that meet the conditions will be harvested. More number of sensor and hardware should be mounted on the agricultural robot to get more information to make the final decision whether the target fruit should be harvested or not. Fruit harvesting robot should take action first in the orchard even for the recognition and identification of which is fruit or not by use of the light reflectance expressed in wave length information in case the fruit is hanging on the tree with leaves. Harvesting robot is mostly used in green factory to harvest vegetables for fresh market use due to comparatively uniform and stable growing under suitable conditions with environment controlled.

C. Green Factory

This type of farming is different from the conventional one using soil, mainly focusing on the mass production of fresh market crops such as lettuce, mini-tomatoes etc., which is similar to the industrial crop cultivation under completely controlled conditions of the environment. This is basically managed on hydroponic system, therefore water is normally used for fertilizing and circulated for saving. Disease infection is tightly controlled. Workers are strictly forced to wear special work clothes, masks and caps, in addition to special boots and gloves. The environment is similar to the clean room of semi-conductor industry plants.

They are also strictly forced to take air showers when getting in and out of the factory. Since vast agricultural land and soil are not required, the completely different kinds of industries which are originally unrelated to agriculture are increasing to join looking for business opportunities because this type of farming uses normally hydroponic system instead of soil. Osaka Prefectural University is leading in this area [12].

Figure 8. Various types of Green Factory

Figure 9 shows the advantages of green factories in terms of materials and costs including future business prospects.

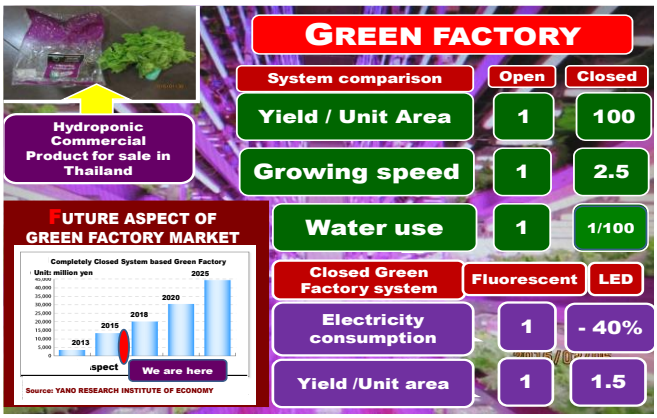


Figure 9. Green Factory, present and future.

D. Drone

Recently, the number of possibilities of drone application to the agricultural sector has grown up rapidly. One of the most important merits of applying drone is to get a wide, bird eye view photo, which enables to find how the crops are growing. Farmers can find what and how they can / should do from the image sent to their device, such as a smart phone. They can make decision regarding how much fertilizer should be applied from the green color image (NDVI, Normalized Difference Vegetation Index).

$$NDVI = (NIR - RED) / (NIR + RED)$$

where NIR–reflection in the near-infrared spectrum
RED–reflection in the red range of the spectrum

According to this formula, the density of vegetation (NDVI) at a certain point of the image is equal to the difference in the intensities of reflected light in the red and infrared range divided by the sum of these intensities. This index defines values from -1.0 to 1.0, basically representing greens, where negative values are mainly formed from clouds, water and snow, and values close to zero are primarily formed from rocks and bare soil. Very small values (0.1 or less) of the NDVI function correspond to empty areas of rocks, sand or snow. Moderate values (from 0.2 to 0.3) represent shrubs and meadows, while large values (from 0.6 to 0.8) indicate temperate and tropical forests. Crop Monitoring successfully utilizes this scale to show farmers which parts of their fields have dense, moderate, or sparse vegetation at any given moment. Simply put, NDVI is a measure of the state of plant health based on how the plant reflects light at certain frequencies (some waves are absorbed, and others are reflected). Chlorophyll (a health indicator) strongly absorbs visible light, and the cellular structure of the leaves strongly reflect near-infrared light. When the plant becomes dehydrated, sick, afflicted with disease, etc., the spongy layer deteriorates, and the plant absorbs more of the near-infrared light, rather than reflecting it. Thus, observing NIR changes compared to red light provides an accurate indication of the presence of chlorophyll, which correlates with plant health. Drone can be used for many other ways, such as seed broadcasting, fertilizer and chemicals application. In these cases, the payload is the problem, namely, how much it can carry and fly. More applications are under consideration; however, the challenges are security, privacy, various regulation and the standards. Hackers can break into seemingly safe remote-controlled engines and networks that control brakes and steering.

III. TECHNOLOGY APPLICATION IN AGRICULTURE

A. Nano Technology

Carbon Nano Fiber is famous and well known for its light weight and strength. It is already used for aviation and car industry. Recently, the bio-based Cellulose Nano Fiber is attracting a lot of interest. It is a newly developed material by Professor Hiroyuki Yano, Humanosphere Research Institute, Kyoto University, Japan from bio-resources equipped with unique physical properties, namely, 5 times stronger and 1/5 lighter than metal, in addition to higher heat resistance. The Cellulose Nano Fiber may take the application area and replace CNF (Carbon Nano Fiber) in the future. In addition, it costs 1/6 of the cost of carbon nano fiber. Cellulose Nano Fiber can be produced not only from trees, but also from various popularly known cellulose materials such as wood, rice straw, cassava and potato.

Various Nano Bubble water provides other hopeful possibilities, such as

- Oxygen Nano Bubbles
- Ozone Nano Bubbles

・Nitrogen Nano Bubbles

Hopefully Applicable industrial sector is shown below.

・Food safety – Vegetable sterilization

・Aquaculture (Fishery) – Oyster sterilization

・Dentistry – Periodontology / Periodontics

・Medical science - Cancer cell control

Oxygen nano bubble water has higher effect of promoting plant growth and shortening the total growth period. Ozone nano bubble water functions effectively for sterilization for various bacteria and fungi. In case of washing out the chemicals attached to agricultural products, 80% are removed by using ozone nano bubble water, whereas only 20% can be removed at one time with ordinary water. Tooth paste water using ozone nano bubble water has been commercialized. Periodontal bacteria in the mouth can be sterilized just by gargling without brushing your teeth.

B. Plasma Technology

Plasma technology can be used to treat waste and change to energy because under high temperature treatment hydrogen can be produced. On the other hand, if the plasma treatment was done under low temperature, waste oil can be changed into fuel. According to the news currently televised, it is said that around 4,000 workers are working everyday at Fukusima nuclear power plant, however it will take 40 years more to remove the debris left in the reactor. The use of plasma is promising for treating highly radioactive debris.

C. Pattern / Face Recognition

The combined technology of image processing, pattern recognition and Artificial Intelligence (AI) is getting popularly applied to recognize and identify the individual person quickly. This technology can be used even for the individual livestock management. Two kinds of memory can be found and considered: one is a tag attached to a part of the body like ear, and the other is the chip type to be embedded in the body. The pedometer, the route traveled, the distance, etc. are automatically recorded and sent to the data center for recording. These data can be used for observing the health status of individual livestock and managing the amount of food to feed.

The net pattern of melons is unique and original to each individual and resembles a human fingerprint. By recording and memorizing this net pattern as an image, the historical background of the melon can be known such as the place and when it was harvested and how it came from the production site. This is one of the areas called Agribiometrics or Bioinformatics.

IV. ROJECTS AND BUSINESS

In this section, some of the ongoing projects and businesses model in commercial base are listed.

A. Blue fin tuna, Kinki University

Blue fin tuna is one of the most popular big thick fishes served at higher class restaurants in Japan. However, to meet the customers demand, fishermen must live away from home for months in remote pelagic fisheries. Sometimes the weather is unseasonable, and they encounter storms and typhoons in case. Some of them will also encounter a fatal accident. If they do not have to live away from home for fishing, it will make them more relaxed and even their family will feel more at ease. Fisheries should be changed from going away and fish for months to keeping put and growing fish. The Kinki University, succeeded in the cultivation of blue fin tuna to grow from the stage of egg up to the final stage for shipment. Currently, blue fin tuna cultivated in this way is delivered to large cities and rural areas and is also served in the cafeteria even on the university campus [13].

B. Osaka Prefectural University

Osaka Prefectural University is one of the first universities in Japan to succeed in researching and commercializing a Green (plant) factory. Just like Kinki University mentioned above, it can be said that this university has demonstrated the industrialization of agricultural products production. The cultivation shelves lined up in an environmentally controlled building are fully covered with LEDs (Light Emission Diodes) of various colors, and the workers working inside seem to be working in the clean room of a semiconductor manufacturing plant, and they are nervous about bringing in pathogens from the outside. Harvested lettuce and other fresh vegetables are delivered not only to university cafeterias, but also to large cities and regions on order. One of the important factors to keep in mind is that producers have a clear and reliable relationship with consumers.

C. EUGLENA Project

This is a joint venture project already launched few years ago. The author does not know how much they have been successful up to now, however, this business model is one of the few successfully launched examples the author knows about. The main product is an algae. Euglena has many characteristic features, but it is noteworthy that it absorbs a large amount of carbon dioxide. Electric power plants discharge a large amount of carbon dioxide, which can be absorbed by Euglena to promote its own growth. The produced Euglena can be sold as a raw material or resource for food, feed and fertilizer, while also contributing to jet plane fuel as bio-fuel. The problem with business operations is that they need technology to produce a large amount of Euglena in a short period of time, and the company founders have also proven this technology. Since it is a bio-based fuel, it contributes a lot to a low carbon and de-carbonized society building from the viewpoint of carbon neutralal concept [14].

Euglena, which efficiently produces polysaccharides, is promising for health foods, cosmetics, bioplastics, biofuel raw materials, and carbon dioxide treatment, but it should be more productive for industrial use. Selection of Euglena strains is required.

Euglena production sufficient to meet the demand basically depends on the expansion of the culture space, but

in order to improve productivity, it is necessary to identify the shape and size of the produced Euglena.

For the purpose of increasing the processing capacity, a method of flowing Euglena in a flow path at high speed and taking an image with a high-speed camera, and an optical observation method using light scattering such as fluorescence imaging, Raman scattering and Mie scattering have been developed [15]. These methods required complex optical systems, and the equipment itself was large and costly. If electrical shape identification becomes possible, complicated optical systems will be eliminated and hyper-compact shape identification device consisting of electrodes and a small circuit board can be realized.

Conventionally, it has been known that there is a correlation between the average diameter of cells between electrodes and across nanopores and impedance characteristics, but it is not suitable for shape measurement of elongated cells such as Euglena. An inclined electrode is added between the normal electrodes, and the position information in the flow path of the target that has entered the electric field is the amplitude of the signal, the magnitude is the signal width, and the shape is the deviation between the rising and falling edges of the signal wave, that is, the eccentricity. As the result of measurement as (Tilt index), information on the cell shape can be obtained only by measuring the electrical impedance. This method of measuring electrical impedance using a device with a tilted electrode made it possible to accurately measure the size and shape of Euglena at a high speed of 1,000 cells per second. By widely applying this method to algae other than Euglena, it is possible to develop useful substances from algae, whose use is expected to expand in the future, from research to commercial-based industrialization.

Figure 10 shows the schematic flow of Euglena cultivation and its various utilizations. The huge amount of carbon dioxide is produced at the electric generation power company, therefore this massive amount of CO_2 can be received by EUGLENA Co. Ltd. and used for growing / cultivating Euglena, because Euglena can be grown up by the absorption of CO_2. Euglena can be used for producing many products such as bio-fuel available for cars and jet plane. In case of application to cars, ISUZU, a famous diesel engine mounted car manufacturer in Japan already has the partnership with EUGLENA Co. Ltd. to produce bio-fuel for diesel engine in which the registered trade name is DEUSEL. This is now being used for public transportation bus and other type of cars. As for the case of bio jet fuel, HONDA business jet had a successful demonstration test flight [16].

EUGLENA Co. Ltd. and HITACHI PLANT Co. Ltd. are jointly working to supply the bio jet-fuel and related plant depending on which product is targeting, bio jet fuel or residue supply for animal feed, organic fertilizer production etc. JX Nisseki Petroleum Co. Ltd plays a role of refining bio-fuel for final product.

Figure 11 shows the difference of the CO_2 absorption capacity between Euglena and various kinds of tree. It can be obviously found from this comparison data that Euglena absorbs 30 times more than cedar and cypress trees. In addition it shows that Euglena can be usefully available for covering the four global issues. Euglena can be categorized as one of the important key resources from this point of view. The author understands that key resource may be defined as the useful important resource to cover the most of four global issues (tetralemma). Most of the agricultural crops are therefore categorized as key resources, however the conditions should be carefully considered to avoid or not to bring in unnecessary problems that conflict with each other among the issues.

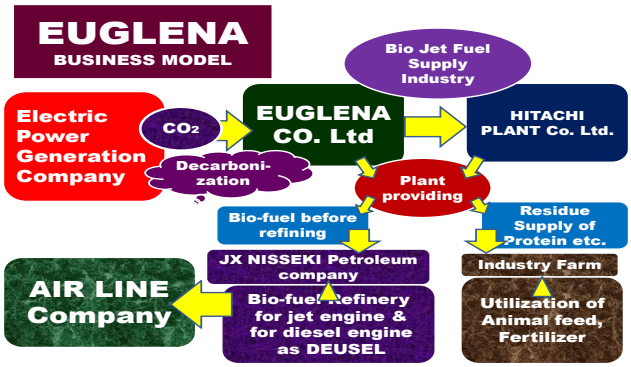


Figure 10. EUGLENA Co. Ltd. business model going on commercial base

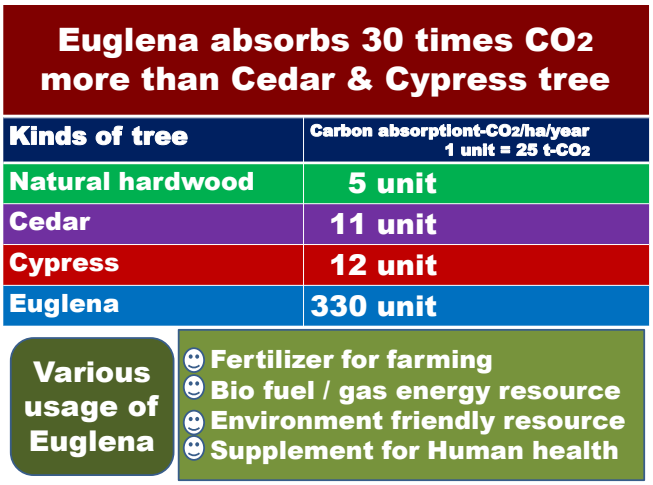


Figure 11. Euglena CO_2 absorption comparison to some of the other trees

D. Good Harvest Plan

This is also a business model proposed by TOYOTA. It is a contract-type business model for small-scale farmers or farmers who have no successor, but own farmland.

The farmer will contact the contracted center and request the dispatch of machine together with driver to carry out the necessary operations timely. The center will respond to the request by dispatching an appropriate driver according to the type of farming operation. The main operation is tillage,

transplanting and harvesting, but the farmer is the owner of the farmland, however, does not physically work directly on farm. The number of such contractor-type farmers is increasing in Japan due to no successors. Agriculture in the high-tech and information era has the potential to significantly increase the entry of industries from completely different original sectors that are not closely related to an original agriculture [17].

This plan is now going under trial for future extension and now being accepted by hundreds of aged farmers who have no next generation successors in some regional area.

In this system, they have one private sector company expressed above as the center or organization to take care of those aged farmers' farmlands not in use and left abandoned for them which is a similar organization to the Farmers Cooperatives. Famers those who are mentioned above have their farmlands, but they can't work physically any more due to aging. However they have no need to sell them. Farmland is one of the precious properties for them. In the future those farmlands may be happily purchased with higher price by the local government in case of new city/rural area development plan is officially announced. This is the reason why most of the aged farmers insist to keep and own the farmland in spite of no successors.

On the other hand, the completely new or originally different industry from agriculture and the related sector would like to join and expand the new market in agricultural sector business. Not only TOYOTA, but also the other industries which are originally different from agricultural sector are actively trying to find and actively join the IT, Information Technology and IoT, Internet of Things based new agriculture business. This is one of the characteristic, but common trends and examples of industries how they can do it easily without owning any farmlands. It can be obviously found from this background that:

1) Any different kind of industry from agriculture can join the agricultural sector business without owning any farmlands, but IT and IoT instead.
2) They can provide the technology and farm management system for whole season farming and production based on the contract.
3) Aging issue of farmers and next era successor growing and development may be solved, however the agriculture promotion must be also achieved along the right way targeting the highly motivated future farmer resource development equipped with higher education

The following shows the framework structure and function of Good harvest plan proposed by TOYOTA.

Farmers can contact and request the company / organization to send the professional operator with machine to complete the requested operation. Of course the farmers pay the equal amount of fee to the company. They have a contract in the beginning of the season. This is one of the contract based farming. Crop can be grown up and managed from the beginning up to the end of the cultivation. Even while the crop is growing, the farmers can monitor and know how their crops are growing through the photo taken by drone and think what they should do. This is one of the examples to see how IT, Information Technology and IoT, Internet of Things make any other various industries to join and find new business opportunity even though they are completely different from the original farmers.

E. Animal Factory

This is an animal version factory of a plant factory. The target livestock are dairy, beef cattle, pig farming, and poultry farming, but unlike growing in a limited space, this is a project to give time to walk around freely in a wide space and to carry out breeding management of livestock with high quality meat. Individual livestock animal management such as beef cattle, daily cow etc. using small device and equipment incorporating information and communication technology is extremely needed. The necessary information is sent as data from sensors attached to dairy cows and beef cattle grazing in the pasture to the central center and used for individual health management.

One micro chip memory or the ID tag is embedded or clipped to each individual livestock animal. It has not only the memory function, but also the one as the sensor for gathering and sending the data as the signal to the central data acquisition station too. While the herd of animals are released in the pasture outside the stable, they walk around freely here and there looking for eating grasses. The embedded micro chip can be used so as to play a role of ID (Identification Card) and sensors, therefore, all the biographical data such as the place of birth, how old it is now, how it was grown up by whom (livestock animal producer) after it was born are installed in there. In addition, the various dynamic health data can be gathered and sent as the telemetry signal to the central data station. Those data are heart pulse, momentum calculated by use of the data obtained from the small sensor similar to the pedometer function, body temperature to find and know the abnormal health condition such as fever, total walking distance and its motion trajectory which can be calculated and processed from GPS signal. On the way back to the stables, the weight of each individual animal can be measured by guiding and passing livestock one by one to the place where the load sensor to measure the weight is buried. The exact three dimensional shape data (weight, length, width) of individual livestock animal can be automatically measured too.

All the gathered data are stored in both of the memories installed in the embedded memory chip and the device set at the central data acquisition station. Those data can be also used in traceability system to ensure the safety and security of food.

F. Beef Traceability

It was already mentioned above how important the mutual liability between producer and consumer. Food

security has four important keywords such as 2QSL consisting of Quantity, Quality, Safety and Liability. One of the most important problems, but difficult one to negotiate, is the mutual liability issue. No matter how famous and well-known companies are, if they do not manage well, they will disperse false information. They can cheat the place of the production, fake the contents of the product, rewrite the expiration date, and do embarrassing acts without hesitation. The other three conditions except mutual liability are relatively easy to clear and satisfy. This is because there is no problem as far as the standard code level is cleared.

As already mentioned above, all the data related to the individual animals are stored in the memory and uploaded on the website / homepage normally prepared by the producer side such as farmers and shops, therefore any consumers, as well as customers seeking beef, can obtain all the information about where the beef in front of them was produced and how it was brought to this place. Standard code varies slightly sometimes depending on the locality and the shop. Barcode and QR code (registered trade number: 4075066) which is the abbreviated expression of Quick Response developed by DENSO Corporation Ltd, Japan is normally mounted on the package of final products. The consumers can obtain the information easily and make the final judge and decision whether to buy or not. The shops and producers should know the standard, because they must be responsible for the product safety especially for food products because it leads to the fatal problem. Even in car manufacturing industry, the product liability is the first priority to consider as well. One of the most important and difficult issues to control is the mutual liability between producers (shops and farmers) and consumers. Some of the shops put the fake label for the purpose of cheating the consumers intentionally caused by the poor management. Once this happened even accidentally, it takes longer time to get the trust and same level of liability as they had before. The issue of product liability should be seriously considered with first priority always. This is because those accidents lead to fatal injuries and even fatal accidents.

The meaning of food security consists of four factors expressed simply with 2QSL. This abbreviated representation shows the capital letter of the four factors of 2Q (Quantity and Quality), Safety and Liability.

Of these four factors, the most difficult to control is 4. There is no problem if the other three meet the set standard values, but it is difficult to express mutual trust between humans numerically, and the level varies from person to person even in case of satisfactory level mutually. Only one of the ways at this moment to secure the safety and to deepen the mutual liability is to accumulate the mutual trust honestly. The name value is also created in this way. Therefore, once the act of losing trust was done, it is difficult to reestablish mutual trust. With the rapid increase of human population in the world, food issue must be one of the most serious issues.

Even now, about 1 billion people in the world are facing food conditions that are insufficient both quantitatively and nutritionally. Traceability is one of the issues that should be considered immediately from the viewpoint of food distribution including the food crisis. It is necessary to ensure sufficient food supply to be reliable more than sufficient production, supply and safety. The author would like to emphasize here the importance of food safety and security in order to make Asia, a resource oriented country, the world's food pantry. The author strongly hopes that by introducing here what kind of situation the effective technology for that is currently in, it will lead to the further importance and solution of the problem.

The reason why the author would like to emphasize the importance of food issue can be summarized again as follows.

1) Food production should be increased due to the rapid increase of human population
2) Asia has a lot of potentiality to make it possible however there are some problems in safety and quality control of the final food products
3) Safe and reliable food must be supplied always
4) Traceability can provide more transparent information related to food safety and liability quickly and exactly.

G. International Collaborations

Asia can be qualified as a world food pantry in production and supply, however, farmers are still working in poor conditions due to various problems such as family labor and low income mainly caused by small scale farming. The ASEAN community based international collaboration is really needed in technology transfer and human resources development. Technology oriented countries should join and actively invest in Asia for further economic promotion and regional peace keeping. The author is making a proposal named FFA (Future Farmers of Asia).

V. PROJECT PROPOSAL

A. Asia Food Project

Asia is one of the huge agricultural production regions and most of the member countries belonging to the ASEAN Economic Community (AEC) are more or less dependent on the individual national economy obtained from agriculture. The agricultural production is extremely high in Asia as a whole, producing an amount comparable to that of the world food pantry, however each farmer has the following characteristics. Most of these characteristics are problematic in reducing the attractiveness of the agricultural industry, this is, their recognition as unattractive industry, and constantly relying on subsidies and support from the central government, which encourages young generation people to leave agriculture and succeeds. The factors that hinder the promotion of agriculture itself are listed as follows.

1) Small-scale farming, 2) Family labor, 3) Low income, 4) Hard work, 5) Aging, 6) Working poor

Figure 12 shows the Asia Food Project scheme and concept. As already explained in Figure 4, this project consists of the collaboration between two types of country groups: one is ASEAN community, which is a resource-oriented country group, will provide resources, and the other group is ASEAN plus 3, which is a technology-oriented country will provide technology. Then both groups will work to create and develop good cooperative and competitive relationships with each other. Technology-oriented countries send experts to this project. Experts should be carefully selected and registered referring to the program contents.

Firstly, the project has to push up Asia to the world's food pantry that enables massive production and supply at a level enough to respond to the world's food supply and demand in a timely manner. Once that goal is achieved, the next step will be a development and creation of the reliable Asian food brand product that can guarantee safety and liability between producer and consumer. As a result, both of the stable agriculture and economy promotion can be achieved, and the sustainable stability of regional peace keeping in Asia can be maintained. In the way of process, technology transfer and human resource development will be carried out in parallel. The main purpose of establishing a community is to deter war between community member countries. Cooperation and competition are necessary for mutual encouragement and prosperity. If the spirit of collaboration and competition among member countries disappears, the interdependence will be increased, which will hinder the operation of the community. This spirit is also necessary to prevent excessive dependence on other countries. If the two promotions of agriculture and economy can be stably managed and maintained, the development of the community will be well promoted, and the mutual prosperity of both technology-oriented and resource-oriented countries will be achieved, and the win-win relationship will be built up as a result.

Figure 12. Asia Food Project scheme

It can be therefore summarized and concluded from the discussion mentioned above that the ultimate goal is the sustainable development under coexistence and co-prosperity of Asian countries. In the process of reaching the final goal, the project has to make Asia the world food pantry firstly and to develop Asian brand food products secondly that can pave safety and security for economic promotion and stable maintenance of regional peace keeping. In the process of reaching this level, technology transfer and human resources development will be carried out in parallel in Asia Techno Farm Initiatives program.

Figure 13 shows the framework of Asia Techno Farm program consisting of three bodies: 1) University, 2) Government and 3) Industry. Those three have their own roles to play jointly to make the program fruitful and successful. They are explained as shown below.

① University

One of the universities offered the project (program) should be approved and appointed as the Asia Food Project organizer to be responsible even for taking care of Asia Techno Farm program management and operation. University farm or experiment station should be provided to be used as the demonstration site, research experiment site and practical cultivation site for growing FFA (Future Farmers of Asia program). FFA program participants are requested to experience the practical rice cultivation for one season from planting to harvesting taking about four months

② Government

The relevant government ministries and agencies will investigate the project proposal applied and submitted by the universities carefully and make a decision to issue the budget allocation, and notify the approved university of the examination result of project application. Some additional comment and advices may be added if they have.

③ Industry

In promoting the project and program, the participation and cooperation by the related industries should be requested, especially for machine trouble shooting and maintenance with the spare parts supply. The related industries are strongly requested to join and provide equipments, and one staff for covering the above mentioned correspondence to various things. Industries are also requested to provide the secondhand equipments as the teaching materials in case. It is desirable if one staff is resident and is always on standby to prepare for problems. Special benefits should be considered instead for the cooperating companies.

The main purpose of the Asia Techno Farm Program is to train farmers who can lead the next-era agriculture using advanced technology. Agricultural income is limited on a small scale, and it is unlikely that it will be able to compete with big scale agriculture in the world market.

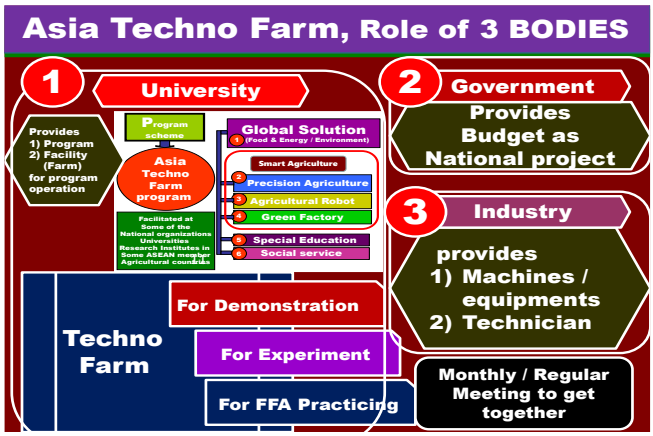


Figure 13. Asia Techno Farm Initiative program

This program focuses on precision agriculture, robotics, and green (plant) factories. It is important to train next-generation farmers in consideration of large-scale farming. Participants are not necessarily limited to Thailand, but considering the acceptance of applicants from ASEAN countries. The mutual collaboration among the member countries of the ASEAN Economic Community can be promoted. English is the common official language aiming to configure a human resources network not only in Asia, but also worldwide toward the future from the viewpoints of internationalization and globalization.

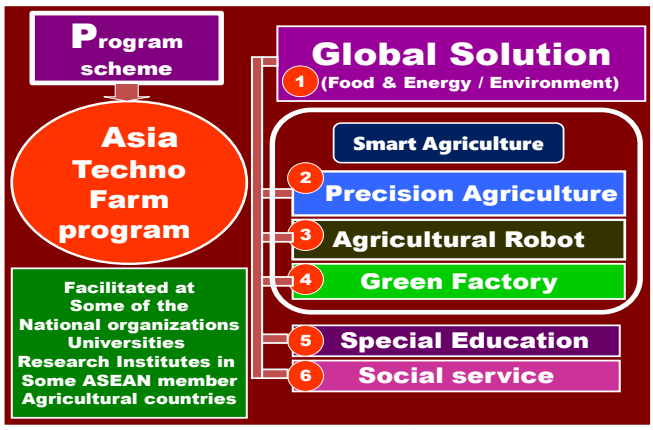


Figure 14. Further detailed program contents of Asia Techno Farm project

The initial goal can be therefore achieved by promoting agricultural mechanization equipped with highly advanced technology considering the farming scale shift from small scale to larger scale due to the upper limit of production from small-scale farming. This is however not to provide the part-time job for small scale farmers to increase the low income, but to achieve the goals shown below fostering large-scale high-tech based farming.

1) Make Asia the world food pantry
2) Create Asian food brand with originality
3) Promote economy and agriculture through the above two processes in addition to maintain the stability of regional peace.

Asia Techno Farm program has several fields and functions as shown in Figure 14. Smart agriculture consisting of three main technologies (precision agriculture, robotics and green factory) are already mentioned, therefore the other three divisions are explained here.

①Global solution

Lectures on global issues are provided and given not only related to agriculture, but also the others such as food, energy, environment caused by poverty and hunger, food residue, agricultural waste, disaster, waste management & treatment, haze issue, ocean pollution due to abandoned plastics, flooding & land slide, release of heavy metals, mass spraying of chemicals (pesticides and herbicides) etc.

⑤Special education

This division is prepared for teaching the additional technique something like the ways how to connect smart phone to the various equipment and sense, measure, monitor, process the signal and gathered data. The skills of regular and final report assembly and presentation material making are taught. Formerly as far as concerned with the training program of machine design for example, the black smith skill was taught first and ends at the level of learning how to use CAD, however, the era changed completely and it is still changing due to new technology innovation day by day. More compact, faster, big data processing capacity handy device development can make it possible. These are now being applied to new era agriculture. The era has come in which even amateurs who has no experience can join and start agricultural business, but some training program should be prepared to educate even the young ones to make them familiar with actual application of knowledge and skill usage for farming. In the information era, advanced knowledge of technology and skills are acquired and it is certain that even amateurs who have nothing to do with agriculture can join the agricultural business. It is however necessary to invest the necessary amount of the capital. In addition, it is really difficult and impossible to provide the technology and financial support to all farmers individually, therefore here introduces one idea on how to change and improve the current situation of Asian agriculture and proposes concrete countermeasures.

1) Most of the farmers can't work physically on farm due to aging. In case of Japan the average age of a farmer is about 65 years old. This background is also an example of the participation of the IT industry such as Good Harvest Plan in the agricultural sector. By the way, who will get more benefit in this case, farmers or industry? For further promotion of agriculture, next era farmers should be grown up, who must be highly interested and motivated to be a farmer equipped with pride and dignity to the profession as the farmer.

2) High technology is developed and innovated day by day, however they don't know how to apply and use them for farming. Even in case of high tech equipment such as smart phone connectable to many of things through internet.. It looks extremely hard for aged farmers to learn and use them on farming, who can't work physically any more.
3) In addition, they require the investment, however, they don't have any extra money to buy those high tech mounted machine and equipment due to low income
4) The new project should be therefore proposed focusing on youth motivation and considering the above mentioned troublesome situation

The Future Farmers of Asia Initiative project proposed here is mainly aiming at the trainer training / development, who can be qualified as the leader enough to transfer those high technologies and apply them to actual farming. The support to all individual farmers should be separately considered because of the reason as shown below.

1) It takes longer time and difficult to consider and take care of all aged farmers.
2) The issue of farmland ownership involves many complicated problems legally under different conditions of individual farmers
3) The highly motivated young generation should be targeted for growing FFA. They are already familiar with those high-tech equipments and devices on a daily basis
4) The target under consideration is neither an individual farmer nor an individual country. That is to train and grow technical leaders first in the region of ASEAN Economic Community.
5) The regional agriculture and economy should be promoted for regional peace keeping by making Asia the world's food pantry as already mentioned before.
The poverty is one of the obstacle factors for regional peace keeping.
6) Agricultural promotion can't be achieved by supporting individual farmers only. By promoting the economy centered on agriculture in the regional community, it is possible to promote cooperation and competition among the participating countries and maintain and stabilize agriculture, economy and regional peace.

Aged farmer has difficulty to follow the high-technology based agriculture, however, the young generation can do it easily with pleasure. On the other hand, in this division, additional curriculum is offered for the farmers especially small scale farmers who would like to actively to change the job from agriculture to the other industry. Active recommendation and guidance to those farmers is also one of the ways of income increase for them. In addition the full-time farmer will get more opportunities to have their farmlands for large scale farming. Reducing the population of small scale farmers makes opportunities for full-time farmers to buy those farmlands for farming scale expansion. Guidance and future possibility on how they should do are advised including the problems to encounter and overcome, the way for negotiation etc. Farmers are however not forced to change the job and the decision making depends on their own will. Only in case they have strong intention to scale-up or change the job from agriculture to some other sector, they are advised and guided.

The statement mentioned above may be misunderstood too. Further description is therefore added here as follows. The situation is different depending on the countries even in Asia. Especially in Thailand the farmers (farmland owner) are not allowed to leave the farmland left abandoned without cultivating any crops. If they do so, they are imposed to pay tax and fine, however, the government controls rice farmers not to grow rice due to the over-production of rice in Japan. The farmers are forced to submit the cultivation plan not to cultivate rice in the beginning of the fiscal year, but they are allowed to cultivate rice just for the purpose of farmland conservation, not to get income from rice production. Farmers are not allowed to harvest rice for this purpose. It should be cut away and treated as "not food", but just for maintaining the land as paddy field. Certain amount of subsidy had been paid instead to cover the income reduction due to this cooperation to the production control. The government officer come to visit farmer and inspect the site one month before the harvest season how the farmer treated rice giving the question of "Was rice treated suitably as "Not rice for food". The purpose of policy enforcement is to keep the income of farmers as the original because overproduction of rice lowers the price of rice. In order to keep farmers' income high, rice production is curtailed, and the government will cover the income that is reduced by production control with taxes. The market principle does not work under this circumstance. Not only that, in addition to criticism from taxpayers for how to use the taxes they paid, farmers' pride in the profession of agriculture and declining motivation have resulted in a situation where the future successors do not grow. Basically, long-term production control policy cannot be continued for a long time by private companies. It is a stupid policy that can be done only by the government using public subsidy funds (tax).

In general, it is common sense not to take production control policy in the situation of overproduction, but to respond by expanding consumption, developing new products, and new markets development. The production control policy has been continued over than a half century, therefore, most of the rice farmers were spiritually discouraged and they have no more hopes to grow next era successors. Almost more than 400 thousand hectares of paddy fields are not cultivated and left abandoned in Japan even though the country is very small and the food self-sufficiency is less than 40 %. The full-time farmers are only 15%. The rest of 85% of the farmers are now part-time farmers. Full-time farmers are increasing their farming scale

by contracting to rent the farmland of small-scale farmers who have difficulty in continuing agriculture due to the aging population and lack of successors. However, since the agricultural land is not accumulated and scattered all over the place, the transportation of machines and the delivery of the agricultural land to the leased agricultural land are problems.

Part-time farmers look mostly satisfied enough because they can get the regular salary monthly from the part-time job, even though they are not paid so much higher, but they can purchase the machine to take care of their farmland maintenance enough. The increase of part-time farmer by providing the part-time job makes the small scale, low income farmer rich enough to buy the machine of their own for taking care of their family such as the education of their children. Most of the part-time farmers are satisfied well with such a daily life, therefore they have no need to sell their farmland and change the job. This situation was caused by the governmental production policy implementation. It can be clearly found from this act that it is important and necessary to support farmers and make them rich financially, however no good results couldn't be brought as expected. No promotion of agriculture couldn't be achieved even they became rich in economy even in the future as far as the same situation is continued. The main point of this paper is not to repeat the same thing again as we learned from this policy enforcement and show the direction which way we should lead and guide the Asian agriculture and grow the Future Farmers of Asia. Fortunately Asia has a lots of natural resources (including food resources) and higher potentiality to produce and supply to all over the world. It is extremely important from the food issue viewpoint due to the world population increase that we should take action immediately to achieve the final goal to transfer the technology which is competitive in the world and to grow human resources (FFA) equipped with higher education.

Japan is well known as one of the technology oriented countries, however, agriculture is not an attractive job and profession for young generation.

The author believes Asia has higher possibility in agriculture, and cooperation and competition between technology oriented and resource oriented countries can maximize its potential towards the future. Time is limited and sooner is better to be ready for taking action to the upcoming food issue caused by the rapid human population growth.

⑥Social service

After the successful completion of the program, the official certificate should be issued. This is the division to take care of those matters. Required documents for official approval, certification and authentication can be prepared and issued if needed.

C. Future Farmer of Asia growing Program

Figure 15 shows the FFA growing program scheme and content. As already mentioned, Asia is one of the huge agricultural production regions in the world. A big amount of food resource is grown up, however, the quality of them is not highly controlled. The value can be added more if it could be controlled in addition to food security. The most important point to keep in mind is to promote the regional agriculture and economy, not to increase the farmer's income.

The summarized content of the FFA growing program is shown in Figure 15. The total length of the training is set around six months including the one month orientation and final wrap-up. Participants are normally accepted not only from Thailand, but also from ASEAN member countries. Acceptance of trainees (participants) from ASEAN member countries basically requires a letter of acceptance from the host country's national government. The cost of acceptance will be borne by the recipient. The target crop of the program is mainly rice since rice is a typical staple food crop, and more than 90% of the world's production is cultivated in Asia. In Thailand, rice can be possibly cultivated two to three times a year, therefore, the stable production and supply can be possible. Even Data collection in rice cultivation tests can be also possible at any time of the year. This is also one of the characteristic merits of rice cultivation in tropical region.

Figure 15. Future Farmers of Asia growing program scheme

Participants are requested to cultivate rice for one season. The purpose is to learn practical skills how to grow it as one season experience. Bi-weekly or monthly report submission is also requested with short time presentation should be done at the regular meeting. Lecturers are invited for delivering the lectures during three to four days per one time invitation. The lecturer has to stay about one week per one time invitation. The rest of two to three days should be spent together with the participants to visit site and discuss on the related topics. This is a valuable opportunity even for the lecturer to know more Asian agriculture and to deepen friendship with trainees. The total number of lecturers will

be around 20 or so. The ratio of the invited lecturers from outside Thailand to inside will be around 60% to 40% or so. Further discussion should be done on this matter referring to the curriculum. The lecturers may be invited from university, research institute, industry, in which they are working actively as the professional, expert, engineer, specialist etc.

D. Workshop holding

Figure 16. Workshop on Precision Agriculture for Thailand 4.0

The workshop on precision agriculture and agricultural machinery industry for Thailand 4.0 was organized and held by Ministry of Science and Technology, on September 19-20, 2017 as shown in Figure 16. The author was requested to contact and arrange the invited speakers. The field demonstration was also conducted by the participated industries using their products

E. Similar training programs in the past

Figure 17. Joint Group training program by JICA and NPO

JICA, Japan International Cooperation Agency, the Government of Japan initiated the group training program focusing on two courses: 1) Rice mechanization and 2) Machine design. Around 15 participants were accepted based on bilateral agreements between partner and host countries. Participants stay almost six months at the accommodation provided by the host country. They were requested to join the lectures and field training of rice cultivation for practical experience and skill-up. Study tour program to visit the various organizations such as university, national research institute & experiment station, government organization related to the course requirement. Two to three days home stay program at Japanese rice farmer was also involved to know not only rice mechanization, but also the daily life of rice farmer. For home stay program at local farmers, the official support was obtained by the city government and agricultural cooperatives.

F. Hopeful possibility for hyper-low cost rice production

There are two planting ways in existing rice farming, however the acceptance ratio of these two planting methods varies depending on the following factors such as climate, locality, labor force hiring costs, and technical liability. In Japan, almost 90% of rice is transplanted, whilst in Thailand the ratio of transplanting to direct sowing is 50% to 50%. The reason why the transplanting is more widely accepted than direct sowing is that the following problems have not been solved yet in direct sowing. They are 1) bird attack to the seed sowed at the ground surface of paddy field, 2) lodging problem by strong wind while growing, because the sowed seed is uncovered with soil and exposed, and 3) weed control. Seed coating by CaO2, calcium peroxide is to provide oxygen to promote germination and improve its percentage under submerged condition in the soil.

Figure 18 shows the coated seed with CaO2, which produces oxygen reacting with water. The ferrite, Fe2O3 can be also used as coating aids for precision seed sowing, because the ferrite coated seed can be used for seed detecting by use of magnetic sensor, therefore the number of seed can be counted per unit time and area. Seed population density sowed per unit area can be easily calculated while the operation is going on and monitored for the planting machine driver to see. It looks that the coating technology has higher potentiality in future as follows. The main purpose of coating using CaO2 shown here is to provide oxygen for promoting the germination and increasing its percentage. The coating with Ferrite (Fe2O3) has the other merits such as additional weight increase of seed to penetrate into the soil, precision sowing by sensing, counting and monitoring the number of seed for the machine driver to know how many seeds are being sowed etc. If the other functional aids like herbicides, pesticides, fertilizer etc. could be possibly coated, the rice cultivation process will be drastically simplified. The direct sowing of those coated seeds by drone will be one of the near future hopeful technologies.

The problem of direct sowing of rice is 1) weed control, 2) bird attack and 3) lodging in order of importance. Just only

two weeks after seed sowing is the most important duration for the farmers to be nervous wondering if the seed is germinated and sprouted enough

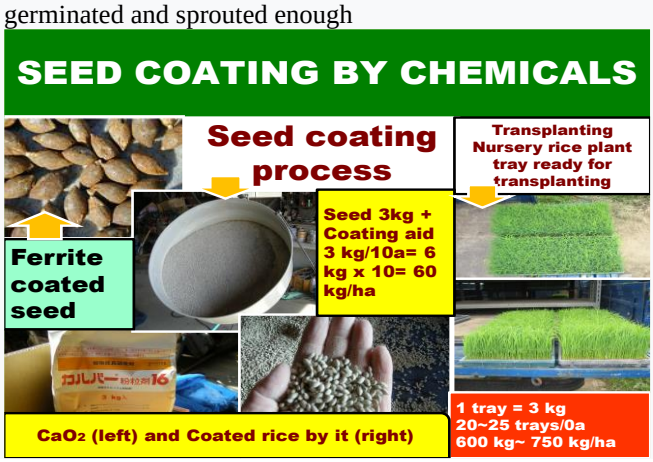


Figure18. Coated rice with chemicals, CaO2

Figure 19 shows the rice combine equipped with direct de-husking function under harvesting operation, which can de-husk the rough rice directly as soon as the rice is harvested and threshed.

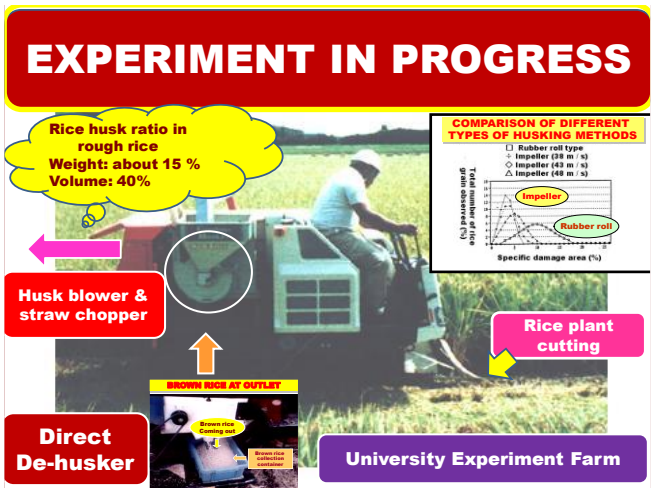


Figure 19. Combine de-husker under harvesting operation

A series of three operations consisting of harvesting, threshing and de-husking can be done in one pass operation simultaneously even high moisture rice. This research was conducted and reported long time ago by the author, however the commercial product is not coming out yet due to the small injury on to the brown rice caused by the impact of impeller rotating at high peripheral speed of 32m/s. The impeller type de-husker is coming to see often at exhibitions mainly focusing on the compactness, maintenance free and simple mechanism and structure as the rice de-husker.

In the future, the possibility of extending these two technologies is extremely high due to the simplification of agricultural operation processes, energy saving, labor saving, and ease of material handling [18].

In the past, the indicators of machine development for agricultural mechanization were 1) energy saving, 2) automation, and 3) safety. There are some overlaps among these three indicators, but basically it starts with mechanizing the basic agricultural operation that is the principle. The general process is to reduce the number of operation process, simplify, perform simultaneous or serial operation, and improve accuracy by automation, and the technology described here is one of the possibilities. In order for technology to be installed in a commercialized machine, it will not be put into practical use or commercialized at once unless some related technologies surrounding it are in the same level of situation and environment. As we already know, it is important to identify problems and investigate their causes. Needless to say, the social needs of the technology are also one of the major factors for technological development, commercialization, and dissemination.

VI. CONCLUSION

Global tetralemma is a common issue not only in Asia, but also in the world, and these issues must be resolved urgently. Asia is qualified as a world food pantry capable of producing and supplying huge amounts of food. Although the production is however large, the quality of the products is not well controlled. From the perspective viewpoint of food security due to the rapid increase in the world population, technology-oriented countries should actively invest in Asia through technology transfer and human resource development. A collaboration in which technology-oriented countries provide technology and resource-oriented countries provide resources is effective in solving mutual common problems. In other words, avoidance of food crisis, escape from hunger and poverty, and promotion of regional economy bring regional peace keeping and its stability. Agricultural policy is not about providing farmers with financial support to increase their income. It is more important to have a policy to strengthen the agriculture rather than supporting farmers. If such a policy is continued, there can be no promotion of agriculture. In Japan, a rice production control policy was implemented for half a century, but, as a result of that, agriculture has declined significantly rather than being promoted. Although the aging of farmers will come sooner or later, looking down on the agriculture industry has brought about a serious situation where young successors are not interested in continuing in that sector. Given that industry is production-based, production control is not an option. Rather, the basic principle is to focus on developing new markets and increasing consumption. Food is an indispensable resource for human survival, and the prosperity of agriculture-based Asia is important and necessary not only for Asia, but also for the world. Rice cultivation system consists of basically three operation processes such as planting, managing rice plant while growing and harvesting. The existing rice

cultivation system can be reconsidered for further improvement to cost down by the replacement of planting method from transplantng to direct sowing and the introduction of combine harvester equipped with rice de-husking functional mechanism that enables the threshing of high-moisture rough rice. This combine has the function of processing the three operations of rice harvesting, threshing, and rice hulling at the harvesting stage, further facilitates material handling and saves labor. There can be seen a high possibility that the shift to a system that aims to save energy and achieve hyper-low cost production of rice is high.

Due to the preference of consumers who dislike brown rice with tiny damage and injury caused in the hulling process of high-moisture rough rice, it has not yet been commercialized. It should be hurry to set and clarify the criteria for the damage rate standard code of brown rice. From this point of view, the machine equipped with this de-husking function has been focusing on the production and utilization of processed rice for a while and still some more time will be needed to be ready for the acceptance environment [19].

IT, Information Technology and IoT, Internet of Things based high-tech agriculture provides the opportunity for many various kinds of industries to join and invest in agriculture. Regional community-based agriculture should be promoted in Asia instead of small scale farming.

ACKNOWLEDGMENT

The author would like to express sincere appreciation to the organizing committee of SIGNAL 2021 for providing invaluable opportunity to deliver the keynote lecture related to his research activity in Asia sustainability at SIGNAL 2021. Thanks to Dr. Ratchatphon Suntivarakorn, Dean of Faculty of Engineering and Dr. Chanoknun Sookkumnerd, Associate Professor, Faculty of Engineering, Khon Kaen University, Thailand for their support and encouragement.

REFERENCES

[1] Nobutaka Ito, Technology for overcoming the Global Tetralemma, SIGNAL2021 The Sixth International Conference on Advances in Signal, Image and Video, Valencia, Spain, May 30 - June 3, 2021 Proceedings. Assigned and registered contribution id: 60011, Registered for main topic: SIGNAL 2021 : Signal process and Registered for secondary topic: SIGNAL 2021 : Signal process

[2] Nobutaka Ito, Asian Agriculture Growth Strategy, Journal of JSAMFE, Japanese Society of Agricultural Machinery & Food Engineers, Vol. 77, No. 4, p. 226~231, 2015

[3] Nobutaka Ito, Asian Agrifuture: How the Asian Agriculture can be achieved, The 7th International Conference of TSAE, Thai Society Agricultural Engineering,, July30, 2020, Suranaree University of Technology, Korat, Thailand

[4] Nobutaka Ito , Asia Techno Farm, The International Conference of Building of Food
Sovereignty through a Sustainable Agriculture: Challenge of Climate Change and Global
Economic Community, University of Jember (UNEJ), East Java, Indonesia, August 1st ~ 3rd, 2017

[5] Nobutaka Ito, Asia Techno Farm Initiative, Thai Society of Agricultural Engineering, Impact
Arena, September 9 ~ 10, 2017

[6] Noutaka Ito, FFA (Future Farmers of Asia) growing program, Proceedings of The
11th TSAE International Conference, 26~27 April, Chulaborn International Convention
Center (Wora Wara Hua Hin Hotel & Convention) Hua Hin, Prachuap Khin Khon, Thailand

[7] Ministry of Agriculture, Forestry and Fisheries, Research Center, Smart Agriculture homepage, https://www.maff.go.jp/j/kanbo/smart/

[8] Front runner, Professr, Dr. Noboru Noguchi, Graduate School of Agriculture, Field Robotics, Hokkaido University https://www.youtube.com/watch?v=bNGkPfJdSDc

[9] Noboru Noguchi, AG / SUM Agsum coverage (Hokkaido: the future brought about by robot tractors) Hokkaido University, Government of Japan, Cabinet Office SIP Secretariat, (AG/SUM means Agricultural Summit and SIP meas Cross ministerial Strategic Innovation Promotion Program respectively)
https://www.youtube.com/watch?v=WllOTRW8Hdw

[10] TV Asahi morning news, One truck driver drives three cars in a platoon on the highway, March 9, 2021
https://news.tv-asahi.co.jp/news_economy/articles/000209325.html

[11] Nobutaka Ito, Robotics & Abenomics for ASEAN Economic Community and Asia Sustainability, MIER 2013, Joint Symposium on Mechanical-Industrial Engineering and Robotics 2013, Chiang Mai, Thailand, Extended abstract, page 18 November 14 – 17, 2013

[12] Osaka Prefectural University, Plant Factory home page
https://www.plant-factory.osakafu-u,ac.jp/ [accessed May 2021

[13] Kinki University Fisheries Laboratory home page:
http://www.flku.jp/english/index.html

[14] EUGLENA-ISUZUproject:
http://www.bloomberg.com/news/articles/2014-06-25/japan-s-isuzu-euglena-to-begin-biodiesel-development-with-algae

[15] Oxford Instruments,
Overview of Various Forms of Light Scattering
https://andor.oxinst.com/learning/view/article/scattering-of-light-an-overview-of-the-various-forms-of-light-scattering

[16] Nikkei Cross tech (X tech) article, June 29, 2021
Micro that electrically determines the shape of algae
cells at high speed ... --Euglena https://www.euglena.jp › news

[17] Toyota Motor Corporation, Ltd., Good Harvest Plan: https://www.toyota.co.jp/housaku.plan.html [accessed May 2021]

[18] Nobutaka Ito, Hyper Low Cost Rice Mechanization System, The 14th TSAE International Conference and the 22nd TSAE National Conference, Paper Code No. 10, May 12 - 13, 2021 organized and hosted by Khon Kaen University, Khon Kaen, Thailand, (Paper Award)

[19] Nobutaka Ito, Asian Agriculture: How the Asian Agrifuture can be achieved, The 7th International Conference of TSAE, Thai Society of Agricultural Engineering, July 30, 2020, Suranaree University of Technology, Korat, Thailand