FATHOM: A Neural Network-based Non-verbal Human Comprehension Detection System for Learning Environments

Fiona. J. Buckingham, Keeley A. Crockett, Zuhair A. Bandar, James D. O'Shea

School of Computing, Mathematics and Digital Technology
Manchester Metropolitan University
Chester Street, Manchester, M1 5GD, UK
fiona.j.buckingham@stu.mmu.ac.uk

Abstract— **This paper presents the application of FATHOM, a computerised non-verbal comprehension detection system, to distinguish participant comprehension levels in an interactive tutorial. FATHOM detects high and low levels of human comprehension by concurrently tracking multiple non-verbal behaviours using artificial neural networks. Presently, human comprehension is predominantly monitored from written and spoken language. Therefore, a large niche exists for exploring human comprehension detection from a non-verbal behavioral perspective using artificially intelligent computational models such as neural networks. In this paper, FATHOM was applied to a video-recorded exploratory study containing a learning task designed to elicit high and low comprehension states from the learner. The learning task comprised of watching a video on termites, suitable for the general public and an interview led question and answer session. This paper describes how FATHOM's comprehension classifier artificial neural network was trained and validated in comprehension detection using the standard backpropagation algorithm. The results show that high and low comprehension states can be detected from learner's non-verbal behavioural cues with testing classification accuracies above 76%.**

Keywords—artificial neural networks; backpropagation; comprehension; FATHOM; non-verbal behaviour

I. INTRODUCTION

Non-verbal behaviour is a form of non-linguistic communication that automatically accompanies verbal conversation. Gestures, facial expressions, and body movement are all examples of non-verbal behaviour [1]. Little work has been done on automatic comprehension detection, yet humans exhibit non-verbal cues consistently while undertaking day-to-day tasks. Thus, the research presented in this paper seeks to examine whether patterns of comprehension and non-comprehension exist within non-verbal behavioural cues.

Previous classroom studies [2-7] have identified non-verbal behavioural indicators of non-comprehension, including facial behaviour, hand and body movements. However, this work has largely relied on subjective human coding [8] with associated inconsistency and upon verbal techniques. Thus there is a role for a non-verbal multichannel, comprehension detection system capable of reliably classifying human comprehension through facial non-verbal behaviour.

Comprehension is often associated with written language [9] and is often defined as "the process of simultaneously extracting and constructing meaning through interaction and involvement with written language" [9]. In this research, we define comprehension as the learner demonstrating through interaction with a tutorial, (via verbal communication and/or non-verbal behaviour), that they understand or grasp the meaning of the tutorial material presented to them at a given point in time. The tutorial in this paper (described in Section V) comprised of each participant watching a factual video and participating in a question and answer (Q&A) session immediately after.

FATHOM [10], is an artificial neural networks (ANN) based system developed specifically to detect levels of comprehension. FATHOM was developed based around an existing physiological profiling system known as Silent Talker [11] and was first trialled during an informed consent assessment process carried out in North-western Tanzania, Africa using a setting similar to that used for a Human Immunodeficiency Virus (HIV)/Acquired Immunodeficiency Syndrome (AIDS) prevention randomized study [10]. The work produced strong evidence [10] that detectable patterns of comprehension and miscomprehension exist within the monitored facial non-verbal multichannels, for the sample of African women with a limited set of non-verbal behavioural features. Initial observations provide grounds to suspect that there will be more, less obvious, micro gestures available for classification.

The aim of the research presented in this paper is to apply FATHOM as a comprehension detection system to a learning task designed to distinguish high and low comprehension states from the learner based on facial non-verbal cues. In order to assess FATHOM's ability, a new exploratory study was designed to capture comprehension levels of adults over the age of 18. The motivation of this work is to ultimately link FATHOM to pedagogical intervention in learner-adaptive online teaching and learning tutorials that could be delivered in 24/7 scenarios to improve the overall learning experience.

This paper continues as follows: Sections II and III review non-verbal behaviour, comprehension and learning. Section IV describes FATHOM - a comprehension detection system using ANNs. Section V presents the experimental study methodology and results. The conclusion and further work can be found in Sections VI and VII.

II. NON-VERBAL BEHAVIOUR

Non-verbal behaviour consists of a variety of signals or cues including visual, audio, tactile and chemical which are exhibited by human beings in order to express themselves [12]. The non-verbal cues are revealed often before a verbal response [13] and potentially can provide early signals to the listener about the sender's state whilst they are formulating the actual verbal response. Typically, non-verbal behaviours are being generated before, during and after the sender articulates a verbal response. A large number of non-verbal behavioural channels are available [14] and research has been undertaken to collect data on individuals to try and identify patterns associated with an individual's state. Knapp and Hall [15] stated that as humans often communicate face-to-face – the face was a source of rich information and should be given higher precedence. Mehrabian [16] found that around 55% of non-verbal messages communicated by an individual came from facial behaviour expressions and debated that even when cues from facial behaviour were not consistent with the verbal response that the listener was most impacted by the facial emotion expressed [17].

Little work has been on the automatic detection and classification of non-verbal behaviour. The traditional way still is to use human judges to code the non-verbal behaviour channels [18,19]. Nonetheless, the judges have to be trained, their assessment is subjective and they can concentrate upon only a limited number of channels at one time [11]. The whole process is time consuming with Johnson [20] reporting it can take one hour to code one channel from one minute of film.

Attempts have been made to automatically detect non-verbal cues using more recent technology using the Microsoft Kinect computer vision algorithm [21], however these attempts have been looking at full body gestures as opposed to fine grained channels which are used by FATHOM. Digital technologies such as camcorders and the Microsoft Kinect have the strength of being able to capture large volumes of continuous non-verbal channels, which can be stored as a multimedia file and used for single or repeated post analysis. On the other hand, digital recordings are limited to environments that facilitate the setup of the technology [22]. Therefore, a niche exists for a computer-based system that is able to automatically monitor multiple channels of non-verbal behaviour from digital video recordings such as FATHOM described in Section IV.

III. COMPREHENSION AND LEARNING

Trying to establish whether or not a learner is comprehending a tutorial as a whole or as independent elements is a non-trivial task. Early work by Dollaghan et al. [23] found that in a classroom environment, whilst a child may not comprehend everything they hear, they should be able to recognise that they do not understand and know way to ask for help. From a teacher perspective, it is important that every child knows how to listen, and asks when he or she does not understand. However, in large classes the ability of an individual to monitor and detect comprehension levels of every child is a very great challenge.

Previous work on detecting comprehension has primarily examined language comprehension [24] from reading, writing and listening aspects. However in [5] children participating in a lesson on electricity containing both easy and hard material were videotaped and individual observers were able to distinguish children comprehending and not comprehending from the non-verbal behaviours alone that they exhibited [5]. Amelsvoort et al. [8] examined if non-verbal cues give out information on how a child perceives the difficulty of an arithmetic problem. Again through audio-visual recordings, children were analysed (after the event) by manually identifying head movements using the Active Appearance Model, which were then used to train a classifier. Empirical experiments undertaken in the study found that it was possible to estimate the difficulty level at above chance levels from non-verbal cues. Other research in classroom environments has used human decoders to analyse recordings to identify non-verbal behavioural patterns associated with non-comprehension [3,25]. Research to date involving classroom based studies shows that that comprehension and non-comprehension patterns do reside within non-verbal behaviour alone but the results are often based on only a few non-verbal channels. This is because human observers cannot process more than 7 ± 2 pieces of information [26] at the same time in immediate memory. In addition, the detection of non-comprehension at a certain point in time occurs after the tutorial has taken place so no remedial action can be put in place by the teacher. The solution proposed in this paper is a comprehension detection system, known as FATHOM which can monitor multiple non-verbal behaviours for comprehension detection and make an automatic decision at a predefined point in time on a person's level of comprehension. This removes the subjectivity of human experts and allows more channels to contribute towards the classification of comprehension. An outline of FATHOM is given in the following section.

IV. FATHOM

FATHOM is computer-based comprehension detection system that uses a collection of neural networks to concurrently monitor multiple channels of human non-verbal behaviour [10]. Fig. 1 shows a screenshot of the FATHOM software automatically processing a learner's facial non-verbal behavioural channels contained within the currently displayed video frame. As the video plays, FATHOM analyses the learner's non-verbal behaviours in every video frame within a predefined period of time measured in seconds. The predefined period of time is known as a timeslot, which can be fixed e.g. every n second(s) or variable. In this paper, FATHOM was configured to use a fixed one second timeslot. At the end of the timeslot, the overall comprehension/non-comprehension classification level is outputted to FATHOM's user interface.
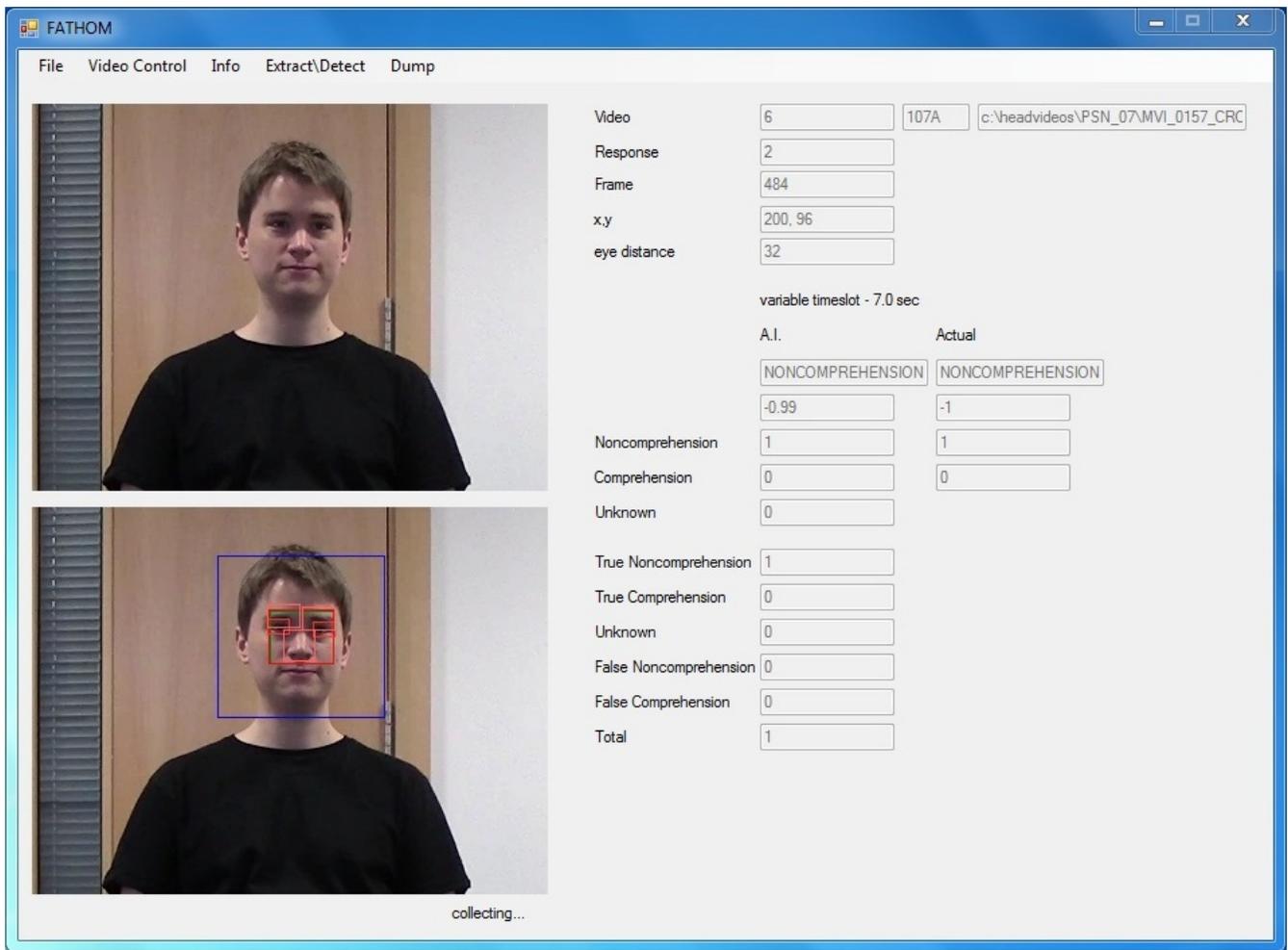
Fig. 1. FATHOM.

FATHOM has forty non-verbal behavioural channels [12], which are categorised as:

- Face (20 channels) e.g. face blushing/blanching.
- Eyes (16 channels) e.g. left/right eye gaze.
- Other (4 channels) e.g. gender.

The neural networks contained within FATHOM are: the object locator ANNs, the pattern detector ANNs and a single comprehension classifier ANN. Firstly, the object locators ANNs identify the location of non-verbal features such as the eyes, eyebrows and the nose. After the object locator ANNs have located their non-verbal feature, the pattern detectors ANNs identify the state of the object such as the right eye gazing to the right. The states of each non-verbal channel is collated for the timeslot and then passed to the final ANN, the comprehension classifier ANN, which outputs a value between +1 and -1, which indicates whether the person was comprehending (+1) or not comprehending (-1) during that period of time.

This research paper is focused upon the training and validation of FATHOM's comprehension classifier ANN with a dataset from an exploratory study (Section V) containing a learning task with associated comprehension assessment questions. FATHOM's comprehension classifier ANN was trained with the standard backpropagation algorithm [27,28] using incremental weight updating and the delta (δ) rule [29]. n-fold cross-validation [30] was used to randomly partition the dataset in to n equally sized subsets. A single subset is retained as the test set and the remaining subsets are used as the training set to form a single fold. The cross-validation process repeats until all n subsets have been used as the test set once (n-folds). The training set is used to train the FATHOM comprehension classifier ANN and the test set was used to determine the error rate of the trained FATHOM comprehension classifier ANN. The advantage of n-fold cross-validation is that all samples within the dataset are used and the results from the n folds can be averaged. Section V describes the experimental study on detecting learner comprehension.

V. EXPERIMENTAL STUDY: DETECTING LEARNER COMPREHENSION

The primary aim of the experimental study was to identify whether high and low human comprehension associated

multichannels of non-verbal behaviour reside within a video-recorded British (UK-based/English speaking) sample of participants. The participants were filmed whilst watching the factual video and during the Q&A session but only the videoed non-verbal cues exhibited during the reading of the questions in the Q&A session were used for post analysis by FATHOM. The exploratory study builds upon lessons learnt in a previous a research study [10] where an African female video-based non-verbal dataset was used to train and validate a backpropagation neural network in the detection of human comprehension. This section will outline the methodology of the experimental study and present the results.

A. Participants

Forty participants were selected to participate in the study from academic and technical staff at the Manchester Metropolitan University (MMU) in the UK. The sample was composed of 20 males and 20 females. The males had a mean age of 41 years old (SD = 14 years) and the females had a mean age of 39 years old (SD = 14 years). Each participant was invited to individually engage in a short learning task followed by a small set of associated assessment questions whilst being video recorded. All participants completed an informed consent form on their participation and the usage of video recorded material for research purposes. Ethical approval was obtained from the MMU Faculty Academic Ethics Committee.

B. Study Procedure

Prior to the study a short learning topic was selected, which was a factual digital video on Termites with a total duration of 8 minutes 40 seconds. The Termite video was targeted at the general public with no age restriction and covered: functional architectural aspects of the termite mounds, roles within the social structure of a termite colony and locations where termite colonies thrive. Two experts (Academic Professors) on the subject area were recruited to develop ten difficult (hard) questions and ten easy questions related to the video content with expert agreement for the participants to answer. The experts were required to devise five open questions and closed questions within each set of hard and easy questions. At the same time, the experts noted down the correct answer(s) for each question, which was later used as the mark scheme.

The study was conducted at MMU in the same room with the same equipment and layout shown in Fig. 2 to ensure consistency in the quality of the digital video recordings. Each participant watched the termite video and then the interviewer followed a script, asking each participant all of the hard and easy questions in a randomised order. If the participant did not respond to a question then the interviewer was instructed to repeat the question. If there was still no response after repeating the question then the interviewer was instructed to ask a neutral probing questions such as "What is your best guess?". The participant was video-recorded whilst watching the termites video and when answering the questions associated with the video content. The digital camcorder was setup to capture the participant's upper body non-verbal cues in to an MP4 multimedia file for post study analysis and extraction by FATHOM.
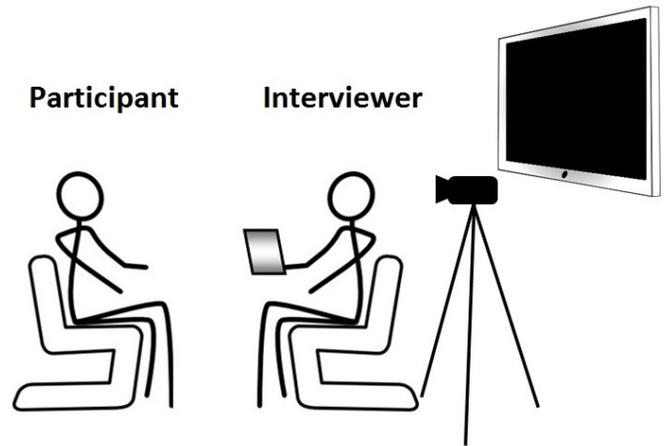

Fig. 2. Study Layout.

To counteract question order effect, the presentation of each subset of open and closed questions were randomised, resulting in four conditions: 1A, 1B, 2A and 2B as shown in Table I. For example, if the participant was randomised in to condition 2A then he/she would receive the easy open questions first, followed by the hard open questions, then the easy closed questions followed by the hard open closed questions. Each participant was randomised in to one of the conditions so that each condition had five males and five females (10 participants per condition). Equally randomising the participants across the conditions in this manner reduces the chance of producing an imbalanced dataset related to gender.

TABLE I. QUESTION ORDER

Condition	Closed Questions		Open Questions	
	Easy	Hard	Easy	Hard
1A	1st	2nd	3rd	4th
1B	2nd	1st	4th	3rd
2A	3rd	4th	1st	2nd
2B	4th	3rd	2nd	1st

C. FATHOM's Comprehension Classifier ANN Training Procedure

After the study, the participant's answers to each of the questions were transcribed from the video recordings and marked using the mark scheme. All forty participant MP4 video files were processed in FATHOM's extraction mode to collate the non-verbal vector-based dataset for cross-validation training of the FATHOM comprehension classifier backpropagation ANN.

Each vector within the dataset represented a 1 second time period and contained the state of all forty non-verbal channels at that point in time, in a normalised format i.e. scaled from +1 to -1. Appended to the end of each vector was a normalised supervisory value, which was used during cross-validation training and testing to determine whether the FATHOM

comprehension classifier backpropagation neural network had correctly classified the input vector. If the supervisory value was +1 then it represented comprehension and if the supervisory value was -1 then it represented non-comprehension.

The extracted dataset was generated from the multichannels of non-verbal behaviours emitted by the participant only when the interviewer was asking the participant each one of the twenty assessment questions. Therefore, if it took 10 seconds for the interviewer to ask the participant an assessment question then successful extraction of all forty non-verbal behaviours would result in ten normalised vectors. Each set of vectors for each question would then be appended with the supervisory value based upon whether the participants answer to that question was marked as correct or incorrect. The purpose was to see if the participants displayed strong non-verbal indicators of comprehension and non-comprehension when the interviewer asked each question.

To ensure that only strongly associated high and low non-verbal indicators of comprehension were extracted from the participant videos for the open questions, the participants marked open answers were all thresholded at 75% e.g. if the participant got $\geq$ 75% of the answer then the question was marked as correct otherwise the question was marked as incorrect. This threshold was determined by the experts in the field.

D. Results

From the forty participant videos, FATHOM extracted 16,951 comprehension vectors and 23,857 non-comprehension vectors. Therefore, the entire dataset was composed of 40,808 vectors with 41.5% in the comprehension class. Table II provides a breakdown of the percentage of correctly answered questions by all participants.

TABLE II.   MARKED QUESTIONS

	Closed Questions		Open Questions	
	Easy	*Hard*	*Easy*	*Hard*
Correct (%)	83	68.5	17	6

In this experiment, FATHOM's comprehension classifier ANN was trained and validated in the detection of human comprehension with the extracted non-verbal dataset using the backpropagation algorithm and 10-fold cross-validation. Each of the cross-validation folds was partitioned as follows: 90% training and 10% testing. The topology of the FATHOM comprehension classifier ANN was: forty inputs, a single hidden layer with twenty neurons and an output layer with one neuron (40:20:1). The maximum number of epochs was 10,000 and the learning rate ( ) was set at 0.005. The checking epochs parameter was 250, which meant that at every 250th epoch the total classification accuracy would be checked and if it had not improved then the backpropagation training would terminate. On commencement of training, the training set was randomised once and the neural network had the weights initialised in the range of $0\pm1/fan$-*in*, where *fan-in* represents the number of inputs entering the neuron. The latter neural network training parameters were determined from previous exploratory cross-validation sessions. Parameters were not optimised for this exploratory experiment.

The 10-fold cross-validation training phase results are in Table III and the testing phase results are in Table IV. The Root Mean Squared Error (RMSE) is a performance metric used to determine the degree of error by squaring the aggregated difference between the neural network output and the supervisory value for the dataset. The classification accuracies (CA) were calculated as follows:

- **Comprehension CA** is the percentage of comprehension vectors classified correctly.

- **Non-comprehension CA** is the percentage of non-comprehension vectors classified correctly.

- **Total CA** is the percentage of comprehension and non-comprehension vectors classified correctly i.e. Total CA = Comprehension CA + Non-comprehension CA.

- **Total Normalised CA** is calculated as the Total CA / 2.

The best performing FATHOM comprehension classifier ANN was obtained in fold 4 with a testing phase total normalised CA of 79.58% in Table IV. In Table III and Table IV all of the neural networks were able to classify non-comprehension vectors more easily than comprehension vectors, which may be due to the entire dataset being composed of 58.5% non-comprehension class. All of the ANNs during 10-fold cross-validation consistently achieved CAs above 67.5%, thus strongly indicating that comprehension and non-comprehension patterns exist within the dataset.

VI. CONCLUSION

Overall, the cross-validation results for the FATHOM comprehension classifier ANN consistently attained total normalised CAs above 76% in the testing phase, which strongly indicates that comprehension and non-comprehension was detectable from the multichannels of non-verbal behaviour emitted by the male and female participants during the questioning phase of the study on the content of the termite's video. The FATHOM comprehension classifier ANN performed better at classifying non-comprehension than comprehension, which may have been caused by there being more non-comprehension vectors than comprehension. This is also likely to have been down to the design of the task in that the two experts will have their own opinions on what constitutes high and low comprehension questions in the field of termites. Consistency in the quality of the video recordings resulted in a large dataset with no discarding of the participant MP4 multimedia files. The application of FATHOM as a comprehension detection system to the learning task has identified that high and low comprehension states can be detected from learner's facial non-verbal cues, thus satisfying the aim outlined in Section I.

TABLE III.   CROSS-VALIDATION TRAINING PHASE RESULTS

Fold	Epoch	RMSE	Total CA	Total Normalised CA	Comprehension CA	Non-comprehension CA
1	750	0.738	81.83	80.85	75.07	86.62
2	1750	0.741	81.72	80.17	71	89.33
3	500	0.75	80.96	79.68	72.12	87.23
4	4750	0.726	82.72	82.38	80.37	84.38
5	3500	0.718	83.09	82.16	76.63	87.69
6	3250	0.721	83.27	82.45	77.62	87.28
7	1750	0.737	82.2	81.77	79.2	84.34
8	2000	0.725	82.55	82.02	78.9	85.14
9	1750	0.73	82.55	81.25	73.53	88.97
10	2750	0.735	82.29	82.04	80.53	83.54
Mean	2275	0.732	82.31	81.47	76.49	86.45

TABLE IV.   CROSS-VALIDATION TESTING PHASE RESULTS

Fold	RMSE	Total CA	Total Normalised CA	Comprehension CA	Non-comprehension CA
1	0.772	79.68	78.68	72.8	84.57
2	0.787	78.75	77.13	67.55	86.7
3	0.787	78.08	76.71	68.61	84.82
4	0.771	80.02	79.58	76.99	82.18
5	0.769	80.02	79.15	73.98	84.32
6	0.768	80.34	79.37	73.62	85.12
7	0.788	79.36	78.63	74.27	82.98
8	0.781	79.12	78.51	74.92	82.1
9	0.771	79.93	78.33	68.9	87.76
10	0.802	78.61	78.24	76.06	80.42
Mean	0.779	79.39	78.43	72.77	84.09

VII. FUTURE WORK

Naturally, further application of FATHOM to previously unencountered, larger representative cultural datasets should enhance its ability at comprehension detection and progressively advance towards answering the following hypothesis:

Is there a general set of non-verbal behaviours that a backpropagation neural network can use to detect high and low patterns of human comprehension across all cultures, genders and age groups?

Further work includes pruning the number of inputs to the FATHOM comprehension classifier ANN and performing cross-validation experiments to empirically determine the best neural network topology i.e. optimising the neural network topology with the minimum number of hidden layers and neurons whilst retaining acceptable classification accuracies. Another significant future direction is to perform a non-verbal channel analysis using alternative, artificially intelligent computational models such as Decision Trees [31,32]. Investigating the performance of a range of machine learning models in non-verbal comprehension detection would enable comparisons and identification of suitability.

The real-world applications of the FATHOM comprehension detection system are numerous. For example, use of FATHOM in the academic world would provide educators with a computerised proxy tool for individually assessing a learner's comprehension level from a non-verbal perspective alongside traditional comprehension assessments methods to facilitate more accurate identification of learner comprehension state in near real-time.

REFERENCES

[1] E. Babad, "Teaching and nonverbal behavior in the classroom". In L. J. Saha and A. G. Dworkin, (Eds.) International Handbook of Research on Teachers and Teaching. Boston, Massachusetts: Springer US, pp. 817-827, 2009.

[2] S. Machida, "Teacher accuracy in decoding nonverbal indicants of comprehension and noncomprehension in anglo and mexican-american children". Journal of Educational Psychology, vol. 78(6), pp. 454-464, 1986.

[3] J. M. Webb, E. M. Diana, P. Luft, E. W. Brooks and E. L. Breenan, "Influence of pedagogical expertise and feedback on assessing student comprehension from nonverbal behavior". The Journal of Educational Research, vol. 91(2), pp. 89-97, 1997.

[4] C. J. Patterson, M. J. Cosgrove and R. G. O'Brien, "Nonverbal indicants of comprehension and non-comprehension in children". Developmental Psychology, vol. 16(1), pp. 38-48, 1980.

[5] V. L. Allen and M. L. Atkinson, "Encoding of nonverbal behavior by high-achieving and low-achieving children". Journal of Educational Psychology, vol. 70(3), pp. 298-305, 1978.

[6] T. P. Mottet and V. P. Richmond, "Student nonverbal communication and its influence on teachers and teaching". In J. L. Chesebro and J. C. McCroskey (Eds.), Communication for Teachers. Needham Heights, Massachusetts: Allyn and Bacon, pp. 47-61, 2002.

[7] J. D. Jecker, N. MacCoby and H. S. Breitrose, "Improving accuracy in interpreting non-verbal cues of comprehension", Psychology in the Schools, vol. 2(3), pp. 239-244, 1965.

[8] M. Amelsvoort, B. Joosten, E. Krahmer and E. Postma, "Using non-verbal cues to (automatically) assess children's performance difficulties with arithmetic problems". Computers in Human Behavior, vol. 29(3), pp. 654-66, 2013.

[9] L. S. Pardo, "What every teacher needs to know about comprehension". The Reading Teacher, vol. 53(3), pp. 272-280, 2004.

[10] F. Buckingham, K. Crockett, Z. Bandar, J. O'Shea, K. MacQueen and M. Chen, "Measuring human comprehension from nonverbal behaviour using artificial neural networks". Proceedings of the IEEE World Congress on Computational Intelligence (IEEE WCCI), Australia, pp. 368-375, 2012.

[11] J. Rothwell, Z. Bandar, J. O'Shea and D. McLean, "Silent Talker: a new computer-based system for the analysis of facial cues to deception". Applied Cognitive Psychology, vol. 20, pp. 757-777, 2006.

[12] J. Rothwell, Z. Bandar, J. O'Shea and D. McLean, "Charting the behavioural state of a person using a backpropagation neural network". Neural Computing & Applications, vol. 16, pp. 327-339, 2007.

[13] V. Manusov and A. R. Trees, ""Are you kidding me?": the role of nonverbal cues in the verbal accounting process". Journal of Communication, vol. 52(3), pp. 640-656, 2002.

[14] P. Ekman and W. V. Friesen, "The repertoire of nonverbal behavior categories, origins, usage, and coding". Semiotica, vol. 1, pp. 49-98, 1969.

[15] M. L. Knapp and J. A. Hall, Nonverbal Communication in Human Interaction, 3rd Ed., Fort Worth, Texas: Harcourt Brace, 1992.

[16] A. Mehrabian, "Communication without words". Psychology Today, vol. 2(4), pp. 53-56, 1968.

[17] A. Mehrabian, Silent Messages, 5th Ed., Belmont, California: Wadsworth Publishing Company, 1971.

[18] P. Ekman and V. W. Freisen, The Facial Action Coding System (FACS), Consulting Psychologists Press, Palo Alto, California, US, 1978.

[19] N. Dael, M. Mortillaro and K. R. Scherer, "The body action and posture coding system (BAP): development and reliability". Journal of Nonverbal Behavior, vol. 36(2), pp. 97-121, 2012.

[20] R. C. Johnson, (1999) "Computer program recognizes facial expressions", Electronic Engineering Times, [Online] [Accessed on 17 June 2014] http://www.eetimes.com/at/news/OEG19990405S0017.

[21] A. S. Won, J. N. Bailenson, S. C. Stathatos and W. Dai, "Automatically detected nonverbal behavior predicts creativity in collaborating dyads". Journal of Nonverbal Behavior, pp. 1-20, 2014.

[22] J. M. Montepare, "Nonverbal behavior in the digital age: meanings, models, and methods". Journal of Nonverbal Behavior, pp. 1-3, 2014.

[23] C. Dollaghan and N. Kaston, "A comprehension monitoring program for language impaired children". Journal of Speech and Hearing Disorders, vol. 51, pp. 264-271, 1986.

[24] K. Pezdek, "Comprehension: it's even more complex than we thought". Advances in Psychology, vol. 39, pp. 215-236, 1986.

[25] E. Skarakis-Doyle, N. MacLellan and K. Mullin, "Nonverbal indicants of comprehension monitoring in language-disordered children". Journal of Speech and Hearing Disorders, vol. 55(3), pp. 461-467, 1990.

[26] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information". Psychological Review, vol. 63, pp. 81-97, 1956.

[27] D. E. Rumelhart, G. E., Hinton and R. J. Williams, "Learning internal representations by error propagation". In D. E. Rumelhart and J. L. McClelland, Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations. Cambridge, Massachusetts: MIT Press, 1986.

[28] D. E. Rumelhart, G. E., Hinton and R. J. Williams, "Learning representations by back-propagating errors". Nature, vol. 323, pp. 533-536, 1986.

[29] M. H. Hassoun, Fundamentals of neural networks. Cambridge, Massachusetts: MIT Press, 1995.

[30] M. Stone, "Cross-validatory choice and assessment of statistical predictions". Journal of the Royal Statistical Society, vol. 36(2), pp. 111-147, 1974.

[31] J. R. Quinlan, "Induction of decision trees". Machine Learning, vol. 1, pp. 81-106, 1986.

[32] J. R. Quinlan, C4.5: programs for machine learning. San Mateo, California: Morgan Kaufmann Publishers Inc, 1993.