PR - PhD - 90 - 777

A RULE-BASED APPROACH FOR SPOTTING CHARACTERS FROM CONTINUOUS SPEECH IN INDIAN LANGUAGES

A thesis
submitted for the award of the degree
of
DOCTOR OF PHILOSOPHY
in
COMPUTER SCIENCE AND ENGINEERING

by

P. ESWAR

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY
MADRAS-600 036

JULY *1990*

CERTIFICATE

This is to certify that the thesis entitled **A RULE-BASED APPROACH FOR SPOTTING CHARACTERS FROM CONTINUOUS SPEECH IN INDIAN LANGUAGES** is the bonafide work of Mr. P. Eswar carried out under my guidance and supervision in the Department of Computer Science and Engineering, Indian Institute of Technology, Madras, for the award of the degree of Doctor of Philosophy in Computer Science and Engineering.

(B. Yegnanarayana)

ACKNOWLEDGMENTS

My research supervisor, Prof.B.Yegnanarayana, has introduced me to the subject of speech recognition. I thank him for the way he has helped me to overcome my initial hesitations with his enthusiasm. I am overwhelmed by his generosity and patience in nurturing my understanding of the problem of speech recognition. Frequent discussions with him helped me to know about man-to-man communication in addition to man-machine communication.

The work was carried out under the quality improvement program at IIT Madras on leave from Anna University, Madras. I thank the generosity of the authorities of Anna University for providing this unique opportunity. I also thank my colleagues in Anna University for the encouragement given.

I thank Prof.R.Nagarajan, Head of the Department of Computer Science and Engineering, IIT Madras, Prof.C.R.Muthukrishnan, Dr.N.Parameswaran and Dr. K. B. Lakshmanan for the opportunities provided and the encouragement given throughout my stay at IIT Madras.

My friends and coresearchers at IIT Madras have helped me in many ways with my frequent struggle with various aspects of computer science and speech recognition. I feel that culmination of the present work is to a large extent due to unfailing support given by Mr.C.Chandra Sekhar, Mrs.Hema A. Murthy, Mr.G.V.Ramana Rao and Mr.R.Sundar. I extend my sincere thanks to them. I extend my thanks to Mr.K.V.Madhu Murthy, Mr.C.P.Mariadassou Mr.M.Prakash, Mr.S.R.Rajesh Kumar and Mr.N.Alwar for the excellent discussions I had with them in matters that are technical and not so technical too. Initially Mr.Sunil Kumar Gupta and I have spent number of sleepless nights in developing the expert system modules. I sincerely acknowledge his help.

This work needed significant inputs from expert acoustic-phonetician in Indian Languages. In this respect I had the good fortune to interact with Dr.(Mrs.)K.Nagamma Reddy of the Centre for Advanced study in Linguistics, Osmania University, Hyderabad, India. I thank her for the patience and interest she had shown in imparting the knowledge which formed the core of the current research effort.

The experimentation and preparation of final thesis should have taken a very long time but for the unstinted support extended by many colleagues in the speech research laboratory in the department. In particular, I would like to record my gratitude to Mr.S.Rajendran, Mr.R.Ramaseshan, Mr.A.S.Madhu Kumar, Mr.A.Ravichandran and Mr.J.Saikumar whc have helped me in the preparation of this thesis. I also record with gratitude the help rendered by Mr.C.Muthuvelu, Mr.Venkatasubramaniam, Mr.A.Rajasekhar and Mr.A.Arul for their help during the preparation of the manuscript.

Finally, I would like to thank my family members for bearing with me and my nocturnal visits home during the period of this research work.

CONTENTS

CHAPTER 2

CHOICE OF UNITS FOR SIGNALTO-SYMBOL TRANSFORMATION 26

CHAPTER3

ACOUSTIC-PHONETIC KNOWLEDGE FOR CHARACTER SPOTTING 36

CHAPTER 4

CHAPTER 5

ABSTRACT

of the thesis on

A Rule-Based Approach for Spotting Characters from Continuous Speech in Indian Languages

Machine recognition of continuous speech involves transforming continuous speech signal into a discrete set of symbols each describing a meaningful speech sound. The objective of this research is to address issues involved in the development of a speech-to-text system for an Indian language. The basic idea is to exploit the nature of the Indian languages for capturing the phonetic information in the speech signal in a symbolic form. A knowledge-based approach for spotting the characters of a language in continuous speech is proposed. The knowledge consists primarily of acoustic-phonetics of speech sounds.

Use of knowledge for character spotting in continuous speech in the Indian language, Hindi is discussed. We first discuss reasons for choosing character as a unit for signal-to-symbol transformation. Acoustic-phonetics of speech in Hindi describe the sounds in a systematic manner in terms of articulatory movements. The manifestation of this acoustic-phonetic knowledge in speech signal is studied with the help of a knowledge expert and speech data analysis. The acoustic-phonetic knowledge for each character of Hindi is then represented in the form of production rules. A significant feature of this knowledge-based spotting approach is that processing of speech signal is done according to description of a given character.

A rule-based implementation of knowledge-based approach for character spotting in continuous speech in Hindi is discussed. While the total number of characters including consonant clusters is estimated to be around 5000, only a subset of about 350 characters consisting of Vowels(V) and Consonant-Vowel(CV) combinations are considered for implementation. The acoustic-phonetic knowledge for all the characters is represented in the form of production rules. Character spotting systems are implemented for each character separately. Inaccuracies in processing the speech signal are represented by assigning confidence measures at every stage of spotting. Fuzzy mathematical concepts are used to relate the character to signal parameters. We demonstrate the flexibility of the system to provide better performance as more knowledge is available for spotting each character, without significantly increasing the overall complexity.

The main contributions of this thesis are:

1) Choice of characters as a symbol
2) Acquisition and representation of acoustic-phonetic knowledge for characters in Hindi
3) A rule-based system implementation of character spotting
4) Use of Fuzzy mathematical concepts to relate characters to signal parameters
5) Performance evaluation of character spotting system.

CHAPTER - I

INTRODUCTION TO THE PROBLEM OF SPEECH RECOGNITION

11 Background to the Problem of Speech Recognition

Speech recognition involves transforming input speech into a sequence of units called symbols and converting the symbol sequence into a text corresponding to the message in the speech signal. The goal of our research is to provide a machine with speech input facility to take dictation of continuous speech. The emphasis is on spotting the characters (symbols) of Hindi, an Indian Language, from continuous speech using knowledge-based approach. The main contributions of this thesis are: (1) Choice of character as a symbol (2) Acquisition and representation of acoustic-phonetic knowledge for characters in Hindi (3) A rule-based system implementation of character spotting (4) Use of fuzzy mathematical concepts to relate characters to signal parameters (5) Performance evaluation of character spotting systems.

Speech recognition systems vary from simple isolated word recognition systems to highly complex continuous speech recognition systems. Continuous speech being the natural mode of human communication, recognition of continuous speech is the ultimate goal in any speech recognition research. Typically, continuous speech recognition is performed in two stages as shown in Fig.1.1. They are: (1) Speech signal-to-symbol transformation and (2) Symbol-to-text conversion. In the signal-to-symbol transformation stage, the input speech is

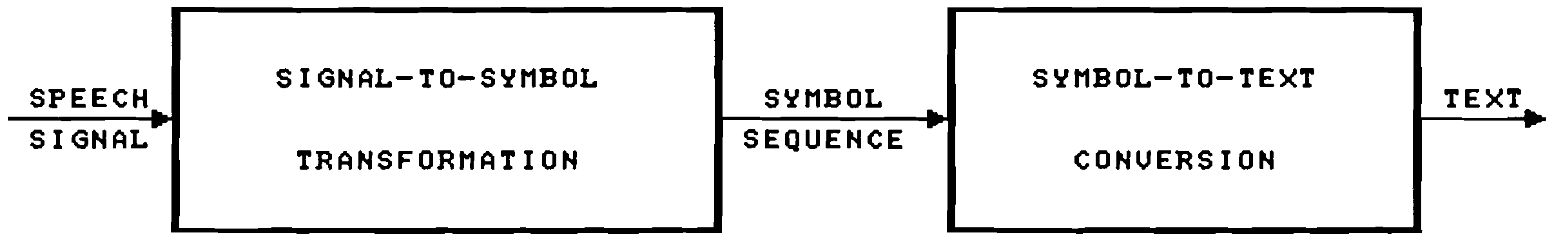


Fig. 1.1. Block diagram of a speech-to-text conversion system

converted into a symbol sequence. Usually these symbols represent some speech units. The symbol sequence is then converted into a meaningful text by the symbol-to-text conversion stage using lexical, syntactic and semantic knowledge sources. In most of the systems the symbols are extracted from the speech signal using a fixed set of parameters, like spectral coefficients. The difficulty in such systems is that any loss of information at the signal-to-symbol transformation stage has to be compensated by the higher level knowledge sources. Another major problem, especially for languages like English, is that, the text consisting of words and sentences has to be expressed as sequences of symbols for all possible pronunciations. Since a given symbol representing a speech unit may be mapped onto different characters (or strings of characters) depending on the context, the symbol-to-text conversion becomes very complex. Moreover, a lot of manual effort is required to generate the pronunciation dictionary in terms of the symbols for new ,vocabularies and tasks. Our approach, in tune with recent trends [106,23,42] is to focus attention on signal-to-symbol transformation stage in order to capture as much information from the signal as possible. The issues involved here are: (1) Choice of the symbols and (2) Methods of processing to be done on the speech signal to derive these symbols. When characters which correspond to unique pronunciation are adopted as symbols, the task of symbol-to-text conversion becomes trivial. But the burden of speech recognition then falls on the signal-to-symbol transformation stage.

The main objective of this thesis is to discuss the issues

Hindi [illegible] ?

in the design of a speech signal-to-symbol transformation module for a speech-to-text conversion system for the Indian language Hindi. The idea is to use to the maximum possible extent, the constraints of speech production and perception and also of the language at the signal level itself, in order to capture the speech information in the signal in a symbolic form. We propose characters of Hindi as symbols, and spotting the characters in continuous speech as the basic approach for signal-to-symbol transformation. A knowledge-based system is proposed for spotting each character.

In this chapter we present a brief review of the attempts being made to realize speech recognition by machine. First, we discuss in the next section various classes of speech recognition systems and also bring out the distinction between speech understanding and recognition. In Section 1.3 we discuss in some detail, the approaches adopted to realize continuous speech recognition. This discussion demonstrates the importance of the acoustic-phonetic block in a speech recognition system. Attempts on developing phonetic engines ? are discussed in Section 1.4. The scope of the current research which focuses primarily on the acoustic-phonetic block of a speech recognition system is given in Section 1.5. This section also dwells briefly on the characteristics of Indian languages relevant for the ideas to be proposed in this thesis. The specific issues of the acoustic-phonetic knowledge of Indian languages are discussed in Chapter III. Since rule-based approach is used to implement character spotting, we present in Section 1.6 a brief review of knowledge-based approach, especially expert systems, for problem solving. Finally, we give an outline of the thesis in Section 1.7 where we discuss the organization of the following chapters in this thesis.

1.2 Classes of Speech Recognition Systems

A number of speech recognition systems [60] were developed in the past twenty years. They can be broadly classified into three categories: (1) Isolated Word Recognition (IWR) systems where words are separated by pauses, (2) Connected Word Recognition (CWR) systems where the basic units are still words but there is no pause between words and (3) Continuous Speech Recognition(CSR) systems where the basic units of recognition are smaller than words.

Most of the IWR systems are speaker dependent and have a limited vocabulary. They consist of two phases: a training phase and a recognition phase. The speaker was first made to utter a word and the corresponding signal is processed and the information is stored as a template. During the recognition phase, the same type of information is obtained from the input utterance and this information is compared with the templates prestored in the memory obtained in the training phase. The match obtained with least distance is labeled as the recognized word. The IWR systems do not have the problems of segmentation and contextual effects. The reference and test patterns may be represented using a spectral filter bank output values [18] or using linear prediction coefficients [70,91] or cepstral coefficients [39] or parameters based on group delay processing [68,101,25] or sometimes by parameters based on auditory models

[35,44,67]. A comparison on the use of some of these parameters in speech recognition is given in [17]. In most of these cases, because of intraspeaker variability, there may not be time alignment between the test pattern and reference template. So some kind of time warping is used to align the templates. IWR systems perform well for speaker dependent and restricted tasks with a limited vocabulary. ok

In connected word recognition (CWR) systems the restriction of pausing between words is removed. However, the speaker is still constrained to speak in a careful manner to minimize the coarticulation effects at word boundaries. In these systems, a set of reference patterns are stored for the words in the vocabulary of the task. Generally the units are isolated words. The connected word pattern is matched to a modified sequence of isolated word patterns taking into consideration the context in which the word occurs and the best matching word pattern is hypothesized as the spoken sentence. Dynamic Time Warping (DTW) algorithms which use dynamic programming to perform time warping are used to provide optimum alignment between the spoken input and the sequence of modified reference word patterns. Various types of dynamic programming (DP) techniques are applied to time warping, like two-level DP matching [87], stochastic DTW [76], one pass DP [7] and embedded training [79]. A Number of techniques, like multiple reference patterns, have been used for speaker independent recognition [81]. To increase the .vocabulary size vector quantization techniques [34] have been proposed. Vector quantization code books give the system the flexibility of speaker independence [92] and large vocabulary [50].

The techniques used for IWR and CWR systems cannot easily be extended to continuous speech recognition (CSR) systems. The main difficulty is with the large number of words that are to be recognized. Also the effects of ~~articulatioat word junctures make the matching process error prone. So, instead of using DTW methods, segmentation and labeling schemes are used for continuous speech recognition. The utterance is normally transcribed in terms of subword units [98,8] like phones[93], phonemes [86], diphones [15,89,90], syllables [73,95,96] and demi-syllables [85]. It is also possible to spot the subword units rather than dividing the utterance into segments and then labeling the segments [88]. Segmentation & labeling → spotting

Speech recognition systems can also be classified by the task they perform. In this, there are two types of systems: (1) speech-to-text systems and (2) speech understanding systems. A speech-to-text system converts input speech into corresponding text, whereas a speech understanding system tries to capture the message in the speech and respond to it. In literature, the word "speech recognition system" is used to denote both these systems. However, in this thesis, when we refer to our system as a speech recognition system, we mean a speech-to-text conversion system and not a speech understanding system.

Although many classes of systems are available, the preferred mode of human-machine communication will be through continuous speech mainly because human beings do not pause between successive words even in such highly restricted tasks such as reading out telephone numbers. We are interested only in continuous speech recognition systems in this thesis.

1 3 Continuous Speech Recognition Systems

A number of CSR systems have been developed since 1960. A major thrust towards this goal was received from ARPA-SUR [53] project in USA in the seventies, the Alvey [104] and ESPIRIT [30] projects in Europe and the fifth generation computing systems project in Japan in eighties. The issues involved in the design of CSR systems are speaker independence, contextual effects in continuous speech, choice of appropriate unit for recognition [93], extraction of relevant parameters and methods of processing the speech signal. Some of the difficulties in designing these systems are: (1) the absence of clear boundaries between speech units in a word or between words, (2) the effect of anticipatory coarticulation which is difficult to model in continuous speech, (3) large-variations in speech sounds uttered by different speakers and sometimes even when the same speaker repeats the same sentence and (4) problem of obtaining precise rules to formalize the relation between the signal parameters and the corresponding symbolic representation due to fuzzy nature of quantities involved. A number of systems were successfully developed for incorporating knowledge sources at various levels to compensate the errors caused by the varying nature of the input speech signal, its extracted parameters and the linguistic knowledge [61].

1.3.1 Brief History of Developed Systems

Reddy [82] demonstrated an initial capability in connected word sequence recognition using a 16 word vocabulary. Four major speech understanding systems which accept continuous speech were developed during ARPA-SUR project [54] in seventies. They were (1)SDC system [46], (2)BBN's HWIM [45], (3) CMU's Hearsay-I1 [29] and (4) CMU's Harpy [9]. All these systems had a limited vocabulary and were designed for a specific task.

Harpy, one of the successful systems during the ARPA-SUR [53] project adopted an integrated network approach for knowledge representation. In this various knowledge sources such as acoustic-phonetics, lexicon and syntax are integrated into a finite state network. To each state in this network is associated a phone template. The network represents allophonic variations of the phonemes occurring in the sentences. The interpretation of an input utterance consists of finding a path through the network with maximum likelihood according to the phone transcription of an utterance. In order to reduce the complexity, a beam search was used.

The fpcus in the development of Hearsay-I1 [29] was to design a framework for experimenting with the representation and cooperation of diverse knowledge sources such as acoustic-phonetic, lexical, syntactic, semantic and pragmatic knowledge sources. The knowledge sources are independent but they are required to cooperate in hypothesizing and in correcting other hypotheses. This cooperation is implemented in a global data structure called blackboard. .The blackboard is partitioned into several levels and each level holds a different representative problem space.

As against this, the BBN's HWIM [45] philosophy is based on perceptual process. It includes an acoustic-phonetic recognizer block, a lexical retrieval block, a word hypothesization block and a syntax block. The HWIM system follows an earlier system SPEECHLIS [99] developed by the same group. The contribution from HWIM system is the integration of syntax and semantics through the design of a parser capable of producing the ~ ~ ~ tree of a sentence and its semantic interpretation. The SDC system follows a strategy that depends on verification of syllables obtained from processing the speech signal using higher level knowledge sources with a uniquely designed mapper that maps the syllables to words.

At the same time when ARPA-SUR project was being undertaken IBM developed a number of CSR systems. The system ARCS [60] uses a hierarchical structure based on segmenting speech into some transitional units and using linguistic information to transcribe the continuous speech. Other systems based on centisecond model and phone model were developed by Jelinek [47] and Bahl [3].

In recent times the importance of using knowledge at the acoustic-phonetic level [106,97] was realized. A few systems have been proposed where the focus was on using the knowledge at the acoustic-phonetic level [13]. In one system by De Mori [20], a set of cooperating expert systems, called expert society, is used for continuous speech recognition based on syllable hypothesization before phoneme recognition is done. It uses a simple structure where acoustic-phonetic and lexical experts are organized hierarchically and communication through them is achieved using a message passing network. Another system proposed by Regel [83] stresses the need for an acoustic-phonetic module which uses vast amount of speech knowledge. The system uses phones as symbols for signal-to-symbol transformation. IBM [48]

TA1760

also announced a 20,000 word speaker dependent system. BYBLOS [16], a continuous speech recognition system developed by BBN, used speech specific knowledge like coarticulation effects to develop context dependent models of phonemes using hidden Markov models [80] which are used by the higher level knowledge sources for speech recognition. REMORA [72] is a speech recognition system aimed at developing the acoustic-phonetic decoding using speech knowledge and expert systems concepts. APHODEX [36] is a continuous speech recognition system developed in France which uses expert system concepts in acoustic-phonetic decoding. SUMMIT [107] system developed at MIT uses knowledge based approach in arriving at a phoneme lattice using multilevel representation of spectrograms. ANGEL [71] is an another system developed at CMU that uses location and classification rules to derive a phoneme lattice from speech signal. SPHINX [62] system developed at CMU uses hidden Markov models to get triphone and function word model for continuous speech recognition.

In the framework of Alvey initiative, VODIS [104], a DTW based continuous speech recognizer was developed. In the same framework experiments on continuous speech recognition systems are being carried out at the Center for Speech Technology and Research in Edinburgh.

In Japan, the work on continuous speech recognition [40] focused on extracting suitable parameters from the speech signal and also developing a suitable language model [55]. At present in the framework of Fifth Generation Computing Systems (FGCS) project work at various universities and research institutes like ATR Interpreting Labs and NTT Human Interface Labs, continuous speech recognition systems that use dynamic programming concepts [94], hidden Markov models [52,51], expert systems [74,41] and artificial neural networks [56,88] are being developed.

The above mentioned systems followed various approaches for recognizing continuous speech. The next section explains some of those approaches.

1.3.2 Techniques used in Continuous Speech Recognition

The systems described earlier used a variety of approaches for recognizing continuous speech. Most of them used a segmentor to divide the signal into phonetic units and a classifier to recognize the individual phonetic segments. Much of the variations in continuous speech due to context, rate of speech are taken care of in the contextual phonological rules used in the labeling process. The sequences of phonetic segments are then processed using a language model. The language model consists of various knowledge sources like lexicon, syntax and semantics. The rules which describe the lexical constraints form the lexical knowledge. Rules in the language which provide information on word order and other grammar rules form the syntactic knowledge base. The subject verb agreement rules and other rules dealing with the meaning of the sentence form the semantic knowledge. These systems were developed keeping in mind the necessity of using speech and linguistic knowledge in one form or other. The techniques used by these systems can be classified as (1) Artificial Intelligence techniques that use pattern recognition and Search, (2) Expert systems, (3) Hidden Markov Modeling(HMM) and (4) Artificial Neural Networks(ANN).

1.3.2.1 Artificial Intelligence Techniques

Methods based on Artificial Intelligence were initiated in 1971 in the frame work of ARPA-SUR [54] project. Here the speech recognition problem was formulated as a search problem and higher level knowledge sources were used to reduce the search effort. Continuous speech recognition is achieved using an acoustic processor followed by a language processor. The acoustic processor acts on the speech signal to produce a phonetic transcription of what has been said. In this stage of processing some fixed parameters like formant frequencies and other parameters are extracted from the digital waveform which are used to derive a phonetic segment lattice. The next step in the recognition process is to find candidate words and word sequences from the phonetic segment lattice using lexical, syntactic, semantic and pragmatic knowledge. This is done by the language processor which searches the phonetic segment lattice for good matching words that can be used as seeds to build longer partial sentences. Thus the output of the language processor is a sentence that satisfies all the constraints imposed by the various knowledge sources. The important contribution to speech recognition in this framework is the extensive use of heuristic search techniques to search through the phonetic segment lattice. Various system configurations came into existence based on how the higher level knowledge sources are modeled and are allowed to communicate with each other. Some of the models are the hierarchical model, blackboard model, integrated network model and locus model [33]. Activation of the knowledge sources is governed by heuristic search techniques and some of the search schemes used are depth first search, breadth first search, probabilistic search and best few (beam)search.

1.3.2.2 Expert System Techniques

In the early eighties, it was realized that knowledge of a task domain can be explicitly obtained and represented. Thus expert system or knowledge-based approach gained prominence in artificial intelligence. The idea in an expert system is that the knowledge available with a human expert can be extracted and used by the system which then performs just like a human expert. Also the knowledge is separated from the control strategy or reasoning mechanism. Knowledge is represented in different forms with provision for assigning confidence to the conclusion arrived at by using the knowledge base. This approach implies that knowledge has to be manually extracted and entered in the system unless some automatic learning procedure is incorporated. Since in most of these systems knowledge is separated from the inference mechanism, updating the knowledge is a simple task. A number of systems have been developed which use the concepts of expert society where a number knowledge sources in the form of expert systems interact mostly in a hierarchical fashion. Expert systems were developed for speech recognition based on the knowledge obtained from spectrogram reading and from various specialists in phonetics and languages.

1.3.2.3 Hidden Markov Model Approach

Soon it was realized that it is difficult to obtain explicit knowledge and update the knowledge using expert system approach for complex tasks like speech recognition, especially for speaker independent and task independent systems. If a large number of training samples are available, then powerful statistical models can be used to accomplish speech recognition. Hidden Markov Model(HMM) is one such statistical technique which is currently used for modeling continuous speech.

As against the DTW approach, where a reference was represented by the pattern itself, the hidden Markov model uses a model to represent the reference. Levinson [64] describes a hidden Markov model as follows: A probabilistic function of a (hidden) Markov chain is a stochastic process generated by two interrelated mechanisms. One is an underlying Markov chain having a number of states and the second one is a set of random functions one of which is associated with each state. At discrete instants of time the process is assumed to be in a unique state and an observation is generated by the random function corresponding to the current state. The underlying Markov chain then changes state according to its transitional probability matrix. The observer sees only the output of the random functions associated with each state and cannot directly observe the states of the underlying Markov chain and hence the term hidden Markov model.

There are distinct advantages of this type of statistical model for speech patterns. It can model the temporal variability of speech data. It can also capture the variability of speech signal in an analysis vector. Moreover there is an established technique for training the model namely the Baum-Welch algorithm. In the experiments described by Levinson [65], which used HMM with vector quantization, the recognition accuracy was comparable to that of DTW recognizers. It is further observed that DTW recognizers have a very simple training phase and a very complicated recognition phase, whereas HMM recognizers have a complicated training phase and a simple recognition phase. To be recognized, the input is compared to a reference model. Hidden Markov Models have been used for different types of system configurations [26].

1.3.24 Artificial Neural Networks Approach

Artificial neural networks models attempt to achieve real time response and human-like performance using many simple processing elements operating in parallel as in biological neural network system. These models have great potential for speech recognition, where many hypotheses are to be pursued in parallel and high computation rates are required. Performance of the current systems are far below the performance of humans [63]. Processing elements or nodes in a neural network are connected with links with variable weights. The simplest node sums a large number of weighted inputs and passes the result through a nonlinearity. A node is characterized by an internal offset or a threshold and by the type of nonlinearity used. The three types of nonlinearity used are hard limiters, threshold logic elements and sigmoidal nonlinearity. More complex nodes may include time dependencies and operations other than simple summation. Neural network models are specified by their topology, node characteristics, and training or learning rules used. These rules specify how weights have to be adapted during use to improve performance. A number of systems based on this approach are being developed [43,84,75,88].

1.3.3 Difficulties with the Existing Systems

While we have said that HMMs and artificial neural networks show good promise, many scientists argue that only by integrating an expert's knowledge with these models one can attain a high recognition performance. In SPHINX, the most successful speech recognition system built so far, it is the integration of speech knowledge with the hidden Markov model formalism that gave it a high recognition rate. Hence the importance of knowledge in continuous speech recognition cannot be discounted.

In most of the speech recognition systems built so far, the speech signal is first segmented into some units which are then labeled. The labeling was done upon subword units mostly based on phonetic characteristics. These systems do not have high accuracy because reliable extraction of phonetic features from parameters of the speech signal is difficult. It is due to the fuzzy nature of relations between features and parameters. Moreover, the signal representation in all these cases was based on fixed parameters irrespective of the phonetic nature of the segment. Hence recent systems are focusing on capturing the phonetic information in the signal in symbolic form using acoustic-phonetic knowledge and processing the signal in a knowledge directed manner.

1.4 Acoustic-Phonetic Block in Speech Recognition: Phonetic Engine

The trend in the last few years is to develop a signal-to-symbol transformation module which captures most of the phonetic information in the speech signal in symbolic form. They use extensively the relevant acoustic-phonetic knowledge to decode the speech signal into phonetic units. This enables one to develop systems for large vocabularies in a speaker independent mode. These high performance front-ends for speech recognition systems are called phonetic engines [69]. Based on the symbols that are being used in the signal-to-symbol transformation, various kinds of phonetic engines are being developed. Most of the earlier systems used some kind of template matching. But the present day phonetic engines use sophisticated knowledge-based techniques for signal-to-.symbol transformation. The phonetic engine concepts are extensively used by De Mori [21], Cole [14] and Haton [42].

(learning) pattern matching → [illegible]
||
connectionist [illegible]

1.5 Scope of the Current Research

Most of the continuous speech recognition systems developed so far were intended to respond to a query in a voice input mode. The systems tend to interpret the input based on the stored knowledge about the vocabulary of the task. They exploit the redundancies of the language and context to correct the errors in the symbol string generated from input speech. Hence these systems perform more of speech understanding than speech recognition. Thus the performance in terms of percentage accuracy of words or sentences is not very relevant. Indeed understanding is a much more complex task than mere recognition, since all the higher levels of knowledge including pragmatics have to be represented and used.

We propose to undertake a well defined task which consists of recognizing the input speech to generate the corresponding text. This is like taking dictation, where the output is unique in the sense that the machine is to reproduce the spoken text without giving any interpretation. The output text is meant for human use and hence a few errors are tolerable. Humans seem to perform this task of taking dictation in an effortless manner even though the subject matter is not well understood and the vocabulary is not completely known.

1.5.1 Speech-to-Text Conversion Systems for Indian Languages

Our main objective is to address the issues relating to the development of a signal-to-symbol transformation system for Indian languages [102]. Ultimately we are interested in developing a speech-to-text system that is speaker independent, vocabulary independent and task independent. Accuracy in the output text is not very important as long as the text is understood by the human user. If necessary, the output text may further be corrected by the user manually.

The organization of our speech-to-text conversion system is shown in Fig.1.2. It consists of four modules: the acoustic-phonetic expert, the lexical expert, the syntactic expert and the semantic expert, which are connected in a hierarchical fashion. The acoustic-phonetic expert converts the input speech signal into a symbol sequence using primarily the acoustic-phonetic knowledge. The lexical expert converts this symbol sequence into a sequence of words using primarily the lexical knowledge. The syntactic expert converts the word sequence into a sequence of

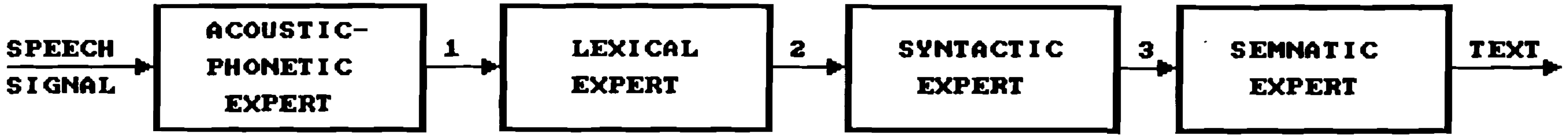


Figure 1..2. Heirarchical model of speech-to-text conversion system

1. Symbol sequence   2.Words   3. Sentences

phrases and clauses forming a sentence using primarily the syntactic knowledge and finally, the semantic expert selects only meaningful sentences to form the output text. Each module may use any type of knowledge it needs to perform its specified function. For example, the acoustic-phonetic expert may use not only acoustic-phonetic knowledge but also any other knowledge needed for signal-to-symbol transformation.

The focus of the present work is on the acoustic-phonetic expert module. The key point is the selection of a symbol and its recognition in continuous speech. In the following subsection we describe some of the characteristics of Indian languages that are exploited in the selection of a symbol for the signal-to-symbol transformation.

1.5.2 *Characteristics of Indian Languages*

We need symbols which can capture all the phonetic variations in speech and at the same time they can be converted easily into a text form [1]. For this purpose we exploit the phonetic nature of Indian languages. By this we mean that in Indian languages like Hindi "we generally speak what we write and we write what we speak". That is why we propose the characters of Hindi as symbols for signal-to-symbol transformation. The text to be obtained in this case will be simply a concatenation of the written characters derived from the speech signal. Thus the complex process of creating a pronunciation dictionary for words is avoided.

case of Hiragana in Japan

Some of the advantages of using characters as symbols are :

(1) The set of characters captures all the permissible

[illegible] SC,V

combinations of consonants and vowels in the language.

(2) The number of characters in Hindi is finite and it is much smaller than the number of words. The number of characters is around 5000.

(3) Characters have a unique relationship to the articulatory description and also to the symbolic representation to which an utterance has to be ultimately converted.

[illegible]

(4) Description of a character captures the necessary coarticulation information relevant for its recognition.

(5) The character set is phonetically structured.

Some of the disadvantages of the character set are :

(1) They are still too large in number to deal with the existing systems.

(2) There are significant coarticulation effects of one character over the adjacent character.

1.5.3 Signal-to-Symbol Transformation for Indian Languages

one para. [illegible]

To recognize a character from speech signal two approaches can be used, namely, (1) segmentation and labeling and (2) spotting. We use spotting because the unit is well defined and contains most of the required information for deriving it from a speech signal. Character spotting requires that the information contained in the speech signal be related to the description of the character. We propose a knowledge-based approach to spotting characters from the speech signal.

Although there are nearly 5000 characters in the language, we consider a subset of about 350 characters which occur about 90% of the time. To spot a character in speech, a description of the character is provided in terms of the speech production mechanism and the parameters of the speech sign1 The description represents the knowledge of the character and this knowledge is used to spot the character in continuous speech. For spotting, a rule-based system is proposed·$0 be used because the rules can also be used to determine the· order in which the signal is to be processed in order to detect the presence of the character. To implement the proposed knowledge-based approach for character spotting, we need the acoustic-phonetic knowledge of the characters of Indian languages. This will be discussed in detail in Chapter III.

1.6 Expert Systems in Automatic Speech Recognition

Automatic speech recognition is characterized by a close interaction between a low level processing, i.e., acoustic-phonetic processing and high level interpretation. Knowledge engineering techniques are helpful at both these levels. expert systems are particularly suitable at both levels because we have human experts who possess explicit knowledge about the domain. They also provide general framework for the architecture of the recognition system, Further, expert systems for speech recognition are particularly suitable because here the problem knowledge is distributed at various levels and an expert system for each level leads to modularity and flexibility in the design of speech recognition systems [106,49,11,22].

The acoustic-phonetic decoding or speech signal-to-character transformation in a speech recognition system constitutes a major bottleneck in the design of practical speech recognition systems especially in multispeaker environments. Prototype-based approaches, like classical pattern matching as explained in the previous chapter, have proved to be insufficient in order to obtain high accuracy. On the other hand a rule-based approach for recognition yields good results but cannot take into account the large amount of training data. This can be done by capturing the expertise accumulated over the years by phonetician and speech scientist, especially in reading spectrograms and other types of parametric representation of character. The .strong motivation for using the expertise can be listed as follows:

(1) Availability of expertise.

(2) Expertise can be formalized by a set of rules.

(3) Expertise is independent of vocabulary.

The hierarchical model proposed for continuous speech recognition system consists of different experts, the first one of this being an acoustic-phonetic expert. A good acoustic-phonetic expert makes the others redundant because the character sequence corresponding to the input utterance is available at the output of the acoustic-phonetic block itself.

The expert system approach enables us to use the knowledge sources a human being uses to perform the task of signal-to-character transformation. In our case the knowledge required for signal-to-symbol transformation is acquired from a trained acoustic-phonetic expert who will be able to spell clearly how the knowledge is used in interpreting speech patterns. This knowledge is used as the rule base for the expert system.

1.7 Organization of the Thesis

The main focus in this thesis is to demonstrate the potential of character spotting approach for speech signal-to-symbol transformation for the Indian language, Hindi. The thesis is organized as follows:

We discuss in the next Chapter considerations in the choice of characters as symbols for signal-to-symbol transformation for continuous speech in Hindi. We also discuss the advantages of character spotting approach and the need for acoustic-phonetic knowledge to implement the approach. Issues in the acquisition of acoustic-phonetic knowledge for Hindi characters and the sources of the knowledge are discussed in detail in Chapter 111. The procedure for acquiring knowledge for each character and representation of this knowledge in the form of rules are described in Chapter IV. The procedure is illustrated in detail for one character. In Chapter V we give implementation details of rule-based expert systems for character spotting and discuss the use of mathematical concepts for interpreting the results of activation of the rules in a knowledge base. In Chapter VI the performance of the character spotting experts are studied through several experiments on groups of characters. The importance of tuning the rule base and entries in the fuzzy table on the performance of the system is emphasized. Finally the results of signal-to-symbol transformation is demonstrated for two utterances. Chapter VII gives a brief summary of the thesis work and discusses issues for further study.

CHAPTER-//

CHOICE OF UNITS FOR SIGNALTO-SYMBOL TRANSFORMATION

The first task in continuous speech recognition is to transform the input speech signal into some symbolic form, which can then be interpreted to determine the corresponding text using higher level knowledge sources such as lexical, syntactic and semantic knowledge. Consideration in the choice of units for symbolic representation for a speech-to-text conversion system for Indian languages form the main subject matter of this chapter. The need for identifying such units and the characteristics these units should possess, are discussed in Section 2.1. Different types of units used in CSR systems together with their relative advantages and disadvantages are presented in Section 2.2. That section also discusses problems with some of these units. In Section 2.3 we discuss phonetic characteristics of alphabet in Indian languages by taking Hindi as an example. This discussion demonstrates the significance of character of the language as a unit. In Section 2.4 the problem of recognizing a character in continuous speech is addressed and the advantages in using a character spotting approach are described. This section also explains the importance of using the acoustic-phonetic knowledge of Hindi character for implementing the character spotting. Details of the relevant acoustic-phonetic knowledge, its acquisition and representation are given in the Chapter III.

2.1 Need for Units/Symbols in Continuous Speech Recognition

The main purpose of symbolic representation of speech signal is to be able to use the higher level knowledge sources to interpret the text corresponding to the speech signal. The units/symbols should represent the phonetic characteristics in the input speech signal, since these characteristics uniquely describe the speech production process which in turn is unique for a given message/text. The significance of proper choice of these units can be seen from the limitations of the system already developed. In most systems where speech signal is represented by units corresponding to a fixed set of parameters, the parameters dictate the amount of phonetic information captured by the symbol sequence. For many systems the recognition accuracy of units corresponding to dynamic sounds such as stops is very poor [100]. Loss of crucial information cannot easily be compensated by the redundancies in the higher level knowledge sources. Moreover these knowledge sources are also incomplete in most cases.

2.2 Types of Units

Speech signal can be represented in terms of units which are derived from the spectral values of a fixed size (10 to 20 msec) segment of speech. In this, each segment of speech is represented by a N-dimensional vector of spectral values. The value of N is typically in the range 16 to 32. The vector space is quantized into a convenient number of levels (typically 100 to 1000). In this case the number of distinct levels form the number of symbols. The main advantage of this scheme is that processing the speech signal is straight forward and no knowledge is used in the processing. However, the symbols are arbitrary and have no relation to the phonetic characteristics of speech sounds. Hence it is extremely difficult to relate the symbol sequence to a text consisting of a sequence of words.

A slightly better characterization of speech signal is through the use of units representing acoustically uniform segments called phones as used in [93]. In this, each segment is represented by the spectral value of the central frame of data (10 to 20 msec) in the segment. There are about 100 such units/symbols used in the systems developed at CMU in 70's [54]. To successfully use this representation, it is necessary to express all possible sequences of utterances in terms of these units, which requires a lot of manual effort for a given task.

Phoneme is a common linguistic unit used to describe speech. There are typically 40 phonemes for a language like English. However phoneme is only an abstraction, and in actual speech it represents several different sounds called allophones. The allophones represent the contextual variations of a phoneme. Utterances are expressed in terms of sequences of phonemes. While the number of phonemes is small and it is relatively easy to describe a text in terms of them, it is very difficult to identify the phonemes in continuous speech. Segmentation and labeling of speech signal becomes a nontrivial task in this case. Recognition accuracy will be very poor, usually in the range 20 % to 40% [61].

Inaccuracies in phoneme labeling occur due to the difficulties in segmentation at the phoneme transitions.

Diphone, which represents transitions of a consonant-vowel sequence, solves the problem of segmentation, because it is relatively easy to determine the central region of a consonant and central region of a vowel. Diphones also capture the allophonic variations caused by coarticulation. But the number of diphones is excessively large. Moreover it is difficult to apply phonological rules on diphones to derive diphone sequences for sentences.

A large unit like syllable is relatively easier to locate and identify in continuous speech. But it is difficult to define precisely all syllables. Also the number of syllables in a language is quite large [73].

Word as a unit, eliminates many problems of segmentation and labeling caused by variability of pronunciation, coarticulation etc. But the number of words in a ianguage is so large that it is impractical to consider it as a unit except for limited vocabulary tasks.

As we have seen, there are tradeoffs in the choice of units for representing the speech signal. The larger the unit, the more are the number of units. Thus at the lower end there are relatively a few phonemes but as we go towards the larger end to diphone, syllable or word levels, the number of units become excessively large. From the point of view of processing complexity also, the identification of a phoneme in continuous speech is difficult because the relevant information for identification of the phoneme is not present in itself due to coarticulation. Further when composing phonemes to form larger units, a great deal of composition rules are necessary because rules have to be used for specifying transition between phonemes and their boundaries. Whereas if the unit is large, then the above difficulties can be overcome in the sense that most of the speech sounds in the units can be easily recognized. This is because the coarticulation information is captured in the unit itself. Moreover the composition rules are simple, since these units are not greatly affected by the environment in which they occur. It ~~it~~ is desirable that there be a natural relationship between the speech production model, the speech signal, the intermediate representation and the ultimate symbol to which the signal has to be converted. If this is the criterion, then most of the units like allophones, diphones and syllables do not have any relationship with the ultimate unit of representation which is normally the orthographic form for the spoken utterance. In most of these cases a pronunciation dictionary is necessary to relate these abstract units to the orthographic characters. It is therefore necessary to have a realistic unit with the following characteristics: (1) The unit is large enough to make detection of such units possible with comparatively easy processing methods, (2) The number of units should not be too large and (3) The units should be able to represent all utterances in the language.

In the next section we discuss the choice of symbol for Indian languages taking these factors into consideration.

23 Choice of Units for Speech Recognition in Indian Languages

In this section we explain our choice of symbol for signal-to-symbol transformation in Indian languages, particularly for

Hindi. We propose the characters of the language, Hindi, as units for the speech signal-to-symbol transformation. The main reason for this choice is the unique relationship between the spoken utterance and the written character which is due to the phonetic nature of the language. The characters are mostly consonants, vowels and combinations of consonants (C) and vowels (V) in the forms CV, CCV, and CCCV. Identification of the segments in the spoken utterance of a character, particularly the stop consonants, is made easier because we already know the acoustic/phonetic environment in which these occur. Since the unit is large, most of the coarticulation information is also captured in the unit itself. Yet, influence of one character over the other exists, and is to be taken into account. But this occurs in a few cases only. Also, the composition of these characters to form higher level representation units like words is a trivial task.

The alphabet for Hindi is given in Fig.2.i. It can be observed that the alphabet are structured according to the place and manner of articulation. In the structure of the stop consonants like /ka/(क), /ca/(च), a), /ta/(त) and /pa/(प), the movement of the articulators is from the velar end to the radiation end. Detailed articulatory description of these is discussed in the next chapter.

The number of distinct units in Hindi is not excessively large. Typically for Hindi the number including all consonant clusters is about 5000. This is calculated as follows. There are totally 33 consonants, 356 CC clusters, 77 CCC clusters and one CCCC cluster. In this calculation we considered only the valid

330
3560
770

Set of characters representing vowel speech sounds

/a/(अ) /a:/(आ) /i/(इ) /i:/(ई)
/u/(उ) /u:/(ऊ) /e/(ए) /ai/(ऐ)
/o/(ओ) /au/(औ)

Set of characters representing CU coabinations uhere C is any plosive and U is /a/ (अ)

PLACE OF ARTICULATION	MANNER OF ARTICULATION				
	UNVOICED		VOICED		NASAL
	UNASP.RATED	ASPIRATED	UNASPIRATED	ASPIRATED	
UELAR	/ka/ (क)	/kha/(ख)	/ga/ (ग)	/gha/(घ)	/ṅa/(ङ)
PALATAL	/ca/ (च)	/cha/(छ)	/ja/ (ज)	/jha/(झ)	/ṅa/(ञ)
RETROFLEX	/ṭa/ (ट)	/ṭha/(ठ)	/ḍa/ (ड)	/ḍha/(ढ)	/ṇa/(ण)
DENTI-ALUEOMR	/ta/ (त)	/tha/(थ)	/da/ (द)	/dha/(ध)	/na/(न)
BILABIAL	/pa/ (प)	/pha/(फ)	/ba/ (ब)	/bha/(भ)	/ma/(म)

Set of other CV coabinaiions where U is /a/(अ)

/ya/(य) /ra/(र) /la/(ल) /va/(व)
/sa/(स) /śa/(श) /ṣa/(ष) /ha/(ह)

The characters in Hindi occur in the follouing forw :

C, U, CU, CCV, CCCU, CCCCU

Pig. 2.1. Structure of the character set for Hindi consonant clusters of **Hindi. Since** each of these clusters can occur with any of the 10 vowels, totally **4660** characters, or roughly 5000 characters, are possible. Though the number appears to be large there is a unique description for each of the character. Because of this unique description, it is much easier to spot the character as a whole in an utterance rather than segmenting and labeling some abstract units, and then forming a symbol from these abstract units. The **difficulty in** our approach is that the number of units is **large.** We have limited our study in the present work to **vowels(V)** and consonant-vowel(CV) combinations. This forms a set of **340 characters** consisting of the 10 vowels and 330 CV combinations. In terms of their frequency, this subset of characters constitute nearly 90 percent of occurences in a text. It is also not difficult to form meaningful words from this subset. For detailed study **only 75 of these 340** characters are used in this thesis. Appendix 1 gives the 340 character subset together with the notation used to represent them in this thesis.

2.4 Our Approach to Signal-to-Symbol Transformation

Having chosen the character as the basic unit, the next step is how to transform a speech signal into the corresponding string of characters. One straight forward method is segmentation and labeling, where a fixed set of parameters are extracted from the speech signal and the characters are identified by matching the descriptions in terms of the parameters. In this thesis we propose a character spotting approach, in which **the description** of a character is given in terms of parameters and features relevant to the character. The speech signal is processed in a manner dictated by the character instead of using a fixed strategy for processing.

The advantages of the character set for Hindi are that they have unique description in terms of speech production apparatus. The characters are also organized phonetically in such a way that a slight change in the articulatory movement is most likely to produce the sound corresponding to the adjacent character. In order to take advantage of the phonetic nature and structure of the character set for spotting a character in speech signal, it is necessary to have a detailed description of the production of the character in isolation and in the context of other characters. The description of each character in terms of speech production parameters and their acoustic manifestation together with the relationship between acoustic features and signal parameters constitutes the acoustic-phonetic knowledge. This knowledge is essential to implement the character spotting approach. We discuss the acoustic-phonetic knowledge of Hindi character set and issues involved in obtaining this knowledge in the next chapter.

Summary

In this chapter we discussed the need for transforming speech signal into discrete units for continuous speech recognition. Different kinds of units used in literature are discussed. Choice of proper units is essential for reducing the complexity in speech-to-text conversion. In this respect the suitability of character as a unit for speech recognition in

Indian languages is brought out. With character as the unit, we proposed an approach based on spotting the characters in an utterance using knowledge-based techniques. In order to implement this approach it is necessary to collect the acoustic-phonetic knowledge which forms the subject matter of the next chapter.

CHAPTER - III

ACOUSTIC-PHONETIC KNOWLEDGE FOR CHARACTER SPOTTING

For spotting the characters in continuous speech, it is necessary to have a complete description of each of the characters in terms of its speech production (articulatory and phonetic description), its acoustic manifestation (acoustic features) and the relation of the acoustic features to the parameters of the speech signal. All this constitutes the acoustic-phonetic knowledge of the character set. The purpose of this chapter is to describe the acoustic-phonetic knowledge relevant for the character set of Hindi. The background needed to describe the acoustic-phonetic knowledge is discussed in Section 3.1. In particular, we give a brief introduction to the speech production mechanism and the categories of speech sounds. We also discuss characteristics of speech waveform to show the relation between the acoustic features and the parameters of the speech signal. In Section 3.2 we give a list of the phonetic features based on speech production. Using this background we provide in Section 3.3 the phonetic description of the character set in Hindi. This section illustrates the phonetic nature of the alphabet in Indian Languages in general, and in Hindi in particular. Acoustic-phonetic knowledge involves relating the phonetic features of speech production to the parameters of speech signal. The sources for acquiring this knowledge are discussed in Section 3.4. The way the acoustic-phonetic knowledge of the character set is derived in our studies is

A-p k (phonetic features
↕
speech signal parameters

illustrated in Section 3.5. The procedure to derive a complete description of the specific knowledge for a character and representation of this knowledge for implementing the character spotting approach are discussed in the following chapters.

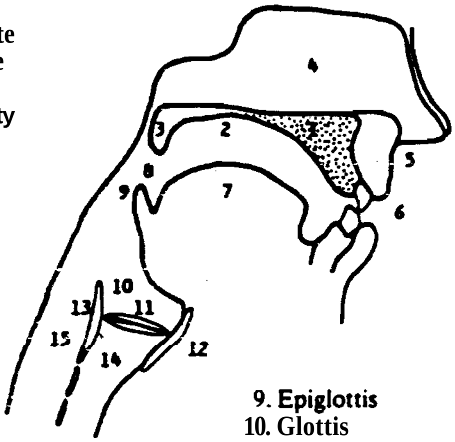
3.1 Speech Production

3.1.1 Speech Production Mechanism (Articulatory Description)

Fig.3.1 shows various parts of the human speech production mechanism. The basic source of power for the production of speech is the respiratory system which moves air in and out of the lungs. As the air is pushed out of the lungs, it goes up the trachea and into the larynx at which point it must pass through the vocal cords. The vocal cords can be kept either apart or close together at will. When the vocal cords are close together, the pressure of the air stream through the narrow opening causes the vocal cords to vibrate resulting in modulation of the air stream. The air passage above the larynx is known as the vocal tract. The vocal tract can be divided into two parts - the oral tract and the nasal tract. The oral tract is formed by the pharynx, oral cavity and the mouth. The principal components of the nasal tract is the nasal cavity and the nostrils.

The muscular flap velum can be either raised to shut off the nasal tract to the air stream from the larynx or lowered to allow the air stream to flow through the nasal tract. The velum, tongue and the lips are the principal articulators in speech production. Characteristics of the speech sounds depend on the configuration of the vocal tract.

I. Hard palate
2. Soft palate
3. Velum
4. Nasal cavity
5. Nostrils
6. Lips
7. Tongue
8. Pharynx



9. Epiglottis
10. Glottis
I I. Vocal cords
12. Thyroid cartilage
13. Cricoid cartilage
14. Trachea
IS. Oesophagus

Fig. 3.1. Vocal tract showing oral and nasal tracts

Figs.3.2 and 3.3 give sections of the vocal tract with the nomenclature of relevant places on the upper and lower surfaces of the vocal tract [59]. The names are used to categorize and describe various speech sounds. With this background of the speech production mechanism and the terminology, let us now look at the categorization of speech sounds in terms of articulatory description.

3.1.2 Categories of Speech Sounds

Speech sounds are characterized by both dynamic and static behavior of articulators. The static behavior of articulators are usually discussed in terms of static positions of articulators as well as on the manner of production (or manner of articulation). The dynamic behavior of articulators are usually discussed in terms of initial position of articulators usually called place of articulation. One may categorize speech sounds into two broad classes - consonants and vowels. Consonants are produced when the air stream through the vocal tract is obstructed in some way by the articulators. The classification of various consonants is based on the place where the obstruction takes place as well as on the manner of articulation in the production of the consonant. Vowels are produced with a relatively free passage of the air stream in the vocal tract with the vocal cords set into strong vibration resulting in sound generation. The quality of the vowel is a function of the shape of the vocal tract and hence vowel classification is based on the positions of the articulators which dictate the shape.

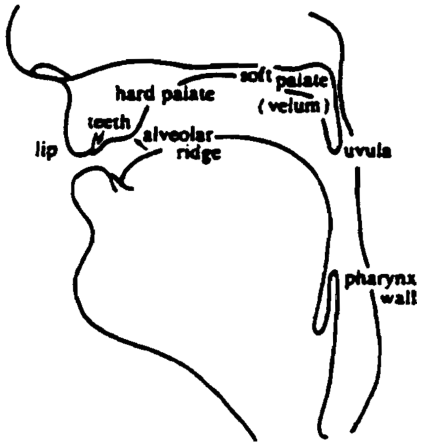


Fig. 3.2. Upper surface of vocal tract

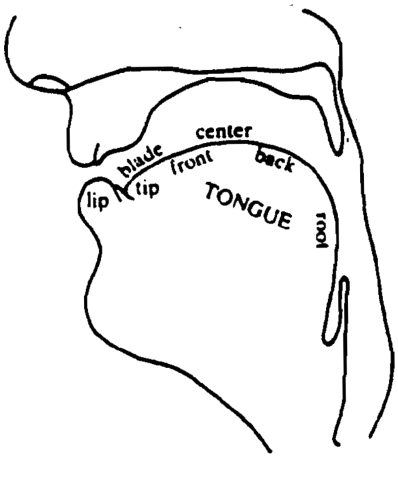
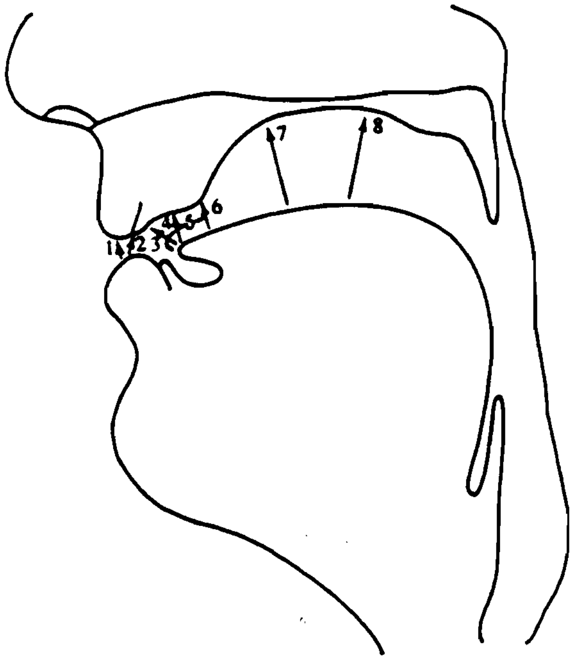


Fig. 3.3. Lower surface of vocal tract

Consonant categorization is usually done in two levels. The first level is based on the ;manner of articulation and the second level is based on the place of articulation. The obstruction in the vocal tract could occur in many ways. The articulation may completely close the oral tract for an instant or a relatively long period. They may narrow the space considerably or they may modify the shape by approaching each other. When complete closure of the oral tract occurs, 'stop' sounds are produced. The complete closure of the oral tract results in pressure building up behind the obstruction. When the articulators come apart, the air stream is released sharply. When the articulators come nearer and the air flow is only partially obstructed, a turbulent air flow occurs at the point of constriction resulting in 'fricative' sounds. The air stream along the center of the oral tract is obstructed when there is incomplete closure of the tract between the tongue and the roof and this produces 'lateral' sounds. 'Trills' are produced When the tip of the tongue intermittently touches the alveolar ridge. 'Tap' sounds are produced when the tongue makes a single tap against the alveolar ridge. 'Affricates' are produced when the tongue tip or blade close on the alveolar ridge as for a stop but instead of coming freely apart separate only slightly. The condition of the vocal cords vibration determines whether the consonant is voiced or unvoiced. The second level classification of consonants based on the place of articulation is dictated by the position of the maximum constriction in the vocal tract. Fig.3.4 shows the points of constrictions in the vocal tract and their names for classifying the places of articulation.



Places of articulation: 1 Bilabial; 2 Labiodental; 3 Dental; 4 Alveolar; 5 Retroflex; 6 Palato-Alveolar; 7 Palatal; 8 Velar.

Fig. 3.4. Position of articulators for different places of articulation

Vowel categorization is based on the shape of the vocal tract and the shape is mostly dictated by the position of the tongue and the shape of the lips. The velum too plays a role in that it dictates whether or not the nasal tract is coupled to the vocal tract. The position of the tongue is categorized in terms of the highest point of the tongue. Fig.3.5 shows the tongue position for different vowels sounds in Hindi. The term 'front' in the figure refers to positions closer to the mouth and 'back' refers to positions inwards away from the mouth. The shape of the lips is usually either rounded or unrounded.

3.1.3 Nature of Speech Signal

Having looked at the characteristics of the speech production mechanism and the categories of speech sounds produced by it let us now look at the speech signal which is the output of the vocal tract system. The speech waveform, which is the time domain representation of the speech signal, depicts the instantaneous amplitude of the signal. The speech waveform is a direct form representation of the sound waves. The nature of the speech signal for different categories of speech sounds will now be discussed.

The condition of the vocal cords dictates the gross characteristics of the speech waveform. The open condition of the vocal cords results in free flow of the air stream and no specific quality is imbibed by the air stream. On the other hand, when the vocal cords are vibrating, the air stream is also modulated and this results in periodicity in the speech signal. These two conditions are characterized as the unvoiced

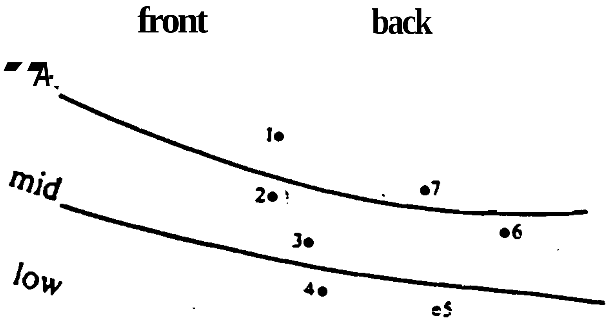


Fig. 3.5. Position of tongue for different vowels

1. /i:/(ई) 2. /i/(इ) 3. /e/(ए) 4. /e:/(ए)
5. /a:/(आ) 6. /u/(उ) 7. /u:/(ऊ)

and voiced nature of the speech signal respectively. The inverse of the base period of the periodic speech signal is characterized as the 'pitch' which directly relates to the frequency of vibration of the vocal cords. The speech waveform for the stop consonants show a burst of signal with no specific periodicity for a short period due to the sharp release of the air stream. The fricatives show a noise like characteristic in the speech waveform due to the turbulent nature of the air stream through the constriction in the vocal tract. The vowels are produced with strong vibrations of the vocal cords and result in high amplitude periodic signal. The quality of the vowel comes out as the fine structure in the waveform shape which is determined by the various resonances of the vocal tract. The resonances of the vocal tract are termed as the formant frequencies. The formant frequencies are not directly apparent in the speech waveform. They are discernible as the peaks in the envelope of the frequency spectrum of the speech waveform. The nasal sounds are produced with the oral tract closed at or near the mouth and the nasal tract opened to the air stream. The nasal sounds are always produced with the vocal cords vibrating. Thus the speech waveform for the nasals show periodicity. The fine structure in the waveform shape reflects the additional resonances due to the nasal cavity. The closed oral tract can absorb energy at certain frequencies and this results in antiresonances or antiformants which appear as valleys or dips in the spectral envelope of the frequency spectrum of the signal. Fig.3.6(a) shows the detailed speech waveform for some categories of speech sound highlighting the points discussed in

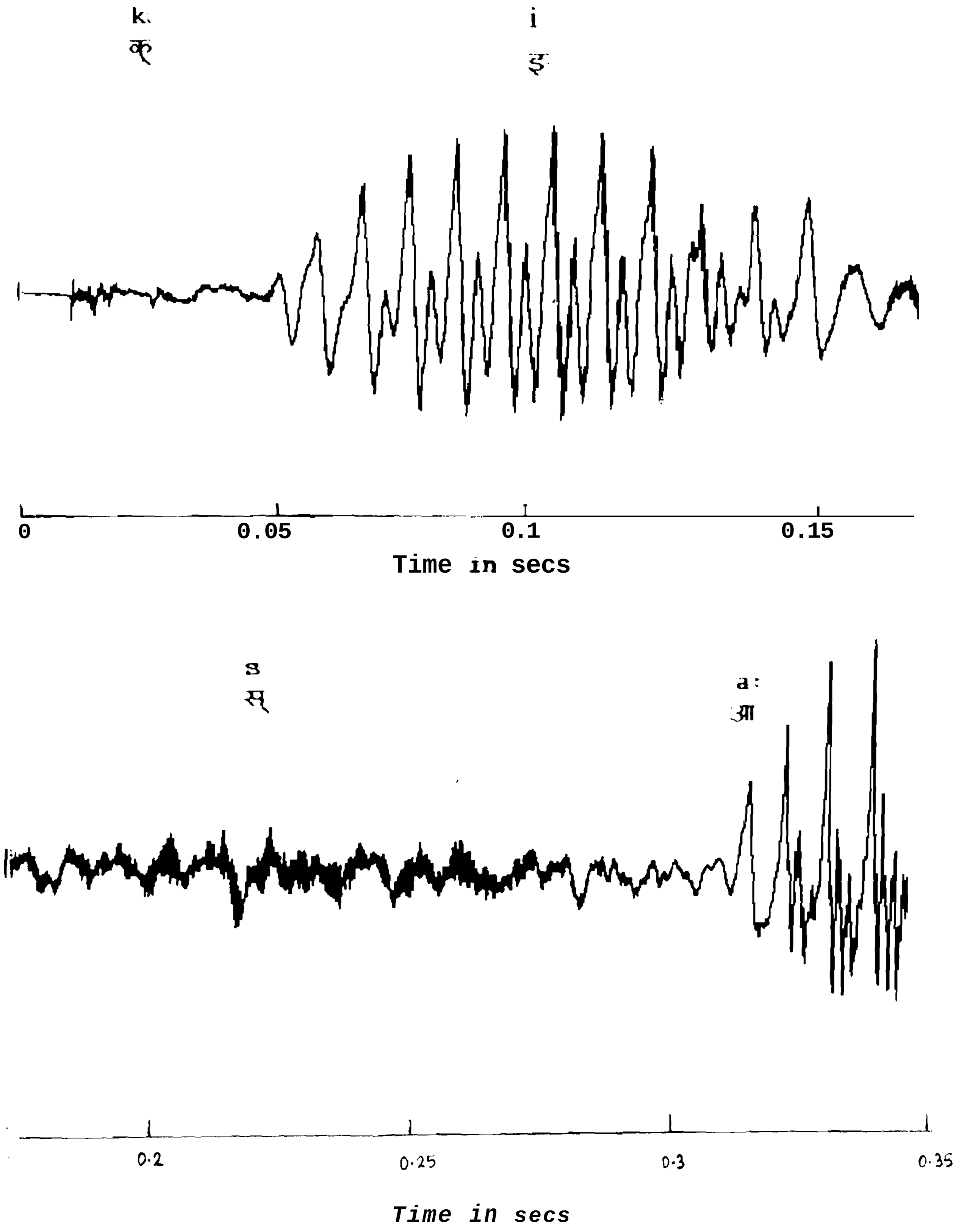


Fig. 3.6(a) Speech waveform

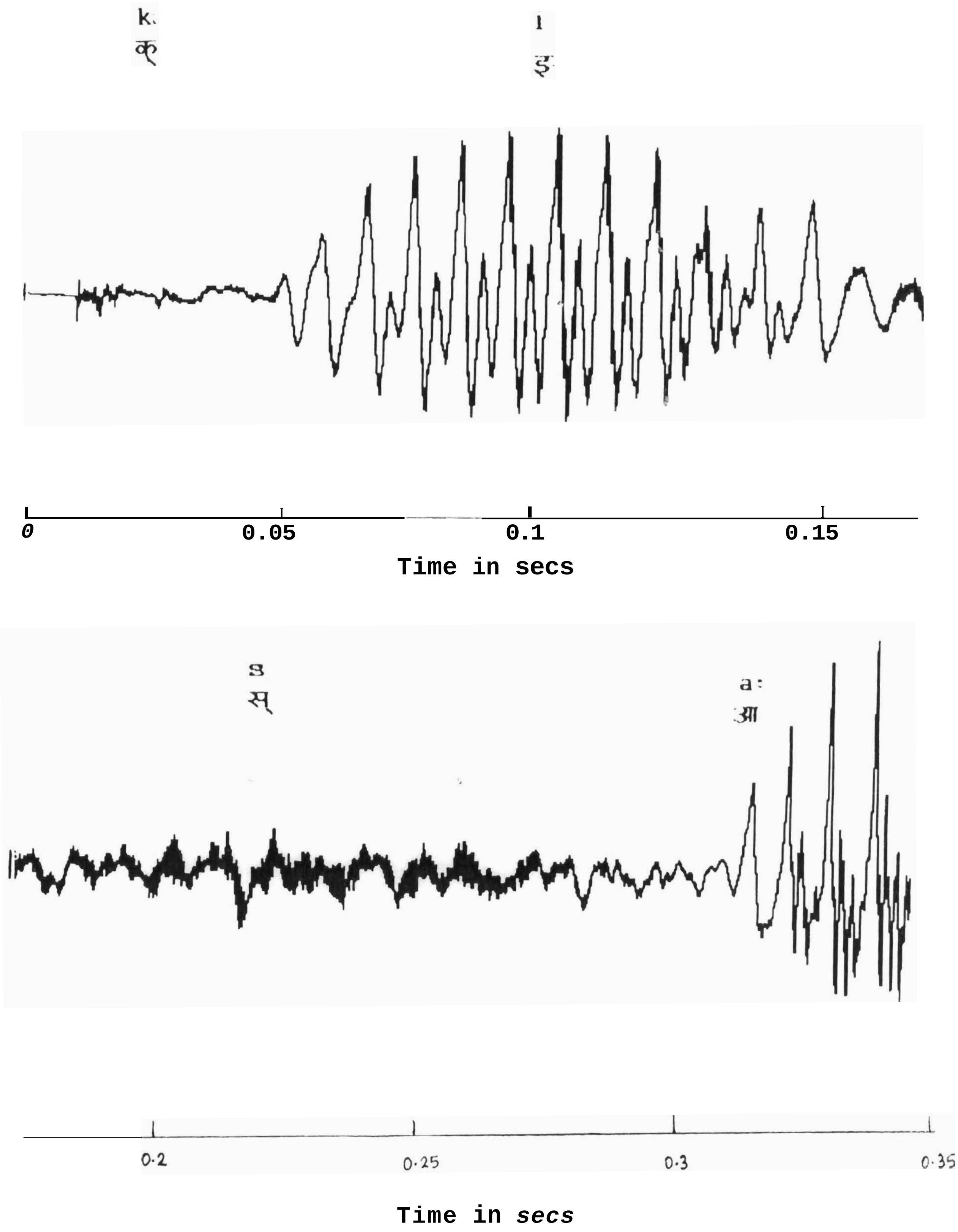


Fig. 3.6 *(a)* Speech waveform

this section. Fig.3.6(b) shows the frequency spectrum of a vowel region of the speech signal identifying the formant peaks.

3.2 Phonetic Features of Speech Production

Phonetics deals with the study of pronunciation of speech sounds of a language. Based on the different categories of speech sounds, the various phonetic features that are used in describing the speech sounds are the following [12]:

1) voiced *2) unvoiced* *3) fricative* *4) affricate* *5) stop*
6) semivowel *7) vowel* *8) aspiration* *9) velar* *10) palatal*
11) labial *12) denti-alveolar* *13) retroflex* *14) front* *15) back*
16) central *17) rounded* *18) unrounded* *19) close* *20) open*
21) half-close *22) nasal* *23) trill* *24) lateral* *25) glide*

A brief description of each of the features with examples for the case of Hindi language are given below.

/ Voiced : Voiced sounds are produced by a mechanism where the vocal cords are in a position such that they will vibrate when there is sufficient air stream. Eg. /b/(ब्), /ḍ/(ड्) and /g/(ग).

1 Unvoiced : Unvoiced sounds are produced by a mechanism in which the vocal cords are opened so wide that there can be no vocal cord vibrations. Eg. /p/(प्), /ṭ/(ट) and /k/(क्).

) Fricatives : Fricative sounds are those sounds that are produced by close approximation of two articulators without complete closure. The turbulent air flow results in noise-like

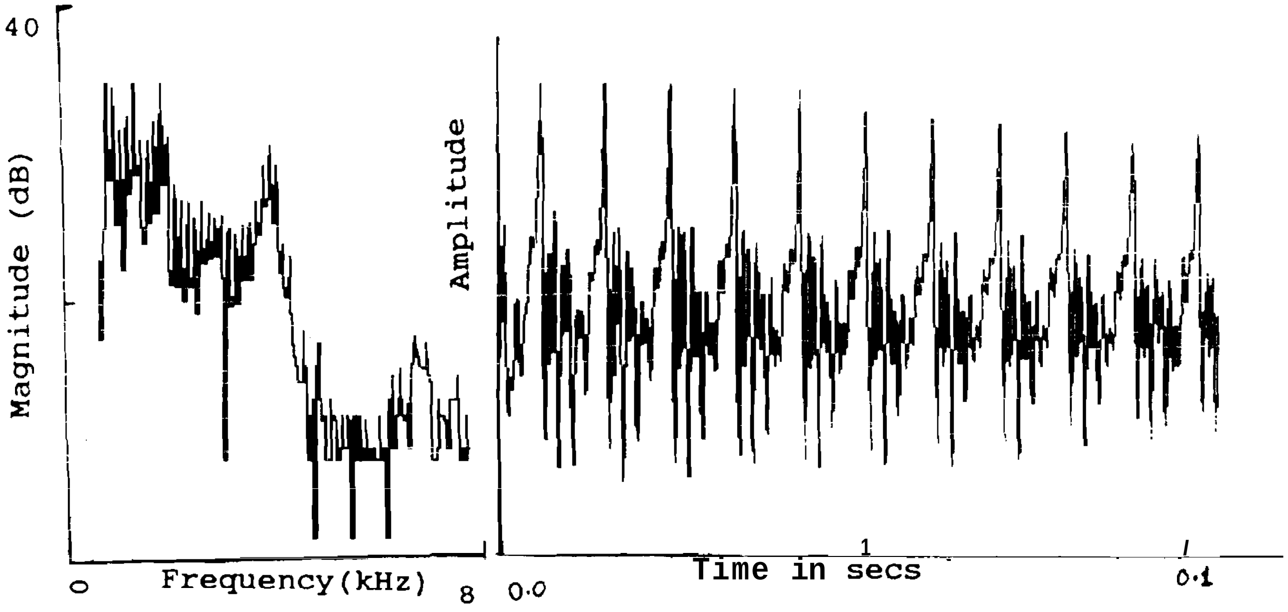


Fig. 3.6(b) FT power spectrum and signal

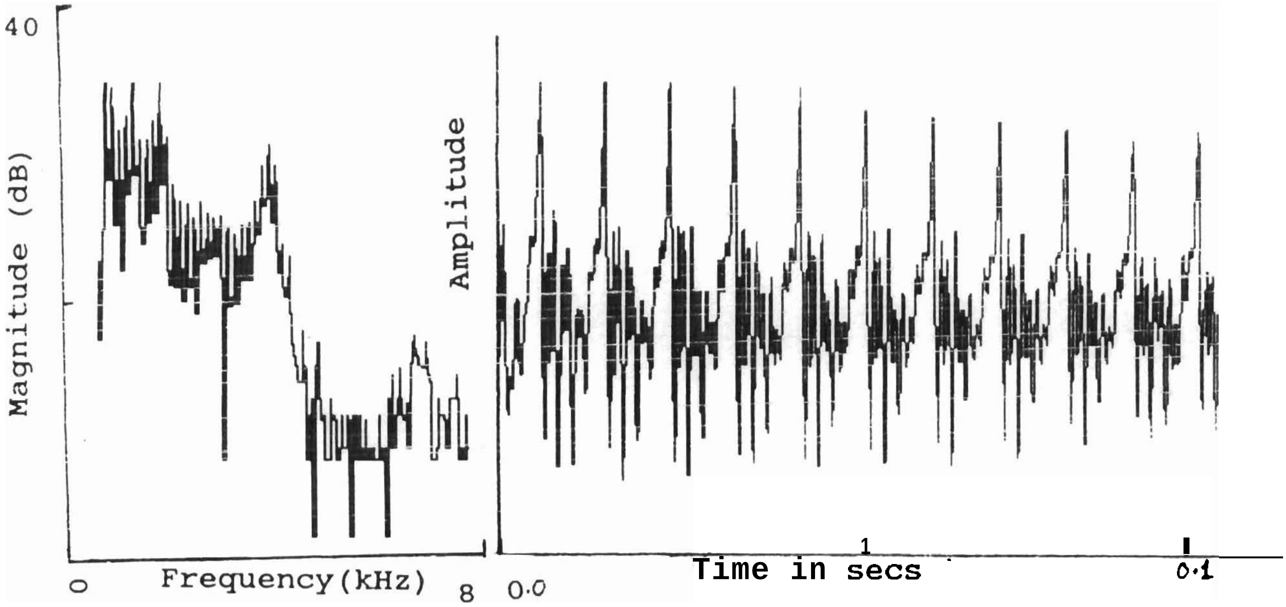


Fig. 3.6(b) FT power spectrum and signal

characteristics in the sound. Eg. /s/(स्) and /ṣ/(ष्).

4 **Stops** : Stop sounds are produced by the complete closure of the vocal tract for a short time by the articulators involved so that air stream cannot escape through the mouth. Eg. /p/(प्), /b/(ब्) and /k/(क्).

5 **Affricate** : These are combination sounds where a stop is followed by a fricative. Eg. /c/(च्) and /j/(ज्).

6 **Semivowel** : Semivowels are vowel-like sounds with more constricted vocal tract than in most vowels. Eg. /y/(य्) and /v/(a).

7 **Vowel** : Vowel sounds are produced with strong vibrations of the vocal cords when none of the articulators come close and the passage of air stream is completely unobstructed. The different categories of vowels are specified in terms of position of the tongue and roundinq of lips. Eg. /a/ (अ), /e/(ए) and /i/(इ).

Aspirated : This sound is produced when forced air stream flows through open vocal tract with very little or no vocal cord vibration. The sound produced is that of a turbulent air flow which implies that it has noise like characteristics but with spectral detail dictated by the vocal tract resonances. This feature is present in certain class of stop consonants. Eg. /ph/(फ्) and /kh/(ख्).

9 **Velar** : This feature specifies one place of articulation of stop consonant. The place of articulation is in the vocal tract at the velum. Eg. /k/(क्) and /g/(ग्).

10 **Palatal** : This feature specifies a second place of articulation. Here the stop consonant is produced with the blade

of the tongue approaching or touching the hard palate. Eg. /c/(च्).

Denti-alveolar : This feature specifies a third place of articulation. Here the stop consonant is produced by the tongue tip approaching or touching a region between dental and alveolar ridge. Eg. /t/(त्).

12 Retroflex : This feature specifies a fourth place of articulation though it resembles closely the manner of articulation. Here the sound is produced by curling the tip of the tongue up and back so that the underside touches the back portion of the ridge. Eg. /ṭ/(ट).

13 Labial : This feature specifies a fifth place of articulation. Here the stop consonant is produced by bringing the two lips together. Eg. /p/(प्) and /b/ (ब्).

14 Front : This feature specifies the type of vowel produced or in a way it specifies the place of articulation for the vowel. This type of vowel is produced when the highest point of the tongue is in the front portion of the oval tract (closer to the mouth). Eg. /i/(इ) and /e/(ए).

15 Back : A vowel of this type is produced when the tongue is close to the upper or back surface of the vocal tract inwards away from the mouth. Eg. /u/(3) and /o/(ओ).

16 Central : This is another feature which specifies the type of vowel produced. A vowel of this type is produced when the position of the highest point of tongue is at the central portion of the vocal tract. Eg. /a/(अ).

17 Rounded : This feature specifies the manner in which the vowels are produced. A vowel of this type is produced with the lips

(18 unrounded)

rounded. Eg. /u/(उ).

19 **Close** : This feature specifies the position of the tongue with respect to its neutral position. When vowels of this type are produced the body of the tongue is raised above its normal position. Eg. /u/(उ).

21 **Half_close** : When vowels of this type are produced the body of the tongue is in its neutral position. Eg. /o/(ओ).

22 **Open** : When vowels of this type are produced the body of the tongue is below its neutral position. Eg. /a/(अ).

23 **Nasal** : Nasal sounds are produced when the air stream is allowed to pass through the nasal tract instead of the vocal tract. Eg. /n/(न्) and /m/(म्).

24 **Trill** : This feature specifies the degree of vibration of an articulator when air stream passes through the vocal tract. Eg. /r/(र्).

25 **Lateral** : These sounds are produced when a large portion of the air stream flows over the side of the tongue. Here the tips or the sides of the tongue touch the roof of the oral cavity. Eg. /l/(ल्).

Though a list of features is given, it is not necessary that the descripticn of a character contain everyone of these features. A character can be described by a subset of these features by studying the speech production mechanism. In the next section we describe the organization of the character set in Hindi using these features.

33 Phonetic Description of the Alphabet in Hindi

The list of characters in Hindi that form the basic alphabet set is shown in Fig.2.1 of Chapter II. The table in the figure gives only the basic set of characters and the actual number of all possible characters is approximately 5000. The number of characters consisting of vowels and consonant vowel combinations only is 340. The characters can be described in terms of the twenty five phonetic features described in the previous section. Appendix 1 gives a description of the characters that are considered in our study.

ok

This description forms one of the key knowledge bases in our approach. Having obtained this description, it is necessary to relate the phonetic features to its acoustic manifestation in the signal. This relationship or in other words the knowledge base is what is considered in the next section.

3.4 Sources of Acoustic-Phonetic Knowledge

Acoustic-phonetic knowledge for the character set has to be collected from various sources. The primary source is an expert phonetician who will be able to describe the speech production process for each character in terms of the articulators and acoustic features. Speech spectrogram is a display of visible speech which is a good tool used by phoneticians to describe the acoustic manifestation of the production process. A detailed analysis of speech signal using signal processing algorithms will enable us to quantify the acoustic features seen in a speech spectrogram. Ultimately, it is our ability to process the speech signal in a desired manner which dictates the utility of the acoustic-phonetic knowledge for speech recognition.

character → phonetic features → acoustic features
(rule) (25) (rules)

3.4.1 Expert Phonetician's Description of Characters

The phonetic features listed in Section 3.2 and the character descriptions based on these phonetic features have evolved through interaction with phoneticians. These categorizations and descriptions help in viewing the character set from a meaningful perspective with an orthogonal set of descriptors.

The integration of the various sets of features is what occurs in the actual acoustic signal. It is necessary to estimate what features in the acoustic signal account for the various descriptors or phonetic features to realize the necessary mapping from the signal to the descriptions. A phonetician can describe a character in terms of its phonetic features and acoustic behaviour.

The expertise in a phonetician is in terms of the understanding of the interaction of various phonetic features and their acoustic manifestation arrived at by careful analysis and interpretation of several cases. This understanding enables a phonetician to interpret the speech signal and perform a reverse mapping from speech signal (or spectrogram) to acoustic features to phonetic features. It is this knowledge that we need to abstract from the phonetician.

acoustic feature
↓
phonetic feature

3.4.2 Spectrogram Reading

Phoneticians use spectrogram as the basis for analysis of speech sounds. Spectrogram is the frequency domain representation of speech signal at various instants of time. The two-dimensional representation of a speech signal consisting of the signal amplitude as a function of time is converted to a three-dimensional representation consisting of the amplitude (actually power expressed in dB) of the various frequency components in the signal as a function of time. The three dimensional space is represented in two dimensions by time and frequency in the x and y axes respectively, with the amplitude represented by the gray levels of the display. Peaks in the envelope of the frequency spectrum of the signal appear as darker regions in the spectrogram. Steadiness of the spectral characteristics of the signal over a certain interval of time is reflected as a steady pattern persisting for some time. Vowel sounds display regular vertical striations in the pattern due to the impulse-like nature of the vocal cord vibrations. A voice bar is also present at the low frequency end due to energy concentration at the pitch frequency. Voiced sounds which in our terminology correspond to the low energy signal with vocal cord vibrations, exhibit only the voice bar in most cases. Noise-like characteristics in the signal result in patterns with randomly distributed dark regions in the frequency axis. Low energy regions in the signal show light patterns while high energy regions exhibit dark patterns. Distribution of signal energy in the frequency spectrum is conveyed by the distribution of the dark regions along the frequency axis. In short, the spectrogram patterns reflect the frequency domain characteristics of the signal as a function of time.

Formants, which appear as dark horizontal bands in the spectrogram, are important features used in spectrogram reading. Due to the relationship of formants to vocal tract resonances, the tracks of formants on the spectrogram provide clues to infer the shape of the vocal tract and the position of the articulators during speech production.

Fig.3.7 shows speech signal and corresponding spectrograms for different vowels of Hindi spoken by a male speaker. The formant tracks for F1, F2, F3 and in some cases F4, F5 can be seen in the figure. It can be clearly seen that differences in the spectrogram for various cases of vowel utterances are in the positions of the formant tracks F1 and F2. Vowels exhibit flat formant tracks as there are no vocal tract changes during the production of steady vowels. Diphthongs, on the other hand, exhibit a change in the shape of the vocal tract in its production. There is a gradual change of the shape of the vocal tract from one vowel .position to another vowel position. This appears in the spectrogram as inclined formant tracks which start from the position corresponding to the initial vowel and gradually change position to that corresponding to the final vowel. Fig.3.8 shows the formant tracks corresponding to the diphthong /ai/(ऐ) and /au/(औ). Semivowels show formant changes in the initial portion of the utterance due to the quick change in the articulatory positions at the beginning of the utterance. Fig.3.9 shows the characteristics of the change in formants for the semivowels /ya/(य), /ra/(र), /la/ (ल) and /va/(व) of Hindi. Fig.3.10 gives the speech signal and the corresponding spectrogram of a sentence spoken in Hindi by a male speaker. A manual transcription of the different regions of the signal in terms of the characters of Hindi is also marked. The sentence contains combinations of CV as well as CCV sounds,

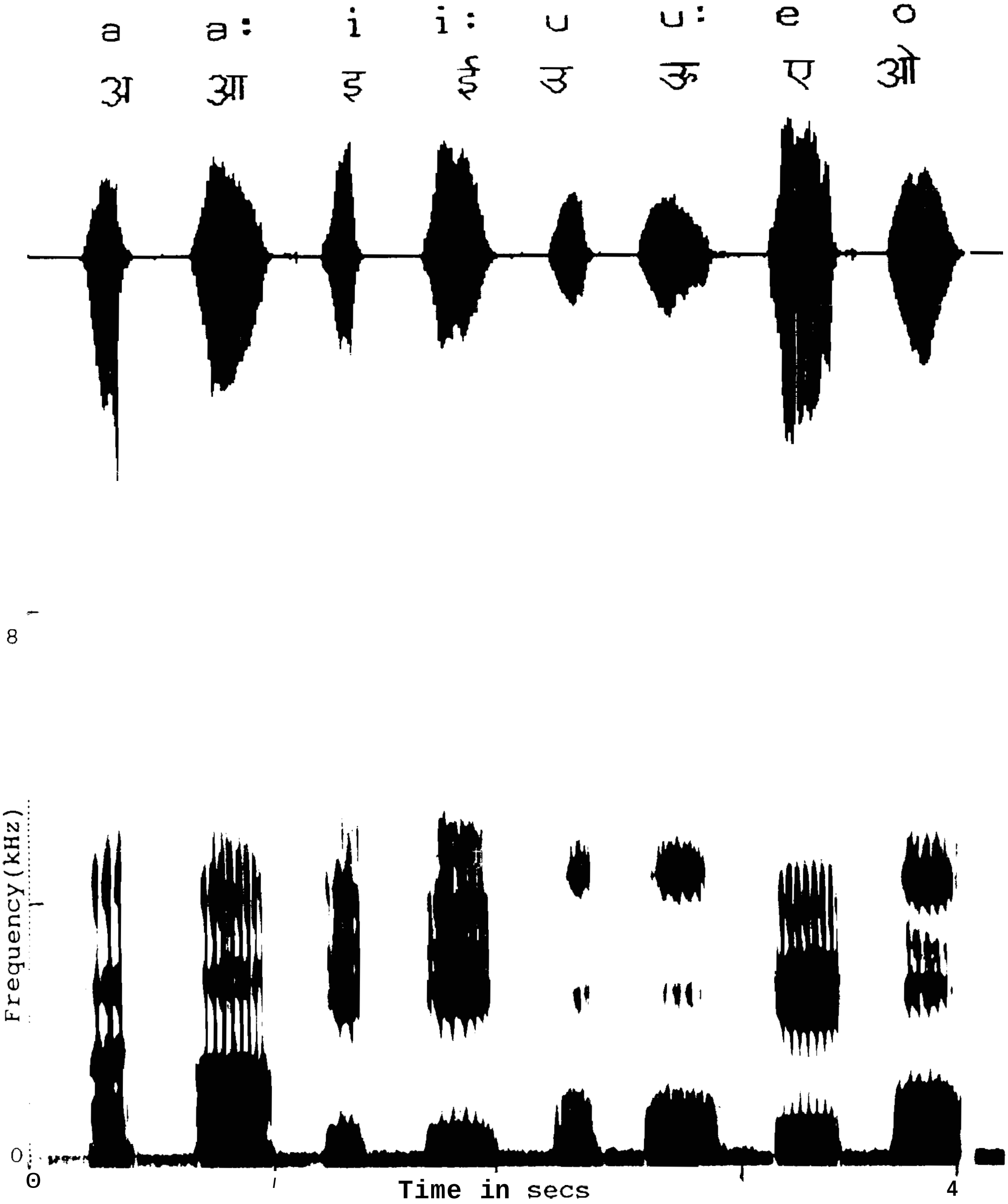


Fig. 3.7. Speech signal and spectrogram for vowels

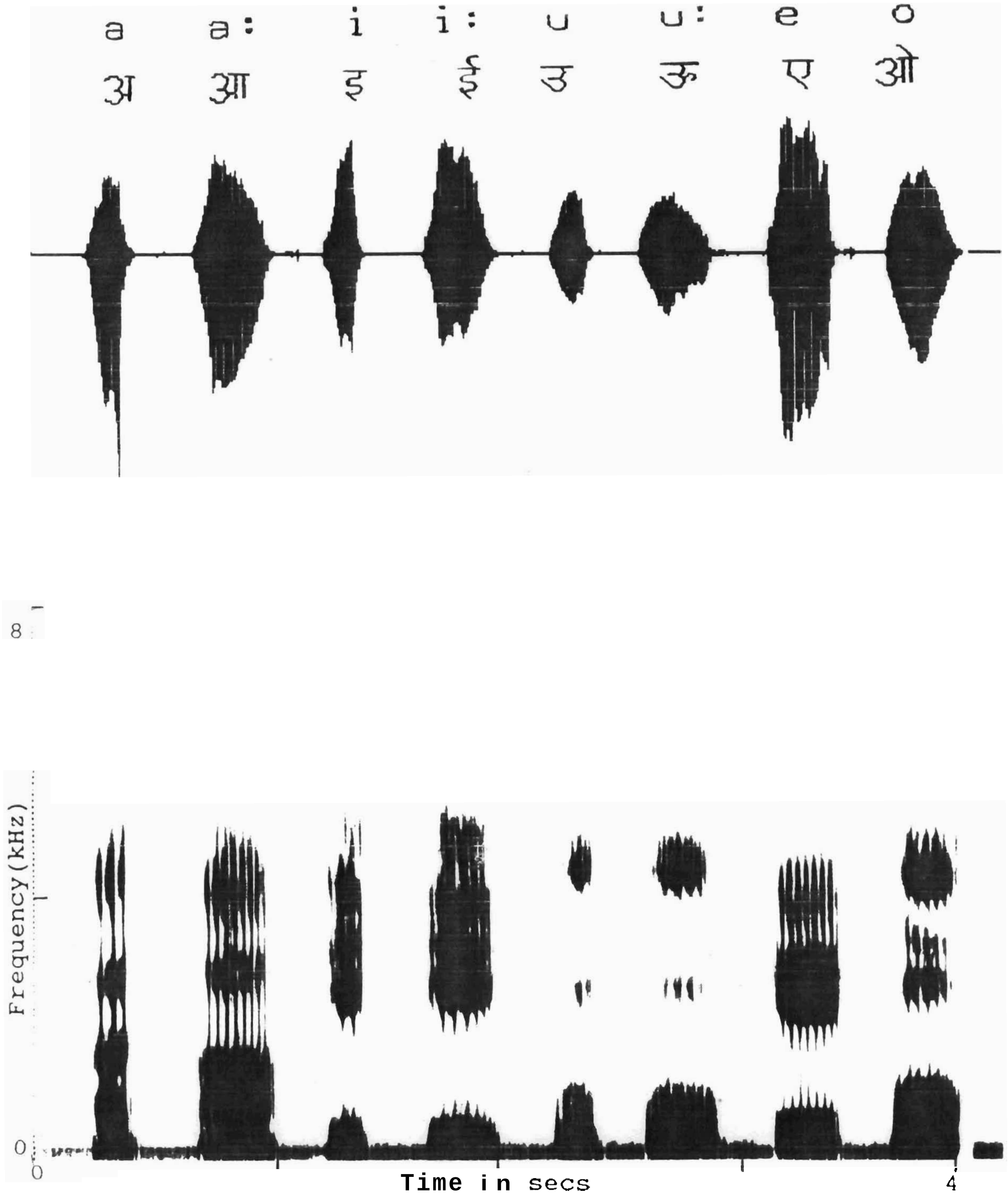


Fig. 3.7. Speech signal and spectrogram for vowels

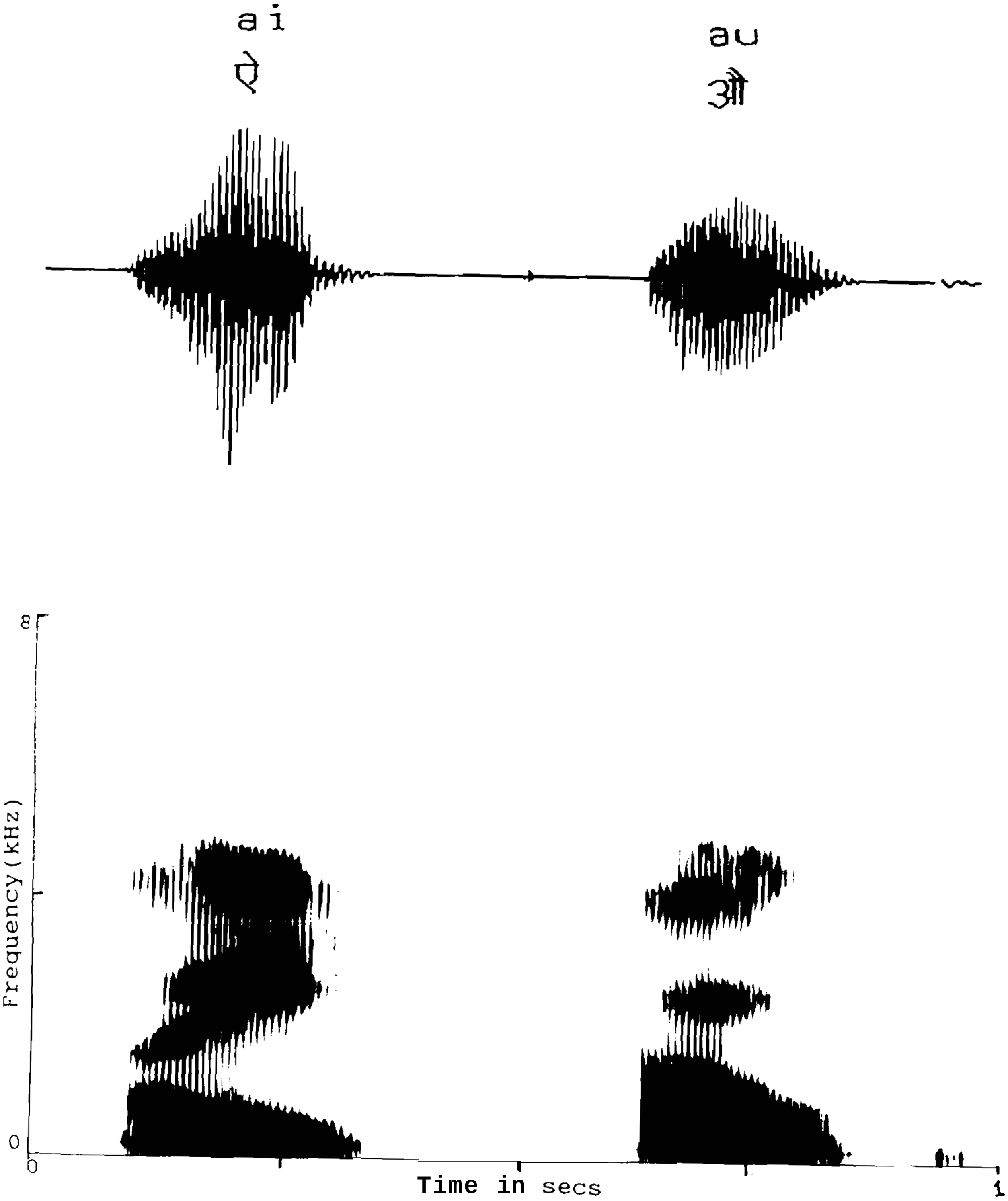


Fig. 3.8. Speech signal and spectrogram for diphthongs

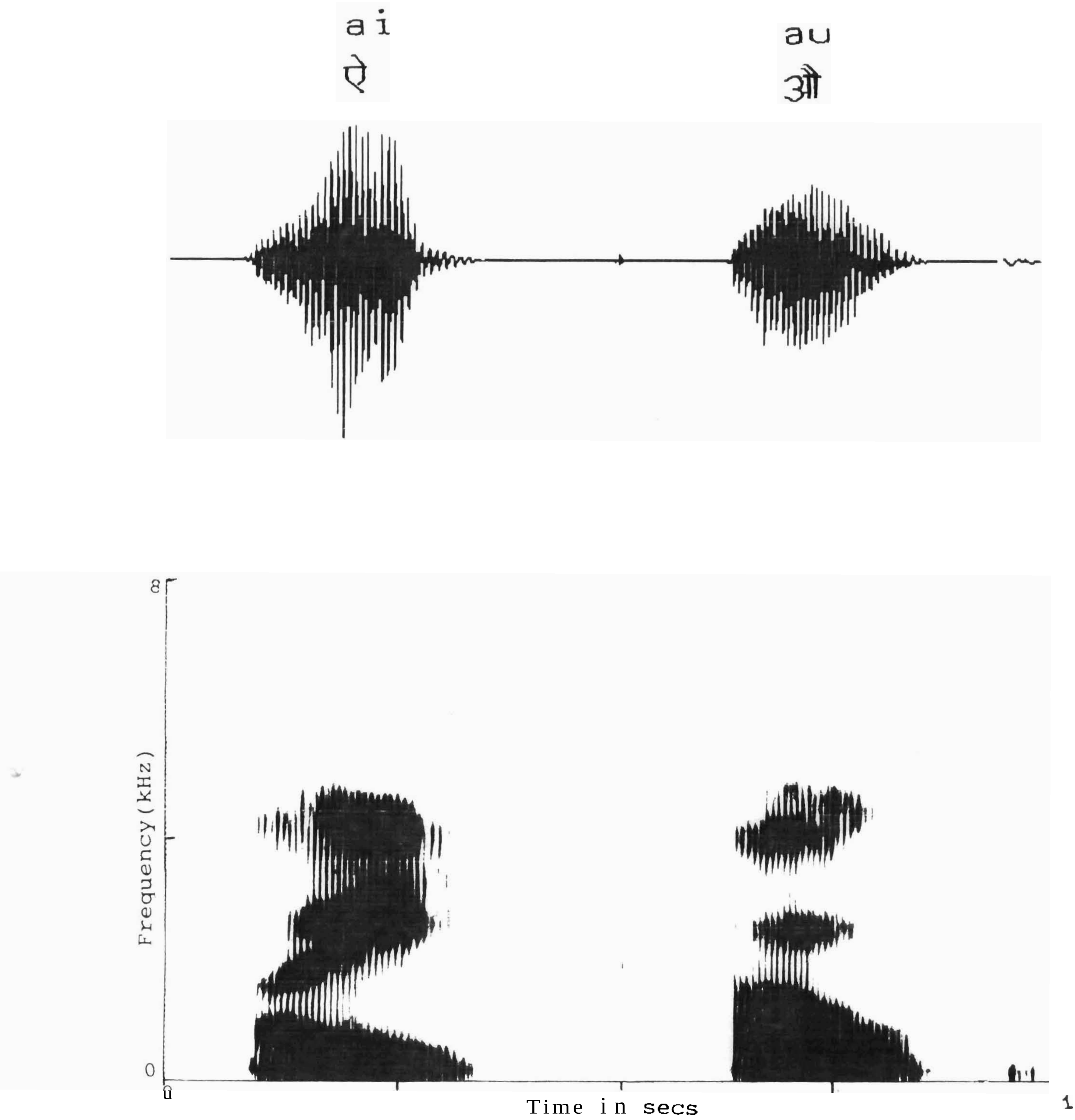


Fig. 3.8. Speech signal and spectrogram for diphthongs

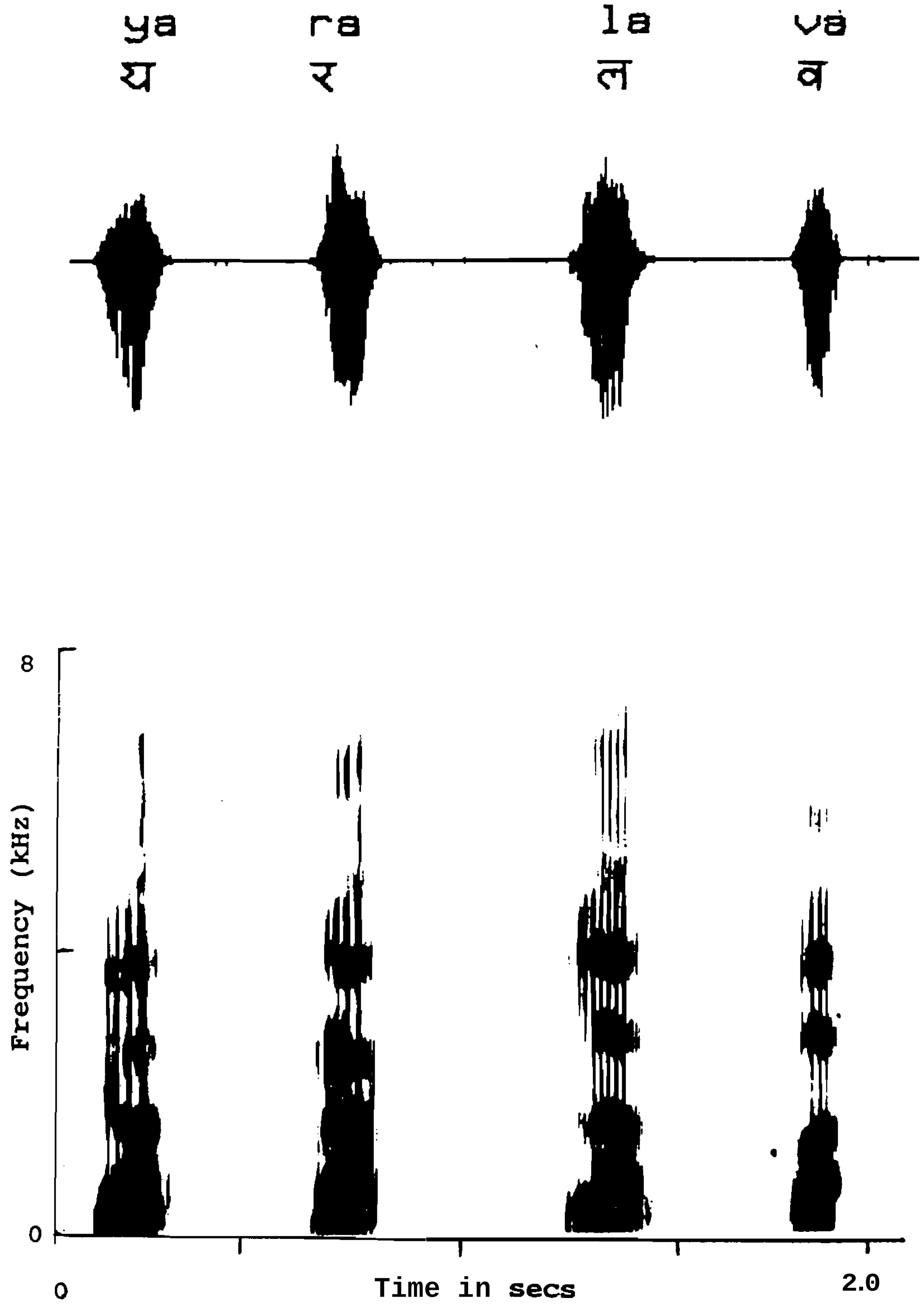


Fig.3.9. Speech signal and spectrogram of sonorants

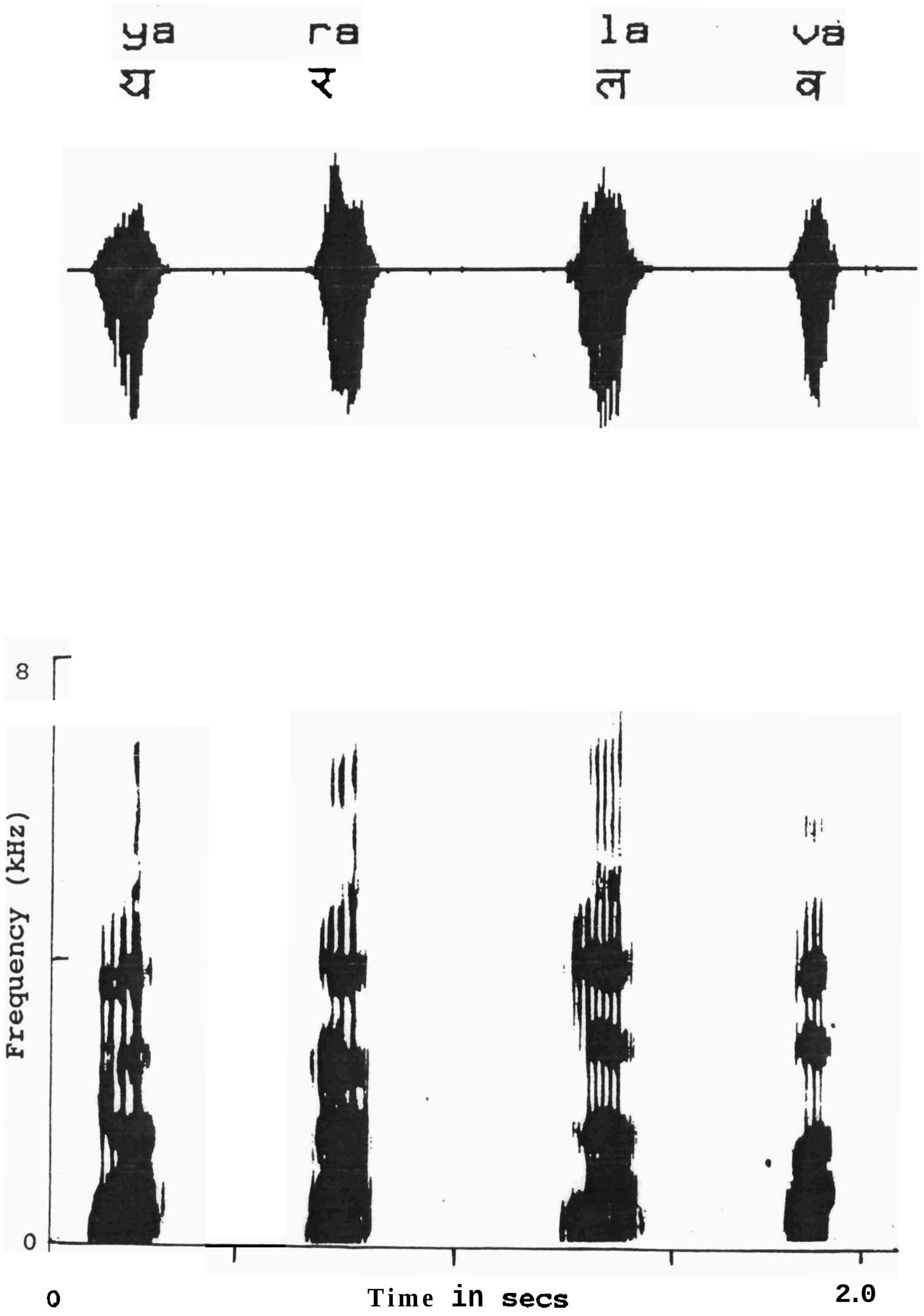


Fig.3.9. Speech signal and spectrogram of sonorants

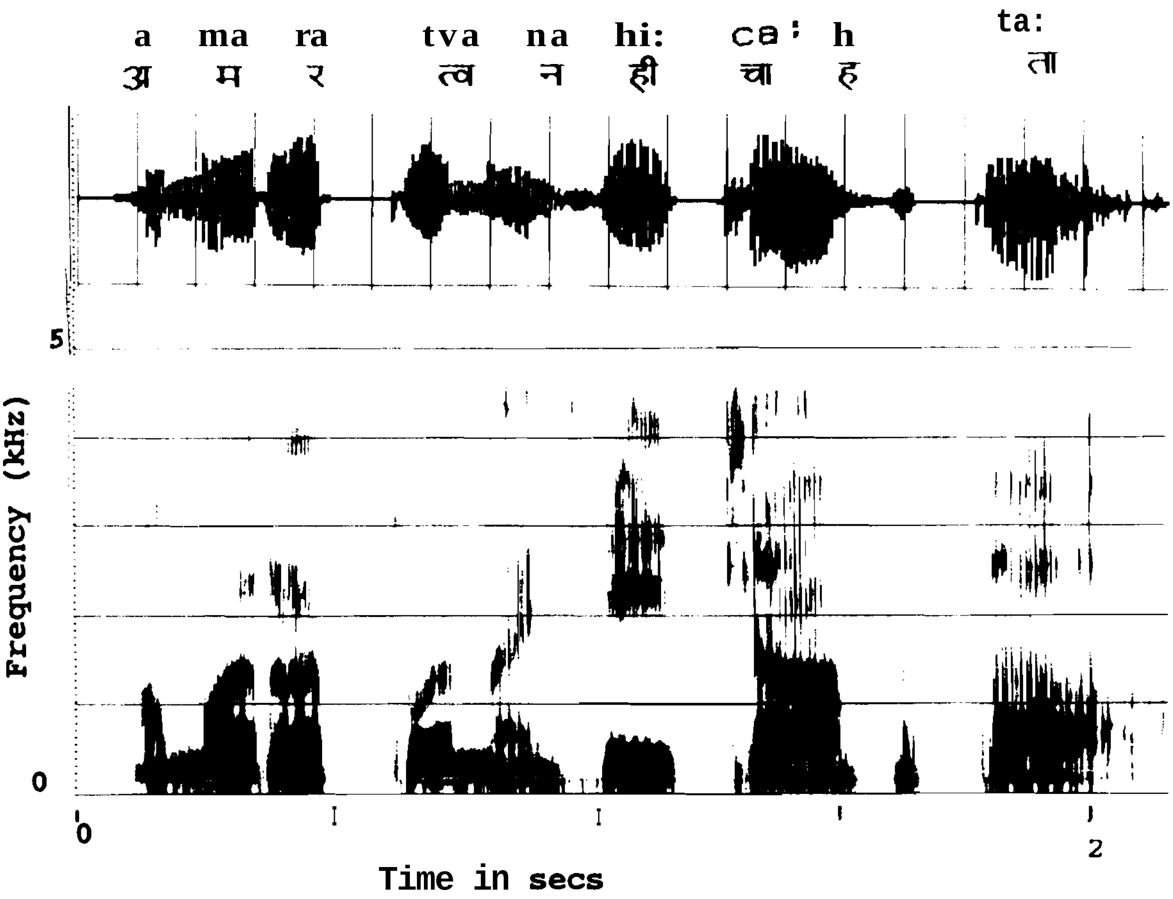


Fig.3.10. Speech signal and spectrogram of the utterance

amaratva nahi: ca:hta:
अमरत्व नही चाहता

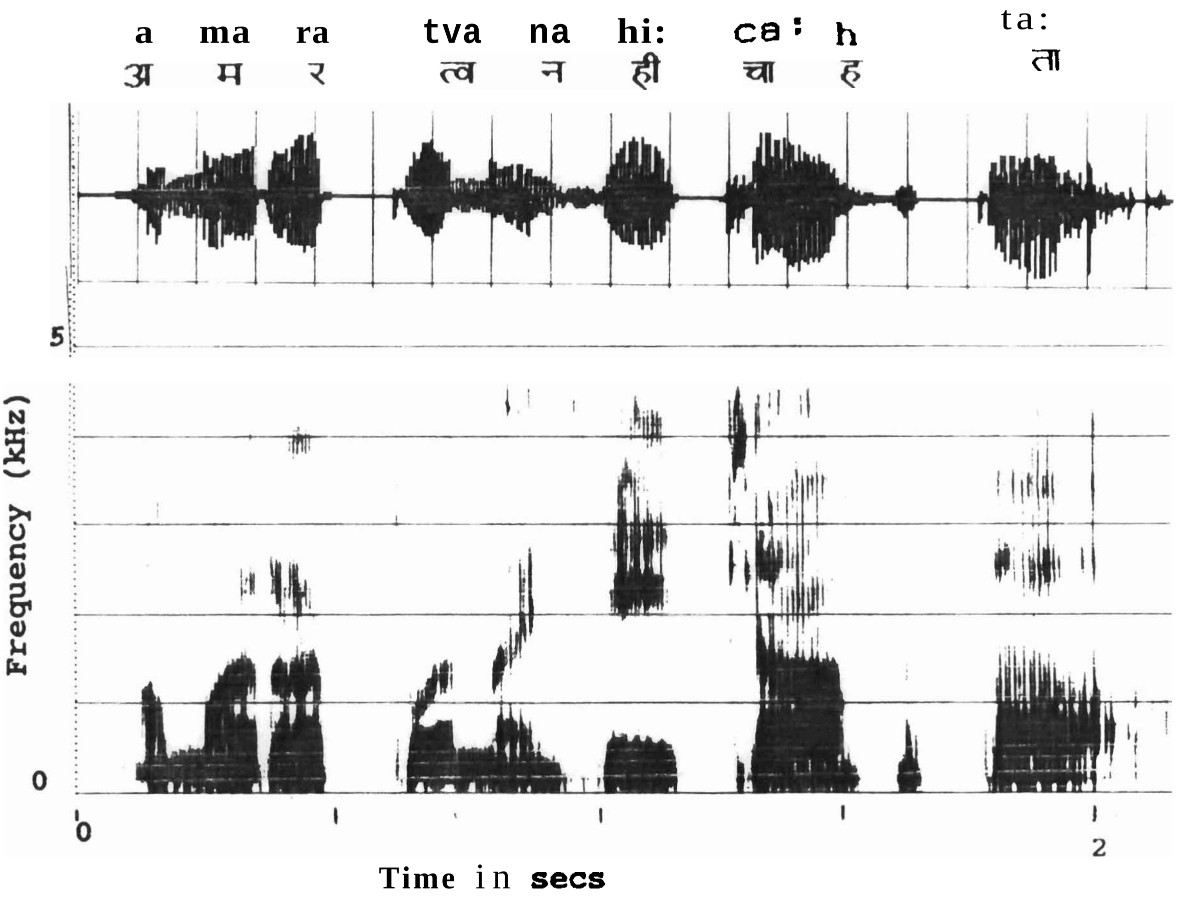


Fig.3.10. Speech signal and spectrogram of the utterance

amaratva nahi: cə: hta:
अमरत्व नही चाहता

where, C is a consonant and V is a vowel. Changes that take place in the speech production mechanism in the course of production of the sentence are evident in the spectrogram.

The pictorial representation of the spectrogram makes it possible for us to see the patterns and detect trends in the signal using our visual capability. The spectrogram thus proves useful in the analysis. One should remember that the spectrograms considered were for the case of speech uttered by a single speaker. For proper analysis, it is necessary to study spectrograms for various speech sounds spoken by several speakers. This comes under the purview of the phonetician.

At this stage, we are in a position to relate the phonetic features to their acoustic behaviour in terms of the spectrogram. The relationship is still qualitative since we discuss in terms of patterns and trends in the spectrogram and their behaviour. These patterns are nothing but descriptions of the acoustic manifestation and hence we can consider this relationship as between phonetic features and acoustic manifestation. In some sense the spectrogram is a faithful representation of the signal. Hence it is possible to quantify some aspects of the acoustic manifestation in terms of values. But this is not enough. We need to quantify the patterns and trends, as well, which is not easily possible. It is better to view the analysis based on spectrogram as providing clues in terms of what to look for in the signal to establish the presence or absence of a feature, and use the values obtained from the spectrogram in 'the detection process. This leads us to the next topic of discussion, namely, parametric analysis of speech signal.

3.4.3 Analysis of Speech Signal

A careful analysis of the speech signal is required to extract features visible on a spectrogram. Certain features can be categorized as gross features and certain others as fine or detailed features. The gross features are the ones that are applicable almost in a universal manner over various categories of speech sounds. The fine features usually relate to a smaller set of categories. The choice of the set of parameters is an important consideration and is based on the techniques available as well as the efficacy of the parameters in representing features of interest. Certain features may warrant the development of special techniques for extracting new parameters.

Some of the parameters that are used in this work to spot gross features are energy (ENR), spectral flatness (SPF), spectral distance or Itakura distance of adjacent frames of signal (SPD), high frequency energy to low frequency energy ratio (HLR), and the first coefficient of linear prediction analysis which reflects the low frequency energy content (LP1). Fig.3.11 shows the plot of these parameters for an utterance. The figure includes the corresponding speech waveform which is time aligned with the parameters. The phonetic transcription of the speech signal is also indicated in the figure. The plot also includes the formant tracks or rather spectral peaks which we shall discuss later. Instants of change in the signal are also identified and the types of changes that occur in various parameters are clearly visible with respect to the identified instants of change in the signal.

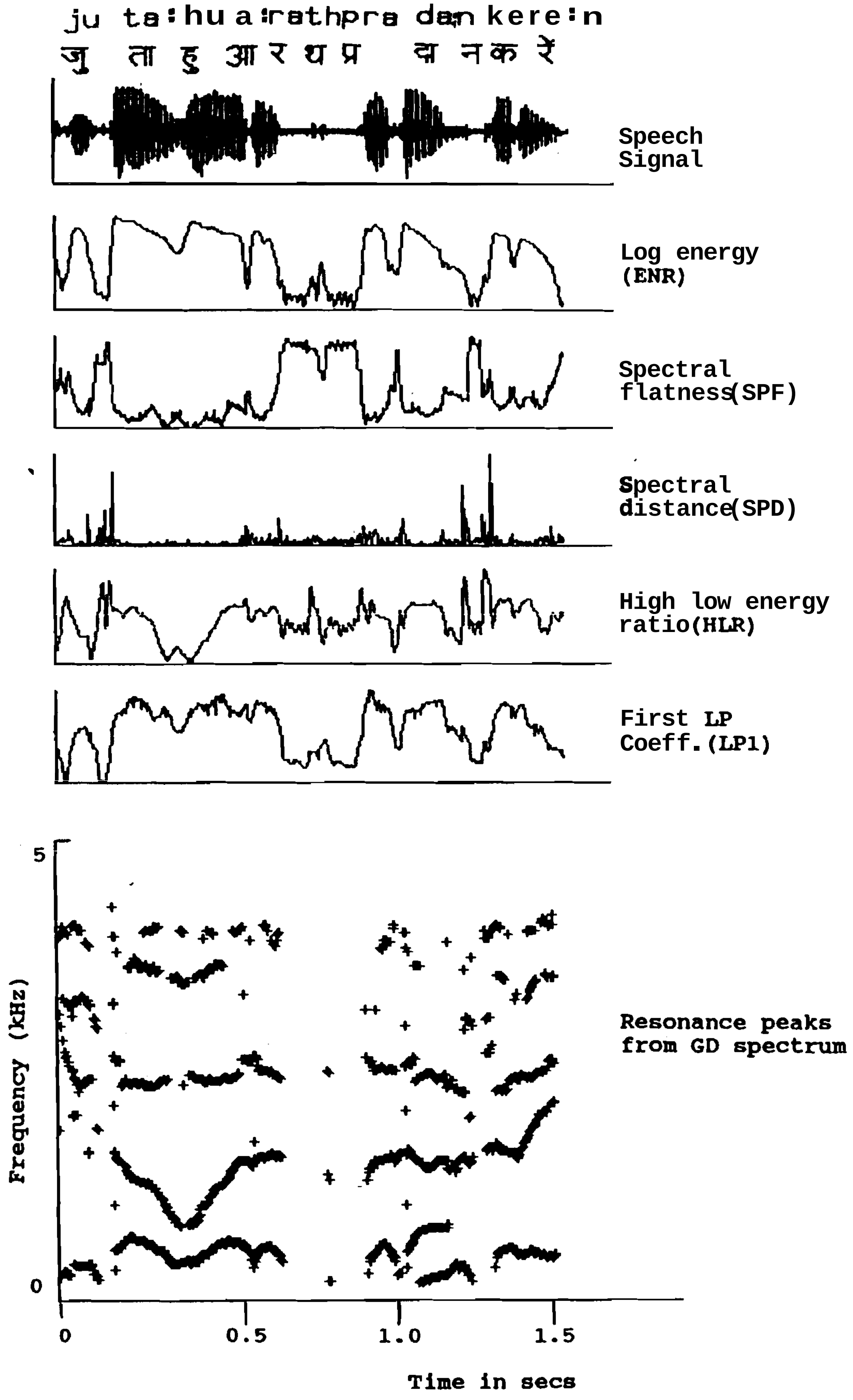


Fig. 3.11. Spectral peaks from Group Delay analysis and other parameters for an utterance

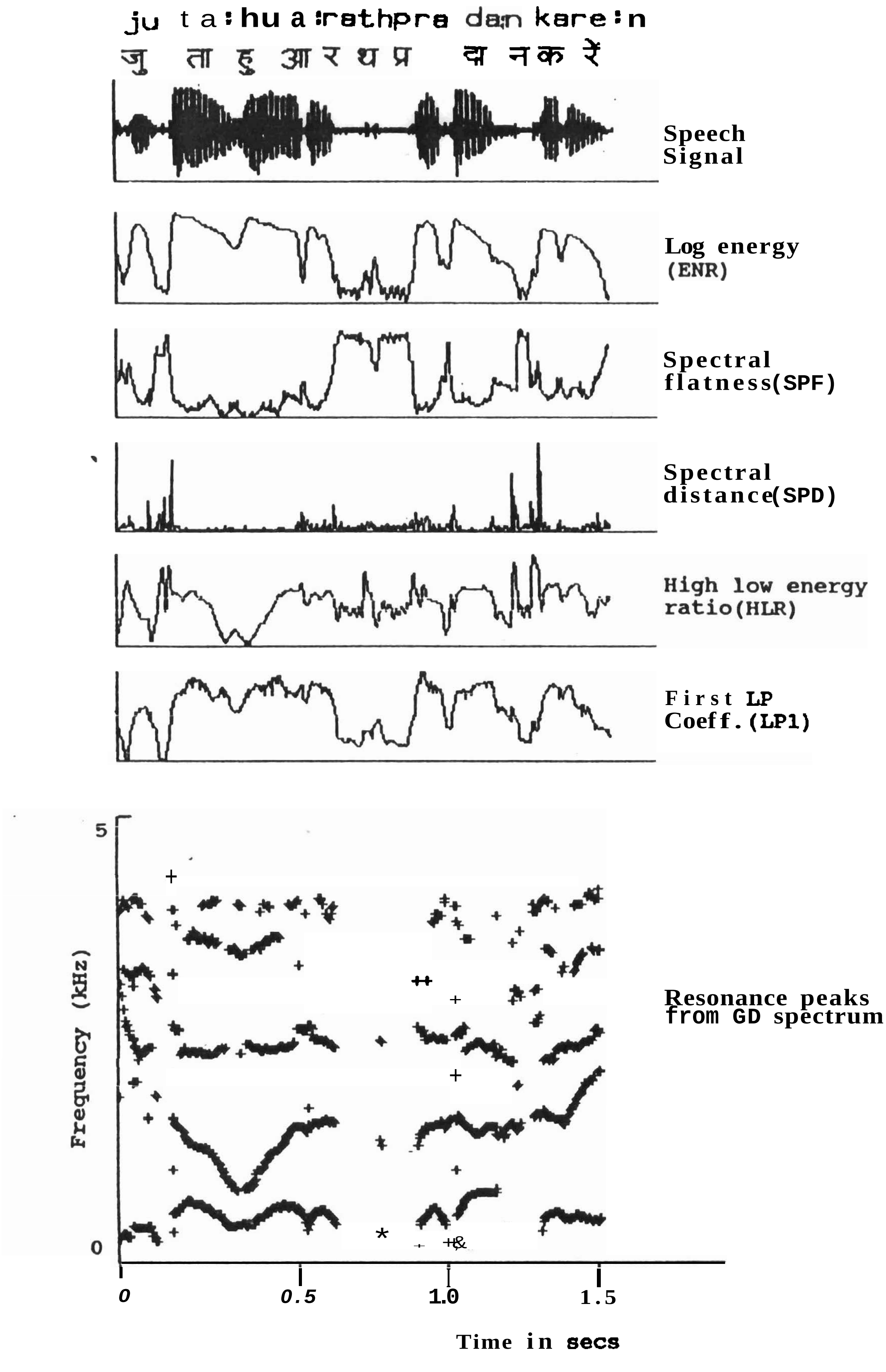


Fig. 3.11. Spectral peaks from Group Delay analysis and other parameters for an utterance

The ENR parameter is chosen due to its direct relationship to the amplitude or the strength of the signal. The SPF parameter provides a gross measure of the shape of the spectrum. It has the ability to characterize the degree of flatness in the spectrum which reflects the degree of noise-like behavior in the signal. This parameter is obtained by linear prediction analysis of the signal. It quantifies the dynamic range of the modeled spectrum. The SPD parameter quantifies the degree of change in spectral characteristics of adjacent regions of the signal. This parameter is also obtained from linear prediction analysis of the signal. This quantifies the spectral difference between the modeled spectrum of the signal in adjacent regions. The HLR parameter refers to the ratio of energies in regions above and below 1.25 kHz. The parameter is computed from the Fourier transform magnitude spectrum. The LP1 parameter quantifies the emphasis of the low frequency in the modeled signal. The parameter has characteristics akin to the inverse of the HLR parameter, but the quantification has differences in emphasis. The former is model based while the latter has an arbitrary point of demarcation between the high and low frequency regions.

Thus, these set of parameters provide a gross characterization of the signal and form the basis for detecting the presence or absence of relevant gross acoustic features. A detailed analysis of these parameters is called for in arriving at the ranges and thresholds for different parameters for various acoustic categories of the signal. This is done by means of plots of the form shown in Fig.3.11 of a large set of utterances.

[illegible] acoustic features

As seen in Section 3.4.2, the detailed acoustic features in the signal are most often inferred from the trends in formant tracks. Hence obtaining formant tracks of the signal should contribute significantly to the detailed acoustic categorization of the signal. A well known technique for obtaining formants is by peak picking the linear prediction modeled spectrum [78]. This gives spurious peaks in many cases. A technique based on group delay spectrum [25] was used in our studies. The formant contours shown in the bottom plot of figure are obtained by this technique. A point to be noted with reference to the frequency verses time plot of the form shown in figure is that it should not be read like a spectrogram. The plot only shows the location of peaks in the spectral envelope. Other details are masked. An important aspect of the technique is that it is nonmodel based and hence peaks obtained from this should reflect the formant information better than that the obtained by model-based techniques. Note that in these plots the peaks for unvoiced segments of data were masked.

Formant contour plots of the form shown are analyzed for a large set of utterances to arrive at the ranges, thresholds and trends of these parameters for various acoustic features.

3.5 Acoustic-Phonetic Knowledge for Character Set

The phonetic features shared by the character set suggests some form of grouping amongst them. The grouping facilitates compilation and organization of the knowledge base. In addition, it helps in working out some common control strategies and rules. The details of the classifications of rule base for the character set based on phonetic features are given in Section 3.5.1. The derivation of acoustic correlates for various phonetic features for each character is discussed in Section 3.5.2. Section 3.5.3 covers the derivation of invariant acoustic cues to overcome variability associated with the signal.

3.5.1 Grouping of Characters

We can group all the character occurring in Hindi into two groups - vowels V and others C CV. The second group consists of consonants occurring on their own and combination of consonants with vowels. In order to achieve this grouping we try to use some gross phonetic features. For example, the feature vocalic[1] can be used to separate vowel regions from other consonantal regions. This feature together with other features like burst and aspiration can be used to separate regions in the utterance into two groups namely vowels and others. A rule that can be used is:

Rule 1: If vocalic region is preceded by stop or burst or aspiration or a combination of any of these, then it is eliminated or rejected as a vowel occurring on it own. If the silence region

----

[1] vowel as a phonetic feature is termed vocalic here to distinguish it from vowel as a character.

preceding the vocalic region is longer in duration, then this region can be considered as a vowel occurring on its own.

In the above classification the speech sounds under the first group are listed as vowels occurring on their own. This list for the case of Hindi is as follows: /a/(अ), /a:/(आ), /i/(इ), /i:/(ई), /u/(उ), /u:/(ऊ), /e/(ए), /ai/(ऐ), /o/(ओ), and /au/(औ). The first rule for all the characters in this list will be the same, namely, Rule 1. This rule when applied on the input utterance will identify regions corresponding to these sounds. This group can be further subdivided into subgroups based on features like front, back and central. This particular feature set is chosen because the number of vowels occurring with central feature is less in number and a large number of vowels are distinguished by front and back features. Based on this classification, the vowels under first group can be subdivided as front vowels, back vowels and central vowels. For example, the subgroup having the common feature front which consists of characters /i/(इ), /i:/(ई), /e/(ए) will have a common rule called front detection rule. Whereas the characters /u/(उ), /u:/(3), /o/(ओ) will have a common rule called back detection rule and the rest of the vowel characters will have a rule called the central detection rule. Each of the sub-groups is further subdivided into smaller groups based on features like high, high-mid, mid, low-mid and low which are the remaining features distinguishing members within each sub group. These remaining features distinguish individual character within each sub group. Hence each character in the subgroup will have the necessary rule to detect the appropriate features.

Thus, in writing the rules for each of the character corresponding to vowel group there can be a set of rules common to a number of characters but there will be at least one rule that distinguishes one character from the other. The same set of rules can be used with slight modifications to account for contextual effects when these sounds are considered in the context of consonants, namely, recognize the vowel part in CV, CCV and CCCV combinations.

The second group corresponding to the other category can be further subgrouped based on suitable distinguishing features. In this group each character has a consonant part followed by vowel part. The consonantal part can be a either a single consonant or a cluster of consonants. Each single consonant can be either a stop, a fricative, an affricate, a lateral, a trill or a glide. A cluster can be a combination of the above. Using suitable features a number of subgroups can be formed. Members of the subgroup can use suitable distinguishing features to either form subgroups or reach individual characters. This can be carried through until each characters has appropriate distinguishing rules.

CV
CCV

The grouping is done based on

(1) The complexity of observing the feature. The less complex it is the easier it is to form the corresponding group. The more complex features are chosen as we form further subgroups. In this manner in the set of rules used for each character the gross features are all chosen first and the finer features

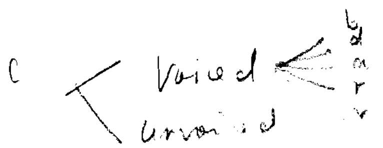
are chosen at the end. This improves the performance of character spotting.

(2) The feature chosen should group the characters based on the ~~the~~ similarity of the corresponding sounds. Among the characters with confusable sounds that are formed into a group more complex features are used to separate out individual characters. The complexity of a feature is based on deriving the feature from the parametric representation of the signal.

Based on these principles the feature voiced is chosen to classify the second group into voiced and unvoiced subgroups. Here the decision is based on the characteristics of the consonant occurring at the beginning of a speech sound. The voiced feature is used because it is easier to identify the presence or absence of this feature in the signal when compared to the features of place articulation.

Further grouping of speech sounds can be obtained by considering each one of the subgroups and their distinguishing characteristic. For example in the case of unvoiced subgroup, the classification can be based on the place of articulation namely, bilabial, dental, alveolar, retroflex and velar. The aspirated feature is moved to the last stage of grouping because when producing such speech sounds it is observed that not many speakers produce this difference in sound though this is a distinguishing feature.

At the final stage of classification, once the initial consonant segment is identified the groupings based on clusters can be done. Here the same knowledge base used for the



consonantal segments can be used to identify the different consonants in the clusters.

Once the consonant region is identified, the groupings developed for the vowel group can be applied to lead to the complete character. It is to be remembered that the groupings are done only to facilitate the listing of rules but ultimately each character will have its own set of acoustic correlates to reflect its individual nature.

phonetic ← acoustic correlates
rule / features
← character

3.5.2 *Acoustic Correlates of Character Set*

Having classified the character set into a taxonomy based on features and putting it in the form of a rule base as a starting point, it is now necessary to determine the acoustic correlates for the phonetic features in the context of each character.

The following list gives some of the parameters and features they determine as obtained from an acoustic-phonetic expert:

Parameter	Description	Feature Indicated
r_0	Total energy 0-5 KHz.	A high value indicates sonorant vowel or voiced consonant. A high value also indicates absence of unvoiced consonant and nasals.
r_1	first auto correlation	A high value indicates lack of high frequency energy (not a fricative)
Zero Crossing	-	A high value indicates fricative, voicelessness, silence or pause.
F_0	Fundamental frequency	Its presence indicates voicing
F_1, F_2 and F_3	First three	These signify the place of articulation of vowels.
Burst Spectrum	Peak of burst spectrum (frequency)	Its value indicates place of articulation for stops and nasals.
Dips	Dips in low frequently energy	It indicates presence of fricative, aspiration and burst.

Based on the above as well as other studies covered in Section 3.4, a general set of parameters which are useful in identifying phonetic feature in the signal are given below.

(1) From the time waveform and autocorrelation function

(a) Fundamental frequency or pitch

(b) Zero crossing rate

(2) From the Fourier transform spectra

(a) Root mean square energy

(b) Total Energy

(c) Very low frequency energy

(d) Ratio of high frequency to low frequency energy

(e) logarithm of energies in nonoverlapping frequency

(3) Autocorrelation coefficients

(a) r_0 - total energy

(b) r_1 - low frequency energy

or

Linear Prediction Coefficients

(a) First LPC - low frequency energy

(b) Spectral Flatness measure

(c) Formants 1, 2 and 3

(d) Nonoverlapping frequency bands of energy.

(4) Changes in parameters described above

(a) Dips in spectral parameters

(b) Sharp changes in intensity

(c) Abrupt drop in energy

(d) Gradual drop in energy

The parameters listed above are not exhaustive and it is not necessary that all these parameters be used in each character expert. The description of the character dictates the type of parameter to be used and also the method of extracting the parameters from the speech signal. For example when vowel segments are being identified in a character it may be necessary to have better spectral resolution whereas when consonantal segments are being analyzed, particularly in the case of stop consonants, it may be necessary to have better temporal resolution. Thus, knowledge-based signal processing can be done to extract relevant parameters using specific signal processing techniques.

In the case of vowels it is generally considered that position of first three formants are sufficient cues for identifying various features like central, back, front, high, high-mid, mid, low-mid, low, rounded and unrounded. Vowel speech sounds or their description in terms of phonetic features are conceived as points in an acoustic vowel space in which the coordinates are first and second formant frequencies. In addition, third formant is used to decide the feature rounded to nrounded. Generally for Hindi vowels it is observed that there is a quality difference as well as durational difference between long and short vowels. These are evident when comparing formant plots for the vowels /a/ (अ) and /a:/ (आ) in the Figs.3.12 and 3.11 respectively.

There are several acoustic cues that contribute to the identification of a stop consonant [2,5,6,8,10,38]. The acoustic subunits that can be used for classification of stops are silence followed by burst followed by presence or absence of aspiration. We use spectral distance measure [37] and spectral flatness

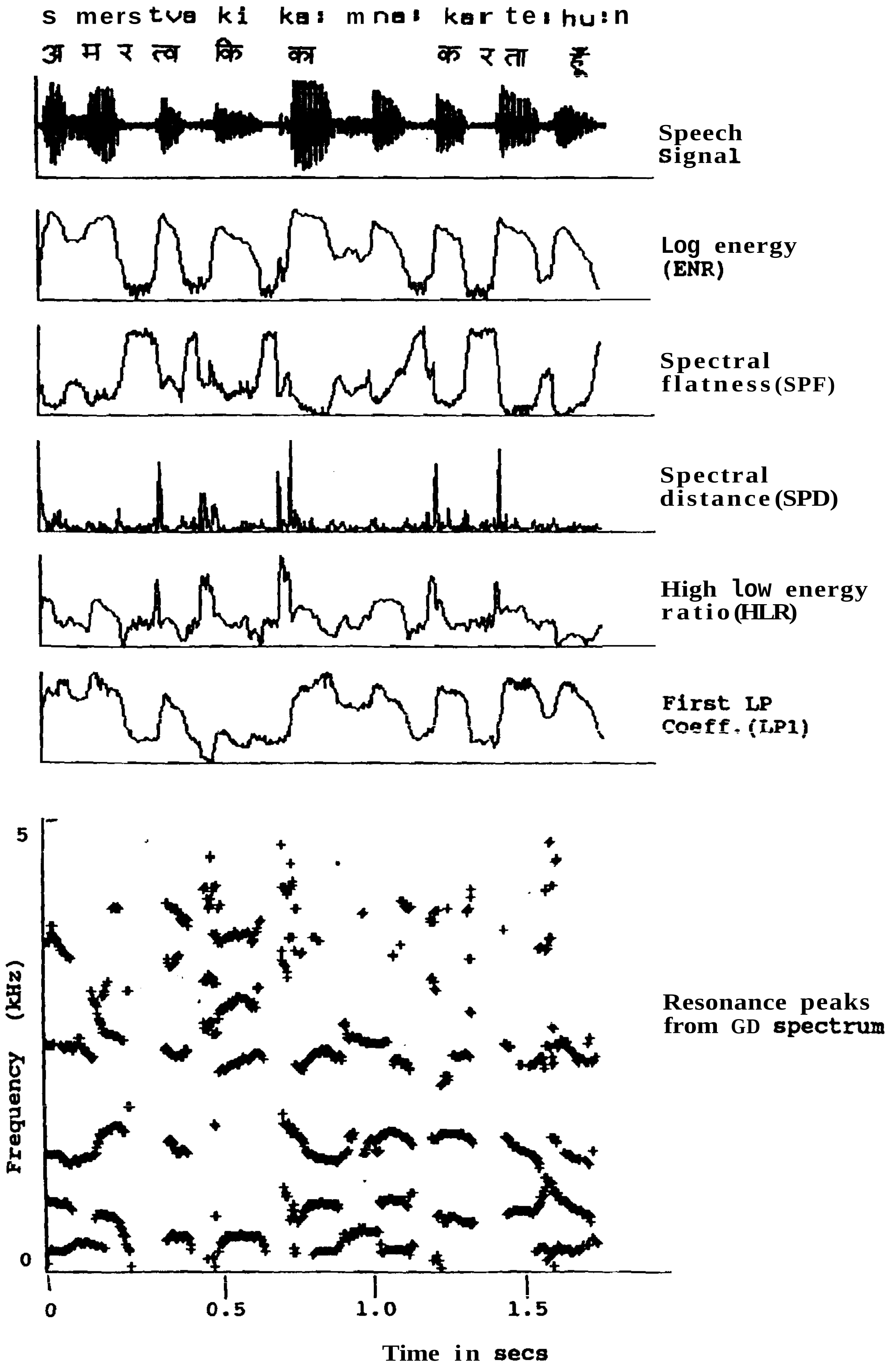


Fig. 3.12. Speech signal and other parameters for comparison of formants for /ka/(क) and /ka:/(का)

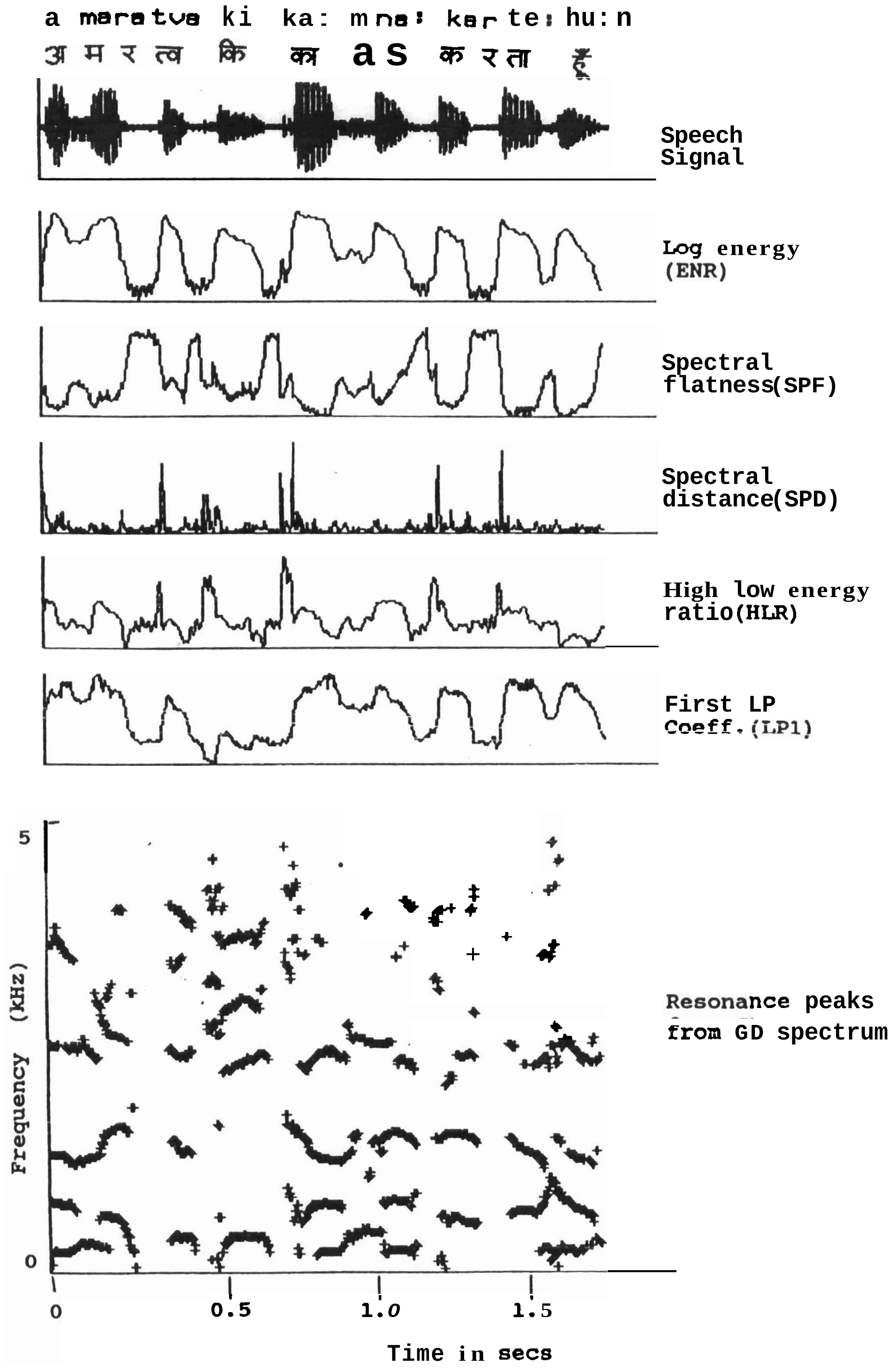


Fig. 3.12. Speech signal and other parameters for comparison of formants for /ka/(क) and /ka:/(का)

measure to take a decision about voiced and unvoiced silence. Location of the burst region can be accomplished by using the condition that there is predominance of high frequency energy over low frequency energy in case of unvoiced stop consonants. In case of voiced stop consonants pitch or voice bar is present and there is effect of low frequency energy in addition to high frequency energy. The abrupt changes in burst release are indicated by spectral distance measure. This is clearly evident in Fig.3.13. The duration of the burst is generally of the order of 5 to 15 milliseconds and that of transition varies from 20 to 70 milliseconds depending on whether the consonant is aspirated or not.

Frequency spectrum of the burst gives some clues to the place of articulation of the stop consonants [4,24]. This is normally obtained by doing special kind of signal processing in the regions which are identified as end of silence or in the burst regions or in the regions which are to the left of vocalic regions when either burst or aspiration is not correctly hypothesized. The location of peak of the burst frequency spectrum gives information about the place of articulation. In addition, the formant transitions into the vocalic region give additional information about the place of articulation of the stop consonant. This clue is better utilized when we consider the stop consonant in context. The advantage of the written characters of a language as symbols in our case is very much evident here. Because most of the characters we consider are CV combinations, the required transition information is captured in the grossly hypothesized region of the character. We look for

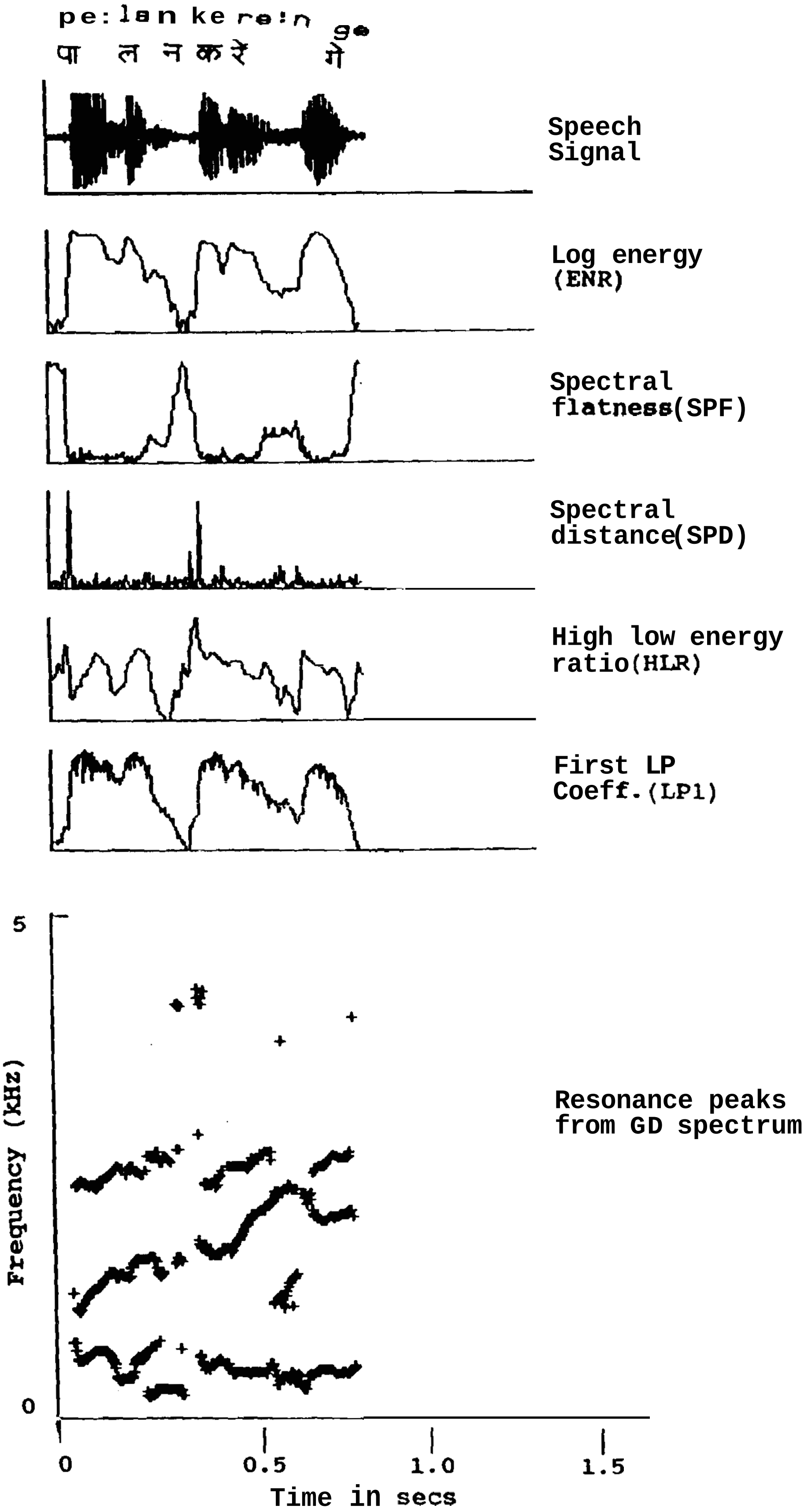


Fig. 3.13. Location of burst using spectral distance measure

place of articulation cues after identifying a particular segment as CV, where C is a stop consonant and V is a vowel. We know the variations of the formants a priori because we know the kind of vowel that follows the consonant. The formant frequency transitions are also known because the target vowel is known.

The peak of the burst spectrum takes on different values in terms of frequency for different places of articulation. If this frequency is very low, then it corresponds to labial place of P
articulation. When it is high it corresponds to dental place of t
articulation. A value in between the two corresponds to velar k
place of articulation.

The voiced and unvoiced nature of the stop consonants can be identified using number of cues [2,27]. These are the voice onset time, transition of first formant, duration of closure and also the duration of the preceding vowel. In case of unvoiced A
stops the duration of closure is found to be smaller--as compared P
to that of voiced closures. Voice onset time is less for voiced stops when compared with that of unvoiced stop consonants. The rising transition of first formant clearly indicates that the manner of articulation of stop is voiced whereas steady nature of the first formant indicates an unvoiced stop consonant. These cues are evident in most of the data analyzed and a few examples are shown in Fig.3.14.

The aspirated and unaspirated nature of the stop consonants characterized by the duration of the stop consonant. In most of the aspirated consonants there is a definite delay in the onset of vowel after the release of the burst. In addition, these regions have formant structure. Formant structure is indicated by

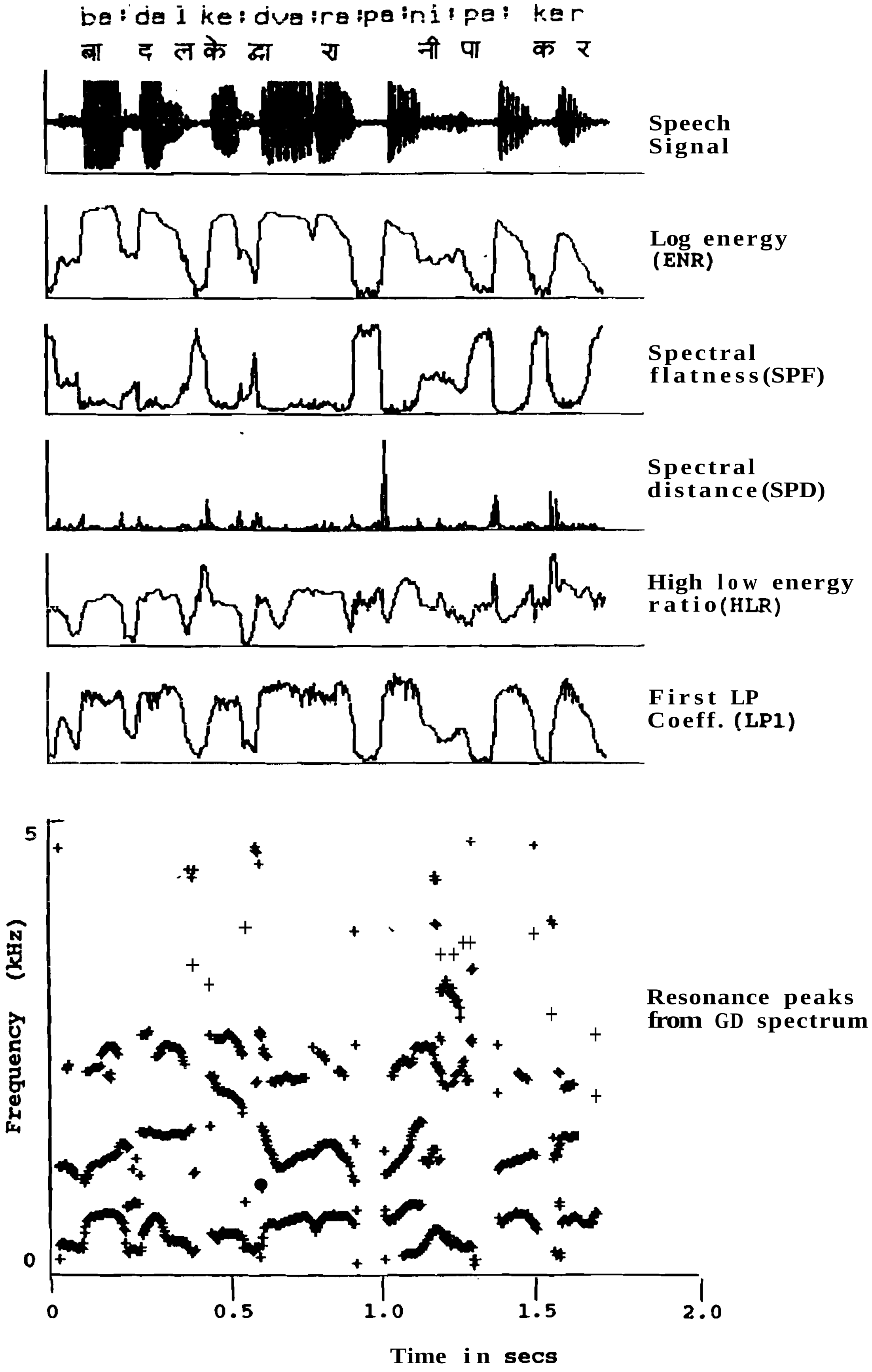


Fig. 3.14. Speech signal and other parameters for studying voiced/unvoiced features

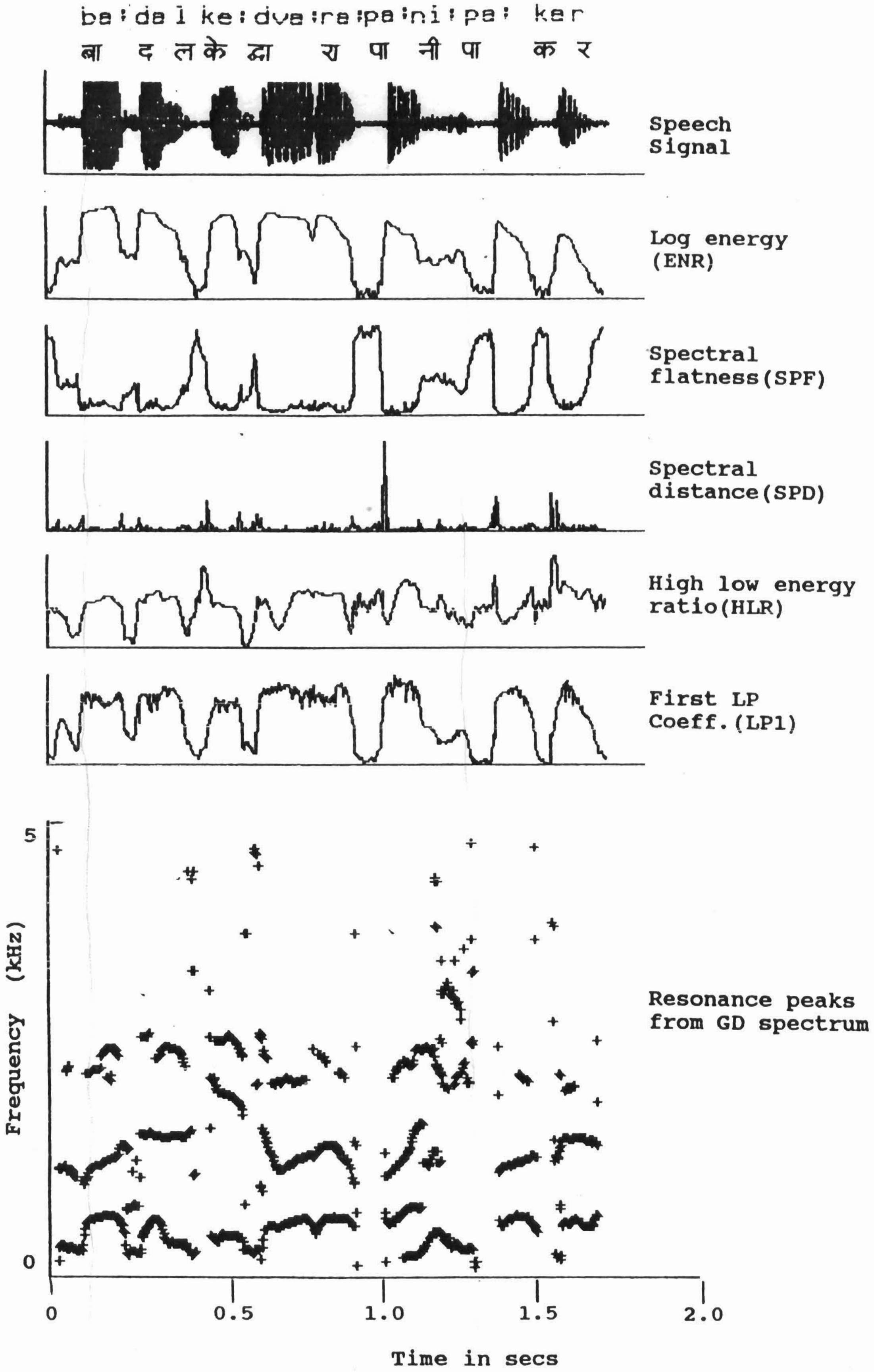


Fig. 3.14. Speech signal and other parameters for studying voiced/unvoiced features

low and mediun. values of spectral flatness (SPF) measure and presence of noise energy is indicated by a high value if high frequency to low frequency energy ratio (HLR). These characteristics are indicated by spectral flatness measure as well as HLR. This is clearly evident as shown by Fig.3.15.

The onset of formant frequencies and formant transitions give the required acoustic correlates for finding the place of articulation of nasal stops [19,31]. In addition, the dips in vocalic regions can be used to locate the nasal stops in case they occur as intervocalic nasals. Irrespective of the place of articulation the nasal consonants indicate presence of a dominant low frequency component around 250 Hz, which gives a gross clue for the location of nasal consonants in addition to the dips in energy. This fact is clearly evident in the Figs.3.16 and 3.17.

Identification of laterals, trills and glides depend on the onset frequencies and direction of formant transitions. For example, if the formant transitions are rising relative to the steady state frequency of the vowel, then it can be identified as /v/(व्). In contrast, if F_2 is falling and F_3 is steady, then it can be identified as /l/(ल्). In all these cases the formant transitions are much longer than in the case of stops. The threshold here is of the order of 60 msec.

Primary acoustic cues for fricatives [32] are their manner of articulation and their place of articulation. These can be voiced or unvoiced fricatives. The onset of noise should be gradual. The distribution of the spectral peaks of the noise contribute to the identification of the place of articulation fricatives.

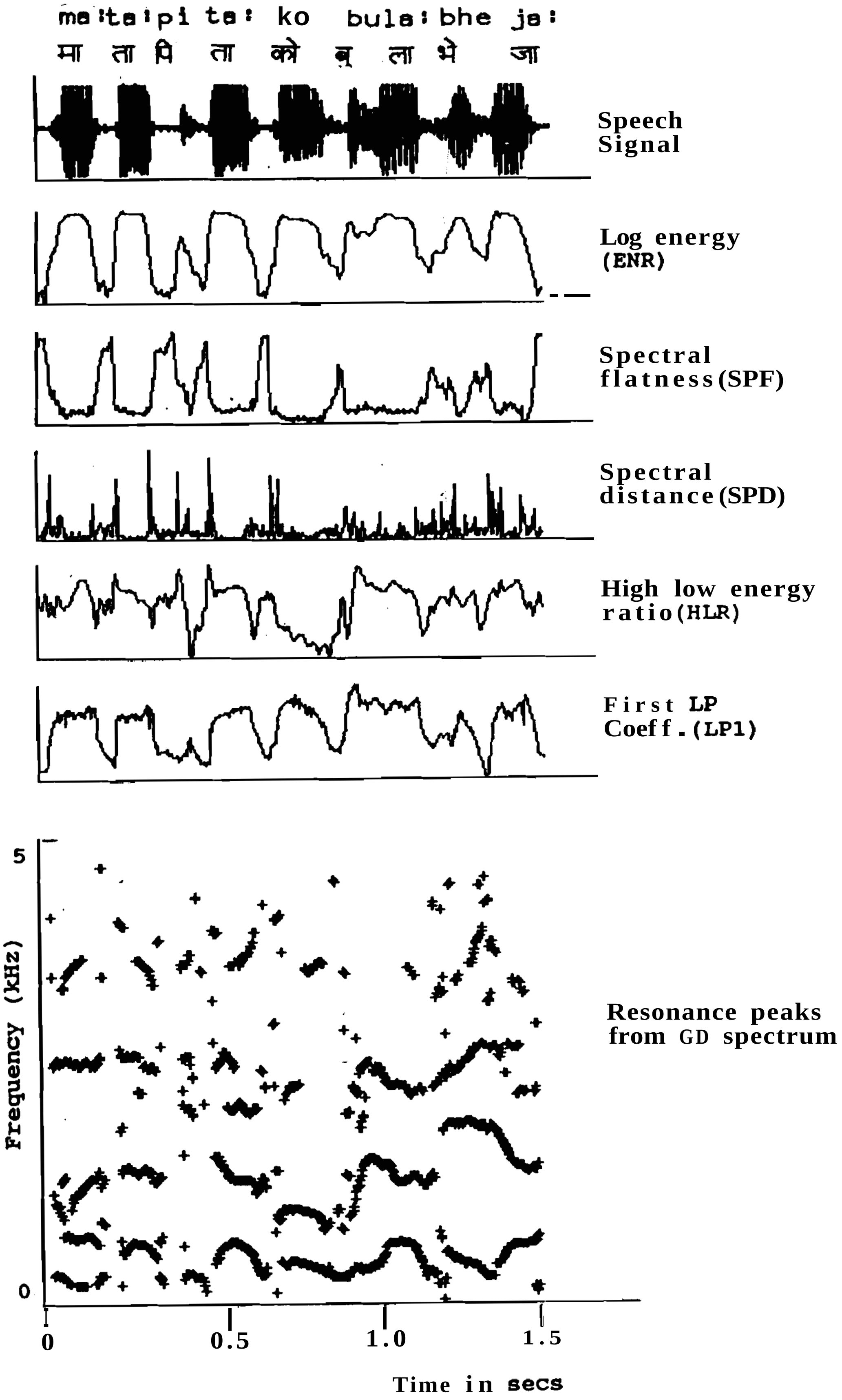


Fig. 3.15. Speech signal and other parameters for studying other consonant sounds

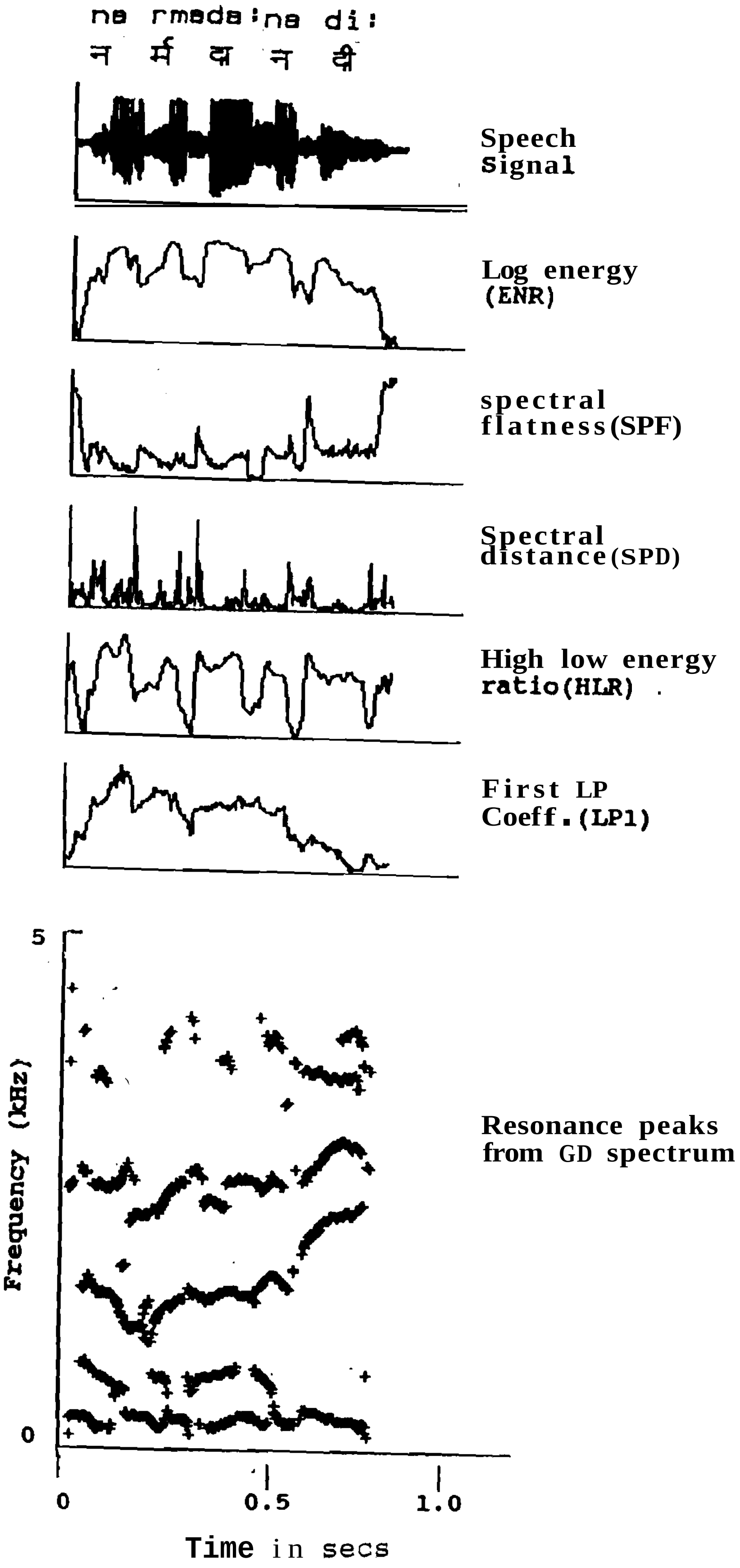


Fig. 3.16. Speech signal and other parameters for studying nasal sounds

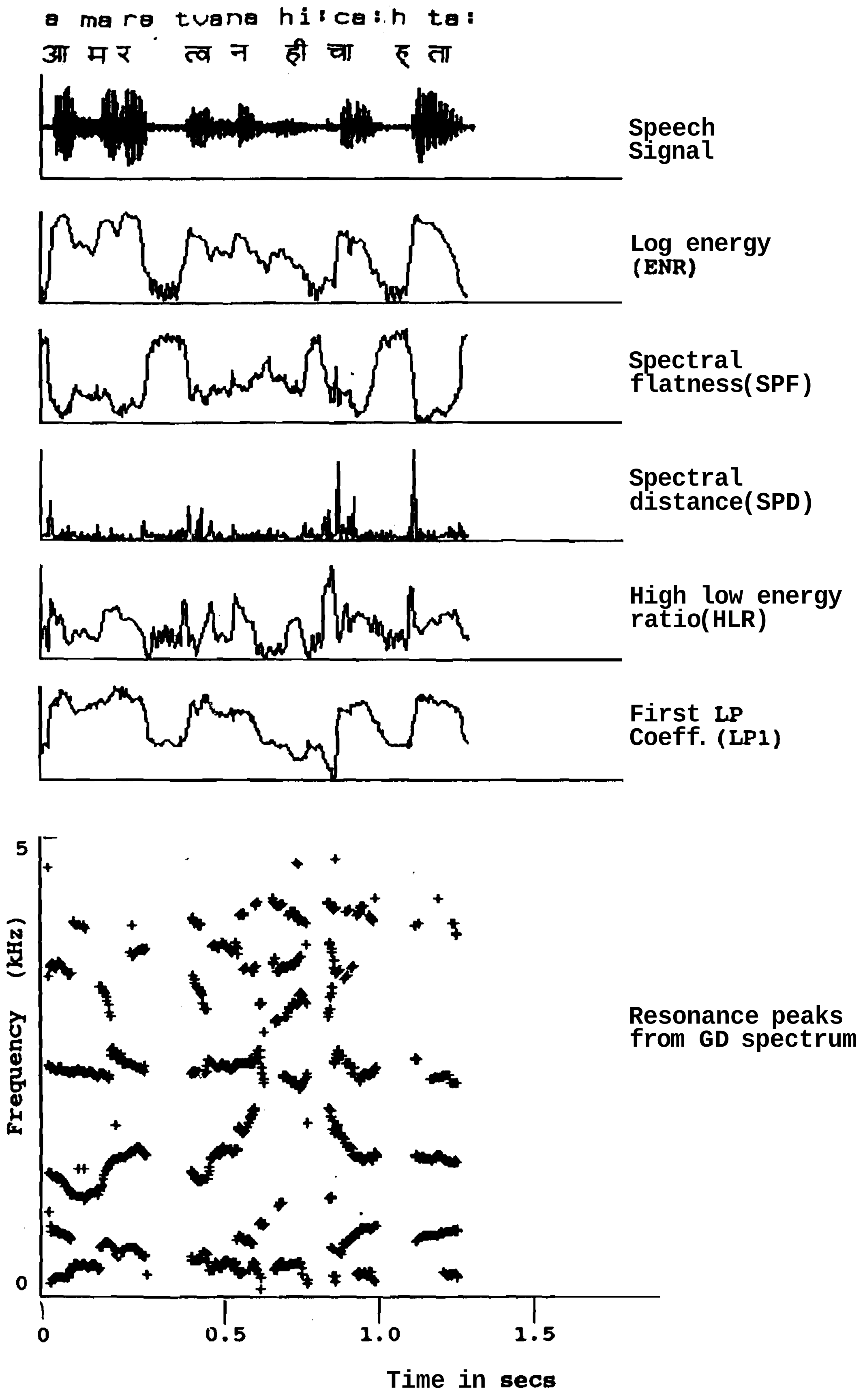


Fig. 3.17. Speech signal and other parameters for studying nasal sounds

3.5.3 Acoustic Invariance and its Use in Character Spotting

Invariant acoustic cues [66] refer to acoustic patterns derived from the signal which remain stable despite the fact that there are many sources of variability affecting the structure of the speech signal. The variations in features arise due to (1) different vocal tract sizes and shapes, (2) different contexts in which the speech sounds occur and (3) rate of speech and stress. In summary, some of the invariant properties that have been identified are:

(1) The generalized patterns in the frequency and amplitude domain. That is spectral patterns which will not be dependent on fine acoustic structure such as formant frequencies say for place of articulation of stop consonants.

(2) Onset properties can be used for discriminating consonant place of articulation and vowel quality.

(3) The amplitude characteristics and the way the acoustic energy changes, and the frequencies at which these acoustic changes occur can be used to discriminate stop consonants, fricatives glides and vowel speech sounds.

(4) Presence of additional peakd in case of nasals and nasalised sounds.

(5) Low frequency periodicity and nature of its change.

(6) Voice onset time and vowel onset time.

Summary

In this chapter the derivation of the acoustic-phonetic knowledge for character spotting is discussed. A brief description of the speech production mechanism is given to provide the necessary background. A set of phonetic features is identified based on the articulatory description and the complete character set is defined in terms of this feature set. Various sources of knowledge are identified to study the phonetic features in terms of acoustic behavior. The choice of parameters derived from the signal and their characteristics are highlighted. The use of various sources of knowledge to implement the character spotting approach is discussed. In the next chapter we shall see the details of illustration of one character expert.

CHAPTER - IV

ILLUSTRATION OF A CHARACTER EXPERT

We propose to develop a rule-based expert system for each character to implement the character spotting approach. For this it is necessary to collect the acoustic-phonetic knowledge for each character and represent the knowledge in a suitable form. In the previous Chapter we have presented the basics of acoustic-phonetics and the sources of this knowledge. In this Chapter we will illustrate how to acquire the knowledge for a character and how to represent the knowledge in the form of rules. We take the character /ka:/(का) for illustration. Speech production mechanism for the character and the description of the character in terms of phonetic features are given in Section 4.1. The description of the character in terms of speech parameters is derived in Section 4.2. Combining the descriptions given in the Sections 4.1 and 4.2, the categories of rules for the character are presented in Section 4.3. This Section also explains the organization of the rules for a character. Representation of the knowledge in the form of IF...THEN rules is illustrated in Section 4.4. Issues in the implementation of a character spotting expert system are discussed in the next Chapter.

4.1 Phonetic Description of the Character /ka:/(का)

The character /ka:/(का) is a consonant-vowel (CV) combination. speech production mechanism of the consonant and vowel parts of the character /ka:/(का) and the phonetic description of the character based on the speech production mechanism are presented in this Section.

The speech production mechanism of the consonant part of the character /ka:/ (का) is as follows [57]: The back part of the tongue is engaged in the formation of a complete barrier which prevents the air stream from passing through the mouth cavity. The velum is raised and the passage of the air stream through the nasal cavity is blocked. The lips as well as the lower jaw are in neutral position. The central and back part of the tongue are raised to the velo-palatal area, filling the arch-like space of the back part of the roof of the mouth, so that air streaming through the lungs is not able to break through. The tip of the tongue lies behind the lower front teeth. A shallow valley is formed in the front part of the tongue and the middle portion of it is raised. The front part and the tip of the tongue are loose and the mid and back part display tensity especially in the area where the contact between the tongue and the roof of the mouth is formed. The vocal cords are silent and the consonant is voiceless. The air stream which is stopped in the laryngo-pharyngeal cavity is not compressed.

The sudden release of the back barrier results in a burst which is, from the point of view of European languages, exceptionally short. It is shorter than the plosive /k/ in English words like /keg/. The burst is very abrupt and clear. The main characteristics of burst of Hindi plosive /k/ (क) is the absence of aspiration unlike the French, Italian and English /k/. The area of articulation for Hindi plosive /k/(क्) in /ka:/( ) is considerably wider. It is largely related to the position of the tongue for the following vowel. If a back vowel follows the sound /k/ (क) as in /ka:/(का) then it will be articulated in the palato-velar region.

During the production of the vowel part of the character /ka:/(का), the jaw angle is open and the tip of the tongue is placed behind the ridge of the lower teeth, not pressing it. The front part of the tongue is slightly hollowed in its mid area and middle part of the tongue is elevated to a slight degree towards the position of the back part of the tongue which is pulled back to a gently sloping position. This vowel is long in duration but is not dipthongised and the quality is stable. There is no rounding of lips during its pronunciation.

The duration of the sound /a:/(आ) in /ka:/(का)varies a great deal. It is usually long, but it may be reduced to the length of a short vowel without losing its quality. The length of the sound depends on the context such as stress, degree of emphasis, position in the word and speaker. The variability of the length of this vowel does not affect its quality. This is similar to the observation made by Ohala[77] that both quality as determined by the formants and length characterize this vowel part in /ka:/(का).

The speech production mechanism as given above the phonetic description of the character /ka:/(का), in terms of the features presented in the previous Chapter, is as follows :

Unvoiced Unaspirated Velar Plosive followed by k (1) Phonetic features

Voiced Long Open Back Vowel a: /ka:/

The relationship between each of these features and the speech signal parameters is derived in the next section.

4.2 Parametric Description of the Character /ka:/(का)

As we have to obtain the features in the phonetic description of a character from the speech signal in order to spot the character, it is necessary to derive the relationship between these phonetic features and some measurable speech signal parameters. This relationship is, in general, one to many because the same feature may be related to speech parameters in different ways in different contexts. Therefore, the parametric description of the phonetic features has to be derived by taking a particular context into account. In this section, we derive the relationship between the phonetic description of the character /ka:/(का) and speech parameters. The parameters used for analysis of speech sounds in the present study are described in the previous chapter.

According to the phonetic description of the character /ka:/(का), the consonant part is an unvoiced plosive. The speech signal of an unvoiced plosive is characterized by a silence region corresponding to the closure part of the plosive, followed by a burst of energy corresponding to the release part. The silence region cf an unvoiced plosive is characterized by low values of the total energy (ENR) and the low frequency energy (represented by the parameter LP1), and a high value of spectral flatness measure (SPF). The burst region of the plosive is: characterized by a high value of the parameter HLR, which is the ratio of the high frequency energy to the low frequency energy, and a mid value of the total energy (ENR).

Ground features of the phonetic feat. (using param

Though the consonant part of the character is described as an unaspirated plosive, according to the expert phonetician there is a slight amount of aspiration present in the production of this particular consonant. This has been confirmed through spectrographic studies. The duration of the aspiration region in the character /ka:/(का) is much less than that of the aspirated consonant in tho, character /kha:/(खा). So, the plosive in /ka:/(का) is described as having a short period of aspiration also. The aspiration region is characterized by the spectral flatness measure (SPF) which is neither high nor low. The high frequency energy (represented by the parameter HLR) in this region is high. The duration of the aspiration region is obtained as the period between the end of the burst region and the voice onset time (VOT) of the following vowel. The VOT is detected by a very high value of the spectral distance (SPD) just before the vowel-like region.

The place of articulation of the plosive, *Velar*, in the context of the character /ka:/(का) is identified by the presence of a peak in the low frequency region of the burst spectrum. This feature is also characterized by the falling transition of the second formant (F2) and coming together of the second and third formant frequencies (F2 and F3) of the following vowel.

Having given the parametric description of the consonant part, we now derive the relationship between the phonetic description of the vowel part of the character /ka:/(का) and the speech parameters. The speech signal in the vowel part is characterized by high values of the total energy (ENR) and the low frequency energy (LP1). These parameters are used to identify vowel-like regions in the speech signal. The phonetic description of the vowel as such is represented mainly by the first and second formant frequencies (F1 and F2) and the difference between F1 and F2. The *Long* feature of the vowel in Hindi is characterized by the duration of the vowel region as well as the quality of the vowel. The quality of the vowel is mainly represented by the formant frequencies. The Back feature of the vowel is identified by a low value of the difference between F2 and F1. The feature *Open* is characterized by a high value of F1 and the feature *Unrounded* by a low value of F2.

Because of the variant nature of the speech signal, it is not possible to fix a range of values for each of the parameters describing a feature such that the range will hold good always. Therefore a range of values is arrived at for each parameter describing a feature after analyzing the speech data for many instances of the character. Fuzzy mathematical techniques (discussed in the next chapter) are used to take care of the cases in which the parameter takes a value just outside the range specified.

The phonetic description of the character and the parametric representation derived for each of the features in the description are used to develop the knowledge base of the character spotting expert system. The method of forming the acoustic-phonetic rules and organization of the rules is the topic of the next section.

rules. = character /ka:/
phonetic feature
acoustic feature
[FUZZY]

43 Categories of Rules for the Character /ka:/(का)

A character spotting expert system activates the acoustic-phonetic rules to estimate the presence of the phonetic description of the character in the speech signal and then hypothesize the presence of the character in the speech signal. In this process, it extracts the necessary parameters from the speech signal. Extraction of parameters like formant frequencies involves computationally intensive processing of speech signal. In order to reduce the total computation time, it is decided to first locate the regions of the speech signal in which the character is likely to be present based on some gross features, Further processing of speech signal is done only in these located regions to extract the parameters necessary for capturing other features in the description of the character and then hypothesize the presence or absence of the character.

taken

The set of gross features for a character consists of features that capture some description of different segments of the character using parameters which can be extracted by simple processing of speech signal. Location rules combine these gross features in a manner governed by the description of the character to identify the regions in which the character is likely to be present. Subsequently, hypothesization rules representing the intrinsic cues, coarticulation cues and context-dependent cues are activated in the located regions to hypothesize the presence or absence of the character [28].

Intrinsic cues capture the invariant properties of the segments in the character. Coarticulation rules capture the variations in the acoustic correlates of speech segments due to the influence of other segments within a character. context-dependent cues capture the variations that the acoustic correlates of a character undergo in different contexts.

The rules for spotting the character /ka:/(का) are organized into different categories as shown in Table 4.1.

Table 4.1 Rules for Spotting the Character /ka:/(కా)

GROSS FEATURES

- * Rules to locate vocalic regions
- * Rules to locate closure regions
- * Rules to locate burst regions
- * Rules to locate aspirated regions

LOCATION RULES

- * Rules to check whether the above gross features conform to the description of the character /ka:/(కా) to hypothesize the possible presence of the character

RULES FOR INTRINSIC CUES

- * Rules to identify the vowel by checking the formants
- * Rules to identify the place of articulation by checking burst spectrum
- * Rules to provide appropriate confidence for identification of character

RULES FOR COARTICULATION CUES

- * Rules to check the effect of vowel on the consonant as given in description
- * Rules to check the effect of consonant on vowel
- * Rules to provide appropriate confidence with which the character is identified

RULES FOR CONTEXT-DEPENDENT CUES

- * Rules to check the effect of adjacent characters as provided by the description to arrive at an overall confidence value with) which the character is identified

4.4 Acoustic-Phonetic Rules for the Character /ka:/(కా)

The gross features, location cues, intrinsic cues, coarticulation cues and context-dependent cues for the character /ka:/(కా) are enumerated in this section.

(1) Gross Features

The gross features for /ka:/(కా) and the parameters used for their identification are as follows :

Silence	Low frequency energy, spectral flatness and total energy
Burst	Total energy and ratio of high frequency energy to low frequency energy
Vocalic	Low frequency energy and total energy
Aspiration	Presence of high frequency energy after burst, spectral flatness and definite delay in the onset of vowel

Parameter plots for two utterances in which the character /ka:/(కా) is present are given in Fig.4.1 and Fig.4.2. The regions where /ka:/(కా) is present are marked in the figures. The parameters log energy (ENR), spectral flatness (SPF), spectral distance (SPD), high frequency to low frequency energy ratio (HLR) and first linear prediction coefficient (LP1) are normalized with respect to their maximum and minimum values in the utterance. Normalized values range from O to 255. The threshold ranges for each parameter used in the rules for

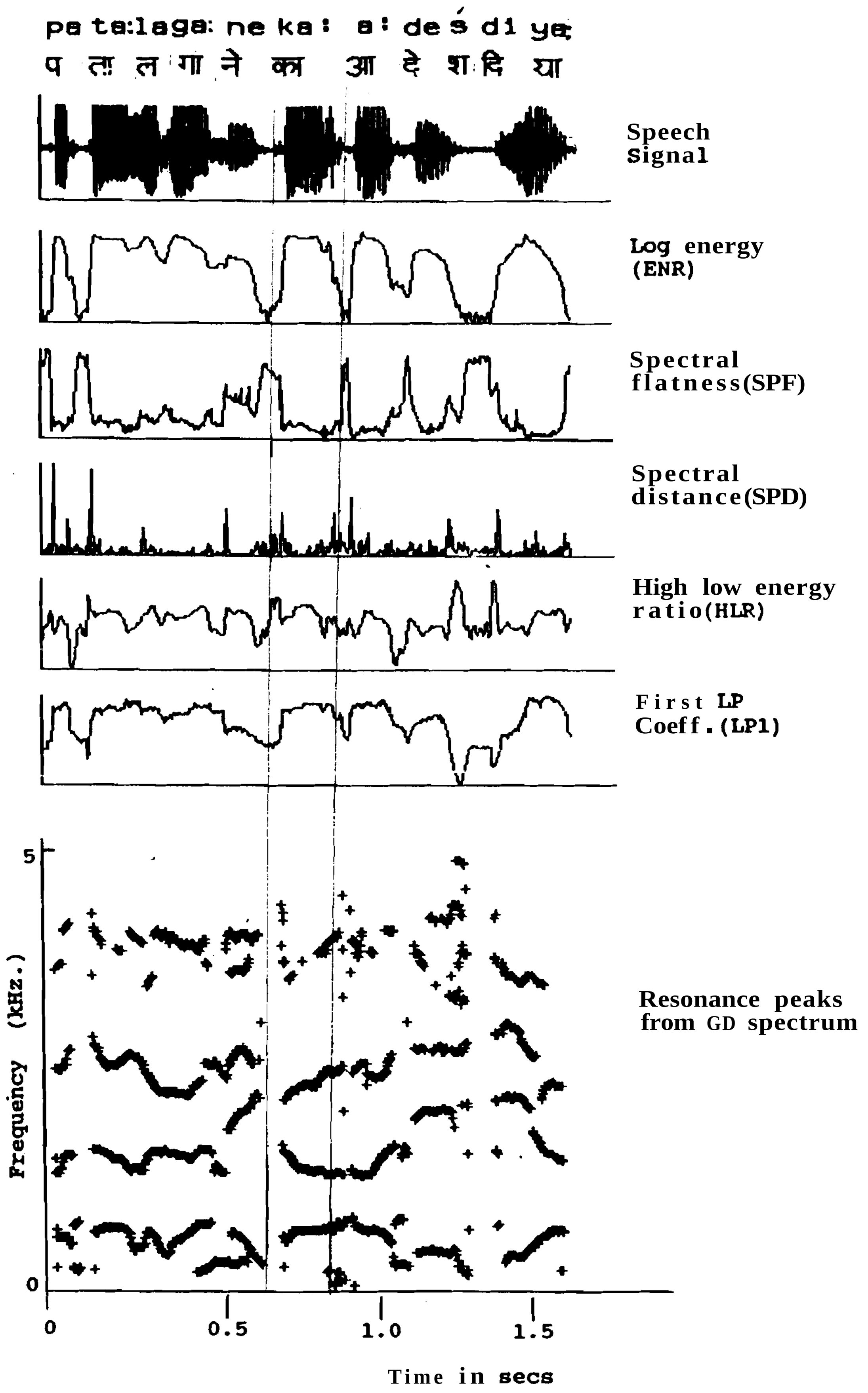


Fig. 4.1. Parameter plots to illustrate the gross features for /ka:/(का)

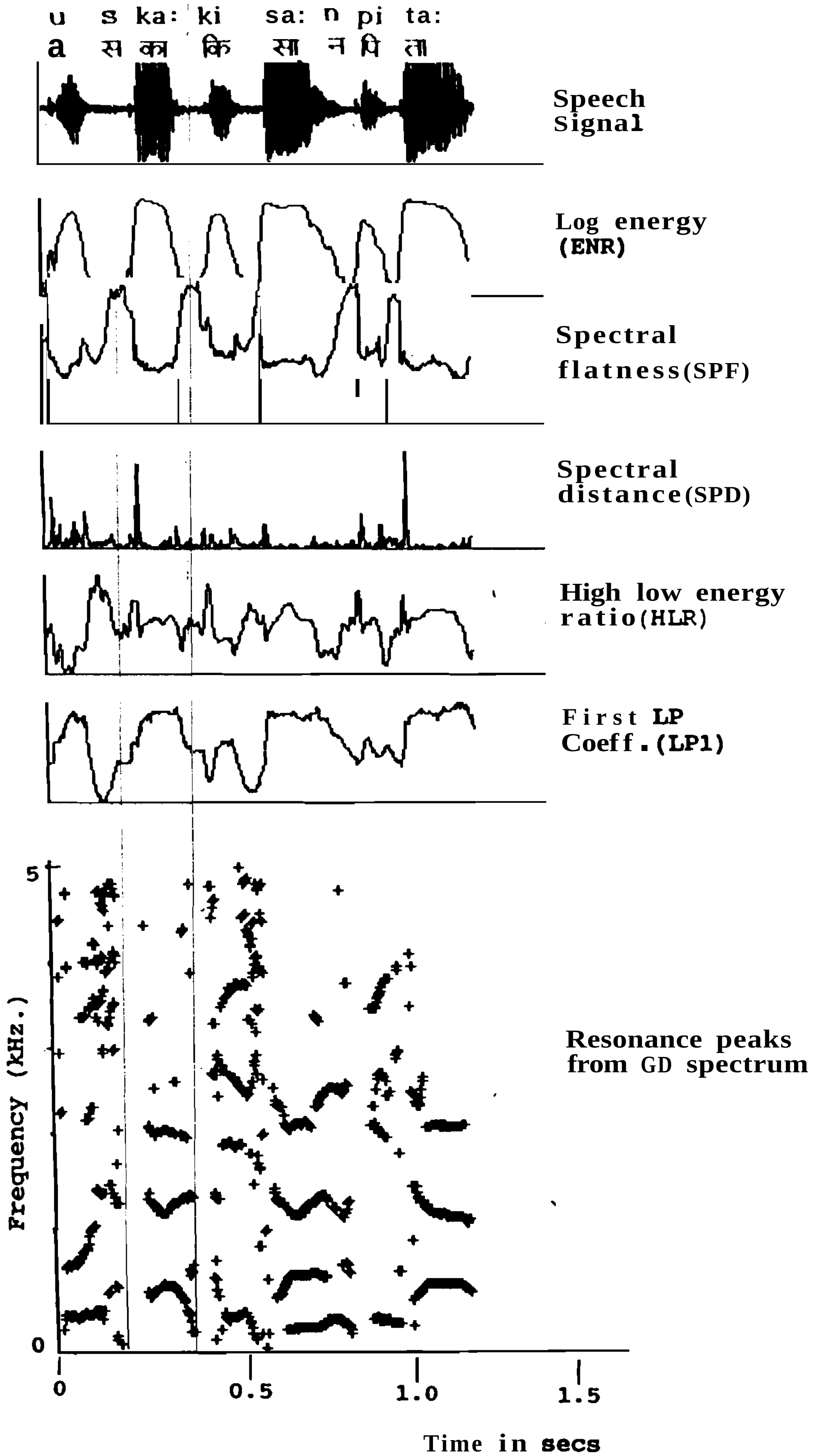


Fig. 4.2. Parameter plots to illustrate the gross features for /ka:/(का)

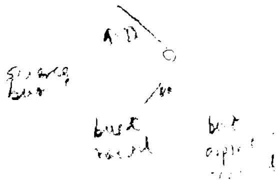
identification of gross features are specified in terms of the normalized values. The rules for identification of gross features are as follows:

```
(1)     IF   ((LP1 is below 100)  and
              (SPF is above 175)  and
              (ENR is below 100))
        THEN locate silence region

(2)     IF   ((ENR is in the range 100  to 150) and
              (HLR is above 150))
        THEN locate burst region

(3)     IF   ((LP1 is above 175)  or
              (ENR is above 200))
        THEN locate vocalic region

(4)     IF   ((SPF is between 125 and 200) and
              (HLR is between 100 and 200))
        THEN locate aspiration region
```

(2) Location Rule

The location rules use the description of character /ka:/(का) to combine the gross features detected and locate the regions where the character is likely to be present. The location rule for /ka:/(का) is as follows:

```
IF    ((silence is followed by burst) and
        ((burst is followed by voiced) or
         ((burst is followed by aspiration) and
        (aspiration is followed by voiced))))
THEN  mark the region as a located  region
```



In the case of /ka:/(का) the gross features are silence, burst, aspiration and vocalic. As per the description of /ka:/(का), silence should be followed by burst before onset of vowel. Sometimes the features detected may not be in sequence. There may be overlapping regions of gross features. At times some features may be completely missing. This is taken care of by the location rules in the way the gross features are combined.

It is observed that the feature aspiration which does not form part of the description of /ka:/ (का)is used in the location rules because from the signal and parameter plots we observe certain amount of aspiration in /ka:/(का). When the location rules are satisfied, the beginning of silence region to the end of vocalic region is marked as a located region of character /ka:/(का).

(3) *Intrinsic Cues*

The rule for capturing the intrinsic cues of the vowel part of /ka:/(का) are based on formant frequency information. F1 and F2 in the rule indicate the first and second formant frequencies.

```
IF    ((F1 is in 600-700 Hz) and
       (F2 is in 1000-1200 Hz) and
       ((F2-F1) is 300-600 Hz))
THEW  hypothesize the vowel of /ka:/ ( का )
```

The invariant feature for place of articulation of the plosive is the presence of burst in the signal and the frequency of the peak of the burst spectrum. The corresponding rule is as follows:

```
IF    (Frequency of the peak of the burst spectrum is in the
       range 1000-1200 Hz)
THEN  hypothesize the place (velar) of /ka:/ ( का )
```

(4) Coarticulation cues (variation)

These are obtained only for the consonant part within the character in a CV context based on formant transitions. The corresponding coarticulation cues are:

i) F2 transition is falling

ii) F2 and F3 come closer

(5) Contat-dependent Cues (T@;&.)

These are obtained for adjacent character context from reading spectrograms. The corresponding context-dependent cues are:

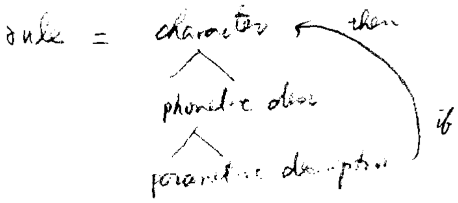
i) Frequency of the peak of the burst spectrum is in the range 900-1200 Hz, if preceded by a back vowel.

ii) Frequency of the peak of the burst spectrum is in the range 1500-2000 Hz, if preceded by a front vowel.

The acoustic-phonetic rules derived from the phonetic and parametric descriptions of the character are represented in the form of IF...THEN rules in the character expert. Issues involved in the implementat —— . of the character expert are discussed in the next chapter.

Summary

In this chapter the concept of character spotting in signal-to-symbol transformation is explained with the help of an illustration. The speech production mechanism of the character /ka:/(का) is presented and the phonetic description of the character is given. The process of deriving the rules from parametric representation of the utterance of a character is discussed. The next chapter gives the details of implementation of character expert systems.



CHAPTER - V

IMPLEMENTATION OF A RULE-BASED SYSTEM

Acquisition and representation of the acoustic-phonetic knowledge is to be followed by activation of the knowledge to accomplish the task of spotting characters in continuous speech. The objective of this chapter is to discuss the issues in implementing a rule-based system for character spotting. Expert system approach is followed in the proposed implementation. In Section 5.1 different methods of activating a rule base are discussed followed by a discussion of the method adopted for activatingthe acoustic-phonetic knowledge. The imprecise nature of the knowledge at the phonetic and parameter levels calls for interpreting rules suitably to provide confidence measures for the outcomes of the rule interpreter. We propose the use of fuzzy mathematical concepts to derive the confidence measures. These concepts are discussed in Section 5.2. With this background the expert system implementation of character spotting is presented in Section 5.3. The section also discusses the design issues involved in the implementation. Finally in Section 5.4 we give the detailed working of a character spotting expert system. Performance of the expert system for several characters is illustrated in Chapter VI.

5.1 Activation of Acoustic-Phonetic Knowledge

The rule-based character expert consists of three components, namely, (1) production rules which describe the relation between the character and speech parameters, (2) a data memory and a working memory, where the data memory consists of speech data and the working memory contains the results of the processing when a rule is fired and (3) an inference engine to interpret the rules. Each production rule is an expression which consists of two parts called antecedents and consequents. They typically take the following forms:

IF (condition) THEN (assertion)
IF (antecedent) THEN (consequent)
IF (condition) THEN (action)

Based on the control strategy that is used to activate the productions, rule-based (or production) systems may be classified as pure production systems or application oriented .(or performance) production systems. In pure production systems a simple recognize-act control strategy is used. In this strategy, all conditions in each production are evaluated, the action specified by one of the applicable productions (which is determined a priori) is executed and the status of the working memory updated. This is repeated until there are no more applicable productions or the goal condition has been satisfied. In contrast, performance production systems are complicated. The rules may be grouped into subclasses and sophisticated conflict resolution strategies may be employed to choose the appropriate rule. In performance production systems, there are two different control strategies that are used to activate the knowledge base, namely, antecedent driven (or data driven) strategy and consequent driven (or expectation driven) strategy.

In the antecedent driven strategy the IF portion of a rule is compared with the current state of the working memory, if the conditions are matched, then the rule is fired, and the contents of the working memory are updated. This process is repeated until a goal is reached. On the other hand, in the consequent driven strategy, the activation starts with the goal element it is wishes to establish. The rule base is then searched to find a rule whose THEN portion is the required goal. The IF portion of this rule provides new subgoals. The rule base is again searched for rules to establish these subgoals which may in turn require that new subgoals be established. This procedure is repeated until all the subgoals are established.

Conflict resolution is an important part of performance production systems. Conflict resolution is a strategy that is employed to choose one rule from a set of rules whose IF portions satisfy the current status of the working memory. Several methods exist for handling conflict resolution [58].

The proposed character expert belongs to the class of performance production systems. The control strategy employed is neither antecedent driven or consequent driven. The strategy is goal oriented, but the goal which is spotting a character is fixed. The organization of the rules is such that they look only for those features (or parameters) in the speech signal that are relevant to the description of the particular character. There is no necessity for a complex conflict resolution strategy in this system. This is because each character is described uniquely. In addition, the productions are ordered as location cues, intrinsic cues, coarticulation cues and context-dependent cues. In each of these groups the rules are again ordered, based on the description of the character. Thus the control strategy is built into the ordering of the production rules based on the description of the character. Once the order of the rules is arrived at, it is a simple operation of sequentially firing each rule until there are no more rules to be fired.

Another important variation from the normal production systems is that at each stage of analysis the hypothesis is not binary. We explain in the next section why it is so and how we interpret the results in the signal-to-symbol transformation.

5 3 Fuzzy Mathematical Concepts Applied to Knowledge Activation

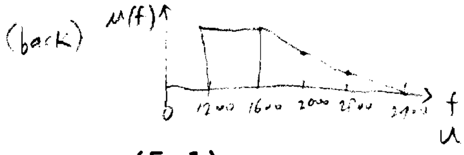
In rule-based decision making using the knowledge obtained from a human expert, it is accepted that the human supplied rules not only have inconsistencies but are also incomplete. In the character spotting expert, in addition to the acoustic-phonetic knowledge which is ambiguous, the parameters that are extracted from the speech signal are also imprecise. This complicates the inferencing mechanism when spotting a character in continuous speech. In order to compensate for these deficiencies, the rules of the expert system for character spotting,instead of taking a binary decision award a confidence measure for the inferences made and sometimes to the rule itself. The awarding of confidence measures is accomplished using Zadeh's theory [105] which proposes the use of fuzzy relations and restrictions to represent the knowledge as correctly as possible.

5.2.1 Relation between the Phonetic Description and Numeric Values

Fuzzy restriction is a " relation which acts as an elastic constraint on the values that may be assigned to a variable". Such restrictions play an important role in speech recognition and particularly in signal-to-symbol transformation because the environment there happens to be fuzzy or uncertain.

Let 'f' denote a numerically valued variable corresponding to the second formant frequency which ranges over the interval 900-2700 Hz. With this interval regarded as the universe of discourse, U, back may be interpreted as the label of a fuzzy subset of U which is characterized by a compatibility function $\mu(f)$. The function $\mu(f)$ may be viewed as the membership function of the fuzzy subset back. The value $\mu_f = \mu(f)$ at the corresponding frequency 'f', represents the grade membership of 'f' in back. For example, let back a fuzzy subset of U consist of the frequencies in the range 1200-1600 Hz. Now $\mu(f)$ for a value 'f' in the range 1200-1600 will be 1.0 while for f taking values 2000, 2500, 2700, $\mu(f)$ might take values of 0.5, 0.2, 0.0 respectively, depending upon the fuzziness desired at the boundaries of the set back.

A fuzzy subset of a universe U consisting of the elements $u_1, u_2, \ldots u_n$ may be expressed as :



$$A = \{<\mu_1,u_1>,\ldots,<\mu_n,u_n>\} \qquad (5.1)$$

where μ_i's represent grade membership of u_i in A. Normally μ_i are assumed to be in the range [0.0 ... 1.0] where 0.0 represents no membership and 1.0 represents full membership. In our case the range of membership indication is chosen to be an integer range [0 .. 127] for ease of computation.

An arbitrary fuzzy subset of the universe U may be expressed as

$$A = \bigcup_{u \in A} <\mu_f(u), u> \qquad (5.2)$$

where μ_f : U → [0.0 ... 1.0] is the membership or compatibility function of A. A denotes the union of fuzzy singletons over the universe U. The points in U where the value $\mu_f(u) > 0.0$ constitutes the support of A and points at which $\mu_f(u) = 0.5$ are the cross over points of A.

For example if the universe of discourse is all the second formant frequencies from 900 to 2700 Hz in steps of 50 Hz, then let

back = (1000, 1050, 1100)

and

backfuzzy = {<0.1,900>, <0.2,950>, <0.5,1000>, <0.7,1050>, <0.5,1100>, <0.2,1200>}

where back and backfuzzy represent nonfuzzy and fuzzy subsets of U.

In a continuous domain if U is the universe of discourse containing all second formant frequencies in the range 900-2700 Hz, then A might be expressed as

$$A = \bigcup_{u \in A} <(1/(1+(\mu_f(u))^2)), u> \qquad (5.3)$$

That is A is a fuzzy subset of the unit interval.

In many cases it is advantageous to compute the membership function of a fuzzy subset of the real line in terms of a standard function whose parameters may be adjusted to fit a specified membership function in an approximate fashion. Three such functions S, S^{-1} and π (Fig.5.1) are defined below:

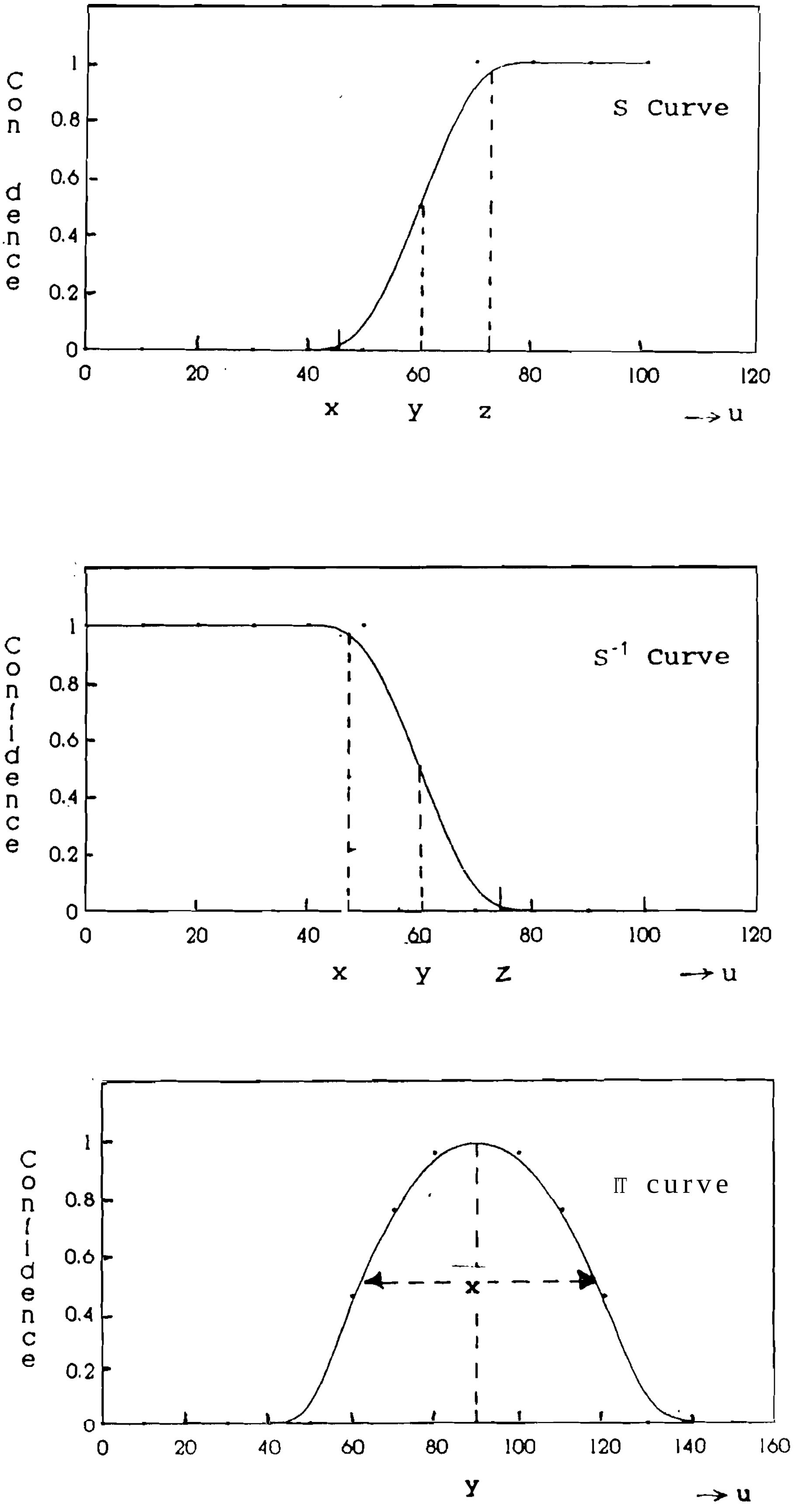


Fig. 5.1. S Curve, S^{-1} Curve and Π Curve

$$
\begin{array}{lll}
S\ (u:\ x,y,z) & = 0 & \text{if} \quad u \le x \\
 & = 2[(u-x)/(z-x)]^2 & \text{if} \quad x \le u \le y \\
 & = 1 - (2[(u-z)/(z-x)]^2) & \text{if} \quad y \le u \le z \\
 & = 1 & \text{if} \quad u \ge z
\end{array}
\qquad (5.4)
$$

$$
\begin{array}{lll}
S^{-1}(u:\ x,y,z) & = 1 & \text{if} \quad u \le x \\
 & = 1 - (2[(u-x)/(z-x)]^2) & \text{if} \quad x \le u \le y \\
 & = 2[(u-z)/(z-x)]^2 & \text{if} \quad y \le u \le z \\
 & = 0 & \text{if} \quad u \ge z
\end{array}
\qquad (5.5)
$$

$$
\begin{array}{lll}
\pi(u:\ x,y) & = S(u:\ (y-x),(y-(x/2)),y) & \text{if} \quad u \le y \\
 & = S^{-1}(u:\ y,(y+(x/2)),(y+x)) & \text{if} \quad u \ge y
\end{array}
\qquad (5.6)
$$

In S curve the parameter y is equal to (x+z)/2 and is the crossover point. In π curve the parameter x is the bandwidth and it is the difference between the crossover points. In the same curve y is the point at which confidence is unity.

In our experimentation the membership function is made to vary from 0 to 127 instead of 0.0 to 1.0 in all the three cases for ease of computation.

5.2.2 Use of Fuzzy Mathematical Concepts in Character Spotting

An example of fuzzy representation of acoustic parameters of the speech signal is given Table 5.1. Table 5.1 gives the confidence measure that a second formant frequency of 1200 Hz exhibits when it is considered as an argument for the different compatibility functions S^{-1}, S and π. The S^{-1}, S and π curves are used as compatibility functions for the features back, front and central respectively. In Table 5.1 μ_{back}, μ_{front} and $\mu_{central}$ represent the corresponding confidence measures obtained.

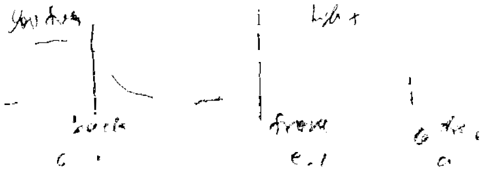


Table 5.1. Confidence measures for a second formant frequency of 1200 Hz for different compatibility functions

Quantitative Value	Fuzzy Membership
Second Formant Frequency is 1200 Hz	$\mu_{back} = 0.7$ $\mu_{front} = 0.0$ $\mu_{central} = 0.2$

To propagate this uncertainty the concept that the degree of uncertainty of a combined proposition is a function of the degree of uncertainty of the component propositions is also used. In order to do this the logical connectives of fuzzy relations are used. For example,

X1 OR X2 = maximum (X1,X2)

X1 AND X2 = minimum (X1,X2)

NOT X1 = NOT(X1) or (127 - X1)

To prevent the exponential blow up of inferences at various stages of signal-to-symbol transformation, the inferences are pruned by using only those whose certainty exceeds a certain threshold. For example let us consider the following rule:

IF (1) Low frequency energy is high or

(2) Total energy is high

THEN vocalic.

Here the vocalic decision will be assigned a confidence or certainty measure which is the maximum of confidences obtained from inferences (1) and (2) but the vocalic decision will completely fail when the confidence of both (1) and (2) fall below a certain predetermined threshold.

5 3 Expert System for Character Spotting

The acoustic-phonetic block is designed as an expert system. The design can follow two different methods. In one method we have an expert system where the acoustic-phonetic knowledge of all the characters can be integrated based on grouping of characters. In the second method an expert system is provided for each character.

In the expert system based on grouping of characters, it is assumed that the characters can be grouped based on the commonality that exists in the descriptions of the characters. The advantage of such a grouping would be that the same parameters need be derived from the speech signal for a particular group. The grouping of characters should be based on the criterion that the confusability within a group is maximum while the confusability across groups is less. Arriving at such a grouping is a difficult as there are a number of issues that need to be addressed. In addition, as the knowledge base increases in size to include the descriptions of all the characters, the complexity of the expert system increases. In the character spotting approach a single expert is used for each character.

5.3.1 Expert System for Each Character

The advantages of having an expert system for each character are:

(1) The description of each character is unique, and it can be represented by a small number of rules. The knowledge base of a character expert can be easily

modified while evaluating the performance of the system.

(2) The knowledge base also dictates the signal processing strategy. It is possible to extract parameters relevant to each character from the signal as determined by its description.

(3) Exception handling is simple, because it is possible to incorporate additional rules without excessively increasing the number of rules.

All the character experts can act simultaneously on the speech signal to spot the respective characters. This structure enables parallel processing techniques to be applied to take care of the large number of character experts. The expert system consists of (1) Acoustic-phonetic knowledge base, (2) Inference engine, (3) Acoustic Processor and (4) Working memory

The proposed model for implementation of individual character experts and combining them to get a symbol sequence corresponding to the speech signal is indicated in Fig.5.2. The output of each character expert gives the confidence level with which the character is hypothesized. It is expected here that each expert is spotting the associated character with maximum confidence, whereas other characters incorrectly spotted by any particular expert have lesser confidence. The integration of the outputs of the experts is done by some simple thresholding scheme. The alternatives at any point of time are limited by rejecting all characters which have confidence level below a certain threshold. The output of the acoustic-phonetic expert

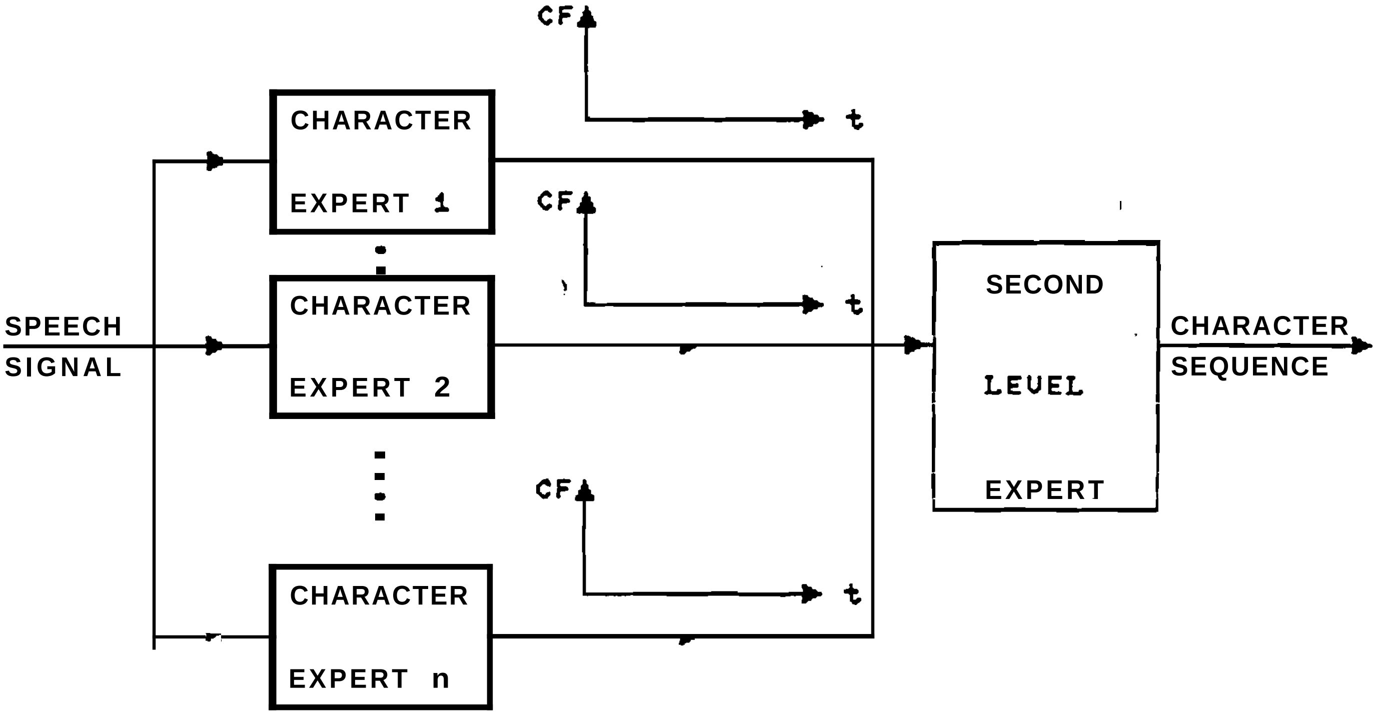


Fig. 5.2. Model of speech signal-to-symbol transformation system based on character spotting approach

bl ck is a sequence of characters with (1) alternatives and their associated confidence levels. (2)

5.3.2 Issues in Designing a Character &pert

The functional block diagram of a character expert is given in Fig.5.3. The main issue in designing a character expert is the integration of numeric and symbolic knowledge.

From Section 5.1 we see that there is a necessity for integrating the qualitative and quantitative data when the character expert is to spot a character in a given utterance. For example the location cues of /ka:/ (का) are described in terms of gross features like silence, vocalic, aspirated and burst. But the features are to be interpreted from the speech signal using numerical data. So in order to hypothesize the location of /ka:/(का) it is necessary to match the description of the character in terms of features which are themselves read from numerical data obtained by signal processing.

Various signal processing techniques are used on the speech signal under the control of the inference mechanism to extract the parameters from the speech signal that are relevant for a particular feature. Algorithms are written to extract parameters from the speech signal. The relevant parameters extracted for that character are stored in buffers. The data in the buffers is analyzed to match the parametric description of a feature. The results of the matching are stored in a special type of data structure, like pointers, and the details available in this data structure are feature name and its attributes like time of beginning, time of ending and the duration. Special procedures

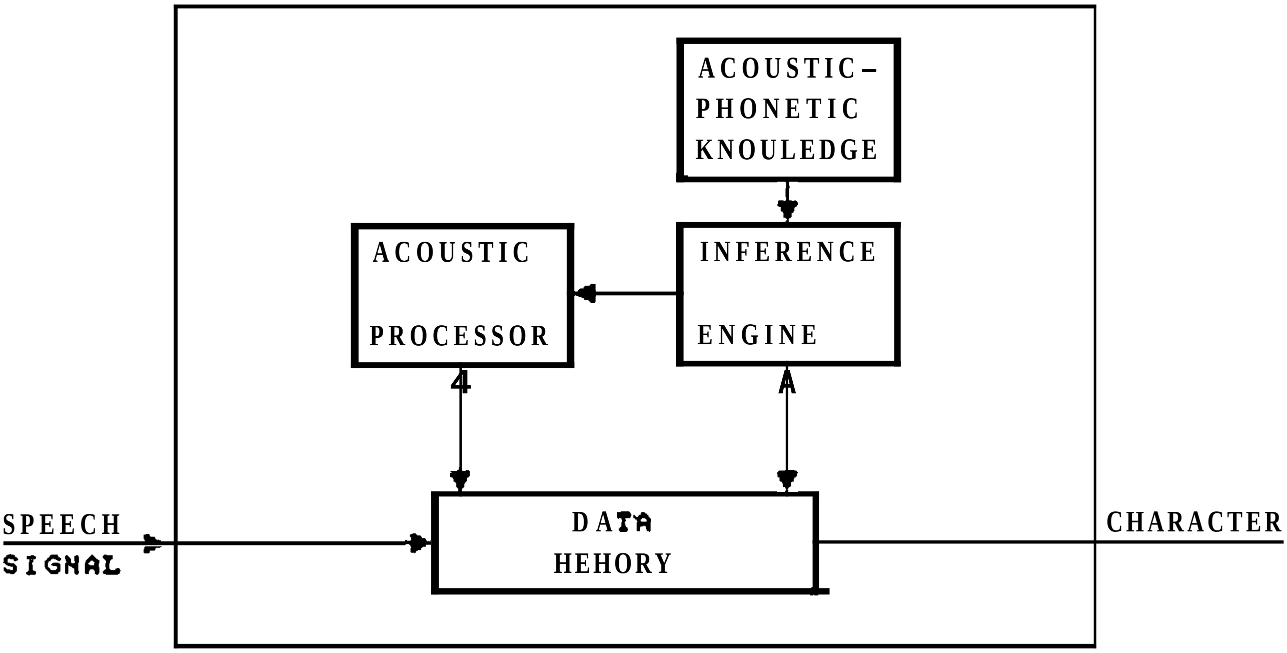


Fig. 5.3. **Functional block diagram of character expert** are written to convert the acoustic description given by the phonetician to a set of IF..THEN rules. The results obtained after a rule has been fired are stored in another data structure which contains a description of the character in terms of its attributes. similarly, procedures are written to take care of the fine features, namely, the intrinsic cues, coarticulation cues and context-dependent cues. The procedures or functions check for the particular parameter or feature to be in the specified range, and based on the results returns a confidence measure. Thus the expert's description of a character which is mostly symbolic in nature and the parametric representation of symbols which involves numerical or quantitative analysis are integrated in the expert system using specific data structures and procedures. The rule base for spotting the character /ka:/(का) and the associated fuzzy table are given in Appendix 2.

5.4 Implementation Details of a Character Spotting Expert System

5.4.1 Description of Working of a Character Spotting Expert System

The operations during signal-to-character transformation are mainly numerical computation, analyzing the parametric data and symbolic manipulation. It is advantageous to encode this operational knowledge as IF..THEN rules. The rule base is divided into number of groups in order that the search in the rule base can be reduced. This also gives structuredness to the rule base. Generally a group is made up of rules which are applicable to a particular context of analysis. So a context name is associated with each of the groups. A simple and flexible grammar is used for the rules, the syntactic and semantic details of which are as follows:

Fuzzy concepts are introduced into the expert system by assigning grade membership to the conditions and predicates(antecedents) in the IF...THEN rules, are evaluated. The grade membership is usually a value between 0.0 to 1.0. However in our case we assign integer values between 0 to 127 to reduce the computational complexity. The action part of a rule can be a predicate. In this case, if the condition part of the rule succeeds, then the maximum of the condition value and the current value of the predicate is assigned to it as its new value. Sometimes a rule cannot be trusted beyond a limit irrespective of its condition being true (i.e., grade membership value of 127). In this case the predicate appearing in the action part should not be assigned more than a maximum limit set for that rule. This is taken care of by an option < const > included in the grammar of the rule base. This sets the above said limit and is called the gramaticality of the rule. Fig.5.4 shows the grammar of the rule base.

There are two types of functions permitted in the rule. One of them, the system functions, are directly implemented into the inference engine. This increases the efficiency of the system when evaluating these functions. Another important characteristic of these functions is that they do not change the external data memory. The other type of function are the user defined functions.

The rules are created as a text 'file. The rules are compiled and internally represented in a form that is suitable

```
Rule Base :  .
          $ <context>




          { <const> ) IF    <cond>
                      THEN  <action> {  , <action> )



               .
           $ <context>


<cond>     -->   func()|func(param,param)|pred|const
<action>   -->   func()|func(param,param)|pred
<param>    -->   func()|func(param,param)|pred|const|v

cond - condition,  func - function,  param - parameter,
pred - predicate,  const - constant, v - variable
```

Fig.5.4. Grammar of the rule base

for the inference engine. The inference engine uses a "single pass" strategy. Successful rules are fired in the order they occur. Conditions are checked first and if the condition value is greater than or equal to the threshold, the rule is triggered i.e., the action part is executed. Facility is provided to remember the value of the evaluated condition function until any other user function is executed. An assumption made here is that execution of the condition in the rule with same argument does not change the state of the system unless a user function is fired.

5.4.2 Procedural Block Diagram of Expert System

The internal block diagram of the expert system is shown in Fig.5.5 The overall design is highly modular. The 1/0 requirements of each of the procedures are clearly specified. All constants, thresholds and fuzzy curves are kept in a look-up table of records of variable size. The inference engine is permitted to communicate with function/action procedure block and the knowledge base. It can also use a section of working memory in case of backtracking. Test functions and action procedure blocks can communicate with other blocks as shown in Fig.5.5.

It is this block which can read/write into the intermediate data structure and which can call the parameter examine routines for feature detection. The parameter examine routines are self contained and general enough to detect all the features in all situations. The routines refer to the tables for various constants and fuzzy curves. Another aspect of the design is to provide interactive tracking and tuning facility which can help

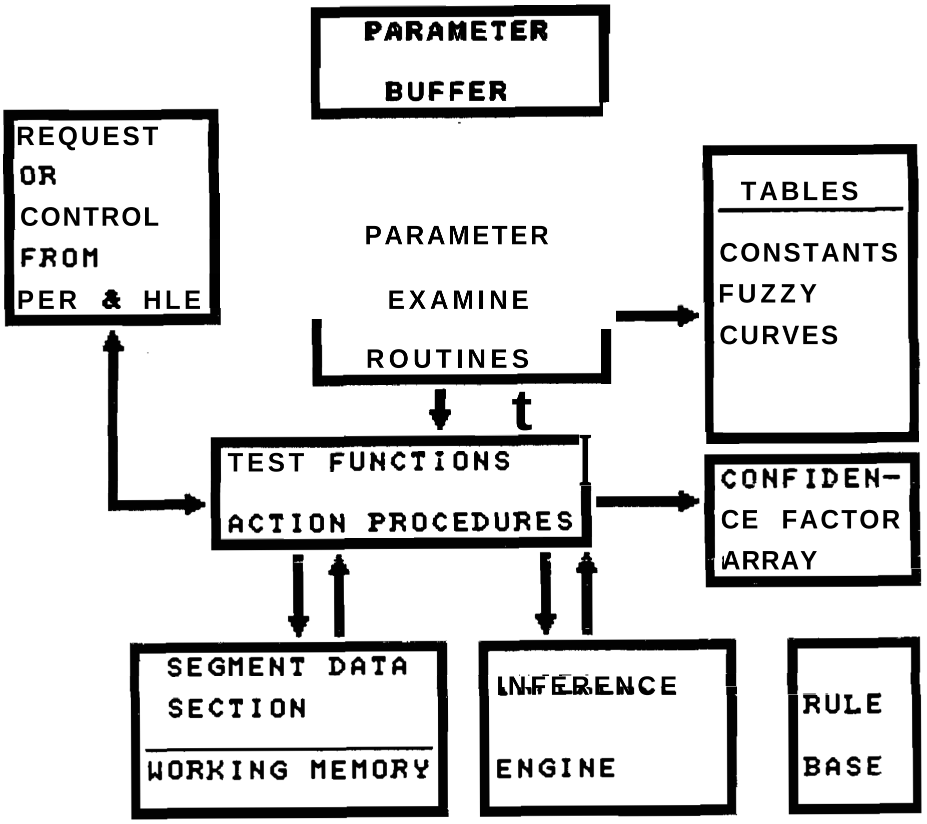


Fig. 5.5. Procedural block diagram of character spotting expert system
PER - Parameter extraction routines
HLE - Higher level experts

in interactive debugging in rule base. This facility provides the setting and resetting of break points, skipping a rule and displaying/modifying a data. This unit can access all state variables of the engine and also data and table memory.

A detailed description of the various blocks in Fig.5.5 may be found in the Appendix 3.

5.4.3 *Inference Flow and Working of the Expert System*

The sequence of steps followed by character expert for spotting a character in a continuous speech recognition are as follows :

(1) Initialization

(2) Hypothesizing the possible location

(3) Intrinsic cues to identify vowel and consonant segments

(4) Coarticulation cues to identify the consonantal segments in a character

(5) Context dependent cues to identify the character.

The inference engine works on the rules of the knowledge base and speech data. On successful matching of premises of a rule, the action specified in the rule is performed and the working memory is updated. The acoustic processor in the character expert processes the speech signal under the control of the inference engine whenever any rule in the knowledge base requires specific signal processing to be performed on the speech signal. The output of the processor is stored in the working memory from where the inference engine can access it whenever required. The activation of rules in the knowledge base is continued until there are no more rules to be triggered. The confidence with which the character is spotted is obtained from the working memory.

Summary

This chapter discussed in detail the different knowledge representation schemes and our choice for representing the acoustic-phonetic knowledge. Organization of this rule base and its activation are also discussed. Reasons for choosing the method of developing an expert system for each character are explained. The issued involved in the design of expert system with particular reference to the use of fuzzy mathematical concepts for knowledge activation are discussed in detail. Working of various blocks of character spotting expert system is explained. The results obtained by testing a large number of character spotting systems on a large data base and the performance evaluation of prototype expert systems are discussed in the next chapter.

CHAPTER-VI

PERFORMANCE EVALUATION OF CHARACTER SPOTTING EXPERTS

The objective of this chapter is to illustrate the performance of the character spotting expert systems on continuous speech of several utterances. Seventy five character experts are chosen for use in this study. 75 Sixty nine sentences 69 collected from a story book in Hindi form the test data for performance evaluation studies. The characters and sentences used in this study are listed in the Appendix 4. The chapter is organized as follows: In Section 6.1 we describe the experimental conditions under which the performance evaluation is carried out. The section also describes the proposed experimental studies to illustrate the working of the spotting expert for different groups of confusable characters. In Section 6.2 we discuss the performance of the system for spotting the gross features used in locating the possible regions of characters. The performance of various stages in the working of a character spotting expert is discussed in Section 6.3 for a few characters. Results for several groups of characters over all the utterances are presented in Section 6.4. In this section we show that with proper tuning of the rules and fuzzy tables it is possible to get good performance for several confusable character sets. Finally, the results of signal-to-symbol transformation using seventy five well tuned character experts on two--utterances are illustrated in Section 6.5. These results show the promise of the approach presented in the thesis, although extensive tuning of the rule base for large number of characters is needed before the approach can be used in practice.

6.1 Experimental Conditions and Proposed Studies

Before we actually start to analyze the results and evaluate the performance of the character spotting systems, we again stress here that our goal is signal-to-symbol transformation to capture the phonetic information in the speech signal as much as possible. This reduces the complexity of the symbol-to-text conversion stage in a continuous speech recognition system. In our case we have chosen the characters of the Indian language, Hindi, as symbols, and the data used is fluent speech spoken by native speakers. The structured nature of the language and the unique relationship between the spoken utterance and the written script of the language allows us to clearly define the rules governing the transformation from speech to characters in Hindi. Our approach to the solution of the problem is unique in the sense that we are attempting here to spot the symbols in a given utterance. The symbols are well defined in terms of articulatory, phonetic and acoustic features. We use a knowledge-based approach wherein the rules relating various characters and their representation in terms of features or parameters of speech data are obtained from different sources. Spotting of the characters is done by activating the knowledge and using fuzzy mathematical concepts to derive confidence levels to the results of spotting. The number of characters is large (about 5000) but the rules that are used for each character spotting system are only a few (about

The spotting systems use a simple inferencing mechanism.

Performance is evaluated by looking at the confidence levels with which a character is spotted in continuous speech.

The speech data consists of utterances of the test sentences (See Appendix 4) in Hindi read by a native speaker. Each utterance is at least 1.0 second long and is spoken normally. The data is recorded in an ordinary laboratory room using a directional microphone. The data is sampled at 10 kHz and digitized using 12-bit precision. The data is processed on a VAXSTATION system (details of the system are given in Appendix 5). For these studies all the parameters are extracted off line. Fuzzy tables were constructed for different character experts. Except for a few gross cues, which help in hypothesizing the possible presence of a character and its boundaries, each character has its own fuzzy table. The fuzzy tables were prepared initially with the help of expert phonetician. These tables were refined based on the results of experiments conducted on a large set of data.

The performance evaluation studies are mainly meant to illustrate how different segments of knowledge in the form of rule-base and fuzzy tables influence the confidence level of character spotting. (1) Our first experiment is on spotting the gross features which are used to locate the possible regions of a character. It is essential that the rules and fuzzy tables related to these rules are tuned in such a way that the features are identified with high confidence. 2) Our second experiment is to illustrate the detailed working of spotting experts for a few characters. The main purpose of this is to show how the ambiguities at the gross feature identification level are resolved using the intrinsic and contextual rules. Our next set of experiments are designed to illustrate performance of the character spotting experts for several confusable character sets over all the test utterances. dur final experiment consists of running all the seventy five character spotting experts on utterances of two sentences to illustrate the overall performance of the proposed signal-to-symbol transformation for continuous speech in Hindi. These experiments are discussed in detail in the following sections.

6 2 Experimental Studies for Spotting Gross Features

Spotting of a character is done in two stages, namely, locating the possible presence of the character based on gross features and identifying the character using intrinsic and contextual cues in the located regions. We describe an experiment to spot the gross features corresponding to the CV characters where C is any of the five unvoiced unaspirated stop consonants, and the V portion is the vowel /a:/(आ). The gross features are vocalic for vowel portion and unvoiced closure (unvoiced silence), burst and aspiration for consonant portion of a character. The parameters used for these features are: energy(ENR), first linear prediction coefficient(LP 1), high-low frequency energy ratio(HLR), spectral distance(SPD) and spectral flatness(SPF). The results of spotting these features are given in terms of the confidence with which each feature is spotted. Only the rules relevant to feature spotting are used on all 'the test utterances. The gross features occur in various contexts in the test data. Table 6.1 shows the results of testing the expert

Table 6.1 Performance characteristics for gross feature identification

GROSS FEATURES TESTED	GROSS FEATURES IDENTIFIED			
	VOCALIC	SILENCE	BURST	ASPIRATED
VOCALIC	420/420 (127)	-	-	-
SILENCE	-	120/120 (120)	10/120 (100)	10/120 (100)
BURST	-	10/120 (100)	120/120 (120)	15/120 (100)
ASPIRATED	-	20/100 (100)	20/120 (100)	120/120 (120)

Notation used in the table : P/Q (CONF)

P indicates the number of times feature is hypothesized

Q indicates the number of times the feature occurred in the data

CONF indicates the confidence with whioh the character is hypothesized. Minimum confidence is indicated for diagonal terms. Maximum confidence is indioated for off diagonal terms systems for locating the features using their parametric description. Entries in the table are of the type P/Q which indicate the number (Q) of occurrences of the feature in all the utterances and the number (P) of times the feature is spotted. The entry also gives in parentheses the lowest confidence level among the occurrences in the case of the correctly identified feature and the highest confidence level in the case of wrongly identified feature. There are a large number of vocalic regions present in the data. This is mainly because most of the characters end with a vowel. The confidence level in spotting the vocalic regions is the maximum possible in almost all the places where these regions are present. This is indicated as 127 in parentheses corresponding to the vocalic feature in Table 6.1. Similarly the features unvoiced closure, burst and aspiration are spotted and their confidence levels are given in Table 6.1. It is seen that there are some misclassifications by the closure, burst and aspiration features, whereas there are no misclassifications by the vocalic features.

6 3 Illustration of Performance of a Character Expert

In this section we discuss the results of applying location rules and hypothesization rules of character experts. We consider the following nine characters for illustration: /ka:/(का), /ca:/(चा), /ṭa:/(टा), /ta:/(ता), /pa:/(पा), /da:/(दा), /ba:/(बा), /ma/(म) and /na/(न). In particular we show how evidences (though vague) from different rules can be effectively combined to arrive at correct decisions. Moreover, the discussion also shows that the performance of a character spotting expert can be improved continuously by modifying the entries in the fuzzy table and also by modifying the rule base as the character expert is run on more and more data.

First we consider the character /ka:/(का) for illustration. The experiment with /ka:/(का) spotting system consists of applying the rules on the speech data corresponding to an utterance of a sentence which has some occurrences of the character. Fig.6.1 shows the confidence levels at various stages in spotting the character. The fuzzy table was initially constructed based on the expert phonetician's knowledge and on the data analyzed for some utterances containing the character. The parameters used to locate the gross features consist of total energy, the first linear prediction coefficient, high-low frequency energy ratio and some transient features based on plots such as spectral flatness and spectral distance. The gross features to be identified here are silence, burst, aspiration and vocalic regions. The expert system was able to spot the gross features wherever they occurred in the utterance. Some other regions which do not correspond to these features were also spotted. We notice that a large number of burst regions are hypothesized. This is because the parameter values have a wide range for burst detection so that they are not missed if present. But spurious regions are taken care of when combining evidence obtained by applying different rules. The figure shows that there are three regions which are located by the rules using gross features. But the final hypothesized region based on intrinsic and context dependent rules is only one and that is the correct region for the character /ka:/(का).

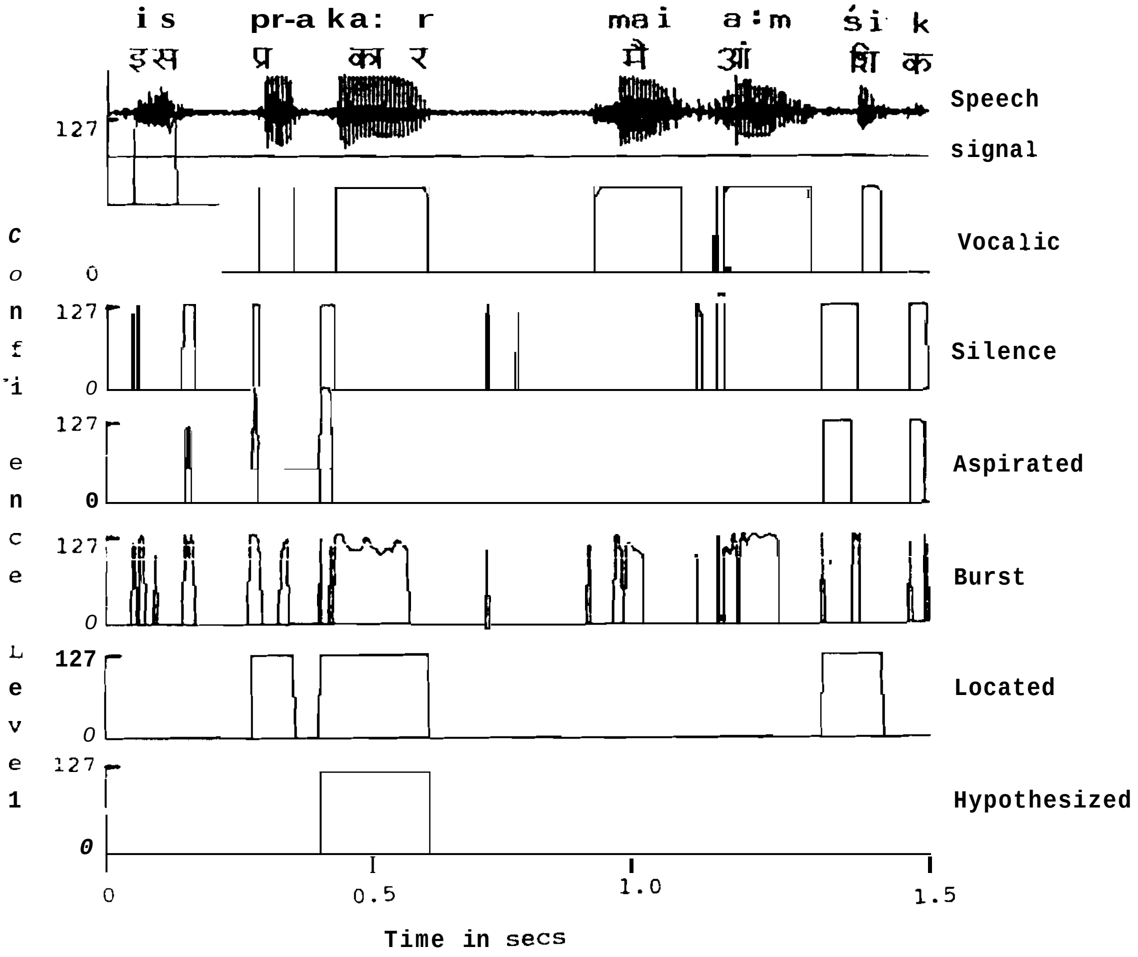


Fig. 6.1. Confidence level plots for the gross features, located regions and hypothesized regions for the character /ka:/ (का)

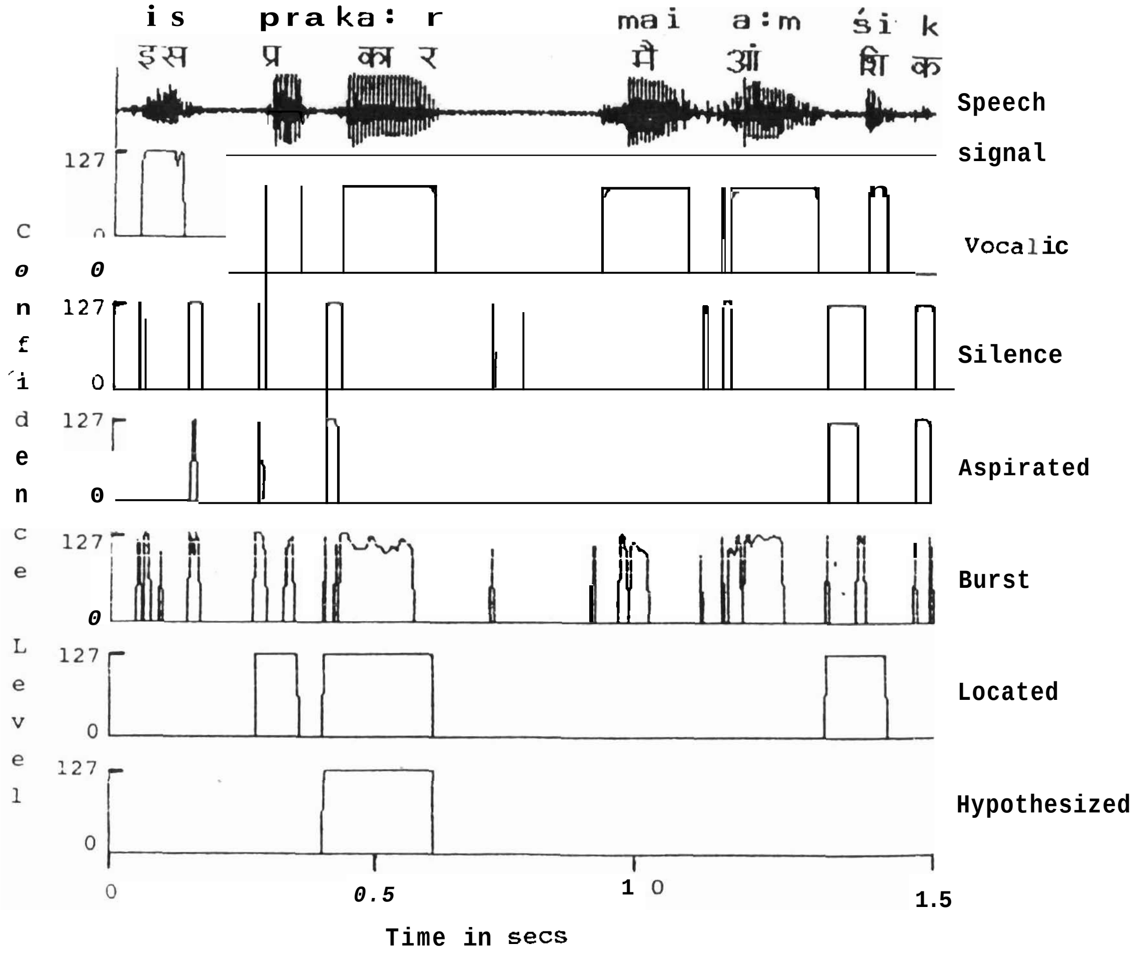


Fig. 6.1. Confidence level plots for the gross features, located regions and hypothesized regions for the character /ka:/ (का)

Next we consider the character /ca:/(चा) where the consonant is described as an unvoiced unaspirated affricate. For an affricate the burst region is followed by a frication region. So frication is included as a gross feature for this character. Fig.6.2 shows the confidence levels with which the gross features of the characters are spotted. This illustrates that gross features need not be same for all the characters and that they are decided by the description of the character.

Similarly, results of character spotting systems for the other unvoiced, unaspirated consonants, namely, /ṭa:/(टा), /ta:/(ता) and /pa:/(पा) are shown in Figs.6.3 to 6.5. It can be seen from these figures and also from Fig.6.1 that aspiration is included as a gross feature, though the consonants of these characters are described as unaspirated. This is done mainly because it is observed experimentally and also from the expert phonetician's knowledge that these consonants contain certain amount of aspiration. The extent of aspiration as measured by the duration is much less than that of the unvoiced aspirated consonants. It is also observed that the duration of the aspiration region is different for each of the unaspirated consonants. This information is used in the location rules when the gross features are combined to locate the character region.

Though the gross features are same, the way they are combined in the location rules need not be same for all characters. This is illustrated by comparing spotting of voiced unaspirated consonant characters with the spotting of unvoiced unaspirated consonant characters. Fig.6.6 and Fig.6.7 illustrate spotting of characters /ba:/(बा) and /da:/(डा), respectively.

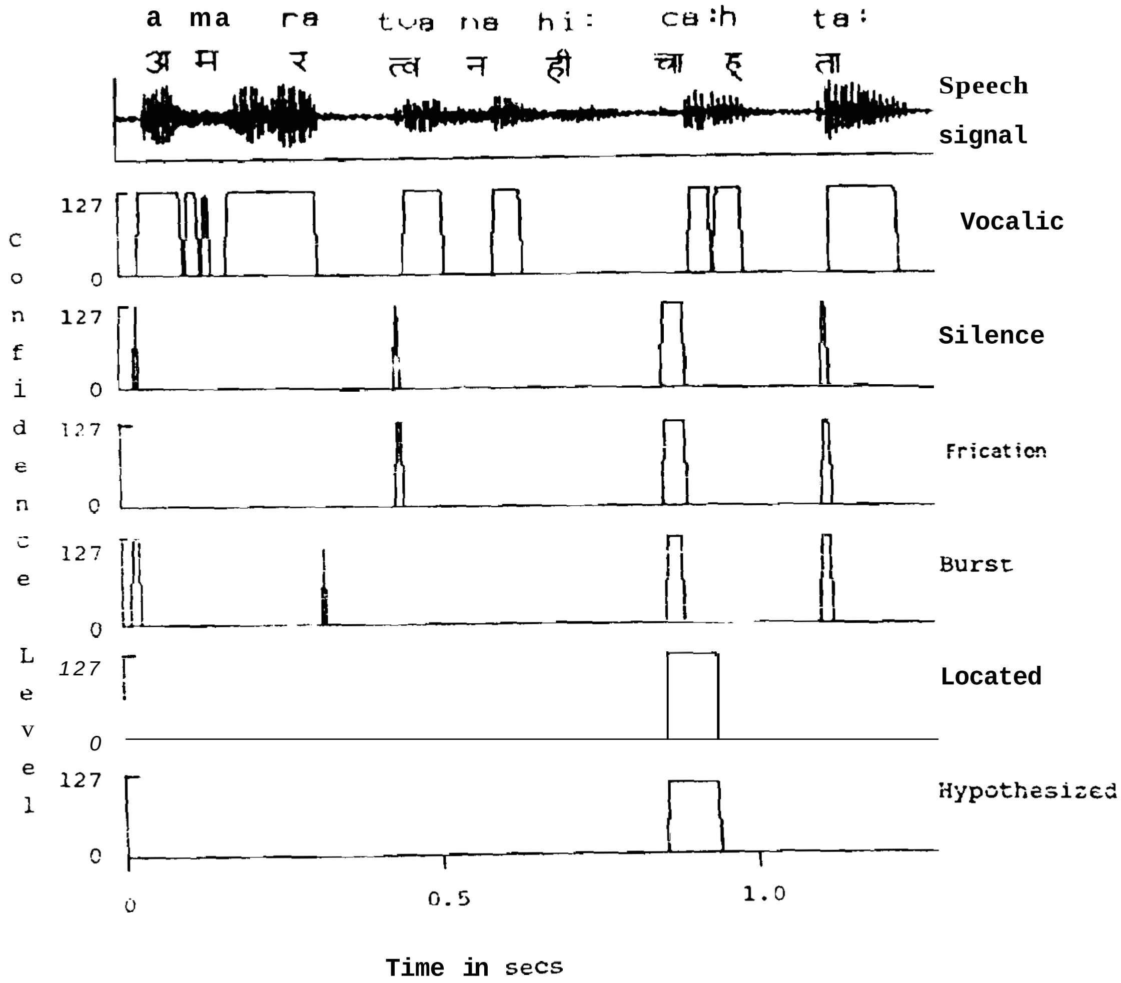


Fig. 6.2. Confidence level plots for the gross features, located regions and hypothesized regions for the character /ca:/ (चा)

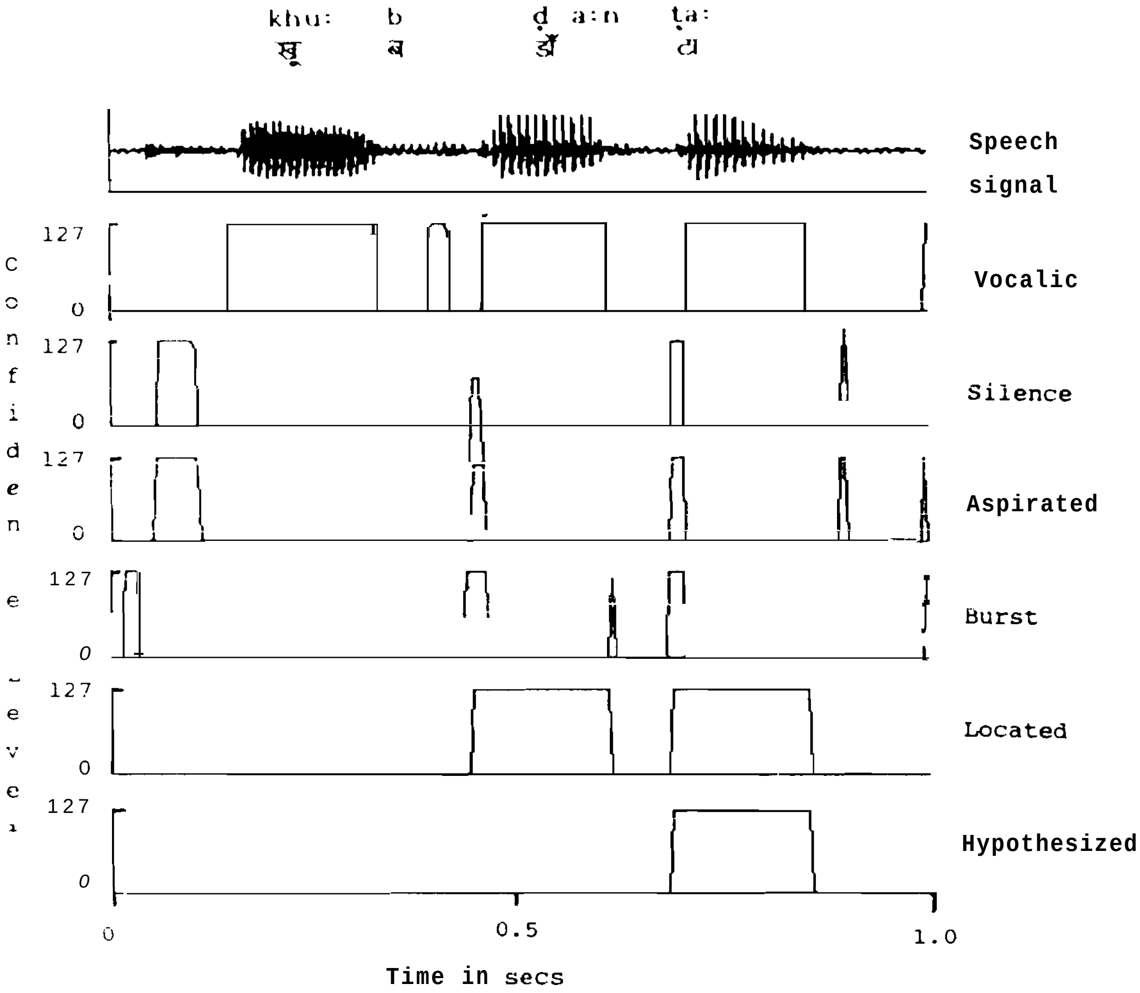


Fig. 6.3. Confidence level plots for the gross features, located regions and hypothesized regions for the character /ṭa:/ (टा)

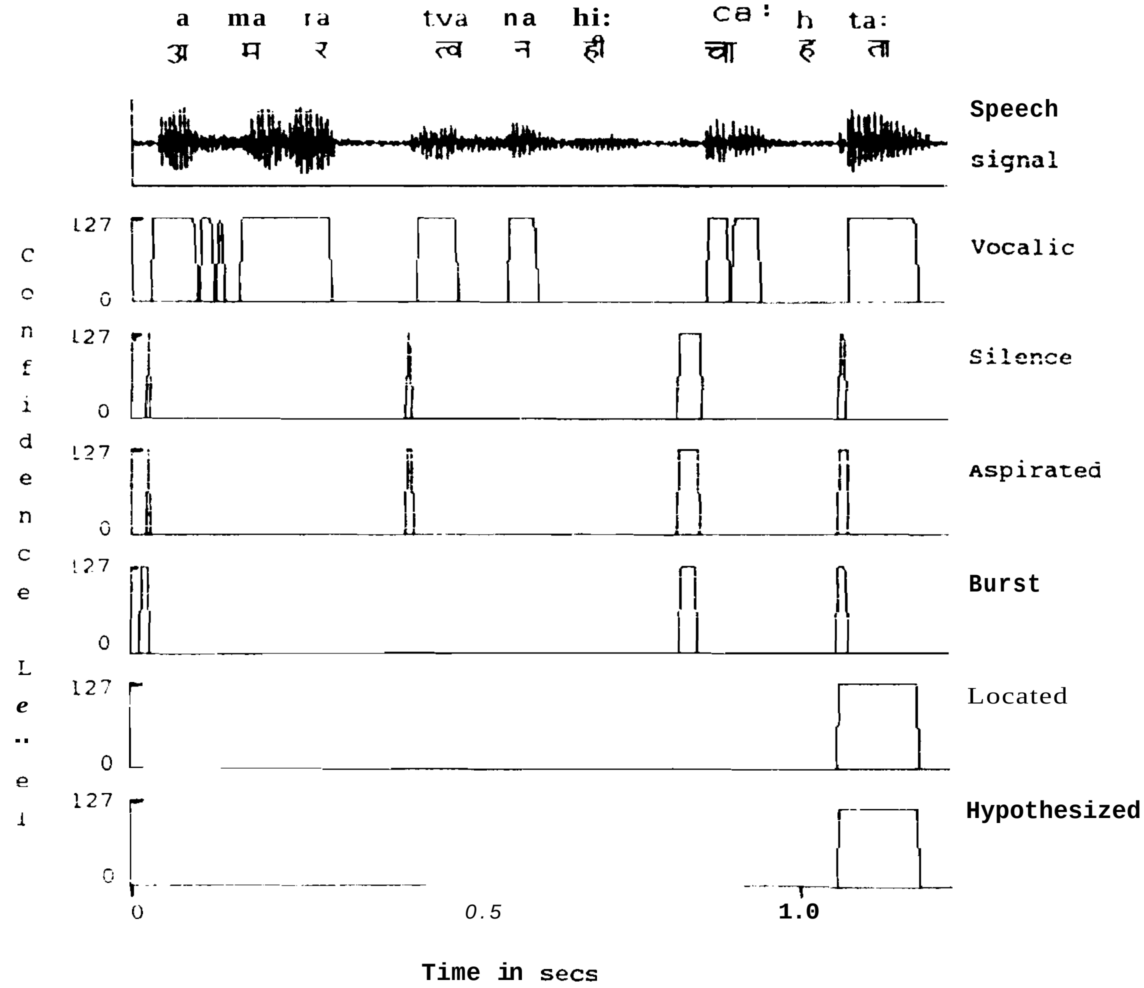


Fig. 6.4. Confidence level plots for the gross features, located regions and hypothesized regions for the character /ta:/ (ता)

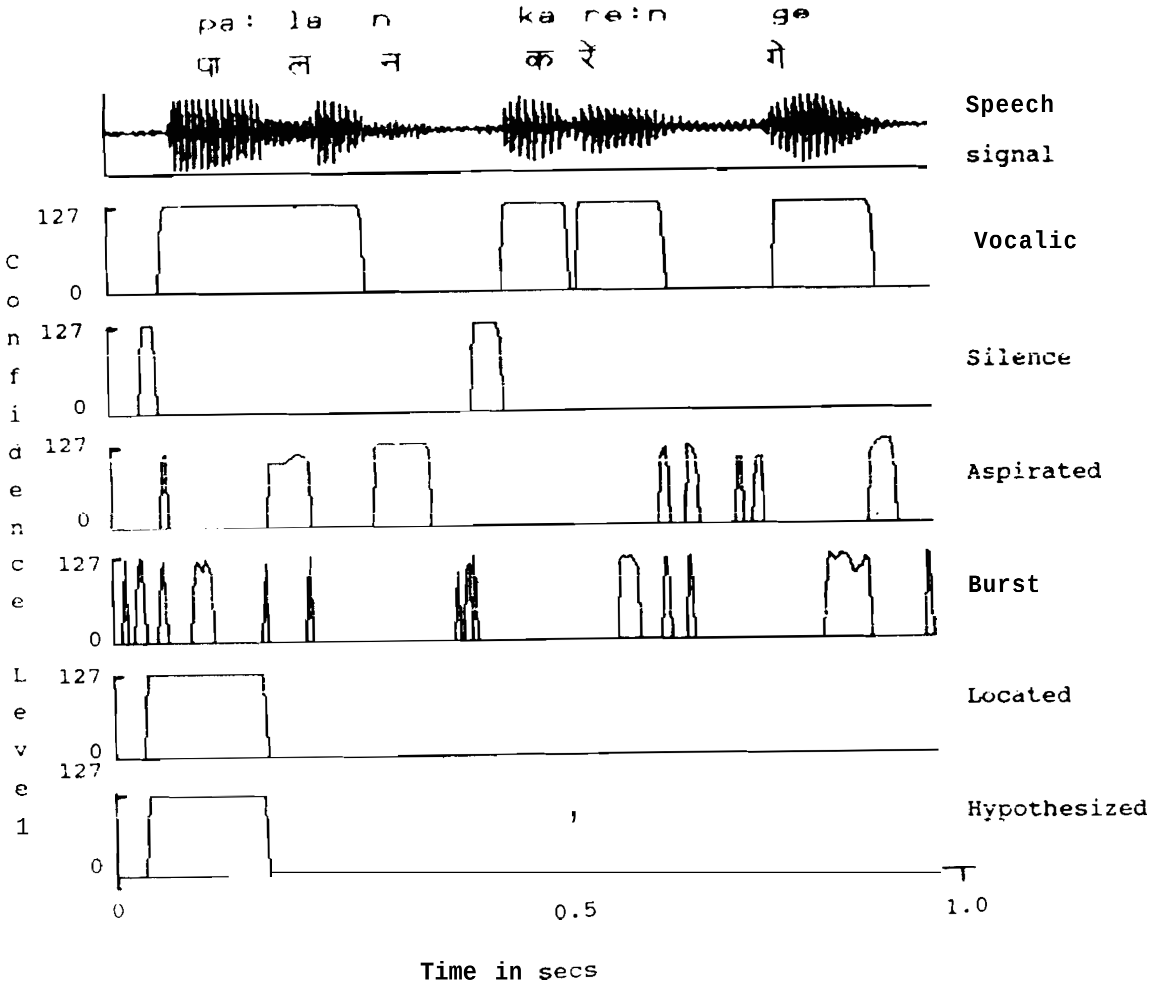


Fig. 6.5. confidence level plots for the gross features, located regions and hypothesized regions for the character /pa:/ (पा)

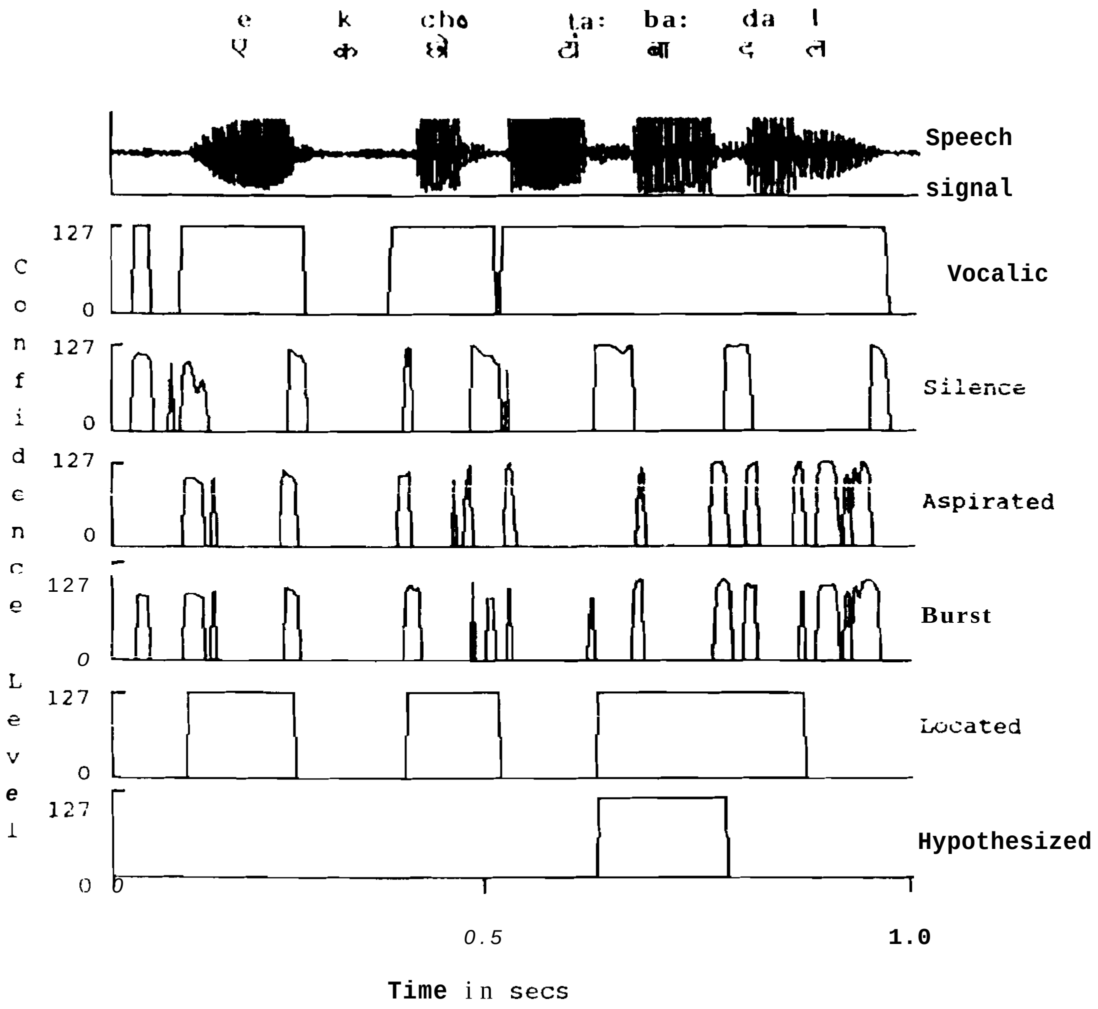


Fig. 6.6. Confidence level plots for the gross features, located regions and hypothesized regions for the character /ba:/ (बा)

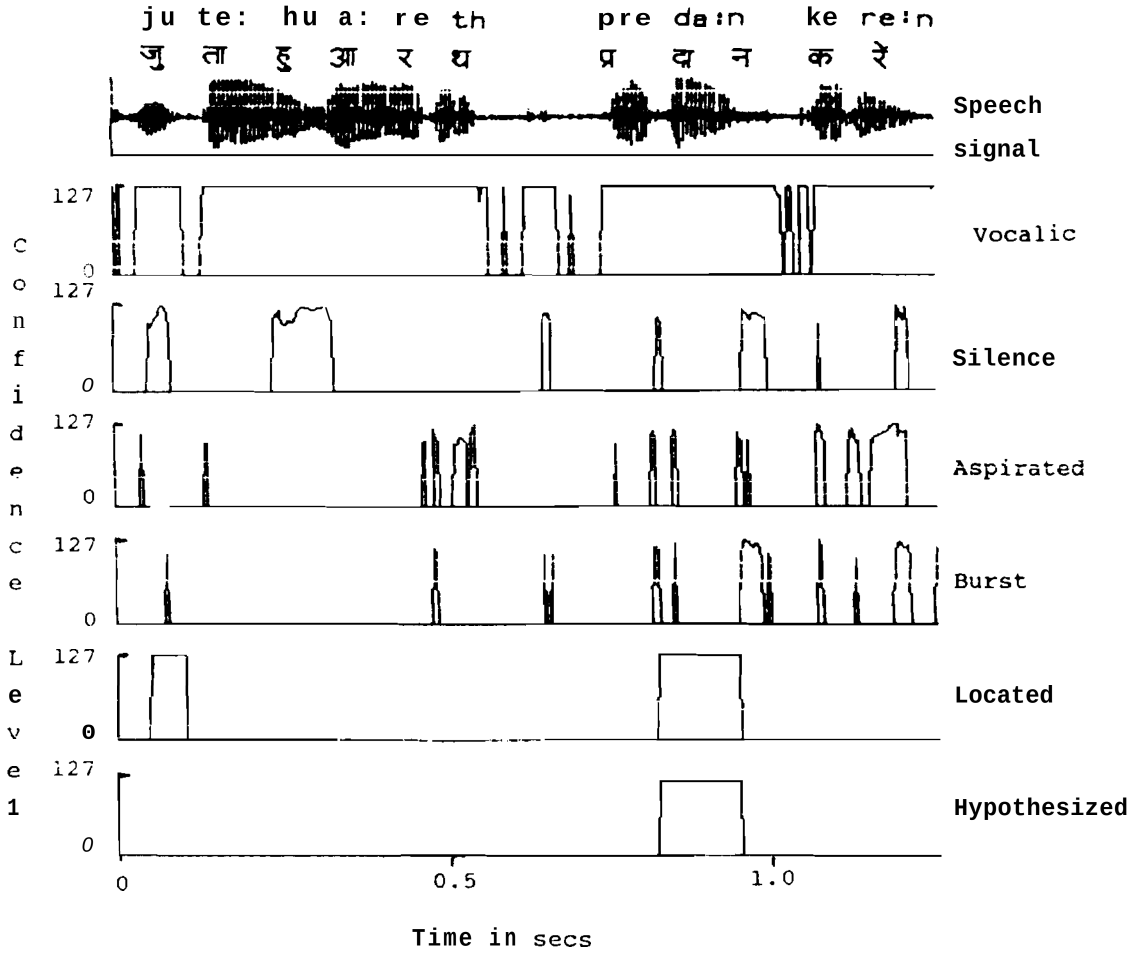


Fig. 6.7 Confidence level plots for the gross features, located regions and hypothesized regions for the character /da:/ (दा) .

In the case of these voiced aspirated characters the gross features silence, burst and aspiration appear within the vocalic region, whereas they generally precede a vocalic region in the case of unvoiced consonant characters. Location rules use this information when the gross features are combined.

Another important point to note here is that the threshold ranges for the parameters used in the rules for identifying a gross feature are not same for all characters. For example, the characteristics of silence and burst regions are different for unvoiced consonants from that of voiced characters. So different thresholds may be used in identifying the gross features for each character.

Hypothesization rules are different for each character. The location rules for a character are tuned such that the character region is not missed. The hypothesization rules are tuned to spot the character with high confidence wherever the character is present. This is illustrated in Figs. 6.1 to 6.7 for different characters. Results of spotting the nasal characters /ma/(म) and /na/(न) are given in Fig.6.8 and Fig.6.9 respectively.

The confidence level plots shown so far in this section indicate the results corresponding to individual character experts where the rules are refined to get the best performance from the character expert. In order to study the effectiveness of character spotting approach it is necessary to obtain the results of running all the character experts on an utterance. In our experiments this is done in stages where we consider sets of confusable characters and evaluate their performance. This study is explained in the next section.

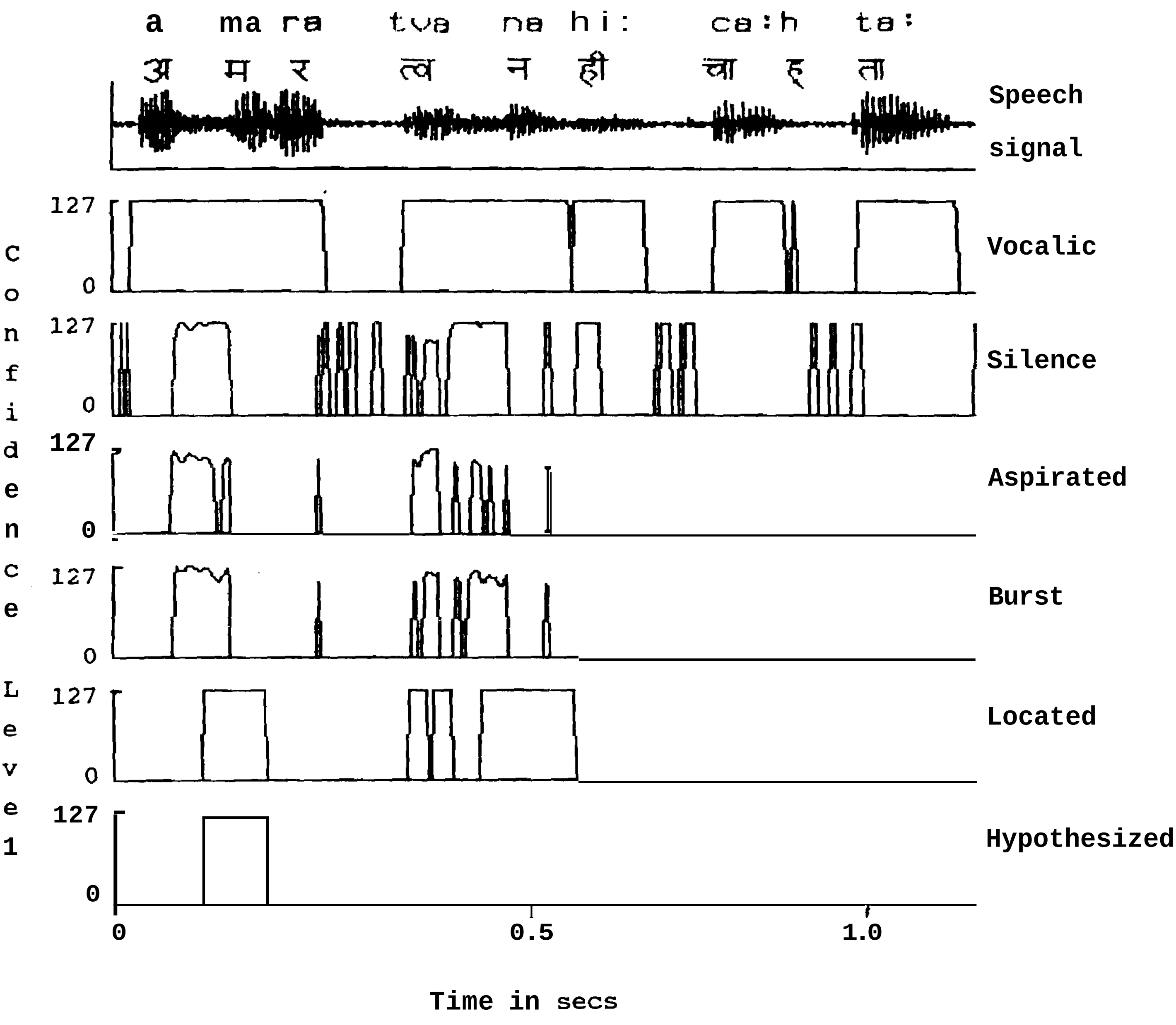


Fig. 6.8. Confidence level plots for the gross features, located regions and hypothesized regions for the character /ma / (म)

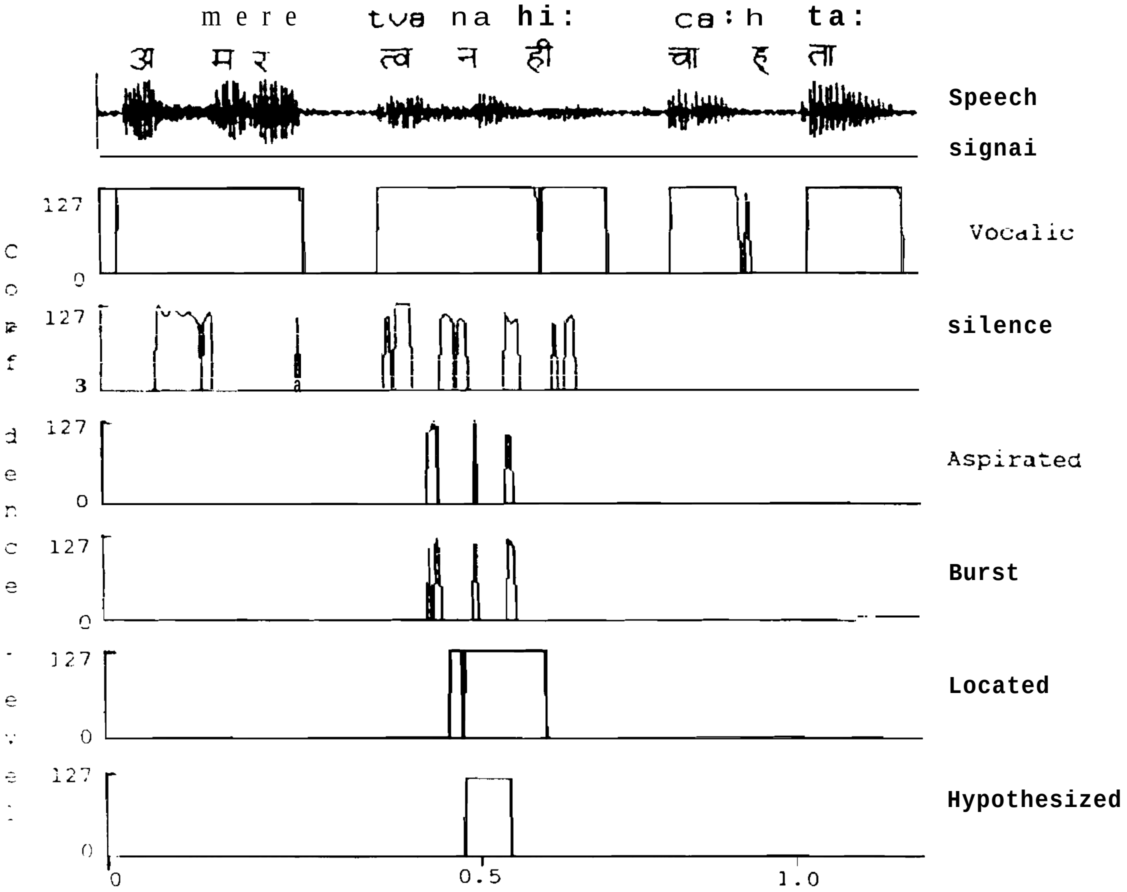


Fig. 6.9. Confidence level plots for the gross features, located regions and hypothesized regions for the character /na / (न)

6.4 Performance Evaluation for Groups of Characters

In this section we study the performance of character spotting experts for groups of characters, mainly to illustrate the results among characters considered to be confusable. In particular, we consider the following groups of characters. The results for these groups are shown in the respective tables.

(1) Table 6.2 : Unvoiced, unaspirated consonants with same vowel ending. {/ka:/(का), a :(चा), /ṭa:/(टा), /ta:/(ता), /pa:/(पा)}

(2) Table 6.3 : Voiced, unaspirated consonants with same vowel ending. {/ga:/(गा), /ja:/(जा), /ḍa:/(डा), /da:/(दा), /ba:/(बा)}

(3) Table 6.4 : Nasals (/ma:/(मा) and /na:/(ना)}

(4) Table 6.5 : Sonorants {/ya:/(या), /ra:/(रा), /la:/(ला), /va:/(वा)}

(5) Table 6.6 : Fricatives /sa:/(सा) and /śa:/(शा).

(6) Table 6.7 : Consonant /k/(क) with different vowel endings. {(/ka/(क), /ka:/(का), /ki/(कि), /ki:/(की), /ku/(कु) (/ku:/(कू), /ke/(के), /ko/(को)}

All these character experts were tuned using large data before they were tried on all the test utterances. The performance is good in most of the cases. Even highly confusable characters such as /ṭa:/(टा), /ta:/(ता) and /pa:/ (पा) have not shown much confusion among themselves.

Table 6.2 Performanoe oharacteristics of experts for consonant-vowel (CV) oombination C is /k/,/o/,/ṭ/,/t/and /p/ and V is /a:/

CHARACTER EXPERTS	IDENTIFIED CHARACTERS				
	/ka:/ (का)	/oa:/ (चा)	/ṭa:/ (टा)	/ta:/ (ता)	/pa: (पा)
/ka:/(का)	10/10 (120)	-	-	3/7 (100)	-
/oa:/(चा)	-	6/6 (115)	-	-	-
/ṭa:/(टा)	-	-	6/8 (115)	-	-
/ta:/(ता)	2/10 (100)	-	-	17/17 (115)	-
/pa:/(पा)	-	-	-	-	6/6 (120)

Notation used in the table :

P/Q (CONF)

P indicates the number of times oharaoter is hypothesized

Q indicates the number of times the oharaote occurred in the data

CONF indicates the confidence with which the character is hypothesized. Minimum confidence is indicated for diagonal terms. Maximum confidence is indicated for off diagonal terms

Table 6.3 Performance characteristics of experts for consonant-vowel (CV) combinations C is /ɡ/,/j/,/ḍ/,/d/and /b/ and V is /a:/

CHARACTER EXPERTS	IDENTIFIED CHARACTERS				
	/ɡa:/ (गा)	/ja:/ (जा)	/ḍa:/ (डा)	/da:/ (दा)	/ba : (बा)
/ɡa:/(गा)	4/4 (115)	-	-	-	-
/ja:/(जा)	-	2/2 (110)	-	-	-
/ḍa:/(डा)	-	-	3/3 (115)	-	-
/da:/(दा)	-	-	-	4/5 (115)	-
/ba:/(बा)	-	-	-	-	2/2 (110)

Notation used in the table : P/Q
(CONF)

P indicates the number of times character is hypothesized

Q indicates the number of times the character occurred in the data

CONF indicates the confidence with whioh the character is hypothesized. Minimum confidence is indicated for diagonal terms. Maximum confidence is indioated for off diagonal terms

Table 6.4 Performance oharaoteristios of experts for consonant-vowel(CU) combinations C is /m/and /n/ and U is /a:/

CHARACTER EXPERTS	IDENTIFIED CHARACTER	
	/ma:/ (मा)	/na:/ (ना)
/ma:/(मा)	4/4 (110)	-
/na:/(ना)	-	5/5 (115)

Notation used in the table : P/Q (CONF)

P indicates the number of times character is hypothesized

Q indicates the number of times the oharacter occurred in the data

CONF indicates the confidence with which the character is hypothesized. Minimum confidence is indiaated for diagonal terms. Maximum confidence is indicated for off diagonal terms

Table 6.5 Performance characteristics of experts for consonant-uowel (CV) combinations C is /y/,/r/,/l/ and /v/ and V is /a:/

CHARACTER EXPERTS			
/ya:/ (या)	/ra:/ (रा)	/la:/ (ला)	/va:/ (वा)
4/4 (108)	6/6 (105)	3/3 (106)	2/2 (100)

Notation used in the table : P/Q (CONF)

P indicates the number of times character is hypothesized

Q indicates the number of times the character occurred in the data

CONF indicates the confidence with which the character is hypothesized.

Table 6.6 Performance characteristics of experts for consonant-vowel (CV) combinations C is /s/and /ś/ and V is /a:/

CHARACTER EXPERTS	IDENTIFIED CHARACTER	
	/sa:/ (सा)	/śa:/ (शा)
/sa:/(सा)	7/7 (110)	-
/śa:/(शा)	-	1/2 (115)

Notation used in the table : P/Q (CONF)

P indicates the number of times character is hypothesized

Q indicates the number of times the character occurred in the data

CONF indicates the confidence with which the character is hypothesized. Minimum confidence is indicated for diagonal terms. Maximum confidence is indicated for off diagonal terms

Table 6.7 PERFORMANCE CHARACTERISTICS OF EXPERTS FOR CONSONANT VOWEL (CV) COMBINATIONS - C IS /k/ AND V IS ANY VOWEL

CHARACTER EXPERTS	IDENTIFIED CHARACTERS							
	/ka/ (क)	/ka:/ (का)	/ki/ कि	/ki:/ (की)	/ku/ (कु)	/ku:/ (कू)	/ke/ (के)	/ko / (को)
/ka/(क)	5/5 (120)	2/10 (100)	-	-	-	-	-	-
/ka:/(का)	1/5 (100)	10/10 (120)	-	-	-	-	-	-
/ki/(कि)	-	-	5/5 (120)	-	-	-	-	-
/ki:/(की)	-	-	-	3/3 (115)	-	-	-	-
/ku/(कु)	-	-	-	-	6/6 (120)	-	-	-
/ku:/(कू)	-	-	-	-	-	(	-	-
/ke/(के)	-	-	-	-	-	-	10/10 (120)	-
/ko/(को)	-	-	-	-	-	-	-	3/3 (120)

Notation used in the table : P/Q (CONF)

P indicates the number of times character is hypothesized

Q indicates the number of times the character occured in data

CONF indicates the confidence with which the character is hypothesized. Minimum confidence is indicated for diagonal terms and maximum confidence is indicated for off diagonal terms.

6.5 Illustration of Signal-to-Symbol Transformation

In order to study the effectiveness of signal-to-symbol transformation, we have considered spotting of all of 75 character experts on two utterances. Limitations due to the systems available made us chose only 75 character experts for this study. The two utterances contain characters from various groups we have considered earlier. The two utterances are

/ma:ta: pita: ko bula: bheja:/(माता पिता को भेजा)
/pita:ji: ko/ (पिताजी को).

The results of these experiments for the two utterances are given in Figs.6.10 and 6.11. All the character spotting systems are tuned in order to reduce the misclassifications. Any misclassifications that have occurred are given as alternate choices for that symbol or character. In the case of the first utterance we see that all the characters present in the utterance were hypothesized with high confidence level. The character /bhe/(भे) was not recognized because it does not form part of 75 character spotting systems. In the second case also the characters contained in the utterance are identified with high confidence level although there were some misclassifications as shown by the character lattice in Fig.6.11.

The results of this experiment suggest that character spotting does help in achieving signal-to-symbol transformation in Indian languages. The main advantage here is the unit chosen for signal-to-symbol transformation and also the spotting approach adopted for signal-to-symbol transformation. It also suggests that a large number of experts to recognize the characters is really not an issue since each expert uses only a few rules and refining the rule base is not a complex task.

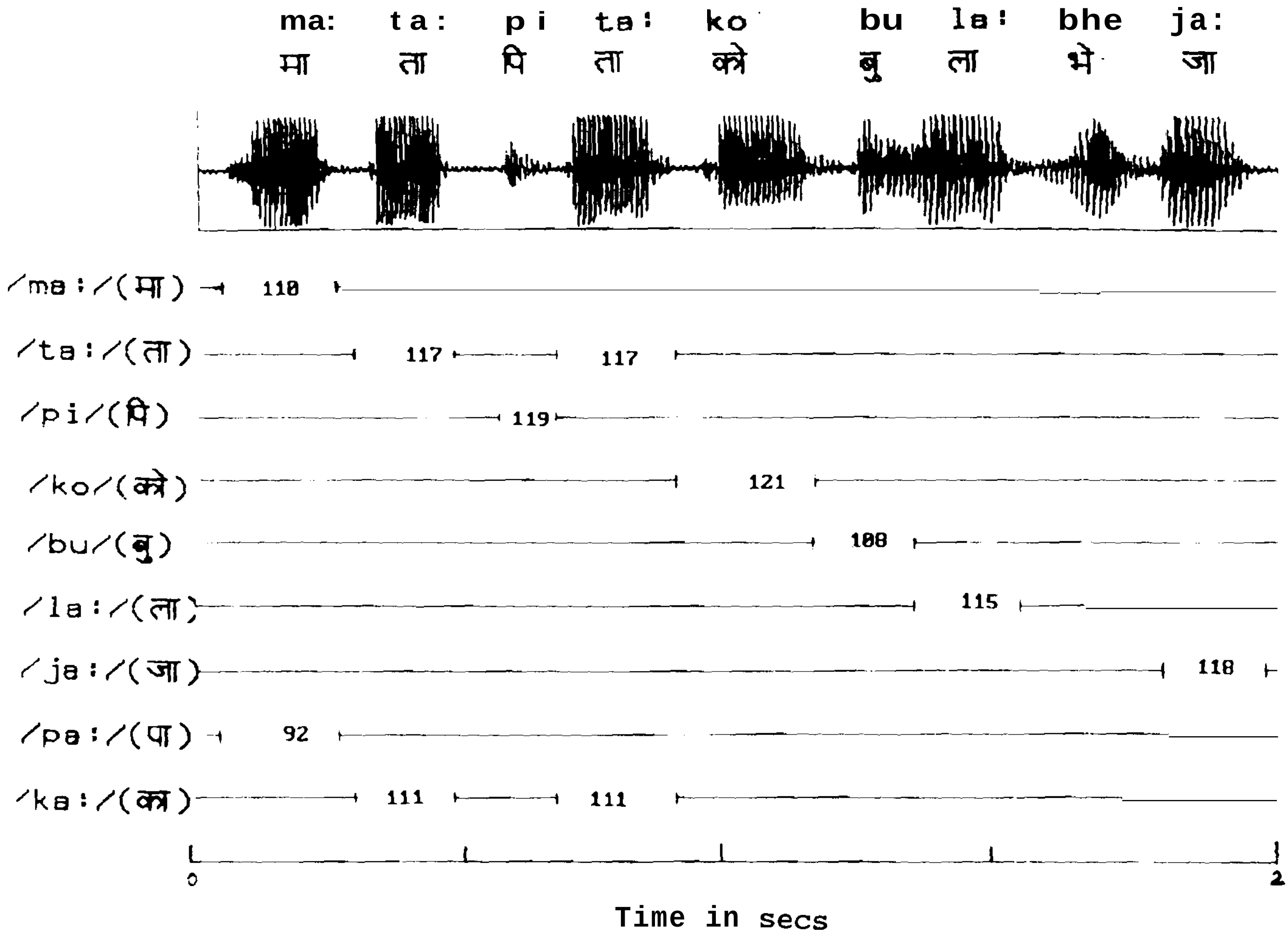


Fig.6.10. Illustration of signal-to-symbol transformation for utterance 1. The confidence level and the region spotted by each character expert are indicated.

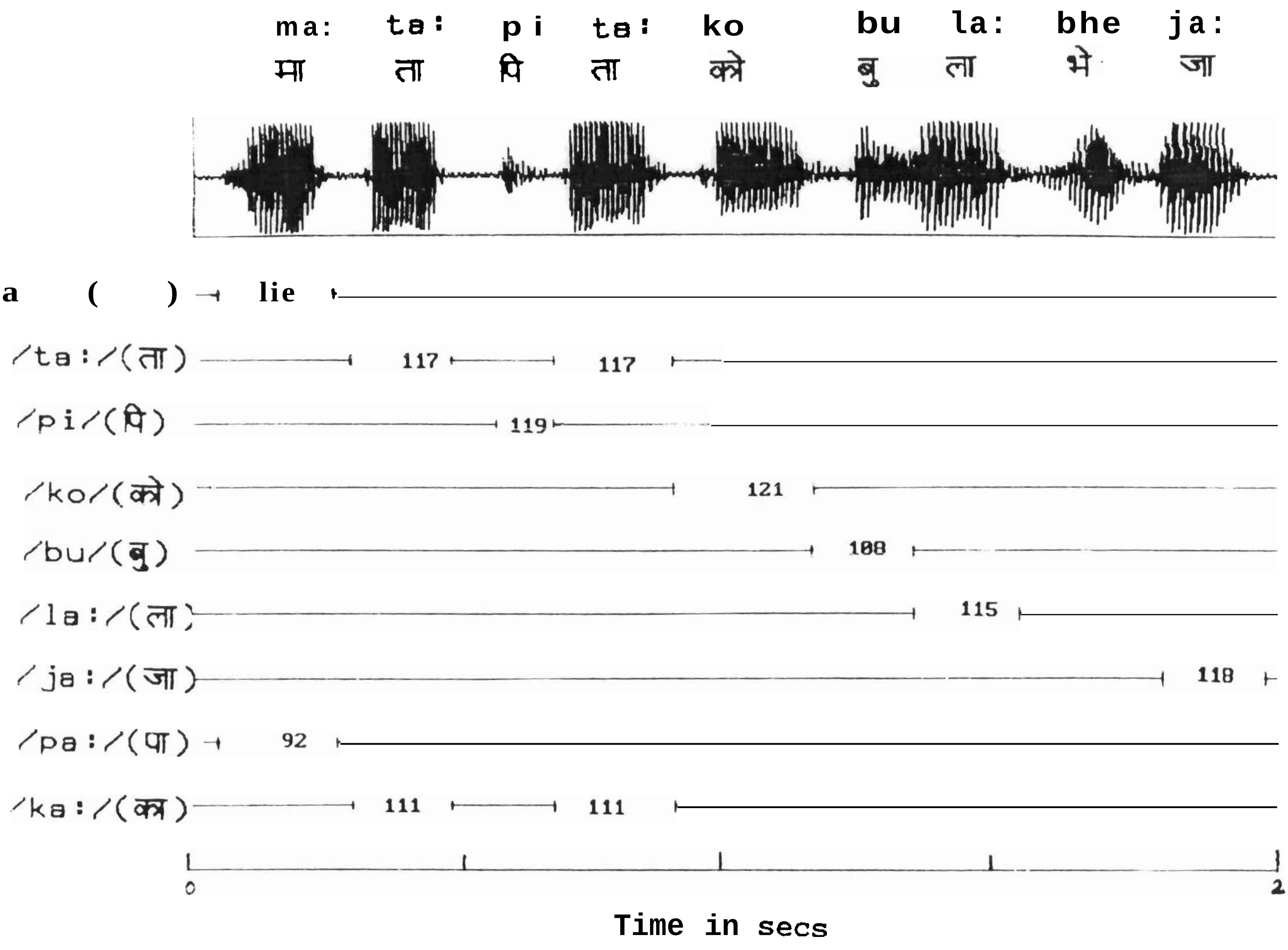


Fig.6.10. Illustration of signal-to-symbol transformation for utterance 1. The confidence level and the region spotted by each character expert are indicated.

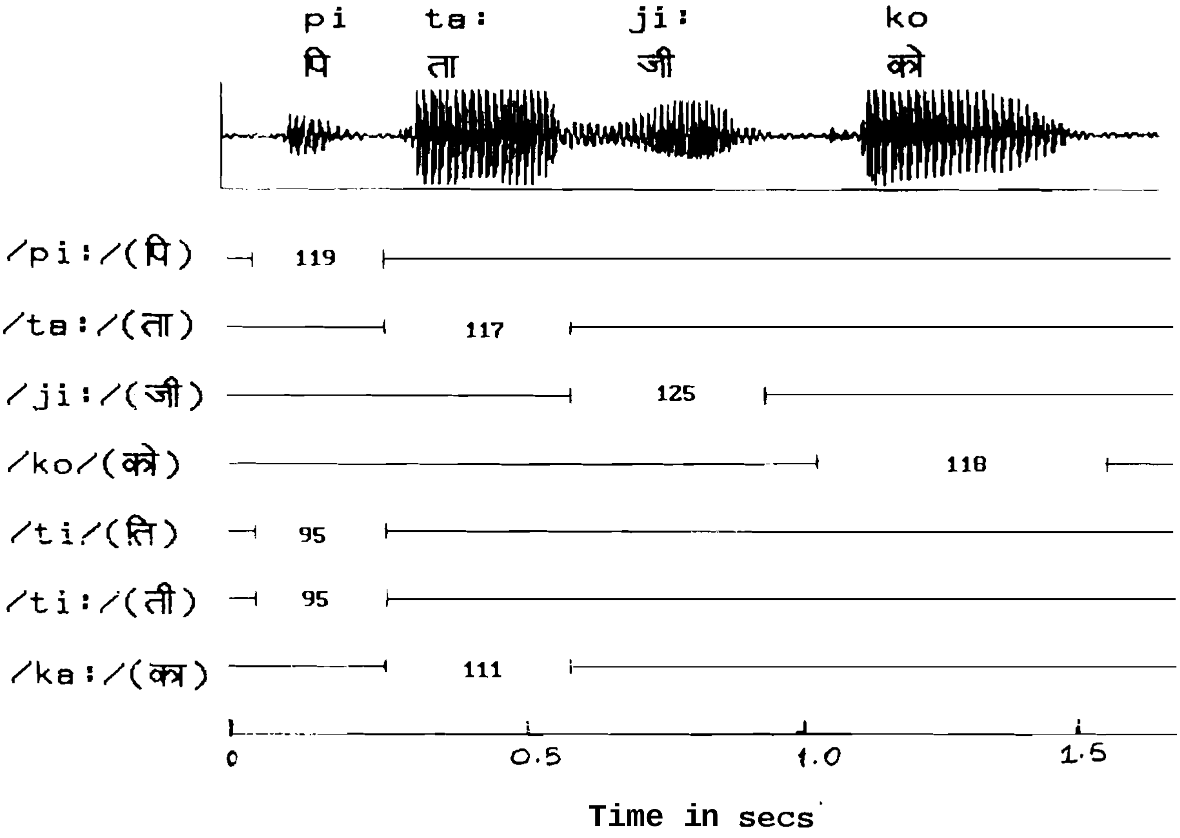


Fig.6.11. Illustration of signal-to-symbol transformation for utterance 2. The confidence level and the region spotted by each character expert are indicated

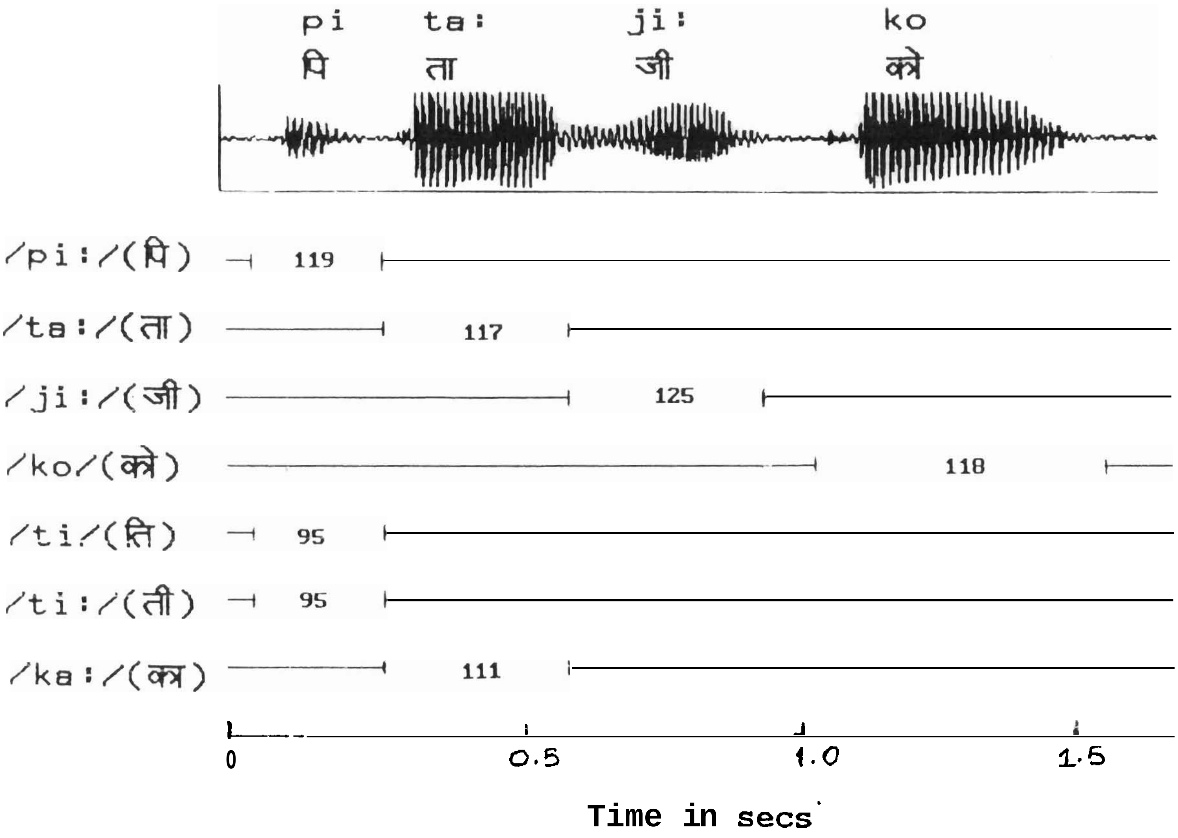


Fig.6.11. Illustration of signal-to-symbol transformation for utterance 2. The confidence level and the region spotted by each character expert are indicated

CHAPTER-VII

SUMMARY AND CONCLUSIONS

This thesis addressed some issues in the development of a speech recognition system for Indian languages. The ultimate goal of this research is to develop a speech-to-text conversion system for Indian languages. The system should give text output that can be understood by a human reader. In this respect the text need not be error free. The system should however be task independent, speaker independent and vocabulary independent. It should accept speech carefully read out from a text in an ordinary office room environment. The objective is to adopt an approach in the design which will eventually accomplish these goals.

Speech-to-text conversion involves two stages, namely, speech signal-to-symbol transformation stage and symbol-to-text conversion stage. Review of literature suggests that the most important block in a continuous speech recognition system is the transformation of speech signal into symbolic form in order to capture the significant phonetic information in the input. Therefore in this thesis we have discussed some issues related to the signal-to-symbol transformation for the Indian language, Hindi. The most important issue is the choice of symbol itself. Proper choice of the symbol significantly reduces the complexity of the symbol-to-text conversion stage. In most Indian languages generally "we write what we speak and we speak what we write". Due to this phonetic nature of these languages we have chosen characters as basic units or symbols. Characters consist of vowels(V), consonants(C), consonant-vowel combinations(CV, CCV, CCCV). Spotting the characters in continuous speech is adopted as the basic approach for signal-to-symbol transformation. Since the number of characters in Hindi is large (about 5000), we have considered in this study only a subset of characters, numbering 340, consisting of the vowels(V) and the consonant-vowel(CV) combinations. This choice was primarily dictated by the fact that these characters occur frequently in Hindi text and also because the design philosophy' can be extended to other characters in a straightforward manner. Moreover, characters consisting of consonant clusters such as CCV and CCCV have much more redundancy and are relatively easier to spot in continuous speech.

To realize the character spotting approach it is necessary to acquire the relevant acoustic-phonetic knowledge for each character and to represent the knowledge in a suitable form. Description of the characters in terms of the speech production mechanism, the acoustic manifestation of speech for each character and the description of acoustic features in terms of the parameters of the speech signal, all these constitute the acoustic-phonetic knowledge. The main source of this knowledge is an expert acoustic-phonetician who can express the features of the characters in terms of the articulatory and acoustic parameters and relate them to the features and parameters derivable from the speech signal. The other sources of knowledge are the literature and experimentation. We have adopted all of these to derive the acoustic-phonetic knowledge for the 340 characters. This knowledge. is represented in the form of production rules. The rules for each character are organized in the form of location rules, intrinsic rules, coarticulation rules and context dependent rules, Gross features such as vocalic, aspiration, burst and silence were used to identify possible regions of the location of the character. Rules for spotting are activated by using an expert system for each character. Due to the ambiguous nature of the features and their relation to speech parameters, fuzzy mathematical concepts were used to infer the results of activation of the rules. The confidence values are derived with tho help of a fuzzy table for the parameter values while activating the rules. The advantage of this approach is that the fuzzy table entries can be refined based on the experimental results on a large set of speech data. The rule base can also be continuously updated by adding, deleting and modifying the rules.

Only 75 of the 350 characters were tested over a large data. The rule base and fuzzy tables for these 75 character experts was refined during experimentation in order to get a good performance for character spotting. The performance of the character spotting experts was studied on utterances of 69 test sentences in Hindi. The performance evaluation studies show that the gross features used for location are spotted with high confidence level. Experiments were also conducted on subsets of confusable characters. The results show that the normally confusable characters such as /ṭa:/(टा), /ta:/(ता) and /pa:/(पा) and /ma:/ (मा) and /na:/(ना) could be spotted with a high level of confidence. Signal-to-symbol transformation of an utterance of a sentence generates a character lattice which suggests that significant phonetic features in the signal can be captured through a string of characters. All these results were demonstrated using only 75 character spotting experts.

The main contributions of this thesis are the following:

(1) Demonstration of the significance of the acoustic-phonetic knowledge to extract the phonetic information in the speech signal in symbolic form.

(2) Importance of the choice of characters as symbols in the development of a speech-to-text conversion system for Indian languages.

(3) Acquisition and representation of acoustic-phonetic knowledge for characters in Hindi.

(4) Development of a rule-based expert system for character spotting in continuous speech in Hindi.

(5) Use of fuzzy mathematical. concepts in interpreting the results of activation of the rules for character spotting.

(6) Demonstration of the potential of character spotting approach for continuous speech recognition in Indian languages.

This thesis is only an attempt to show the possibility of using a character spotting approach for continuous speech recognition in Hindi. The scope of the study was restricted to a few characters occurring in Hindi. But it requires a lot of manual efforts to collect the acoustic-phonetic knowledge and to tune the rule base and the entries in the fuzzy tables. However, this is only a one time effort. Once it is done, it helps to develop a speech-to-text system that is task independent and vocabulary independent. Tuning the rules and tables is also

[illegible]

✓ what kinds of Systematic errors k.s.

needed to accomplish a good recognition performance independent of speaker.

Basically the whole set of character experts have to be implemented and the results of these systems acting simultaneously on the speech signal have to be studied from various angles like better performance and real-time response. It is observed that in a few spotting systems 'the fuzzy tables have to be refined based on speech from various speakers. This refinement is done manually at present. ~~It has to be automated.~~ This means that the system has to be provided with the learning capability to refine the rules automatically. A system that combines the learning capabilities of neural networks at the lower stage (speech parameters to acoustic feature transformation) with the knowledge-based approach in feature to character conversion can be implemented. It is possible to achieve real-time response from the system if all the character spotting experts are implemented in parallel. It is also possible to explore grouping the characters so that any common rules need be applied only once.

V

[illegible] ref

There is scope to improve the performance of the character spotting by using other approaches like the hidden Markov model (HMM) and artificial neural networks(ANN). It appears that a combination of these methods may have to be developed to deal with the variations and ambiguities in speech. The main thrust in this development should be to provide methods to process the signal in a manner suitable to spot a given character. While the character spotting approach seems to be promising for phonetic languages like Indian languages, it is not clear at this stage how successful this will be for languages like English.

Japanese, Swedish, Arabic, Spanish

Appendix I

DESCRIPTION OF CHARACTER SET

In this appendix, we list all the 340 characters occuring as consonant-vowel combinations. For each Hindi character, the notation we follow to represent it and the equivalent Computer Ponetic Alphabet (CPA)[1] notation are provided. The description of the character is given in detail. Some of the characters could not be represented using Computer Phonetic Alphabet. We have used the nearest fit.

The notation which we follow for the description of the character in terms of phonetic and articulary features is explained below. In this notation we have used the diacritic 'h' to represent aspirated sounds. For some sounds in Hindi where there is no equivalent in CPA, nearest diacritics are used to represent the equivalent CPA notation.

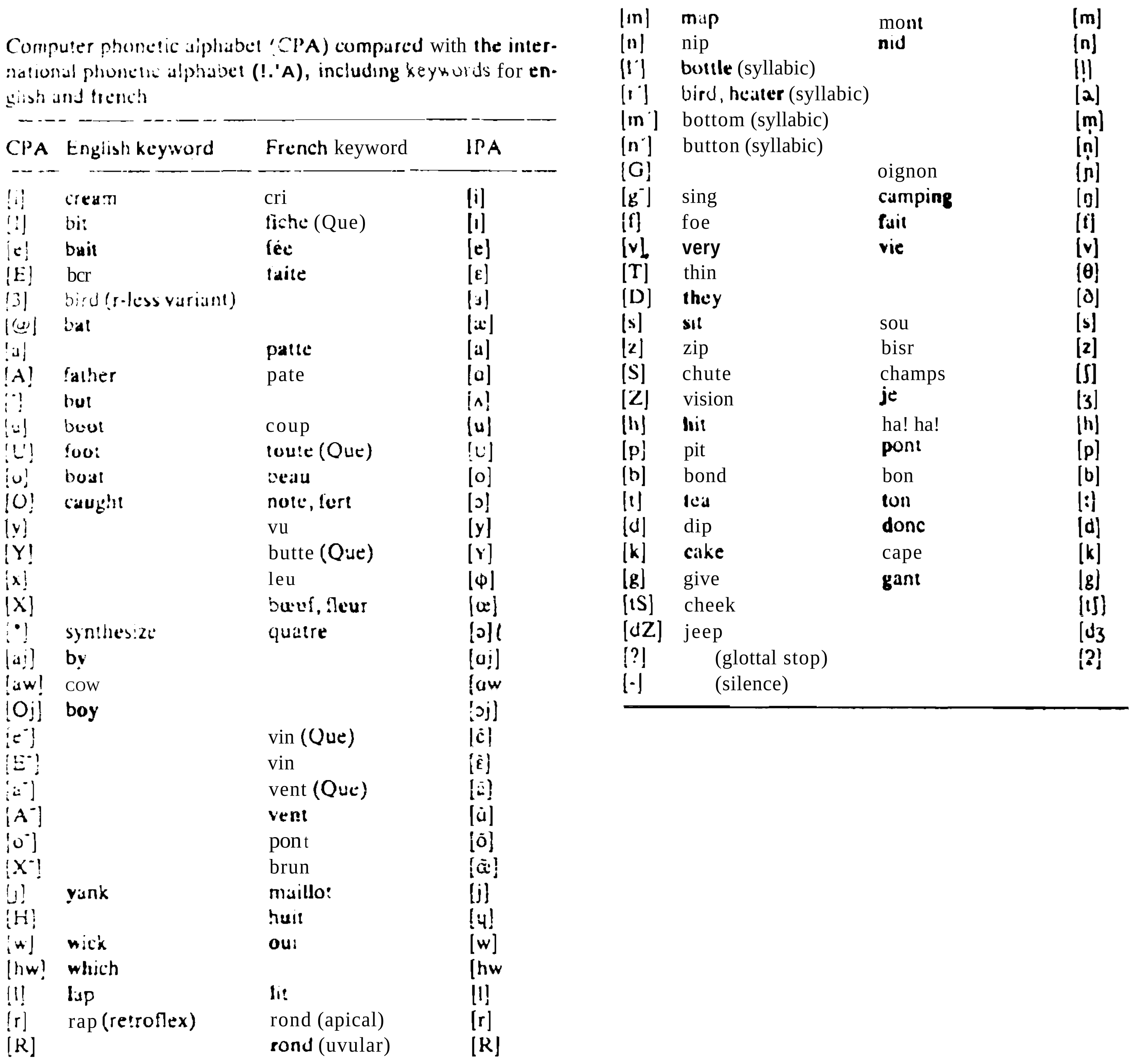
In this report, all noncluster characters in Hindi are described in terms of the following set of features:

(1)Front
(2)Back
(3) Central
(4)Open
(5)Close
(6)Halfclose
(7)Rounded
(8)Unrounded
(9)Short
(10)Long
(11)Diphthong
(12)Voiced
(13)Unvoiced
(14)Aspirated
(15)Unaspirated
(16)Velar
(17)Palatal
(18)Alveolar
(19)Dental
(20)Bilabial
(21) Labio-dental
(22)Retrofiex
(23)Fricative
(24)Lateral
(25)Trill
(26)Semivowel
(27)Nasal

Notation:

{ } Beginning and end of the description. || 'Followed by'.

(, ,) All features within parentheses separated by commas should be present simultaneously,



Computer phonetic alphabet (CPA) compared with the international phonetic alphabet (I.'A), including keywords for english and french

CPA	English keyword	French keyword	IPA
[i]	cream	cri	[i]
[I]	bit	fiche (Que)	[ɪ]
[e]	bait	fée	[e]
[E]	bcr	faite	[ɛ]
[3]	bird (r-less variant)		[ɜ]
[@]	bat		[æ]
[a]		patte	[a]
[A]	father	pate	[ɑ]
[^]	but		[ʌ]
[u]	boot	coup	[u]
[U]	foot	toute (Que)	[ʊ]
[o]	boat	beau	[o]
[O]	caught	note, fort	[ɔ]
[y]		vu	[y]
[Y]		butte (Que)	[ʏ]
[x]		leu	[φ]
[X]		bœuf, fleur	[œ]
[*]	synthesize	quatre	[ɔ](
[aj]	by		[ɑj]
[aw]	cow		[ɑw
[Oj]	boy		[ɔj]
[e~]		vin (Que)	[ẽ]
[E~]		vin	[ɛ̃]
[a~]		vent (Que)	[ã]
[A~]		vent	[ɑ̃]
[o~]		pont	[õ]
[X~]		brun	[œ̃]
[j]	yank	maillot	[j]
[H]		huit	[ɥ]
[w]	wick	oui	[w]
[hw]	which		[hw
[l]	lap	lit	[l]
[r]	rap (retroflex)	rond (apical)	[r]
[R]		rond (uvular)	[R]
[m]	map	mont	[m]
[n]	nip	nid	[n]
[l']	bottle (syllabic)		[l̩]
[r']	bird, heater (syllabic)		[ɚ]
[m']	bottom (syllabic)		[m̩]
[n']	button (syllabic)		[n̩]
[G]		oignon	[ɲ]
[g~]	sing	camping	[ŋ]
[f]	foe	fait	[f]
[v]	very	vie	[v]
[T]	thin		[θ]
[D]	they		[ð]
[s]	sit	sou	[s]
[z]	zip	bisr	[z]
[S]	chute	champs	[ʃ]
[Z]	vision	je	[ʒ]
[h]	hit	ha! ha!	[h]
[p]	pit	pont	[p]
[b]	bond	bon	[b]
[t]	tea	ton	[t]
[d]	dip	donc	[d]
[k]	cake	cape	[k]
[g]	give	gant	[g]
[tS]	cheek		[tʃ]
[dZ]	jeep		[dʒ
[?]	(glottal stop)		[ʔ]
[-]	(silence)		

Computer Phonetic Alphabet(CPA) reproduced from

Ref. 1. **Matthew Lennig and Jean Paul Brassard, "Machine-Readable Phonetic Alphabet for English and French", Speech Communication, Vol.3, pp 166, 1984.**

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
1	/a/	अ	[^]	((Short, Open, Central/Back, Unrounded)
2	/a:/	आ	[A]	{(Long, Open, Back, Unrounded)}
3	/i/	इ	[I]	((Short, Close, Front, Unrounded))
4	/i:/	ई	[i]	{(Long, Close, Front, Unrounded)}
5	/u/		[U]	{(Short, Close, Back, Roundedj}
6	/u:/	ऊ	[u]	{(Long, Close, Back, Rounded)}
7	/e/	ए	[E]	{(Short, Half-close, Front, Unrounded))
8	/ai/	ऐ	[aj]	{((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
9	/o/	ओ	]	{(Short, Half-close, Back, Rounded))
10	/au/	औ	[aw]	{((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
11	/ka/	क	[k^]	((Unvoiced, Unaspirated, Velar) || (Short, Open, Back, Unrounded))
12	/ka:/		[kA]	((Unvoiced, Unaspirated, Velar) || (Long, Open, Back, Unrounded))
13	/ki/	कि	[kI]	((Unvoiced, ~naspirated,Velar) || (Short, Close, Front, Unrounded))
14	/ki:/	की	[ki]	((Unvoiced, Unaspirated, Velar) || (Long, Close, Front, Unrounded))
15	/ku/	ु	[kU]	{(Unvoiced, Unaspirated, Velar) || (Short, Close, Back, Rounded))
16	/ku:/	, ू	[ku]	((Unvoiced, Unaspirated, Velar) || (Long, Close, Back, Rounded))
17	/ke/	के	[kE]	((Unvoiced, Unaspirated, Velar) || (Short, Half-close, Front, Unrounded)
18	/kai/	कै	a	{(Unvoiced, Unaspirated, Velar) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
19	/ko/	को	[kO]	{(Unvoiced, Unaspirated, Velar) || (Short, Half-close, Back, Rounded))
20	/kau/	कौ	[kaw]	((Unvoiced, Unaspirated, || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
21	/ma/	ख	[k^h^]	((Unvoiced, Aspirated, Velar) || (Short, Open, Back, Unrounded))
22	/kha:/		[k^{h}A]	((Unvoiced, Aspirated, Velar) || (Long, Open, Back, Unrounded))
23	/khi/	खि	[k^{h}I]	((Unvoiced, Aspirated, Velar) || (Short, Close, Front, Unrounded))
24	/khi:/	खी	[k^{h}i]	((unvoiced, Aspirated, Velar) || (Long, Close, Front, Unrounded))
25	/khu/	खु	[k^{h}U]	((Unvoiced, Aspirated, Velar) || (Short, Close, Back, Rounded))
26	/khu:/	खू	[k^{h}u]	((Unvoiced, Aspirated, Velar) || (Long, Close, Back, Rounded))
27	/khe/	खे	[k^{h}E]	((Unvoiced, Aspirated, Velar) || (Short, Half-close,Front, Unrounded))
28	/khai/	खै	[k^{h}aj]	((Unvoiced, Aspirated, Velar) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
29	/kho/	खो	[k^{h}o]	((Unvoiced, Aspirated, Velar) || (Short, Half-close, Back, Rounded))
30	/khau/	खौ	[k^{h}aw]	((Unvoiced, Aspirated, Velar) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
31	/ga/	ग	[g^]	{(Voiced, Unaspirated, Velar) || (Short, Open, Back, Unrounded))
32	/ga:/		[gA]	((Voiced, Unaspirated, Velar) || (Long, Open, Back, Unrounded))
33	/gi/	गि	[gI]	((Voiced, Unaspirated, Velar) || (Short, Close, Front, Unrounded))
34	/gi:/	गी	[gi]	((Voiced, Unaspirated, Velar) || (Long, Close, Front, Unrounded))
35	/gu/	गु	[gU]	((Voiced, Unaspirated, Velar) || (Short, Close, Back, Rounded))
36	/gu:/	गू	[gu]	((Voiced, Unaspirated, Velar) || (Long, Close, Back, Rounded))
37	/ge/	गे	[gE]	((Voiced, Unaspirated, Velar) || (Short, Half-close, Front, Unrounded)
38	/gai/	गै	[gaj]	((Voiced, Unaspirated, Velar) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
39	/go/	गो	[gO]	((Voiced, Unaspirated, Velar) || (Short, Half-close, Back, Rounded))
40	/gau/	गौ	[gaw]	((Voiced, Unaspirated, Velar) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phon- etic Code	Hindi Symbol	CPA Code	Description
41	/gha/	घ	[g^h^]	((Voiced, Aspirated, Velar) || (Short, Open, Back, Unrounded)}
42	/gha:/		[g^hA]	((Voiced, Aspirated, Velar) || (Long, Open, Back, Unrounded))
43	/ghi/	घि	[g^hI]	((Voiced, Aspirated, Velar) || (Short, Close, Front, Unrounded))
44	/ghi:/	घी	[g^hi]	((Voiced, Aspirated, Velar) || (Long, Close, Front, Unrounded)}
45	/ghu/	घु	[g^hU]	{(Voiced, Aspirated, Velar) || (Short, Close, Back, Rounded)}
46	/ghu:/	घू	[g^hu]	{(Voiced, Aspirated, Velar) || (Long, Close, Back, Rounded))
47	/ghe/	घे	[g^hE]	((Voiced, Aspirated, Velar) || (Short, Half-close, Front, Unroundedj
48	/ghai/	घै	[g^haj]	((Voiced, Aspirated, Velar) || ((Short, Open,Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
49	/gho/	घो	[g^hO]	((Voiced, Aspirated, Velar) || (Short, Half-close, Back, Rounded)}
50	/ghau/	घौ	[g^haw]	((Voiced, Aspirated, Velar) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
51	/~a/	च	[tS^]	((Unvoiced, Unaspirated, Palatal) || (Short, Open, Front, Unrounded))
52	/ca:/	चा	[tSA]	((Unvoiced, Unaspirated, Palatal) || (Long, Open, Back, Unrounded))
53	/ci/	चि	[tSI]	((Unvoiced, Unaspirated, Palatal) || (Short, Close, Front, Unrounded))
54	/ci:/	ची	[tSi]	{(Unvoiced, Unaspirated, Palatal) || (Long, Close, Front, Unrounded))
55	/cu/	चु	[tSU]	((Unvoiced, Unaspirated, Palatal) || (Short, Close, Back, Rounded))
56	/cu:/	चू	[tSu]	((Unvoiced, Unaspirated, Palatal) || (Long, Close, Back, Rounded)}
57	/ce/	चे	[tSE]	((Unvoiced, Unaspirated, Palatal) || (Short, Half-close, Front, Unrounded)
58	/cai/	चै	[tSaj]	((Unvoiced, Unaspirated, Palatal) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
59	/co/	चो	[tSo]	((Unvoiced, Unaspirated, Palatal) || (Short, Half-close, Back, Rounded))
60	/cau/	चौ	[tSaw]	((Unvoiced, Unaspirated, Palatal) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Pho-- etic Code	Hindi Symbol	CPA Code	Description
61	/cha/	छ	[tsʰ^]	{(Unvoiced, Aspirated, Palatal) || (Short, Open, Front, Unrounded)}
62	/cha:/	छा	[tsʰA]	{(Unvoiced, Aspirated, Palatal) || (Long, Open, Back, Unrounded)}
63	/chi/	छि	[tsʰI]	{(Unvoiced, Aspirated, Palatal) || (Short, Close, Front, Unrounded)}
64	/chi:/	छी	[tsʰi]	{(Unvoiced, Aspirated, Palatal) || (Long, Close, Front, Unrounded)}
65	/chu/	छु	[tsʰU]	{(Unvoiced, Aspirated, Palatal) || (Short, Close, Back, Rounded)}
66	/chu:/	छू	[tsʰu]	{(Unvoiced, Aspirated, Palatal) || (Long, Close, Back, Rounded)}
67	/che/	छे	[tsʰE]	{(Unvoiced, Aspirated, Palatal) || (Short, Half-close, Front,Unrounded)}
68	/chai/	छै	[tsʰaj]	{(Unvoiced, Aspirate d,Palatal) || ((Short, Open, Cen tnl, Unrounded) || (Short, Close. Fro nt Unrounded), Diphthong)}
69	/cho/	छो	[tsʰo]	{(Unvoiced, Aspirated, Palatal) || (Short, Half-close, Back, Rounded)}
70	/chau/	छौ	[tsʰaw]	{(Unvoiced, Aspirated, Palatal) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong)}

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
71	/ja/		[dZ^]	((Voiced, Unaspirated, Palatal) || (Short, Open, Front, Unrounded))
72	/ja:/	जा	[dZA]	((Voiced, Unaspirated, Palatal) || (Long, Open, Back, Unrounded))
73	/ji/	जि	[dZI]	((Voiced, Unaspirated, Palatal) || (Short, Close, Front, Unrounded)}
74	/ji:/	जी	[dZi]	((Voiced, Unaspirated, Palatal) || (Long, Close, Front, Unrounded))
75	/ju/	जु	[dZU]	((Voiced, Unaspirated, Palatal) || (Short, Close, Back, Rounded))
76	/ju:/	जू	[dZu]	((Voiced, Unaspirated, Palatal) || (Long, Close, Back, Rounded)}
77	/je/	जे	[dZE]	((Voiced, Unaspirated, Palatal) || (Short, Half-close, Front, Unrounded)
78	a	जै	[dZaj]	{(Voiced, Unaspirated, Palatal) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong)}
79	/jo/	जो	[dZO]	((voiced, Unaspirated, Palatal) || (Short, Half-close, Back, Rounded)}
80	/jau/	जौ	[dZaw]	{(Voiced, Unaspirated, Palatal) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

<table>
<tr><th>No.</th><th>Phon-
etic
Code</th><th>Hindi
Symbol</th><th>CPA
Code</th><th>Description</th></tr>
<tr><td>81</td><td>/jha/</td><td>झ</td><td>[dz^{h}^]</td><td>((Voiced, Aspirated, Palatal) || (Short, Open, Front, Unrounded))</td></tr>
<tr><td>82</td><td>/jha:/</td><td>झा</td><td>[dz^{h}A]</td><td>((Voiced, Aspirated, Palatal) || (Long, Open, Back, Unrounded))</td></tr>
<tr><td>83</td><td>/jhi/</td><td>झि</td><td>[dz^{h}I]</td><td>((Voiced, Aspirated, Palatal) || (Short, Close, Front, Unrounded))</td></tr>
<tr><td>84</td><td>/jhi:/</td><td>झी</td><td>[dz^{h}i]</td><td>((voiced, Aspirated, Palatal) || (Long, Close, Front, Unrounded)}</td></tr>
<tr><td>85</td><td>/jhu/</td><td>झु</td><td>[dz^{h}U]</td><td>((voiced, Aspirated, Palatal) || (Short, Close, Back, Rounded))</td></tr>
<tr><td>86</td><td>/jhu:/</td><td>झू</td><td>[dz^{h}u]</td><td>{(Voiced, Aspirated, Palatal) || (Long, Close, Back, Rounded)}</td></tr>
<tr><td>87</td><td>/jhe/</td><td>झे</td><td>[dz^{h}E]</td><td>((Voiced, Aspirated, Palatal) || (Short, Half-close, Front, Unrounded)</td></tr>
<tr><td>88</td><td>/jhai/</td><td>झै</td><td>[dz^{h}aj]</td><td>((Voiced, Aspirated, Palatal) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong)}</td></tr>
<tr><td>89</td><td>/jho/</td><td>झो</td><td>[dz^{h}o]</td><td>((Voiced, Aspirated, Palatal) || (Short, Half-close, Back, Rounded))</td></tr>
<tr><td>90</td><td>/jhau/</td><td>झौ</td><td>[dz^{h}aw]</td><td>((Voiced, Aspirated, Palatal) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong)}</td></tr>
</table>

'No.	Phonetic Code	Hindi Symbol	CPA Code	Description
91	/$a/	ट	[t^]	((Unvoiced, Unaspirated, Retroflex) || ((Short, Open, Central, Unrounded))
92	/ṭa:/	टा	[tA]	((Unvoiced, Unaspirated, Retroflex) || (Long, Open, Back, Unrounded))
93	/ṭi/		[tI]	((Unvoiced, Unaspirated, Retroflex) || (Short, Close, Front, Unrounded)}
94	/ṭi:/	टी	[ti]	((Unvoiced, Unaspirated, Retroflex) || (Long, Close, Front, Unrounded))
95	/tu/	टु	[tU]	((unvoiced, Unaspirated, Retroflex) || (Short, Close, Back, Rounded))
96	/tu:/	टू	[tu]	((Unvoiced, Unaspirated, Retroflex) || (Long, Close, Back, Rounded))
97	/ṭe/	टे	[tE]	((Unvoiced, Unaspirated, Retroflex) || (Short, Half-close, Front, Unrounded)
98	/tai/	टै	[taj]	((Unvoiced, Unaspirated, Retroflex) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
99	/ṭo/	टो	[tO]	{(Unvoiced, Unaspirated, Retroflex) || (Short, Half-close, Back, Rounded))
100	/ṭau/	टौ	[taw]	((Unvoiced, Unaspirated, Retroflex) || ((Short, Open, Back, Unrounded) || (Short, close, Back, Rounded), Diphthong))

<table>
<tr><th>No.</th><th>Phon-
etic
Code</th><th>Hindi

Symbol</th><th>CPA

Code</th><th>Description</th></tr>
<tr><td>101</td><td>/ṭha/</td><td>ठ</td><td>[t^h^]</td><td>((Unvoiced, Aspirated, Retroflex) ||
(Short, Open, Central, Unrounded))</td></tr>
<tr><td>102</td><td>/ṭha:/</td><td>ठा</td><td>[t^hA]</td><td>((Unvoiced, Aspirated, Retroflex) ||
(Long, Open, Back, Unrounded))</td></tr>
<tr><td>103</td><td>/ṭhi/</td><td>ठि</td><td>[t^hI]</td><td>((Unvoiced, Aspirated, Retroflex) ||
(Short, Close, Front, Unrounded))</td></tr>
<tr><td>104</td><td>/ṭhi:/</td><td>ठी</td><td>[t^hi]</td><td>((Unvoiced, Aspirated, Retroflex) ||
(Long, Close, Front, Unrounded))</td></tr>
<tr><td>105</td><td>/ṭhu/</td><td>ठु</td><td>[t^hU]</td><td>((unvoiced, Aspirated, Retroflex) ||
(Short, Close, Back, Rounded))</td></tr>
<tr><td>106</td><td>/ṭhu:/</td><td>ठू</td><td>[t^hu]</td><td>(Unvoiced, Aspirated, Retroflex) ||
(Long, Close, Back, Rounded))</td></tr>
<tr><td>107</td><td>/ṭhe/</td><td>ठे</td><td>[t^hE]</td><td>((Unvoiced, Aspirated, Retroflex) ||
(Short, Half-close, Front, Unrounded)</td></tr>
<tr><td>108</td><td>/ṭhai/</td><td>ठै</td><td>[t^haj]</td><td>((Unvoiced, Aspirated, Retroflex) ||
((Short, Open, Central, Unrounded) ||
(Short, Close, Front, Unrounded),
Diphthong))</td></tr>
<tr><td>109</td><td>/ṭho/</td><td>ठो</td><td>[t^ho]</td><td>((Unvoiced, Aspirated, Retroflex) ||
(Short, Half-close, Back, Rounded))</td></tr>
<tr><td>110</td><td>/ṭhau/</td><td>ठौ</td><td>[t^haw]</td><td>((Unvoiced, Aspirated, Retroflex) ||
((Short, Open, Back, Unrounded) ||
(Short, Close, Back, Rounded),
Diphthong))</td></tr>
</table>

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
111	/ḍa/	ड	[d^]	{(Voiced, Unaspirated, Retroflex) || (Short, Open, Central, Unrounded))
112	/ḍa:/	डा	[dA]	((Voiced, Unaspirated, Retroflex) || (Long, Open, Back, Unrounded))
113	/ḍi/	डि	[dI]	{(Voiced, Unaspirated, Retroflex) || (Short, Close, Front, Unrounded))
114	/ḍi:/	डी	[di]	((Voiced, Unaspirated, Retroflex) || (Long, Close, Front, Unrounded))
115	/ḍu/	डु	[dU]	((Voiced, Unaspirated, Retroflex) || (Short, Close, Back, Rounded))
116	/ḍu:/	डू	[du]	{(Voiced, Unaspirated, Retroflex) || (Long, Close, Back, Rounded)}
117	/ḍe/	डे	[dE]	{(Voiced, Unaspirated, Retroflex) || (Short, Half-close, Front, Unrounded)
118	/ḍai/	डै	[daj]	((Voiced, Unaspirated, Retroflex) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong)}
119	/ḍo/	डो	[dO]	((Voiced, Unaspirated, Retroflex) || (Short, Half-close, Back, Rounded))
120	/ḍau/	डौ	[daw]	{(Voiced, Unaspirated, Retroflex) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
121	/ḍha/	ढ	[d^{h}^]	((Voiced, Aspirated, Retroflex) || (Short, Open, Central, Unrounded))
122	/ḍha:/	ढा	[d^{h}A]	((Voiced, Aspirated, Retroflex) || (Long, Open, Back, Unrounded))
123	/ḍhi/	ढि	[d^{h}I]	((Voiced, Aspirated, Retroflex) || (Short, Close, Front, Unrounded))
124	/ḍhi:/	ढी	[d^{h}i]	((Voiced, Aspirated, Retroflex) || (Long, Close, Front, Unrounded))
125	/ḍhu/	ढु	[d^{h}U]	((Voiced, Aspirated, Retroflex) || (Short, Close, Back, Rounded))
126	/ḍhu:/	ढू	[d^{h}u]	((Voiced, Aspirated, Retroflex) || (Long, Close, Back, Rounded))
127	/ḍhe/	ढे	[d^{h}E]	((Voiced, Aspirated, Retroflex) || (Short, Half-close, Front, Unrounded)
128	/ḍhai/	ढै	[d^{h}aj]	((Voiced, Aspirated, Retroflex) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
129	/ḍho/	ढो	[d^{h}o]	((voiced, Aspirated, Retroflex) || (Short, Half-close, Back, Rounded))
130	/ḍhau/	ढौ	[d^{h}aw]	((Voiced, Aspirated, Retroflex) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
131	/ta/	त	T	((Unvoiced, Unaspirated, Denti-alveolar) || (Short, Open, Central/Back, Unrounded))
132	/ta:/	ता	[TA]	((Unvoiced, Unaspirated, Denti-alveolar) || (Long, Open, Back, Unrounded))
133	/ti/	ति	[TI]	((unvoiced, Unaspirated, Denti-alveolar) || (Short, Close, Front, Unrounded))
134	/ti:/	ती	[Ti]	((Unvoiced, Unaspirated, Denti-alveolar) || (Long, Close, Front, Unrounded))
135	/tu/	तु	[TU]	((Unvoiced, Unaspirated, Denti-alveolar) || (Short, Close, Back, Rounded))
136	/tu:/	तू	[Tu]	((Unvoiced, Unaspirated, Denti-alveolar) || (Long, Close, Back, Rounded))
137	/te/	ते	[TE]	((unvoiced, Unaspirated, Denti-alveolar) || (Short, Half-close, Front, Unrounded))
138	a i /	तै	[Taj]	((Unvoiced, Unaspirated, Denti-alveolar) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
139	/to/	तो	[TO]	((Unvoiced, Unaspirated, Denti-alveolar) || (Short, Half-close, Back, Rounded))
140	/tau/	तौ	[Taw]	((unvoiced, Unaspirated, Denti-alveolar) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
141	h a /	थ	[T^{h}^]	((Unvoiced, Aspirated, Denti-alveolar) || (Short, Open, Central/Back, Unrounded)}
142	/tha:/	▪	[T^{h}A]	((Unvoiced, Aspirated, Denti-alveolar) || (Long, Open, Back, Unrounded)}
143	/thi/	थि	[T^{h}I]	((unvoiced, Aspirated, ent ti-alveolar) || (Short, Close, Front, Unrounded)}
144	/thi:/	-	[T^{h}i]	((Unvoiced, Aspirated, ent ti-alveolar) || (Long, Close, Front, Unrounded))
145	/thu/	थु	[T^{h}U]	((Unvoiced, Aspirated, Denti-alveolar) || (Short, Close, Back, Rounded)}
146	/thu:/	थू	[T^{h}u]	((Unvoiced, Aspirated, Denti-alveolar) || (Long, Close, Back, Rounded)}
147	/the/	थे	[T^{h}E]	((Unvoiced, Aspirated, Denti-alveolar) || (Short, Half-close, Front, Unrounded))
148	/thai/	थै	[T^{h}aj]	((Unvoiced, Aspirated, Denti-alveolar) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong)}
149	/tho/	थो	[T^{h}o]	((Unvoiced, Aspirated, Denti-alveolar) || (Short, Half-close, Back, Rounded)}
150	/thau/	थौ	[T^{h}aw]	((Unvoiced, Aspirated, Denti-alveolar) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong)}

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
151	ha/	द	[D^]	((Voiced, Unaspirated, Denti-alveolar) || (Short, Open, Central, Unrounded))
152	/da:/	दा	[DA]	((Voiced, Unaspirated, Denti-alveolar) || (Long, Open, Back, Unrounded))
153	/di/	दि	[DI]	((Voiced, Unaspirated, Denti-alveolar) || (Short, Close, Front, Unrounded))
154	/di:/	दी	[Di]	((Voiced, Unaspirated, Denti-alveolar) || (Long, Close, Front, Unrounded))
155	/du/	दु	[DU]	((Voiced, Unaspirated, Denti-alveolar) || (Short, Close, Back, Rounded))
156	/du:/	दू	[Du]	((Voiced, Unaspirated, Denti-alveolar) || (Long, Close, Back, Rounded)j
157	/de/	दे	[DE]	((Voiced, Unaspirated, Denti-alveolar) || (Short, Half-close, Front, Unrounded))
158	/dai/	दै	[Daj]	((Voiced, Unaspirated, ent ti-alveolar) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
159	/do/	दो	[DO]	((Voiced, Unaspirated, Denti-alveolar) || (Short, Half-close, Back, Rounded))
160	/dau/	दौ	[Daw]	((Voiced, Unaspirated, d en ti-alveolar) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
161	/dha/	ध	[D^{h}^]	((Voiced, Aspirated, Denti-alveolar) || (Short, Open, Central, Unrounded))
162	/dha:/	धा	[D^{h}A]	((Voiced, Aspirated, Denti-alveolar) || (Long, Open, Back, Unrounded))
163	/dhi/	धि	[D^{h}I]	((Voiced, Aspirated, Denti-alveolar) || (Short, Close, Front, Unrounded))
164	/dhi:/	धी	[D^{h}i]	((Voiced, Aspirated, ent ti-alveolar) || (Long, Close, Front, Unrounded))
165	/dhu/	धु	[D^{h}U]	((Voiced, Aspirated, Denti-alveolar) || (Short, Close, Back, Rounded))
166	/dhu:/	धू	[D^{h}u]	((Voiced, Aspirated, Denti-alveolar) || (Long, Close, Back, Rounded)j
167	/dhe/	धे	[D^{h}E]	((Voiced, Aspirated, Denti-alveolar) || (Short, Half-close, Front, Unrounded)
168	/dhai/	धै	[D^{h}aj]	((Voiced, Aspirated, Denti-alveolar; || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
169	/dho/	धो	[D^{h}o]	((Voiced, Aspirated, Denti-alveolar) || (Short, Half-close, Back, Rounded))
170	/dhau/	धौ	[D^{h}aw]	((Voiced, Aspirated, Denti-alveolar) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phon-etic Code	Hindi Symbol	CPA Code	Description
171	/pa/	प	[p^]	((Unvoiced, Unaspirated, Bilabial) || (Short, Open, Central, Unrounded))
172	/pa:/		[pA]	((Unvoiced, Unaspirated, Bilabial) || (Long, Open, Back, Unrounded))
173	/pi/	पि	[pI]	((Unvoiced, Unaspirated, Bilabial) || (Short, Close, Front, Unrounded))
174	/pi:/	पी	[pi]	((Unvoiced, Unaspirated, Bilabial) || (Long, Close, Front, Unrounded))
175	/pu/	ु	[pU]	((Unvoiced, Unaspirated, Bilabial) || (Short, Close, Back, Rounded))
176	/pu:/	पू	[pu]	((Unvoiced, Unaspirated, Bilabial) || (Long, Close, Back, Roundedjj
177	/pe/	पे	[pE]	((Unvoiced, Unaspirated, Bilabial) || (Short, Half-close, Front, Unrounded)
178	/pai/	पै	p a	{(Unvoiced, Unaspirated, Bilabial) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
179	/po/	पो	[pO]	((Unvoiced, Unaspirated, Bilabial) || (Short, Half-close, Back, Rounded))
180	/pau/	पौ	[paw]	((Unvoiced, Unaspirated, Bilabial) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

<table>
<tr><th>No.</th><th>Phon-
etic
Code</th><th>Hindi

Symbol</th><th>CPA

Code</th><th>Description</th></tr>
<tr><td>181</td><td>/~ha/</td><td>फ</td><td>[p^{h}^]</td><td>{(Unvoiced, ~~piratedBilabial) ||
(Short, Open, Central, Unrounded))</td></tr>
<tr><td>182</td><td>/pha:/</td><td>फा</td><td>[p^{h}A]</td><td>{(unvoiced, Aspirated, Bilabial) ||
(Long, Open, Back, Unrounded))</td></tr>
<tr><td>183</td><td>/phi/</td><td>फि</td><td>[p^{h}I]</td><td>{(Unvoiced, Aspirated, Bilabial) ||
(Short, Close, Front, Unrounded))</td></tr>
<tr><td>184</td><td>/phi:/</td><td>फी</td><td>[p^{h}i]</td><td>{(Unvoiced, Aspirated, Bilabial) ||
(Long, Close, Front, Unrounded))</td></tr>
<tr><td>185</td><td>/phu/</td><td>फु</td><td>[p^{h}U]</td><td>{(Unvoiced, Aspirated, Bilabial) ||
(Short, Close, Back, Rounded))</td></tr>
<tr><td>186</td><td>/phu:/</td><td>[illegible]</td><td>[p^{h}u]</td><td>{(Unvoiced, Aspirated, Bilabial) ||
(Long, Close, Sack, Rounded)}</td></tr>
<tr><td>187</td><td>/phe/</td><td>फे</td><td>[p^{h}E]</td><td>{(Unvoiced, Aspirated, Bilabial) ||
(Short, Half-close, Front, Unrounded)</td></tr>
<tr><td>188</td><td>/phai/</td><td></td><td>[p^{h}aj]</td><td>{(unvoiced, Aspirated, Bilabial) ||
((Short, Open, Central, Unrounded) ||
(Short, Close, Front, Unrounded),
Diphthong)}</td></tr>
<tr><td>189</td><td>/pho/</td><td>फो</td><td>[p^{h}o]</td><td>{(Unvoiced, Aspirated, Bilabial) ||
(Short, Half-close, Back, Rounded))</td></tr>
<tr><td>190</td><td>/phau/</td><td>फौ</td><td>[p^{h}aw]</td><td>{(Unvoiced, Aspirated, Bilabial) ||
((Short, Open, Back, Unrounded) ||
(Short, Close, Back, Rounded),
Diphthong)}</td></tr>
</table>

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
191	/ba/	ब	[b^]	((Voiced, Unaspirated, Bilabial) || (Short, Open, Central, Unrounded))
192	/ba:/	बा	[bA]	((Voiced, Unaspirated, Bilabial) || (Long, Open, Back, Unrounded))
193	/bi /	बि	[bI]	((Voiced, Unaspirated, Bilabial) || (Short, Close, Front, Unrounded))
194	/bi:/	बी	[bi]	((Voiced, Unaspirated, Bilabial) || (Long, Close, Front, Unrounded))
195	/bu/	बु	[bU]	((Voiced, Unaspirated, Bilabial) || (Short, Close, Back, Rounded))
196	/bu:/	बू	[bu]	((Voiced, Unaspirated, Bilabial) || (Long, Close, Back, Rounded))
197	/be /	बे	[bE]	((Voiced, Unaspirated, Bilabial) || (Short, Half-close, Front, Unrounded)
198	/bai/	बै	[baj]	((Voiced, Unaspirated, Bilabial) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
199	/bo /	बो	[bo]	((Voiced, Unaspirated, Bilabial) || (Short, Half-close, Back, Rounded))
200	/bau/	बौ	[baw]	{(Voiced, Unaspirated, Bilabial) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
201	/bha/	I?	$[b^{h}\wedge]$	{(Voiced, Aspirated, Bilabial) || (Short, Open, Central, Unrounded))
202	/bha:/	भा	$[b^{h}A]$	{(Voiced, Aspirated, Bilabial) || (Long, Open, Back, Unrounded))
203	/bhi/	भि	$[b^{h}I]$	{(Voiced, Aspirated, Bilabial) || (Short, Close, Front, Unrounded))
204	/bhi:/	भी	$[b^{h}i]$	{(Voiced, Aspirated, Bilabial) || (Long, Close, Front, Unrounded))
205	/bhu/	भु	$[b^{h}U]$	{(Voiced, Aspirated, Bilabial) || (Short, Close, Back, Rounded))
206	/bhu:/	भू	$[b^{h}u]$	{(Voiced, Aspirated, Bilabial) || (Long, Close, Back, Rounded))
207	/bhe/	भे	$[b^{h}E]$	{(Voiced, Aspirated, Bilabial) || (Short, Half-close, Front, Unrounded)
208	/bhai/	भै	$[b^{h}a\ j]$	{(Voiced, Aspirated, Bilabial) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
209	/bho/	भो	$[b^{h}o]$	{(Voiced, Aspirated, Bilabial) || (Short, Half-close, Back, Rounded))
210	/bhau/		$[b^{h}aw]$	{(Voiced, Aspirated, Bilabial) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
211	/?a/	ण	[ṇ^]	{(Retroflex, Nasal) || (Short, Open, Central, Unrounded))
212	/ṇa:/	णा	[ṇA]	{(Retroflex, Nasal) || (Long, Open, Back, Unrounded))
213	/ṇi/	णि	[ṇI]	{(Retroflex, Nasal) || (Short, Close, Front, Unrounded))
214	/ṇi:/	णी	[ṇi]	{(Retroflex, Nasal) || (Long, Close, Front, Unrounded)}
215	/ṇu/	णु	[ṇU]	{(Retroflex, Nasal) || (Short, Close, Back, Rounded))
216	/ṇu:/	णू	[nu]	{(Retroflex, Nasal) || (Long, Close, Back, Rounded)}
217	/ṇe/	णे	[ṇE]	{(Retroflex, Nasal) || (Short, Half-close, Front, Unrounded)
218'	/ṇai/		[ṇaj]	{(Retroflex, Nasal) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong)}
219	/ṇo/	णो	[ṇO]	{(Retroflex, Nasal) || (Short, Half-close, Back, Rounded))
220	/ṇau/	णौ	[ṇaw]	{(Retroflex, Nasal) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
221	/na/	न	n	((Alveolar, Nasal) || (Short, Open, Central, Unrounded))
222	/na:/	ना	[nA]	{(Alveolar, Nasal) || (Long, Open, Back, Unrounded))
223	/ni/	नि	[nI]	((Alveolar, Nasal) || (Short, Close, Front, Unrounded))
224	/ni:/	नी	[ni]	{(Alveolar, Nasal) || (Long, Close, Front, Unrounded))
225	/nu/	नु	[nU]	((Alveolar, Nasal) || (Short, Close, Back, Rounded))
226	/nu:/	नू	[nu]	{(Alveolar, Nasal) || (Long, Close, Back, Rounded))
227	/ne/	ने	[nE]	{(Alveolar, Nasal) || (Short,Half-close,Front, Unrounded))
228	/nai/	नै	[naj]	((Alveolar, Nasal) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
229	/no/	नो	[nO]	{(Alveolar, Nasal) || (Short, Half-close, Back, Rounded))
230	/nau/	नौ	[naw]	((Alveolar, Nasal) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phon-etic Code	Hindi Symbol	CPA Code	Description
231	/ a /	म	[m^]	((Bilabial, Nasal) || (Short, Open, Central, Unrounded))
232	m a :	मा	[mA]	((Bilabial, Nasal) || (Long, Open, Back, Unrounded))
233	/mi/	मि	[mI]	((Bilabial, Nasal) || (Short, Close, Front, Unrounded))
234	/mi:/	मी	[mi]	((Bilabial, Nasal) || (Long, Close, Front, Unrounded))
235	/mu/	मु	[mU]	{(Bilabial, Nasal) || (Short, Close, Back, Rounded))
236	/mu:/	मू	[mu]	((Bilabial, Nasal) || (Long, Close, Back, Rounded)}
237	/me/	मे	[mE]	((Bilabial, Nasal) || (Short, Half-close, Front, Unrounded)
238	/mai/	मै	[maj]	((Bilabial, Nasal) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong)}
239	/mo /	मो	[mO]	((Bilabial, Nasal) || (Short, Half-close, Back, Rounded))
240	/mau/	मौ	[maw]	((Bilabial, Nasal) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phon- etic Code	Hindi Symbol	CPA Code	Description
241	/ya/	य	[j^]	((Voiced, Palatal, semivowel) || (Short, Open, Front, Unrounded))
242	/ya:/		[jA]	((Voiced, Palatal, Semivowel) || (Long, Open, Back, Unrounded))
243	/yi/	यि	[jI]	((Voiced, Palatal, Semivowel) || (Short, Close, Front, Unrounded))
244	/yi:/	यी	[ji]	((Voiced, Palatal, Semivowel) || (Long, Close, Front, Unrounded))
245	/yu/	यु	[jU]	((Voiced, Palatal, Semivowel) || (Short, Close, Back, Rounded))
246	/yu:/	यू	[ju]	((Voiced, Palatal, Semivowel) || (Long, Close, Back, Rounded))
247	/ye/	ये	[jE]	((Voiced, Palatal, Semivowel) || (Short, Half-close, Front, Unrounded)
248	/yai/	यै	[jaj]	((Voiced, Palatal, Semivowel) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
249	/yo/	यो	[jo]	((Voiced, Palatal, Semivowel) || (Short, Half-close, Back, Rounded))
250	/yau/	.	[jaw]	((Voiced, Palatal, Semivowel) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phon- etic Code	Hindi Symbol	CPA Code	Description
251	Ia /	र	[r^]	{(Voiced, Alveolar, Trill) || (Short, Open, Central, Unrounded))
252	/ra:/	रा	[rA]	((Voiced, Alveolar, Trill) || (Long, Open, Back, Unrounded))
253	/ri /	रि	[rI]	((Voiced, Alveolar, Trill) || (Short, Close, Front, Unrounded)}
254	/ri:/	री	[ri]	((Voiced, Alveolar, Trill) || (Long, Close, Front, Unrounded))
255	/ru /	रु	[rU]	{(Voiced, Alveolar, Trill) || (Short, Close, Back, Rounded))
256	/ru:/	रू	[ru]	{(Voiced, Alveolar, Trill) || (Long, Close, Back, Rounded! }
257	/re/	रे	[rE]	(Voiced, Alveolar, Trill) || (Short, Half-close, Front, Unrounded)
258	/rai/	रै	[raj]	((Voiced, Alveolar, Trill) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
259	/ro/		O	((Voiced, Alveolar, Trill) || (Short, Half-close, Back, Rounded)}
260	/rau/	रौ	[raw]	{(Voiced, Alveolar, Trill) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
261	/la/	ल	[l^]	((Voiced, Alveolar, Lateral) || (Short, Open, Central, Unrounded))
262	/la:/	ला	[lA]	((Voiced, Alveolar, Lateral) || (Long, Open, Back, Unrounded))
263	/li/	लि	[lI]	((Voiced, Alveolar, Lateral) || (Short, Close, Front, Unrounded))
264	/li:/	ली	[li]	((Voiced, Alveolar, Lateral) || (Long, Close, Front, Unrounded))
265	/lu/	लु	[lU]	((voiced, Alveolar, Lateral) || (Short, Close, Back, Rounded))
266	/lu:/	लू	1	((Voiced, Alveolar, Lateral) || (Long, Close, Back, Rounded; }
267	/le/	ले	[lE]	((Voiced, Alveolar, Lateral) || (Short, Half-close, Front, Unrounded)
268	/lai/	लै	[laj]	{(Voiced, Alveolar, Lateral) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong)}
269	/lo/	लो	[lO]	((voiced, Alveolar, Lateral) || (Short, Half-close, Back, Rounded))
270	/lau/	लौ	[law]	((Voiced, Alveolar, Lateral) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phon- etic Code	Hindi Symbol	CPA Code	Description
271	~a/	व	[v^]	((Voiced, Labio-dental, semivowel) || (Short, Open, Front, Unrounded))
272	/va:/	वा	[vA]	((Voiced, Labio-dental, Semivowel) || (Long, Open, Back, Unrounded))
273	/vi/	वि	[vI]	((Voiced, Labio-dental, Semivowel) || (Short, Close, Front, Unrounded))
274	/vi:/	वी	[vij	((Voiced, Labio-dental, Semivowel) || (Long, Close, Front, Unrounded))
275	/vu/	वु	[vU]	((Voiced, Bilabial, Semivowel) || (Short, Close, Back, Rounded))
276	/vu:/	वू	[vu]	((Voiced, Bilabial, Semivowel) || (Long, Close, Back, Rounded))
277	/ve/	वे	[vE]	((Voiced, Labio-dental, semivowel) || (Short, Half-close, Front, Unrounded)
278	/vai/	वै	[vaj]	((Voiced, Labio-dental, Semivowel) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
279	/vo/	वो	[vO]	((Voiced, Bilabial, Semivowel) || (Short, Half-close, Back, Rounded))
280	/vau/	वौ	[vaw]	((Voiced, Labio-dental, Semivowel) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
281	/śa/	श	[ś^]	((Unvoiced, Palatal, Fricative) || (Short, Open, Front, Unrounded))
282	/śa:/	शा	[śA]	((Unvoiced, Palatal, Fricative) || (Long, Open, Back, Unrounded))
283	/śi/	शि	[śI]	((Unvoiced, Palatal, Fricative) || (Short, Close, Front, Unrounded))
284	/śi:/	शी	[śi]	((Unvoiced, Palatal, Fricative) || (Long, Close, Front, Unrounded))
285	/śu/	शु	[śU]	((Unvoiced, Palatal, Fricative) || (Short, Close, Back, Rounded))
286	/śu:/	शू	[śu]	((Unvoiced, Palatal, Fricative) || (Long, Close, Back, Rounded))
287	/śe/	शे	[śE]	((Unvoiced, Palatal, Fricative) || (Short, Half-close, Front, Unrounded)
288	/śai/	शै	[śaj]	((Unvoiced, Palatal, Fricative) || ((Short, Open, Central, Unrounded) || (Short, close, Front, Unrounded), Diphthong))
289	/śo/	शो	[śo]	((Unvoiced, Palatal, Fricative) || (Short, Half-close, Back, Rounded))
290	/śau/	शौ	[śaw]	((Unvoiced, Palatal, Fricative) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
291	pa/	ष	[Z^]	((Unvoiced, Retroflex, Fricative) || (Short, Open, Central, Unrounded))
292	/ṣa:/	षा	[ZA]	((Unvoiced, Retroflex, Fricative) || (Long, Open, Back, Unrounded))
293	/ṣi/		[ZI]	((Unvoiced, Retroflex, Fricative) || (Short, Close, Front, Unrounded))
294	/ṣi:/	षी	[Zi]	((Unvoiced, Retroflex, Fricative) || (Long, Close, Front, Unrounded))
295	/ṣu/	षु	[ZU]	((Unvoiced, Retroflex, Fricative) || (Short, Close, Back, Rounded))
296	/ṣu:/	षू	[Zu]	((Unvoiced, Retroflex, Fricative) || (Long, Close, Back, Rounded))
297	/ṣe/	षे	[ZE]	((Unvoiced, Retroflex, Fricative) || (Short, Half-close, Front, Unrounded)
298	/ṣai/	षै	[Zaj]	((Unvoiced, Retroflex, Fricative) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
299	/ṣo/	षो	[ZO]	((Unvoiced, Retroflex, Fricative) || (Short, Half-close, Back, Rounded))
300	/ṣau/	षौ	[Zaw]	((unvoiced, Retroflex, Fricative) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phon- etic Code	Hindi Symbol	CPA Code	Description
301	/ ~ a /		P A]	{(Unvoiced, Alveolar, Fricative) || (Short, Open, Central, Unrounded))
302	/sa:/	सा	[sA]	{(Unvoiced, Alveolar, Fricative) || (Long, Open, Back, Unrounded))
303	/si /		[sI]	{(Unvoiced, Alveolar, Fricative) || (Short, Close, Front, Unrounded))
304	/si:/	सी	[si]	{(Unvoiced, Alveolar, Fricative) || (Long, Close, Front, Unrounded))
305	/su/	सु	[sU]	('(unvoiced, Alveolar, Fricative) || (Short, Close, Back, Rounded))
306	/su:/	सू	[su]	{[Unvoiced, Alveolar, Fricative) || (Long, Close, Back, Rounded)}
307	/se/	से	[sE]	{(Unvoiced, Alveolar, Fricative) || (Short, Half-close, Front, Unrounded)
308	/sai/	सै	[saj]	{(Unvoiced, Alveolar, Fricative) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
309	/so/	सो	[sO]	{(Unvoiced, Alveolar, Fricative) || (Short, Half-close, Back, Rounded))
310	/sau/	सौ	[saw]	{(Unvoiced, Alveolar, Fricative) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
311	/ha/	ह	[h^]	{(Unvoiced, Glottal, Fricative) || (Short, Open, Central, Unrounded))
312	/ha:/	हा	[hA]	{(Unvoiced, Glottal, Fricative) || (Long, Open, Back, Unrounded))
313	/hi/	हि	I	{(Unvoiced, Glottal, Fricative) || (Short, Close, Front, Unrounded))
314	/hi:/	ही	[hi]	{(Unvoiced, Glottal, Fricative) || (Long, Close, Front, Unrounded))
315	/hu/	हु	[hU]	{(Unvoiced, Glottal, Fricative) || (Short, Close, Back, Rounded))
316	/hu:/	हू	[hu]	{(Unvoiced, Glottal, Fricative) || (Long, Close, Back, Rounded))
317	/he/	हे	[hE]	{(Unvoiced, Glottal, Fricative) || (Short, Half-close, Front, Unrounded)
318	/hai/	है	[haj]	{(Unvoiced, Glottal, Fricative) || ((Short, Open, Central, Unrounded) || (Short, Close, Front, Unrounded), Diphthong))
319	/ho/	हो	[ho]	{(Unvoiced, Glottal, Fricative) || (Short, Half-close, Back, Rounded))
320	/hau/	हौ	[haw]	{(Unvoiced, Glottal, Fricative) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

No.	Phonetic Code	Hindi Symbol	CPA Code	Description
321	/kga/	क्ष	[kZ^]	{(Unvoiced, Unaspirated, Velar) || (Unvoiced, Retroflex, Fricative) || (Short, Open, Central, Unrounded))
322	/kṣa:/	क्षा	[kZA]	{(Unvoiced, Unaspirated, Velar) || (Unvoiced, Retroflex, Fricative) || (Long, Open, Back, Unrounded))
323	/kṣi/	क्षि	[kZI]	((Unvoiced, Unaspirated, Velar) || (Unvoiced, Retroflex, Fricative) || (Short, Close, Front, Unrounded))
324	/kṣi:/	क्षी	[kZi]	((Unvoiced, Unaspirated, Velar) || (Unvoiced, Retroflex, Fricative) || (Long, Close, Front, Unrounded)}
325	/kṣu/	क्षु	[kZU]	((Unvoiced, Unaspirated, Velar) || (Unvoiced, Retroflex, Fricative) || (Short, Close, Back, Rounded))
326	/kṣu:/	क्षू	[kZu]	((Unvoiced, Unaspirated, Velar) || (Unvoiced, Retroflex, Fricative) || (Long, Close, Back, Rounded))
327	/kṣe/	क्षे	[kZE]	{(Unvoiced, Unaspirated, Velar) || (Unvoiced, Retroflex, Fricative) || (Short, Half-close, Front, Unrounded)
328	/kṣai/	क्षै	[kZaj]	{(Unvoiced, Unaspirated, Velar) || (Unvoiced, Retroflex, Fricative) || ((Short, Open, Central, Unrounded) || (Short,Close,Front,Unrounded),Diphtho
329	/kṣo/		[kZO]	((Unvoiced, Unaspirated, Velar) || (Unvoiced, Retroflex, Fricative) || (Short, Half-close, Back, Rounded))
330	/kṣau/	क्षौ	[kZaw]	((Unvoiced, Unaspirated, Velar) || (Unvoiced, Retroflex, Fricative) || ((Short, Open, Back, Unrounded) || (Short, Close, Back, Rounded), Diphthong))

Appendix 2

RULE BASE AND FUZZY TABLE FOR THE CHARACTER /ka:/(का)

In this appendix we give a list of the rule base used for spotting the character /ka:/(का) and the associated fuzzy table. The rule base with the associated fuzzy table uses the various parameters and the thresholds used locating different gross features and the character. The fuzzy table gives an idea as to how these parameter values are used to compute the confidence measures. For example, in the detection of vocalic region we use Log Energy (ENR) or first linear prediction coefficient (LP1) and this is indicated in the antecedents of the rule for voicing. The number shown along with the parameter indicates the entry in the fuzzy table that has to be used for calculation of confidence measure. In order to find the confidence level the second arguement in the fuzzy table indicates the type of fuzzy curve to be used. This table gives the threshold to be used for calculation of confidence. For example the rule corresponding to voicing says that it should consider ENR and LP1. The function chk_enr_limit(ENR 4) means'that this function uses the thresholds provided by the 4 th row in the fuzzy table. This has 4 arguements represented by argl, arg2, arg3 and arg4 as indicated in the fuzzy table. The first arguement says that S curve be used to obtain the ccnfidence and the other three arguements provide the necessary thresholds for S curve. The values provided in the 4th row of fuzzy table are 0,200,210 and 220. This shows that the confidence measure is obtained using a S curve represented by the first zero and the limits on S curve are given by 200, 210 and 220. This shows that any value of energy above 220 will have a maximum confidence and any value below 200 will have minimum confidence. The intermediate values have a confidence ranging from maximum to minimum. The maximum and minimum values in our study are 127 and 0 respectively. Similalrly other functions and predicates of a rule can be evaluated. The 'd's indicated in the column corresponding to number of arguements relate to the number of arguements used by a function when computing confidence factors. The number of arguements a function can take when using the fuzzy table differ and this is indicated by the number of 'd's in the second column.

Rule Base for spotting character /ka:/ (का)

```
$start
IF Max
THEN  init1(),init2().
IF Max
THEN initialize(),init_no(Cur_smpl_no),CHNG_CNTXT($silence)
$silence
IF  check-end-file()
THEN  initialize(),CHNG_CNTXT( $unaspirated).
IF chk_param_limit1(LE 15)
THEN  confl.
IF chk_param_limit(HLR  14)
THEN  conf2.
IF chk_param_limit1(LP1 25)
THEN  conf3.
IF   AND(conf3 AND(conf2  confl))
THEN  conf.
IF Max
```

```
THEN  load_cf(conf  1),CHNG_CNTXT( $silence).
$unaspirated
IF check-end-file()
THEN initialize(), CHNG_CNTXT( $burst).
IF chk_param_limit1(LE  9)
THEN confl.
IF chkgaram_limit1(LP1  8)
THEN conf2.
IF chk_param_limit(HLR 28)
THEN conf3.
IF AND(confl AND ( conf3 conf2))
THEN conf.
IF Max
THEN load_cf(conf 2),CHNG_CNTXT( $unaspirated).
$burst
IF  check-end-file()
THEN  initialize(), CHNG_CNTXT( $voiced).
IF chk_param_limit1(LP1 11)
THEN  confl.
IF chk_param_limitl(LE 12)
THEN conf2.
IF chk_param_limit(TRN  13)
THEN conf3.
IF chk_param_limit(HLR  24)
THEN conf4.
IF OR(conf4  AND(confl AND(conf3 conf2 )))
THEN conf.
IF Max
THEN load_cf(conf  3),CHNG_CNTXT( $burst).
$voiced
```

```
IF check-end-file()
THEN initialize(), CHNG_CNTXT( $find_ka).
IF chk_param_limit(LP1 5)
THEN confl.
IF  chk_param_limit(LE  4)
THEN conf2.
IF  OR(confl conf2)
THEN conf.
IF Max
THEN load_cf(conf 0),CHNG_CNTXT( $voiced).
$find_ka
IF Max
THEN  find_ka_region().
IF Max
THEN  init3(),load_cf2(127 4),CHNG_CNTXT( Sloop).
Sloop
IF check_ka_cnt()
THEN  emit2(), exit(1).
IF  check-formants()
THEN confl.
IF  check_burst()
THEN conf2.
IF  AND(confl  conf2)
THEN conf.
IF Max
THEN  load_cf1(conf 5),emit(conf), CHNG_CNTXT( $loop).
```

LE - Log energy   LP1 - first Linear predection coefft.

HLR - ratio of high frequency energy to low frequency energy

SPD - spectrl distance SPF - Spectral fnatness conf- confidence

Fuzzy Table

S. No.	Number of Arguements	Arg 1	Arg 2	Arg 3	Arg 4	Arg 5	Arg 6
1	ddd	250	170	2			
2	dddddd	1	120	0	140	60	250
3	dddd	0	240	245	250		
4	dddd	0	150	160	170		
5	dddd	0	230	240	250		
6	dddd	0	30	80	130		
7	dddd	0	120	150	180		
8	dddd	0	180	190	200		
9	dddd	0	120	125	130		
10	dddd	0	120	160	200		
11	dddd	0	150	165	180		
12	dddd	0	120	125	130		
13	dddd	0	150	160	170		
14	dddd	0	130	140	150		
15	dddd	0	120	125	130		
16	dddd	0	100	130	160		
17	dddd	0	120	130	140		
18	ddddd	1	120	140	60	500	
19	dddd	0	10	15	20		
20	dddd	0	60	70	80		
21	dddd	0	5	10	15		
22	dddd	1	0	80	160		
23	dddd	1	135	165	195		
24	dddd	0	200	230	250		
25	dddd	1	100	160	220		
26	dddd	1	0	50	100		
27	dddd	0	150	175	200		

Arg Arguement

Appendix 3

MODULES OF THE CHARACTER SPOTTING EXPERT

The modules in the character spotting expert system are the following :

(1) Rule preprocessor

(2) Inference Engine

(3) Look-up table

(4) Segment data and working memory

(5) Parameter examine routines

(6) Test functions and action routines

(7) Interactive tracing and tuning unit (ITT)

A brief explanation of each of the above modules is given in the following sections.

A3.1 Rule Preprocessor

This module converts the rule text into a form suitable to the inference engine. The rules are written following the syntax mentioned previously. This module generates a coded rule object file, which contains exact triplet matching of the text file. Each triplet consists of three bytes. First byte consists information about the the type of triplet (function, context, predicate etc.) and the other two contain a unique number associated with each distinct entity. For example, predicate 1 might be stored as shown in Fig.A3.1.

1	2	3
P	0	1

Fig. A3.1. Structure of a triplet for a predicate

The second file generated by this module is meant for 'C' compiling and linking to the other modules of the expert system. It stores the addresses of all the functions and variables and also the offset address of each context group. This module also generates a file which contains all the names used in the rules. The names are grouped together according to their type and are arranged in alphabetical order in each group. Each name in a group will accordingly be given a number and these numbers are stored along with the names. This is to be used in tracing and debugging the whole system. The processor performs simple syntactic analysis and uses symbol table look-up to generate the table files.

A3.2 Inference Engine

The function of this module is to execute the rule base according to the triplets in the coded rule object file. In the cyclic operation it first reads the current triplet in the ruie base, advances the current triplet pointer to the next and takes action according to the type of triplet just read. By the syntax of the rule itself, it knows whether it is evaluating a condition part of the rule or it is firing an action procedure. During the condition evaluation it stores the identity of the function in the hash table. The stored value is used if the same function is being used again before any user written action is fired. It clears the hash table as soon as a user written action function or procedure is found. The arguments are made available to the user written functions by a global array.

A number of system functions like AND, OR, PLUS, ..., ASSIGN are implemented directly into the engine thereby acquiring faster execution of frequently used functions. The hash table is not affected by an action procedure having only system functions.

Condition is evaluated for the current rule. If the condition value is greater than the current threshold then the rule is fired, otherwise the next rule is processed. If the rule is fired and consequent action is triggered, it then goes to the next immediate rule or the first rule of the next context. The above procedure is repeated for every rule that is marked as a current rule. In the training and debug mode it transfers control to the ITT unit after reading each triplet, so that ITT can take appropriate action at any point of execution.

A3.3 Look-up table

It is a collection of record structures of variable size. All the necessary constants, thresholds and fuzzy curves are stored in this table using appropriate record. Any of the records can be referenced by specifying the table-entry number corresponding to it. An array returns pointer to that record by indexing with table-entry number. The contents of the record are read from a file in the first initialization process. At the same time the array is also initialized. The table file is created using a text editor in a prescribed format. The table is referred to the ITT unit so that it can be verified or modified at any time.

A3.4 Segment Data Section and Working Memory

Context information is necessary to hypothesize a symbol for a given segment. The segment data section is meant to store all the information about a segment which has been analyzed. This stores a concise history of the analysis made on any segment. It is also useful when the higher level expert wants additional analysis to be done on a particular segment. Working memory includes all the temporary and intermediate features. Several global parameters are also stored in this memory.

A3.5 Parameter Examine Routines

The parameter examine routines are written to extract any of the descriptive features from any of the parameters specified. These routines are very general in nature taking care of all the cases. The main routines written are computation of level, tracking for a constant level with specified tolerances and location of regions with particular characteristics.

Each parameter number, table entry number and others tells how the analysis has to be performed. For level routine the table entry number determines the mapping to be performed after computing the absolute values. For example, the absolute value can be mapped using a particular fuzzy curve to determine a grade membership to a linguistic modifier say very high energy level.

A function is written which takes the actual value and one of the fuzzy curves and returns the mapped grade membership value. For example, the procedure ER_level(param_no, table-entry-no, 1-span, r_span, cur-smpl-no) will find the average absolute level in the range between cur-smpl-no - l_span and cur-smpl-no + r_span for the indicated parameter number and then the average value will be transformed by the curve indicated by the table_entry-no. It returns the transformed integer. The track function is meant for tracking a parameter curve for a constant level specified till it violates the indicated tolerances.

A3.6 Test Functions and Action Routines

These are collection of functions and procedures which appear in the rules. These are written by user and linked to the engine. Test functions perform tests on various data and return an integer value in the range between 0 - 127 while the action procedures modify or create these data values and also communicate with other modules. The test functions make use of examine routines to verify the presence or absence of certain properties:n the parameter.

A3.7 Interactive Tracking and Tuning

The unit provides interactive facility for tracking and tuning the system during the execution of the rule base. It provides the facilities (1) to evaluate and stop at next rule or action procedure (2) to skip and stop at next rule or action procedure (3) to display data values (4) to modify data values and (5) to set and reset break points and (6) to execute rule functions.

This uses a straight forward implementation. All the data, variables, tables, system status variables are given access tc it. In case of display/modify data enough information (name, address of field in record) is stored and/or asked by the user about the data to calculate its address. Then that particular location is accessed and displayed or modified. In break point mode the ITT unit keeps track of the current rule number and as soon as it becomes equal to one of the set points it stops execution and waits for user response.

Appendix 4

LIST OF CHARACTERS AND SENTENCES USED IN PERFORMANCE EVALUATION OF SIGNALTO-SYMBOL TRANSFORMATION

A4.1 List of characters

01. / ka / (क)
02. / ka: / (का)
03. / ki / (कि)
04. / ki: / (की)
05. / ke / (के)
06. / ko / (को)
07. / ca / (च)
08. / ca: / (चा)
09. / ci / (चि)
10. / cu / (चु)
11. / cu: / (चू)
12. / ṭa / (ट)
13. / ṭa: / (टा)
14. / ti / (टि)
15. / ṭi: / (टी)
16. / ṭu / (टु)
17. / ṭu: / (टू)
18. / ṭe / (टे)
19. / ṭo / (टो)
20. / ta: / (ता)
21. / ti / (ति)
22. / ti: / (ती)
23. / tu / (तु)
24. / tu: / (तू)
25. / to / (तो)
26. / pa / (प)
27. / pa: / (पा)
28. / pi / (पि)
29. / pu: / (पू)
30. / ga / (ग)
31. / ga: / (गा)
32. / gi / (गि)
33. / gi: / (गी)
34. / ja: / (जा)
35. / ji: / (जी)
36. / ju / (जु)
37. / ḍa / (ड)
38. / da: / (डा)
39. / ḍu: / (डू)
40. / ḍi / (डि)
41. / ḍi: / (डी)
42. / da / (द)
43. / da: / (दा)
44. / ba / (ब)
45. / ba: / (बा)
46. / bu / (बु)
47. / bo / (बो)
48. / kha / (ख)
49. / kha:/ (खा)
50. / khi / (खि)
51. / khi:/ (खी)
52. / khe / (खे)
53. / kho / (खो)
54. / tha / (थ)
55. / pha / (फ)
56. / pha:/ (फा)
57. / phi / (फि)
58. / pho / (फो)
59. / na / (न)
60. / na: / (ना)
61. / ne / (ने)
62. / ni: / (नी)
63. / ma / (म)
64. / ma: / (मा)
65. / me / (मे)
66. / mo. / (मो)
67. / mi / (मि)
68. / mi: / (मी)
69. / ya: / (या)
70. / ra: / (रा)
71. / la / (ल)
72. / la: / (ला)
73. / va / (व)
74. / va: / (वा)
75. / sa: / (सा)

A4.2 List of sentences

01 yadl: aisa: ha1
यदी ऐसा हैं

02 to mai bhi: pu:rṇ
तो मैं भी पूर्ण

03 amaratva nahi: ca:hta: hu:n
अमरत्व नहीं चाहता a

04 mai yah ca:hta: hu:n
मैं यह चाहता a

05 ki yagn karte samay
कि यज्ञ करते समय

06 agni mujhe . ghoḍon se
अग्नि मुझे घोडों से

07 juta: ·hua: rath prada:n karen
जुता हुआ रथ प्रदान करें

08 yahi: var mai a:pse
यही वर मैं आपसे

09 ca:hta: hu:n
चाहता a

10 is praka:r mai a:msik
इस प्रकार मैं आंशिक

11 amaratva ki: ka:mna: karta: hu:n
अमरत्व की कामना करता हूँ

12 ra:vaṇ bhi: saparivа:r
रावण भी सपरिवार

13 narmada: nadi: ke:
नर्मदा नदी के

14 kina: re pahu: nca:
किनारे पहूँचा

15 ra:vaṇ sada: apne sa:th
रावण अपने साथ

16 ek sone ka: sivaling rakha:
एक सोने का शिवलिन्ग रखा

17 karta: tha:
करता था

18 narmada: nadi:
नर्मदा =|a

19 pu:rvi: disa: se
पूर्वी दिशा से

20 paśc im ki: or bahti: hai
पश्चिम की ओर बहती है

21 aur paśc:.im sa:gar men
और पश्चिम सागर में

22 gir ja:ti: hai
गिर जाती है

23 aca:nak yah nadi:
अचानक यह नदी

24 paśc im se pu:rab ki: or
पश्चिम से पूरब की ओर

25 bahne lagi:
बहने लगी

26 ra:vaṇ ne unhen
रावण ने उन्हें

27 is ba:t ka:
इस बात का

28 pata: laga:ne ka: a:des diya:
पता लगाने का आदेश दिया

29 usne ek din'
उसने एक दिन

30 donon ko bula:kar
दोनों को बुलाकर

31 khu:b ḍa:ṇṭa:
खूब डाँटा

32 pita: ji: ko
पिताजी को

33 santust karne kelie
करने केलिए

34 mai ne tumhen ek upa:y
मैं ने तुम्हें एक उपाय

35 bata:ta: hu:n
बताता हूँ

36 bata:o krisn canara bola:
बताओ कृष्ण चन्द्र बोला

37 pratyek ha: r ki:
प्रत्येक हार की

38 viseṣata: bata: ne laga:
विशेषता बताने लगा

39 da:roga: ne
दारोगा ने

40 sacca:i: ka: pata: laga:ne
सच्चाई का पता लगाने

41 ke lie in donon ke
के लिए इन दोनों के

42 ma:ta: pita: ko bula: bheja:
माता पिता को बुला भेजा

43 ek choṭa: ba:dal
एक छोटा बादल

44 varṣa: karke
वर्षा करके

45 cupca:p lauṭ a:ya:
चुपचाप लौट आया

46 are beta:
अरे बेटा

47 tu:' ne yah kya: kiya:
तू ने यह क्या किया

48 yah ca:hta: tha: ki:
a की

49 praśansa: ka: pa:tr ko
प्रशंसा का पात्र को

50 uska: kisa:n pita:
उसका किसान पिता

51 bi: ma: r pada:
बीमार पडा

52 pha:yda: uṭha:kar ghar lauṭa:
फायदा उठाकर घर लौटा

53 aisi: . sthiti men
ऐसी स्थिति में

54 kya: karna: ca: hiye
क्या करना चाहिये

55 din bhar is prasn 'par
दिन भर इस प्रश्न पर

56 vica:r karti: rahi:
विचार करती रही

57 tulasi: ne apne so:ne ki:
तुलसी ने अपने सोने की

58 cu:ḍiyo se paisa: juṭa:ya:
चूडियो से पैसा जुटाया

59 krisṇ candr ne
कृष्ण चन्द्र ने

60 mu:nh se bol phu:ṭa:
मुंह से बोल फूटा

61 indr ne apa:r
इन्द्र ने अपार

62 para: kram. . pradarsit kiya:
पराक्रम प्रदर्शित किया

63 tumhare a:deśon ka:
तुम्हारे आदेशों का

64 pa:lan karenge
पालन करेंगे

65 koi: aisa: upa:y bata:o
कोई ऐसा बताओ

66 nahi:n juṭa: pa:ya: hu:n
नहीं जुटा पाया . हूँ

67 thoda: sa: pa:ni: barsa:ne
थोडा सा , पानी बरसाने

68 ki: śara:rat nahi:n karta:
नहीं करता

69 ba:dal ke dva:ra: pa:ni: pa:kar
बादल के द्वारा पानी पाकर

Appendix 5

VAXSTATION SYSTEM DETAILS

This appendix describes the hardware and software support in the VAXSTATION II/GPX system, on which the character spotting expert systems are implemented. The VAXSTATION system provides an environment for performing signal processing work. The VAXlab system is a combination of hardware and software components that creates the environment that the LabStar software requires. The VAXlab system can be used to control the real time hardware which consists of the A/D converter, the D/A converter and a real time clock. But the LabStar software actually provides a set of routines to perform real time 1/0 using the VAXlab hardware. The following two sections describe the VAXlab hardware and the LabStar software.

A5.1 VAXlab Hardware for 1/0 Support

The AAV11-D is a two-channel 250-kHz digital-to-analog (D/A) converter with direct memory access (DMA). ADV11-D is a 50-kHz analog-to-digital (A/D) converter with programmable gain and DMA. The KWV11-C clock module is used as a steady frequency source for the A/D and D/A devices. File I/O, a LabStar module device, moves data to a disk file using Queued Input Output (QIO). In QIO the user program queues buffers to the device for continuous processing of data. The device moves the data directly to disk using block I/O. As each file is read or written in blocks of 512 bytes each, the transfer is done very fast.

When the A/D and the D/A devices are set to do continuous Direct Memory Access (DMA), the DMA hardware runs continuously instead of stopping at the end of each buffer. The DMA can run at top speed without interruptions because it is confined to a 64K-byte block of memory that it wraps around. All the software has to do is to keep filling or emptying the buffers as fast as the DMA empties or fills them. We have used continuous DMA for the analog to digital conversion.

Al.2 LabStar Software for 1/0 Support

The LabStar Input Output (LIO) routines provide two types of interfaces: (a) synchronous read/write 1/0 and (b) asynchronous queued I/O. Synchronous 1/0 enables the user program to transfer a set of values to the device with one routine call. The routine call stops the program until the 1/0 completes. Asynchronous 1/0 enables the user program to queue several sets of values to be transferred. The program continues execution during 1/0 operations, enabling 1/0 operations to continue on one or more devices simultaneously. Asynchronous 1/0 has been used in the speech editor package.

Each asynchronous 1/0 device has a device queue and a user queue. The user program puts a buffer in the device queue to send it to the device. The device processes the buffer and puts the buffer in the user queue to return it to the program. LIO$ENQUEUE and LIO$DEQUEUE are the routines for accomplishing this. With devices set for asynchronous I/O, a program can set a device to forward completed buffers to another device. When the first device completes a buffer it immediately enqueues the buffer to the second device.

The LabStar Graphics Package (LGP) is a set of routines that can plot both real time data as well as data produced by calculations. These routines use the GKS software for plotting.

REFERENCES

1. Allen, J., "A Perspective on Man-Machine Communication by Speech," Proc. IEEE, vol.73, no.11, pp.1541-1550, November 1985.

2. Atal,B. and Rabiner, L., "A Pattern Recognition Approach to Voiced/Unvoiced/Silence Classification with Application to Speech Recognition," IEEE Trans. Acoust., Speech and Signal Processing, vol. 24, no. 3, pp.201-212, June 1976.

3. Bahl, L.R., Jelinek,F. and Mercer,R., "A Maximum Likelihood Approach to Continuous Speech Recognition," IEEE Trans. Pattern Anal. and Machine Intell., vol.PAMI-5, no.2, pp.179-190, March 1983.

4. Blumstein, S. and Stevens,K., "Property Detectors for Burst and Transition in Speech Perception," J. Acoust. Soc. of Amer., vol.61, no.1, pp.1301-1313, 1977.

5. Blumstein, S. and Stevens,K. , "Accoustic Invariance in Speech Production : Evidence from Measurements of Spectral Characteristics of Stop Consonants," J. of Acoust. Soc. of Amer., vol.66, no.4, pp.1007-1017, 1979.

6. Blumstein, S. and Stevens,K., "On Some Issues in the Pursuit of Acoustic Invariance in Speech : Reply to Lisker," J. of Acoust. Soc. Amer., vol.77, no.3, pp.1203-1205, 1985.

7. Bridle, J.S. ,Brown, M.D. and Chamberlain, R.M., "An Algorithm for Speech Recognition," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Paris, May 1982,

pp.899-902.

8. Brown, K. and Algazi, "Characterization of Spectral Transition with application to sub-word unit segmentation and Automatic Speech Recognition," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Glasgow, 1989, pp.104-107.

9. Bruce Lowerre and Raj Reddy, D., "The Harpy Speech Understanding System," in Lea, W.A., (Ed.), Trends in Speech Recognition, pp.340-361, Prentice Hall, Englewood Cliffs, 1980.

10. Bush, M.A. and Kopec,G.E., "Selecting Acoustic Features for Stop Consonant Identification," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Boston, 1983, pp.742-745.

11. Carbonnel, N., "An Expert System for Automatic Reading of French Spectrograms," Proc. IEEE Int. Conf. Acoust., Speech and Signal processing , San Diego, 1984.

12. Chandra Sekhar,C., Singh, S.K., Eswar, P. and Yegnanarayana, B., "Feature Spotting Approach for Speech Signal-to-Symbol Transformation for Continuous Speech in Indian Languages," SPEECH 88, Edinburgh, 1988.

13. Cole, R.A., "Feature Based Speaker Independent Recognition of Isolated English Letters," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing , Boston, 1983, pp.731-734.

14. Cole, R.A., "CMU-Phonetic Classification System," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Tokyo, 1986, pp.2255-2258.

15. Colla, A.M., "Automatic Diphone Bootstrapping for Speaker

Adaptive Continuous Speech Recognition," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, San Diego, 1984.

16. Chow Y.L., Dunham, M.O., Kimball, O.A., Krasner, M.A., Kubala, G.F., Makhoul, J., Price, P.J., Roucos, S. and Schwartz,R.M.,"BYBLOS: The BBN Continuous Speech Recognition System," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Dallas, 1987, pp.89-92.

17. Davis ,S.B. and Mermelstein, P., "Comparison of Parametric Representation for Monosyllabic Word Recognition in Continuously Spoken Sentences," IEEE Trans. Acoust., Speech and Signal Processing, vol.28, no.4, pp.357-366, August 1980.

18. Dautrich, B., Rabiner, L. and Martin, T., "On the Effect of Varying Filter Bank Parameters on Isolated Speech Recognition," IEEE Trans. Acoust., Speech and Signal Processing, vol.31, pp.793-807, 1983.

19. De Mori, R., Gubrynowicz,A. and Laface, P., "Inference of a Knowledge Source for the Recognition of Nasals in Continuous Speech," IEEE Trans. Acoust., Speech and Signal Processing, vol.27, no.5, pp.538-549, August 1979.

20. De Mori, R., Giordana,A., Laface, P., and Saitta, L., "Parallel Algorithms for Syllable Recognition in Continuous Speech," IEEE Trans. Pattern Anal. Machine Intell., vol.PAMI-7, no.1, pp.56-68, January 1985.

21. De Mori, R. and Giordana,A., "Phonetic Feature Hypothesization in Continuous Speech," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Boston, 1983, pp.316-

319.

22. De Mori, R., Giordana,A., Laface, P. and Saitta, L., "An Expert System for Speech Decoding," Proc. AAAI Conference, Pittsburgh, pp.107-110, 1982.

23. De Mori, R.(Ed.), New Svstems and Architectures for Speech Recoqnition and Svnthesis , NATO ASI Series, New York, 1985.

24. Diane Kewley Port, "Time Varying Features as Correlates of Place of Articulation in Stop Consonants," J. Acous. Soc. Amer., vol.73, no.1, pp.322-335, 1983.

25. Duncan, G., Hema A. Murthy and Yegnanarayana, B., "A Non-Parametric Method' of Formant Estimation using Group Delay Spectra," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Glasgow, April 1989, pp.572-575.

26. Dutoit, D., "Evaluation of Speaker-Independent Isolated Word Recognition over telephone network," Proc. European Conference on Speech Tech., Edinburgh, 1987, pp. 241-244.

27. Dutta Majumdar, "On Automatic Plosive Identification using Fuzziness in Property Sets," IEEE Trans. Systems, Man and Cybernetics, vol.SMC-8, no.4, pp.302-308, 1979.

28. Eswar, P., Chandra Sekhar, C., Gupta, S.K., Yeganarayana, B. and Nagamma Reddy, K., "An Acoustic-Phonetic Expert for Analysis and Processing of Continuous Speech in Hindi," Proc. European Conference on Speech Technology, vol.1, pp.369-372, Edinburgh, 1987.

29. Erman, L.D., Lesser, V.R. and Raj Reddy, D., "The HEARSAY-I1 Speech Understanding System," Computing Surveys, vol.12, no.2, pp.213-253, 1980.

30. Fissore, L., Gioclia, E., Laface, P., Mica, G., Pieraccini,

L. and Rullent,C., "Experimental Results on Large Vocabulary Continuous Speech Recognition and Understanding," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, New York, 1988, pp. 414-417.

31. Fuzimura, K. "Analysis of Nasal Consonants," J. Acoust. Soc. Amer., vol.34, no.12, pp.1865-1875, 1962.

32. Fuzisaki, H. and Kunisaki, O., "Analysis, Recognition and Perception of Voiceless Fricative Consonants in Japanese," IEEE Trans. Acoust., Speech and Signal Processing, vol.26, no.1, pp.21-37, February 1978.

33. Garry Goodman and Raj Reddy, D., "Alternative Control Structures for Speech Understanding System," in Lea, W.A., Trends in Speech Recoanition, pp.234-246, Prentice Hall, Englewood Cliffs, New York 1980.

34. Gersho, A. and Cuperman, V., "Vector Quantization : A Pattern Matching Technique for Speech Coding," IEEE Communications Magazine, December 1983.

35. Ghitza, O., "Auditory Neural Feedback as a Basis for Speech Processing," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, New York, 1988, pp.91-94.

36. Gong, V.F. and Haton, J.P., "A specialist society for Continuous Speech Understanding," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, New York, 1988, pp.627-630.

37. Gray, A.H. and Markel, J.D., "Distance Measures for Speech Processing," IEEE Trans. Acoust., Speech and Signal Processing, vol.24, no.5, pp.380-391, October 1976.

38. Halle and Stevens., "Acoustic Properties of Stop

Consonants," J. Acoust. Soc. Amer., vol.29, no.1, pp.107-116, 1981.

39. Hanson, B. and Wakita, H, "On the use Bandpass Liftering in Speech Recognition," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Tokyo, 1986, pp.768-771.

40. Hashi Wakita and Shozo Makino, "Recent Work in Speech Recognition in Japan," in Lea, W.A., (Ed.), Trends in Speech Recognition, pp.483-497, Prentice Hall, Englewood Cliffs, 1980.

41. Hatazaki, K. and Komoroi. Y, "Phoneme Segmentation using Spectrogram reading Knowledge," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing , Glasgow, 1989, pp.393-396.

42. Haton, J.P., "Knowledge Based Approach in Acoustic-phonetic Decoding of Speech," in Niemann, H., Lang, M. and Segrer, G. (Eds.), Recent Advances in Speech Understandins and Dialoq Systems, NATO-AS1 Series, vol.46, pp.51-69, 1988.

43. Huang, C., "Neural Net Approach to Speech Recognition," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, New York, 1988, pp.99-103.

44. Hunt, M.J. and Lefebvre, C., "Speaker Dependent and Independent Speech Recognition Experiments with an Auditory Model ," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, New York, 1988 , pp.91-94.

45. Jared J. Wolff and W.A. Woods, "The HWIM Speech Understanding System," in Lea, W.A., (Ed.), Trends in Speech Recognition, pp.316-338, Prentice Hall, Englewood Cliffs, 1980.

46. Jeffrey Barmett, Martin I Benskin, Richard Gilman and Iris Karmey, "The SDC Speech Understanding System," in Lea, W.A., (Ed.), Trends in Speech Recognition, pp.272-290, Prentice Hall, Englewood Cliffs, 1980.

47. Jelinek, F., "Continuous Speech Recognition by Statistical Methods," Proc. IEEE, vol.64, no.4, pp.532-556, April 1976.

48. Jelinek, F., "Experiment with the Tangora 20,000 Word Speech Recognizer," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Dallas, 1987, pp.701-704.

49. Johannsen, J., "A Speech Spectrogram Expert," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Boston, 1983, pp.746-749.

50. Kavelar, R., "A DTW Integrated Circuit for a One Thousand Word Recognition System," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, San Diego, 1984, pp.700-703.

51. Kawabata, T. and Shikano, K., "Island Driven Continuous Speech Recognisers using Phoneme based HMM Word Spotting," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Glasgow, 1989, pp.461-464.

52. Kita, K., Kawahati, T. and Hatiajawa, T., "HMM Continuous Speech Recognition using Stochastic Models," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Albuquerque, 1990, pp.581-584.

53. Klatt, D.H., "Overview of ARPA Speech Understanding Project," in Lea, W.A., (Ed.), Trends in Speech Recognition, Prentice Hall, Englewood Cliffs, pp.249-270, 1980.

54. Klatt, D.H., "Review of ARPA Speech Understanding Systems ," J. Acoust. Soc. Amer., vol.62, no.5, pp.1345-1366, 1978.

55. Kobayashi, Y. and Niimi, Y., "An Overview of Linguistic Decoding for a Voice-Input Question/Answering System," JIETE Spl. Issue on Speech Processing, vol.34, no.1, pp.96-101, Jan-Feb 1988.

56. Koinori, Y., Hatazaki, K., Tanaka, T. and Kawabata, T., "Combining Phoneme Identification Neural Networks into an Expert System using Spectrogram Reading Knowledge," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing , Albuquerque, 1990, pp.505-508.

57. Kostic .A , A Short Outline of Hindi Phonetics, Indian Statistical Institute, Calcutta, 1977.

58. Kowalik, K., Knowledqe Based Problem Solving, Elsevier Science, 1986.

59. Ladefoged, B., A Course in Phonetics, Hartcourt, Brace, Jovanovich, New York, 1975.

60. Lea, W.A. , "Speech Recognition : Past, Present and Future" in Lea, W.A.,(Ed.), Trends in Speech Recoqnition, Prentice Hall, Englewood Cliffs, pp.39-90, 1980.

61. Lea, W.A., (Ed.), Trends in Speech Recoanition, Prentice Hall, Englewood Cliffs, 1980.

62. Lee, K.F., "Large Vocabulary Speaker Independent Continuous Speech Recognition -- The SPHINX sustem ," CMU Report, CMU\CS 88-148. April 1988.

63. Leung, H.C., "Some Phonetic Recognition Experiments using Artificial Nets," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing , New York, 1988 , pp.422-426.

64. Levinson, S.E., Rabiner, L.R. and Sondhi, M.M., "An Introduction to the Applications of the Theory of

Probabilistic Functions on a Markov Process to Speech Recognition System," Bell Sys. Tech. Jou., vol.62, pp.119-137, 1983.

65. Levinson, S., "Structural Methods in Automatic Speech Recognition," Proc. IEEE, vol.73, no.11, pp.1625-1650, November 1985.

66. Lisker, L., "The pursuit of Invariance in Speech Signals," J. Acoust. Soc. Amer., vol.77, no.3, pp.1195-1202, 1985.

67. Loeb, E.P. and Lyon, R.F., "Experiments in Isolated Digit Recognition with a Cochlear Model," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing , Dallas, 1987, pp.1131-1134.

68. Madhu Murthy, K.V. and Yegnanarayana, B., "Effectiveness of Representation of Signals Throgh Group Delay Functions," Signal Proceesing, vol.17, pp.141-150, 1989.

69. Mangione, P.A., "SSI's Phonetic Engine," Speech Technology, vol.3, no.2 pp.84-86, March/April 1986.

70. Markel, J.D., and Gray, A.H., Linear Prediction of Speech Springer-Verlag, 1976.

71. Mariani, J. "Recent Advances in Speech Processing," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Glasgow, 1989, pp.429-440.

72. Martelli, T., "REMORA : A software Architecture for the Collaboration of Different Knowledge Sources in Phonetic Decoding of Continuous Speech," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Dallas, 1987, pp.387-390.

73. Mermelstein, P., "Automatic Segmentation of Speech into

Syllabic Units," J. Acous. Soc. of Am., vol.58, pp.880-883, January 1975.

74. Mijoguchi, R.and Sujino, K., "A Continuous Speech Recognition System based on Knowledge Engineering Techniques," Proc. IEEE Int. Conf. Acoust.,Speech and Signal Processing, Tokyo, 1986, pp.1221-1224.

75. Montacie, C., Choukri, K. and Chollet, G., "Speech Recognition using Temporal Decomposition and Multilayer feed forward Automata," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing', Glasgow, 1989, pp.409-412.

76. Nakagawa, S. and Nakanishi, H., " Speaker Independent English Consonant and Japanese Word Recognition by a Stochastic Dynamic Time Warping Method," JIETE Spl. issue on Speech Processing, vol.34, no.1, pp.87-95, Jan-Feb 1988.

77. Ohala, M., Aspects of Hindi Phonoloqy, Motilal Banarsidass, New Delhi, 1983.

78. Rabiner, L.R. and Schaffer, R.W., Diaital Processins of Speech Siqnals, Englewood Cliffs, Prentice-Hall, New York, 1980

79. Rabiner, L.R., Bergh, A. and Wilpon, J., "An Embedded Word Training Procedure to Connected Word Recognition," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Paris, 1982, pp.1621-1624,.

80. Rabiner, L.R. and Juang, B.H., "Introduction to Hidden Markov Model," IEEE Acoust., Speech and Signal Processing Magazine, vol.3, no.1, pp.4-16, January 1986.

81. Rabiner, L.R. and Levinson, S.C., Rosemberg, A.E. and Wilpon, J.C., "Speaker Independent Recognition of Isolated

Words using Clustering Techniques," IEEE Trans. Acoust., Speech and Signal Processing, vol.27, August 1979.

82. Reddy, D.R., "Computer Recognition of Connected Speech," J. Acoust. Soc. Amer., vol.42, no.5, pp.329-432, 1967.

83. Regel, P., "Module for Acoustic-Phonetic Transcription of Fluently Spoken German Speech," IEEE Trans. Acoust., Speech and Signal Processing, vol.30, no.3, pp.440-450, June 1982.

84. Renals, S. and Rohwer, R., "Learning Phoneme Recognition using Neural Networks," Proc. IEEE Int. Conf. Speech and Signal Processing , Glasgow, 1989, pp.113-116.

85. Rosenberg, A.E., Rabiner, L.R., Levinson, S.E. and Wilpon, J., "A Preliminary Study on the use of Demi-syllables in Automatic Speech Recognition," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Florida, 1981, pp.967-970.

86. Roucus, S.O., "A Stochastic Segment Model for Phoneme Based Continuous Speech Recognition," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Boston, 1987, pp.73-76.

87. Sakoe, H., "Two Level DP Matching - A Dynamic Programming Based Pattern Matching Algorithm for Connected Word Recognition," IEEE Trans. Acoust., Speech and Signal Processing, vol.27, no.6, pp.585-595, December 1979.

88. Sawai, H. and Waibel, A., "Spotting Japanese CV Syllables and Phonemes using Time Delay Neural Networks," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Glasgow, 1989, pp.25-28.

89. Scagliola, C., "The use of Diphones as Basic Unit in Speech

Recognition," 4 th FASE Symposium on Acoustics and Speech, Venice, April 1981.

90. Scagliola, C., "Continuous Speech Recognition by Diphone Spotting : A Preliminary Implementation," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Boston, 1982, pp.2008-2011.

91. Schroeder, M.R., "Predictive Coding of Speech - Historical Review and Directions for Future Research," Proc. IEEE Int. Conf. Acous., Speech and Signal Processing, Tokyo, 1986, pp.3157-3164.

92. Shikano, K., Lee, K.F. and Raj Reddy, "Speaker Adaptation through Vector Quantization," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Tokyo, 1986, pp.2643-2646.

93. Shoup, J.E., "Phonological Aspects of Speech Recognition," in Lea, W.A. (Ed.), Trends in Speech Recoanition, pp.125-150, 1980.

94. Sugumara, N., "Continuous Speech Recognition using Large Vocabulary Word Spotting and CV Spotting," Proc. IEEE Int. Conf. Acoust.,Speech and Signal Processing, Albuquerque, 1990, pp.121-124.

95. Watanabe, T., "Segmentation free Syllable Recognition in Continuously Spoken Japanese," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Boston, 1983, pp.320-323.

96. Watanabe, Y., "Syllable Recognition for Continuous Japanese Speech Recognition," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Tokyo, 1986, pp.2295-2298.

97. Weinstein ,C., McCandless, S., Mondshein, L. and Zue, V.W.,

"A System for Acoustic Phonetic Analysis of Continuous Speech," IEEE Trans. Acoust., Speech and Signal Processing, vol.23, no.1, pp.54-67, February 1975.

98. Wilpon, J.G. and Jeung, B.H., "An Investigation on the use of Acoustic Subword Units for Automatic Speech Recognition," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Dallas, 1987, pp.1310-1313.

99. Woods, W. A. , "Motivation and Overview of SPEECHLIS: An Experimental Prototype for Speech Underatanding Research," IEEE Trans. Acoust., Speech and Signal Processing, vol.23, pp.2-10, February 1975.

100. Yanagida, M., Yamashita, Y. and Kakusho, O., "Detection and Identification of Plosive Sounds in Japanese Words," JIETE Spl. issue on Speech Processing, vol.34, no.1, pp.82-87, Jan-Feb 1988.

101. Yegnanarayana,B., Saikia, D.K. and Krishnan, T.R., "Significance of Group Delay Functions in signal Reconstruction from Spectral Magnitude or Phase," IEEE Trans. Acoust., Speech and Signal Processing, vol.32, no.3, pp.610-623, June 1984.

102. Yegnanarayana, B., "Voice Input/Output Systems for Indian Languages ," Proc. Annual Convention of Computer Society of India, pp.3-23, January 1988.

103. Yegnanarayana, B., Chandra Sekhar, C., Ramana Rao, G.V., Eswar, P. and Prakash, M., "A Continuous Speech Recognition System for Indian Languages," Proc. of Regional Workshop on Computer Processing of Asian Languages, Bangkok, 1989, pp.347-356.

104. Young, S.R., "Speech Recognition in VODIS II ," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, New York, 1988, pp.441-444.

105. Zadeh, L., Fuzzy Systems and Their Applications to Cognitive Processes, Academic Press, 1985.

106. Zue, V.W., "Use of Speech Knowledge in Automatic Speech Recognition," Proc. IEEE, vol.73, no.11., pp.1602-1615, November 1985.

107. Zue, V.W., Glass, J., Phillips, M. and Seneff, S., "Acoustic Segmentation and Phonetic Classification in the SUMMIT System," Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, Glasgow, 1989, pp.389-392.

LIST OF FIGURES

Fig.6.10. Illustration of signal-to-symbol transformation for utterance 1

Fig.6.11. Illustration of signal-to-symbol transformation for utterance 2

LIST OF TABLES

LIST OF PUBLICATIONS

1. Eswar, F., Gupta, S.K., Chandra Sekhar, C. and Yegnanarayana, B., "An Acoustic Phonetic Expert for Analysis and Processing of Continuous Speech in Hindi," Proc. European Conference on Speech Technology, vol.1, pp. 369-372, Edinburgh, 1987.
2. Eswar, P., Chandra Sekhar, C. and Yegnanarayana, B., "Spotting Orthographic Characters from Continuous Speech in Indian Languages using Expert Systems Approach," accepted (to be revised and submitted) for publication in Speech Communication Journal.
3. Eswar, P., Chandra Sekhar, C. and Yegnanarayana, B., "Use of Fuzzy Mathematical Concepts in Character Spotting for Automatic Recognition of Continuous Speech in Hindi," accepted (to be revised and submitted) for publication in the Journal of Fuzzy Sets and Systems.
4. Yegnanarayana, B., Sekhar, C.C., Rao, G.V.R., Eswar, P. and Prakash, M., "A Continuous Speech Recognition System for Indian Languages," Proc. of Regional Workshop on Computer Processing of Asian Languages, pp.347-356, September 1989.