Proceedings of the 23rd Annual Convention of the Computer Society of India, January 6-9, 1988, Madras

Voice Input/Output Systems for Indian Languages

B. Yegnanarayana
Dept. of Computer Science and Engineering
Indian Institute of Technology
Madras *600* 036

Abstract

In this paper an overview of problems and prospects of voice input/output to a computer are discussed. Current attempts to provide speech input/output facilities to a computer are described. The scope of speech recognition problem is defined. Issues involved in the design of text-to-speech and speech-to-text systems are discussed. Since any sophisticated voice input/output system uses several language speclfic features, a case is made for a different approach for designing these systems for Indian languages.

1. INTRODUCTION TO MAN-MACHINE COMMUNICATION BY VOICE

Our thirst for more and more computing power, and for better 'natural' interaction with computers is ever increasing. While the developments in computer technology appear to be not so much need based, it is very easy for one to see the need for these developments in the field of Artificial Intelligence, especially, when we want to communicate with a machine in the most natural way, as we do with other human beings. The objective of this paper is to discuss one aspect of this natural communication, namely, man-machine communication by voice [1].

While the applications of such a natural interface appear to be vague at present, the one single application which justifies our effort is dictation. In particular, if we are able to produce a reasonable draft of text of a spoken input, it would completely transform the existing communication and office transactions. Likewise, the conversion of text into a spoken message will revolutionize the concept of education and teaching, especially to less privileged lot in a multilingual society like ours in India. In this paper we will discuss issues involved in the design of a voice input and output system for a computer for Indian languages.

A voice input/output system consists of a speech recognition part and a speech synthesis part. The primary objective of the voice input system is to produce a speech-to-text transcription in a given language. Human beings seem to do this task effortlessly, whereas if a machine were to do the same task, it seems necessary to equip it with the knowledge of phonetics and linguistics. For simple isolated word recognition systems, it may be adequate to use a pattern matching approach. But such systems are speaker-dependent and have very limited application. For a system to be acceptable for common use, it should be speaker-independent, task-independent and even vocabulary-independent.

Speech synthesis systems could likewise be considered at several levels of complexity. Ideally, one would like to have a conversion from printed text to a natural spoken form of speech output. Here again a person familiar with the language of the text has no difficulty in speaking it out. But for a machine to do the same task, the text has to be converted to a phonemic string using letter-to-sound rules, and then the phonemic string is converted to speech using additional prosodic information in the synthesis. Voice output systems could be classified into two broad categories depending on the nature of the synthesis process. Speech coding systems use a fixed set of parameters to produce a relatively intelligible and acceptable speech at the cost of flexibility in terms of the range of utterances that could be produced. In contrast, synthetic speech produced by rule provides less intelligible and less natural sounding speech, but these systems have the capability of automatically converting unrestricted text in ASCII format into speech. Over the last few years, significant improvements in the design of text-to-speech systems have begun to eliminate the advantage of simple coded-speech voice response systems over text-to-speech systems.

The speech signal carries, besides the Hnguistic message, information about the speaker's identity, his language, his physical and emotional state, and his geographical and societal background. Speech is perceived and understood by humans due to certain auditory hints sparked by the input signal. It is our linguistic comprehension which plays an important role in enabling us to create a message from the auditory hints. Therefore the structure of the language and the semantic context of the verbal communication play an important role, besides the acoustic signal. The discreteness of words of a spoken language, as we perceive it, is absent in the speech signal. It is obscured by the levels of encoding between the words and the spoken utterance, but is reconstructed by the listener through his Hnguistic competence. An automatic speech recognition system faced with the same input signal must be capable of doing a similar reconstruction. It Is also interesting to note that the linguistic message Is perceived by a human irrespective of the speaker and the rate at which he speaks.

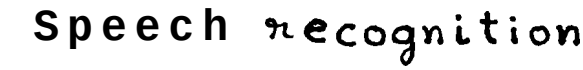
The basic problem in speech recognition, therefore, consists of (1) determining the

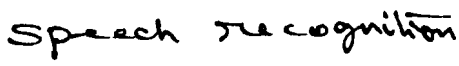
auditory hints from the speech signal, and (2) using the linguistic constraints to understand the message from the auditory hints.

Viewed in this fashion, it is not clear whether the auditory hints correspond to the basic linguistic units, called phonemes. Even if they do, it is not clear whether all the phonemes are needed to determine the message. It is this kind of information that is ultimately needed for an artificial recognition system to work for an unrestricted task for any speaker.

Methods for speech recognition have been developed on the assumption that a speech utterance consists of distinct phonemes. The nature of the phonemes has been studied in order to recognize them in an unknown utterance. However, the extension of this procedure to the recognition of continuous speech leads to serious difficulties. Basically the difficulties lie in the nonuniqueness of such phonemes, and also in their distortions in continuous speech.

Continuous speech recognition is considered difficult because the word boundaries are difficult to determine. But the fact is that the word boundaries do not exist at all in many cases. In other words, even if one were tn find a method of placing a boundary between two words in continuous speech, the techniques of isolated word recognition are not always applicable. The distinction between isolated word and continuous speech recognition can be appreciated from the analogy of hand written text. A text can be written either with letters in isolation as in example (a) or with letters in a continuous manner as in example (b) below:

Example: a) 

b) 

Although the words and the message are clear in both the cases, the individual characteristics of each letter are grossly distorted in (b), even if one were to place a distinct boundary between letters. Moreover the variability in continuous writing is so wide that many times it may be impossible to determine a boundary between letters.

The above example clearly illustrates the need for different approaches for different types of speech recognition systems. Each type of system has its' own intrinsic difficulties. Also, clues to speech recognition are sometimes at a suprasegmental level. The sequence of occurrences of these features must be related in order to interpret the message information in the utterance. Redundancies in a language help overcome the ambiguities present in the feature sequence.

The features mentioned above may not have any meaning for an isolated word recognition

system. Each word must be considered as a pattern class. The relevant features, that are invariant to several distortions, speaking rates, speaker variations, etc., should be identified and extracted. The difficulty in this type of systems is lack of adequate redundancy to correct for ambiguous utterances.

In an effort to unify the approaches to speech recognition for different tasks, the acoustic signal is analysed on a short time basis, as for parametric representation. The parametric representation is then used suitably for isolated or continuous speech recognition. Obviously such a unified approach cannot provide satisfactory results.

The current status of speech recognition is discussed briefly in the next section. The section also discusses the processes involved in speech recognition and synthesis systems. Some specific proposals for speech recognition and synthesis for Indian languages are made in Section III. Preliminary results of our studies on speech recognition of isolated utterances of Hindi alphabet and signal-to-symbol transformation based on character spotting approach for continuous speech in Hindi are discussed in Section IV.

2. CURRENT STATUS OF SPEECH RECOGNITION

2.1 Speech Recognition and Speech Understanding

Literally speech recognition means identifying the speech sounds in the input acoustic waveform. But in most cases of practical importance, the objective is to determine the intended meaning from the input speech. In other words, speech recognition by machine consists of transforming the continuous speech signal into a representation which the machine can interpret (using built-in rules) as one of the allowable messages for which some predetermined response has already been built into the system. Such a machine is called speech understanding system. A large effort had gone into building such systems during the period 1971-1976 in USA with the support of Advanced Research Project Agency (ARPA) [21. Simultaneously several attempts have been made towards building speech recognition systems for isolated words and for string of words.

2.2 Spectrum of Possible Recognition Systems

A spectrum of speech recognizers is possible to cover a wide variety of practical applications. These recognizers include:

(1) Isolated word recognizers, which independently handle words that are preceded and followed by pauses.

(2) Recognizers of sequences of isolated words which use sequence constraints ("syntax") to limit the alternative words at each stage in the sequence.

(3) Word spotting systems, which detect occurrences of key information carrying words in the context of free-flowing continuous speech.

(4) Digit string rscognizers, which handle uninterrupted sequences of spoken digits.

(5) Word sequence recognizers, which identify uninterrupted (but strictly formatted) sequence of words.

(6) Restricted speech understanding systems, which handle total sentence relevant to a specific task.

(7) Task-independent continuous speech recognizers, which identify wording of sentences without restriction to a specific task.

For each speech input facility, one must ask which of these recognition capabilities is needed. It is then appropriate to ask what has been learned from previous attempts to develop and use each such type of recognition system.

2.3 Processes Involved in Speech Recognition

A native speaker uses, subconsciously, his knowledge of the language, the environment, and the context in understanding a sentence. These are called sources of knowledge which are needed to limit the alternative words that have to be selected at each point in an input utterance. The different processes which reflect the use of knowledge sources are:

(1) Extracting important acoustic parameters ("acoustic analysis").

(2) Identifying vowels and consonants in the speech ("phonetic analysis").

(3) Matching sequence of speech sounds to expected pronunciations of words ("word matching").

(4) Using stress, intonation, and the timing of speech to identify aspects of the structure of the sentence ("prosodic analysis").

(5) Verifying the grammaticality of hypothesized word sequences and predicting the possible identities of unidentified words by contextual constraints ("syntactic analysis").

(6) Testing the meaningfulness of apparent word sequences and hypothesizing other meaningful and semantically related words that might extend partial interpretations of the sentence ("semantic analysis").

(7) Determining the plausibility of hypothesized word sequences, based on the discourse context and the task being performed ("pragmatic analysis").

2.4 Gaps in Current Technology

After the ARPA speech understanding effort, and the developments in the realization of other speech recognition systems, it has'been found that the following are among the top-priority aspects of recognition that need attention (Listed in descending order of priority):

(1) Acoustic-phonetic analysis

(2) Prosodic cues to linguistic structures

(3) Using linguistics to constrain ambiguities

(4) Phonological rules

(5) Fast or near real-time processing

(6) Performance evaluation

It is interesting to note that significant progress has been made so far in higher level processing such, as representation and utilization of knowledge sources, whereas very little progress can be claimed in the front end acoustic-phonetic analysis. Obviously, no amount of sophistication at later stages can compensate for the loss of vital information at the front end analysis stage.

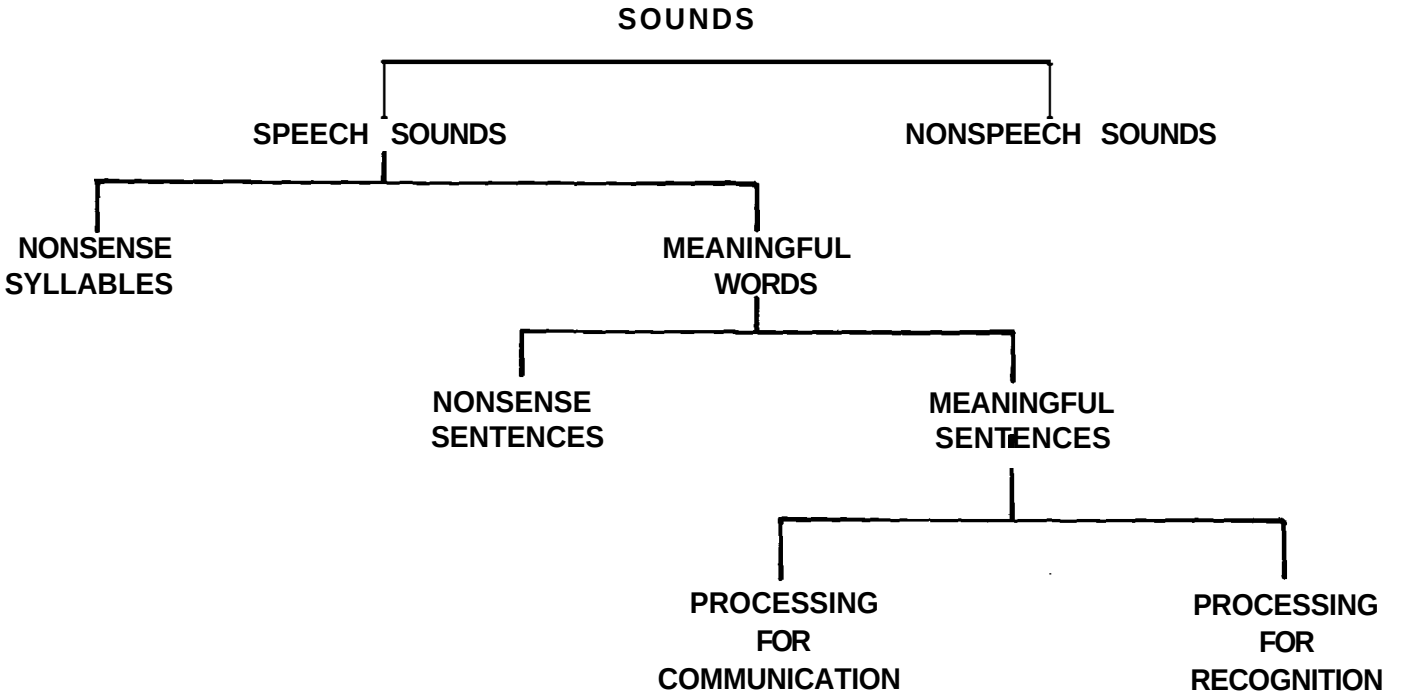


Fig- 1. CATEGORIES OF SOUNDS

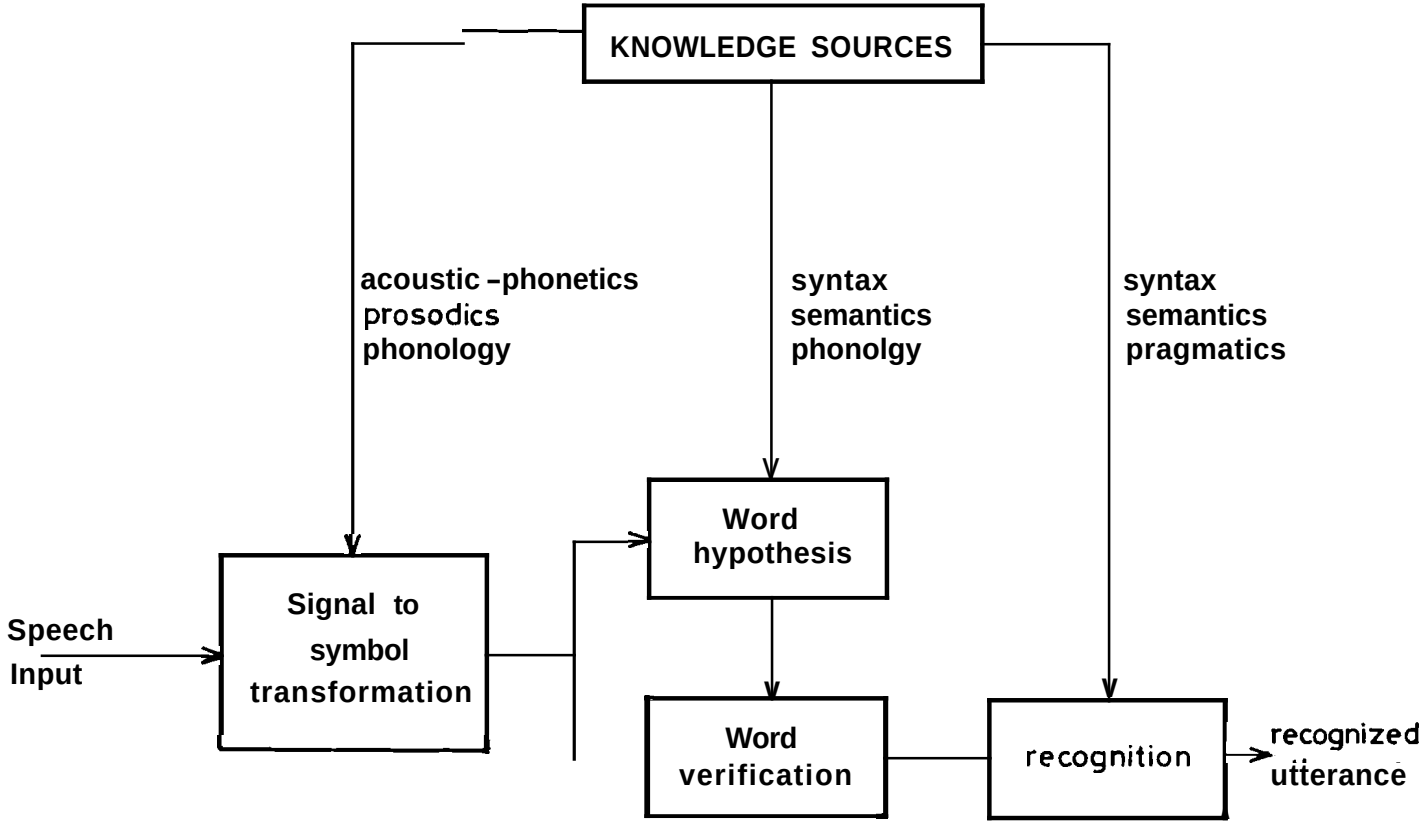


Fig. 2. BLOCK DIAGRAM INDICATING SEQUENCE OF STEPS FOR A SPEECH RECOGNITION SYSTEM.

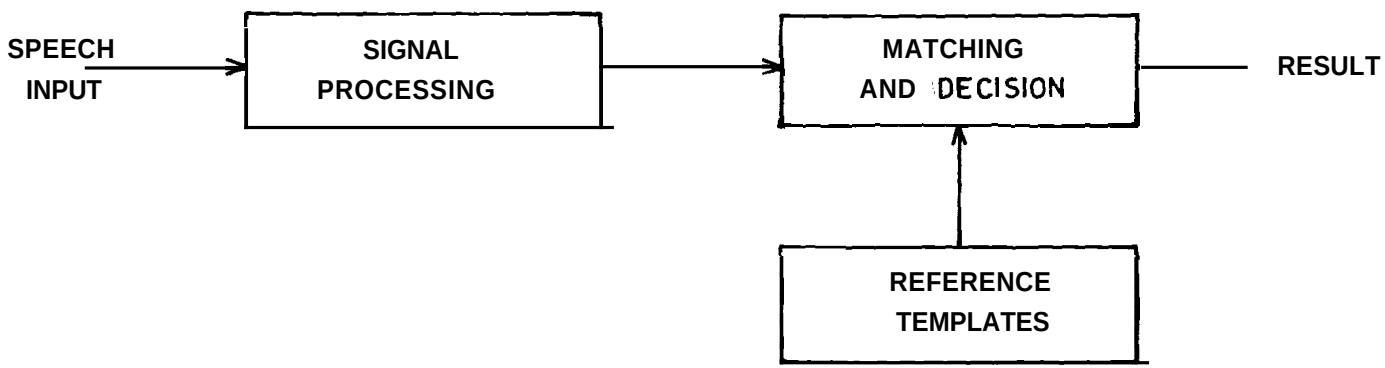


Fig-3. BASIC PROCESSES IN AN ISOLATED WORD RECOGNITION SYSTEM

2.5 Scope of Speech Recognition

The scope of the speech recognition problem is illustrated in Fig.1 where the different categories of input acoustic signals are given. Obviously, we are not interested in recognizing all types of speech-like sounds. In most practical situations, we are interested in the recognition of elther isolated words or continuous speech corresponding to meaningful sentences. Since only meaningful sentences are involved, use can be made of the different knowledge sources like syntax, semantics, etc., to distinguish the input speech from other sounds. Knowledge sources can also be effectively used to deal with ambiguities in the acoustic waveform.

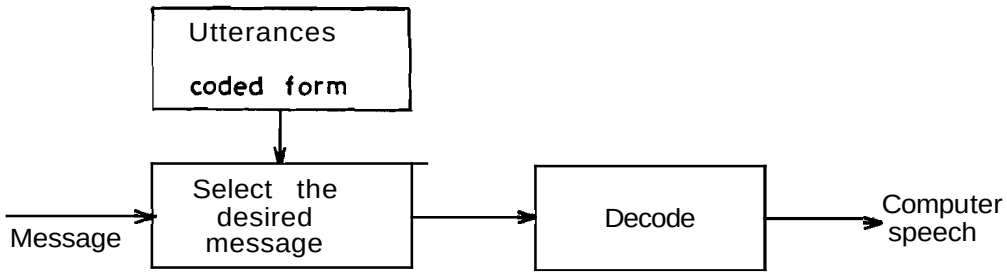
2.6 Levels of Speech Recognition Systems

The basic steps involved in continuous speech recognition are shown in Fig.2. The speech signal is transformed into some predefined symbols, such as phones, phonemes, syllables or even words. From the sequence of symbols and use of syntax, semantics and phonology, certain possible words are hypothesised. The hypothesised words are verified by matching with the symbol sequence to determine the exact word. Any ambiguity in the sequence of words is resolved at the final recognition stage by using the knowledge of syntax and pragmatics. The 'ambiguity at various stages arises because of the uncertainity at the signal level regarding the features of symbols and segmentation boundaries. These uncertainties arise due to human nature in producing variations in speech and also the variation caused by environmental factors such as background noise.

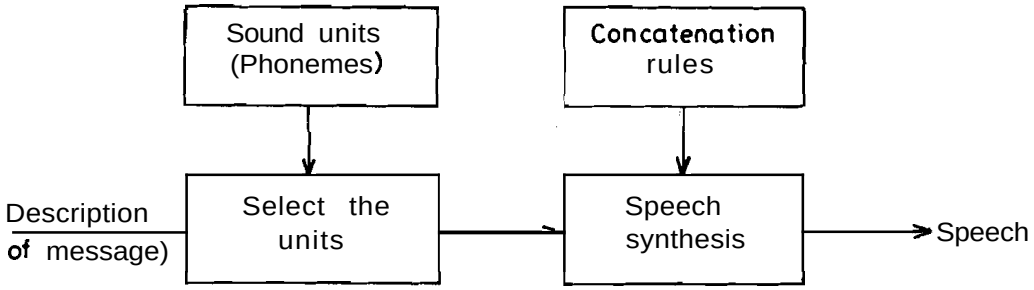
The basic steps involved in the recognition of isolated words or a sequence of isolated words are shown in Fig.3. Here the input speech is segmented into separate words and then the recognition of each word takes place by signal processing and matching. In signal processing, the input speech is compressed to represent the significant speech information by a few parameters or features. The presence or absence of a particular word in an unknown utterance is determined by matching the extracted features with stored templates in a predetermined fashion.

2.7 Speech Synthesis Systems

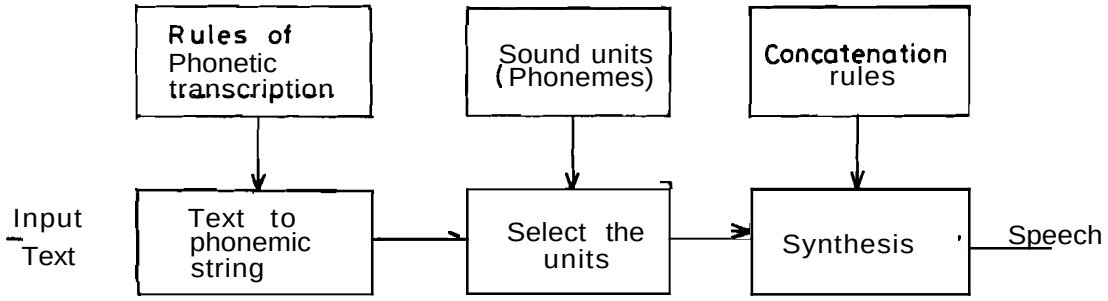
Compared to voice input systems, voice output systems appear to be relatively feasible with the current technology [3],[4]. However, producing a natural sounding synthetic speech



(a) Voice response system (Speech units are utterances of phrases or sentences. The synthetic speech is a concatenation of stored coded waveform)



(b) Speech synthesis from phonemic string (The message is given as a string of symbols from a preselected set of phonetic units like phonemes)



(c) Unrestricted text-to-speech synthesis (The input text is first converted to phonetic symbols and then the sound units corresponding to these symbols are obtained from the system)

Fig. 4. CATEGORIES OF SPEECH SYNTHESIS SYSTEMS

from an unrestricted text is still a far-off dream. The text might be entered from a keyboard, or optical character recognizer, or obtained from a stored data base. Depending on the nature of the input, synthesizers of different complexities can be built as shown in Fig.4. Use of large speech units, such as phrases and sentences, can give high quality output speech, but requires much memory. Voice response systems (Fig.4a) handle input text of limited vocabulary and syntax, while text-to-speech systems (Fig.4c) accept an unrestricted input text. Text-to-speech systems construct speech from text using small stored speech units and extensive linguistic processing, while voice response systems reproduce speech directly from previously coded speech, using signal processlng techniques. Issues in text-to-speech conversion are the choice of basic unit for synthesizing unrestricted text, the rules for deriving these units from input text and the rules for concatenating these units to produce natural sounding synthetic speech.

3. VOICE INPUT/OUTPUT SYSTEMS FOR INDIAN LANGUAGES - A PROPOSAL

From the above discusslon, it is obvious that any voice Input/Output system should use the phonetic and linguistic features of a given language. In this respect a speech system developed for one language may not easily be adaptable for another language. Moreover, there may be some specific features of a given language which may aid in developing certain blocks of a voice input/output system. It is quite possible that the parallel between written form and pronunciation as in Indian languages, may significantly influence the design of a speech recognition system. We described brlefly here the outline of our proposed systems for speech recognition and speech synthesis for Indian Languages.

3.1 Speech-to-Text system

The main objective of our speech project is to provide a limited dictation machine capability into a computer. The system should be independent of speaker, vocabulary and task. The system should accept speech in normal environments and should degrade gracefully wlth deterioration of quality. The system should provide only a text output corresponding to input speech, i.e., only speech recognition. At present, the system is not expected to understand the input speech.

Basically, speech-to-text systems consist of two stages. The first stage converts the analog speech input into some symbolic form using signal processlng and some knowledge of acoustic-phonetics. The symbolic form is later converted, In the second stage, into meaningful text using higher ievel sources of knowledge such as lexical, syntactic, semantic, etc. Earlier attempts used crude approaches to perform the slgnai-to-symbol transformation, which were

mostly based on some simple signal processing algorithms. The symbols are usually some arbitrary units corresponding to acoustlcaily unlform segments. The number of symbols and the parameter pattern corresponding to each symbol varied from system to system. The systems were highly dependent on high level knowledge sources to disambiguate the symbol sequence into meaningful text. The main disadvantages of these systems are :

(a) The complexity of representation of knowledge sources grew with the size of the task and vocabulary.

(b) The systerns were highly speaker dependent, task dependent, vocabulary dependent and environment dependent.

(c) The signal-to-symbol transformation used crude techniques. Any significant information lost at this stage could not be recovered in the later stages.

Consequently all these systems remained as laboratory demonstration pieces, wlth very limited practical utility.

Current systems seem to lay lmportance on the signal-to-symbol transformation stage, and thus the emphasis in most of these systems is on the design of a phonetic engine. For nonphonetic languages like English, the main problem of choice of symbols remains. The arbitrariness in the symbol choice creates another problem–that of providing a description of the vocabulary in terms of these symbols. Usually this symbolic description is a laborious process involving several man-months of effort for any practical task.

We propose to exploit the phonetic nature of Indian languages to design our speech recognition system [5],[6]. In particular, we choose the written characters of one language as symbols. We expect that most of the meanlngful speech sounds of the language can be transformed into a unique symbol (character), as we feel that, typically in Indian languages "we write what we speak and we speak what we write". This also takes care of variability of pronunciation, because for a different pronunciation of a given word, there will be a different symbol sequence. The signal-to-symbol transformation stage is expected to capture most of the information available in the input.

Since speech contains much more information than its wrltten equivalent, it Is possible to capture and store suprasegmental information along wlth the symbol(character) sequence . Even in the character sequence, several possible close alternatlves can be stored in case of ambiguities in recognition.

The symbol(character) string can be transformed into a meaningful text by using higher level knowledge sources in the second stage. The proposed three blocks in this stage are: lexical, syntactic and semantic expert units. The objective of the lexical expert is to determine word boundaries based on the rules of the language. The input data to this expert is in the form of strings of characters with alternatives and suprasegmental information. The syntactic expert block ensures that the output text contains meaningfully correct sentences of the language. The various blocks are proposed to be implemented using rule-based expert systems. While the purpose of each block is clear, it may use in its rules knowledge from other sources as well. For example, the acoustic-phonetic knowledge to convert speech into symbolic form. Likewise each of the other blocks might use several sources of knowledge.

Each of the blocks can be made very sophisticated by incorporating large number of rules. It is also possible to provide interaction among blocks. It is interesting to note that the complexity of any block increases, if it has to compensate for the poor performance of other blocks. Moreover, the system still produces a reasonably good speech-to-text conversion even with partial knowledge in each of the blocks. In other words, the system does not demand complete knowledge in each block. The system can be refined by incorporating more and more rules. Thus the system degrades gracefully with degradation of either the input speech quality or the nonavailability of complete knowledge from different sources.

3.2 Text-to-Speech System

The phonetic nature of the Indian languages can also be exploited for designing a text-to-speech conversion system. The text-to-speech conversion system would require the conversion of input characters into phonetic symbols and the rules for the synthesis of speech from the given symbol string.

A simple version of the system consists of direct conversion of character string into speech by appropriately concatenating the waveforms corresponding to each character. This is possible, and it produces intelligible speech because of the phonetic nature of the languages. Improvements of this system could be in the direction of incorporating suitable concatenation rules and also use of suprasegmental features. A systematic approach is needed for designing a versatile and sophisticated system.

4. STUDIES ON SPEECH RECOGNITION OF HINDI ALPHABET

As part of our overall Voice Input/Output system development, we have taken up initially

FIG. 5 THE WAVEFORMS OF DIFFERENT UTTERANCES OF HINDI STOP CONSONANTS

Table 1- Hindi stop consonants and their characteristics

classification based on articulation (row-wise)		Unaspirated K-Pset	Aspirated KH-PHset	Unaspirated G-bset	Aspirated GH-BH set	Nasals
		classification based on excitation (column-wise)				
Velar	(K-set)	क (k)	ख (k^h)	ग (g)	घ (g^h)	ङ (ng)
Palatal	(tS-set)	च (ts)	छ (ts^h)	ज (dz)	झ (dz^h)	ञ (gn)
Retroflex	(t-set)	ट (t)	ठ (t^h)	ड (d)	ढ (d^h)	ण (nn)
Dental	(T-set)	त (T)	थ (T^h)	द (D)	ध (D^h)	न (n)
Bilabial	(p-set)	प (p)	फ (p^h)	ब (b)	भ (b^h)	म (m)

studies on two independent tasks related to speech recognition of Hindi alphabet. The first task is the development of a signal-dependant approach for recognition of isolated utterances of Hindi alphabet. The second task is the development of a knowledge-based approach for spotting characters of Hindi in the continuous speech of an utterance.

In Hindi alphabet, the stop consonants are the most confusable subset, and special techniques are needed for recognition of these sounds. We discuss briefly the characteristics of isolated utterances of Hindi stop consonant's. We show that the performance of these isolated letter recognition can be improved significantly by using signal-dependent analysts. We also consider the case of continuous speech in Hindi, and discuss the performance of character spotting using knowledge-based approach.

4.1 Speech Recognition of Hindi Stop Consonants

The stop consonants are produced by completely closing the oral cavity and then releasing the built-up air pressure. The different consonants have different points of closure in the oral cavity. The innermost point of closure in the mouth is at the glottis. The other points of the closure are where the back, middle and front parts of the tongue press against the appropriate regions of the upper palate. Finally, beyond the teeth there is closure at the lips. Thus the five rows in Table-1 represent five different classes of the consonants for the vowel ending ॅ (/a/). These correspond to the five places of articulation. The first four consonants in each row of Table-1 belong to the nonnasalised category. the first two are unvoiced and the next two are voiced. The fifth one is nasal. These members can also be grouped on the basis of the aspiration. The first and the third are of the unaspirated type, whereas the second and the fourth are aspirated. The row-wise and column-wise arrangement of the consonants correspond to the classification according to the place of articulation and manner of production, respectively.

In contrast with the different excitation features such as voicing and aspiration associated with the "manner", the features associated with the different "places " are weak. Hence discrimination among the column-wise consonants is difficult. initially we considered the design of a system for the voiceless stop consonants given in the first column, with the vowel ending ॅ (/a/). Acoustically they differ among themselves in their short leading consonantal part and at the interface with the succeeding vowel part. If the distinguishing primary features are not captured in these short duration, there is almost no chance of recovery from error in the recognition based on the remaining portions of the utterance. Fig.5 shows the waveforms for utterances of Hindi stop consonants. The plots show that the waveform along any given column appear nearly same. Conventional approach of isolated word speech recognition using nonlinear time warping performs very poorly on this confusable set of alphabet. We have developed a

signal-dependent parameter extraction and matching strategy [7] for such a vocabulary. This strategy improves the performance to nearly 80% correct recognition, compared to about 30% obtained by the conventional methods. We have developed a hierarchical approach for recognition of the entire subset of Hindi stop consonants for the different types of vowel endings [4]. The overall recognition performance is over 70%.

4.2 Knowledge-based Approach for Character Spotting

The first stage of a recognition system for continuous speech is signal-to-symbol transformation. As mentioned earlier, we have chosen the written characters of Hindi as symbols. By this choice, we propose to exploit the phonetic nature of Indian languages to design the speech-to-symbol transformation stage. We have decided to realise this using an expert system approach for spotting each written character. The advantages of this approach are :

1. The speech signal can be processed in a manner dictated by the requirements for spotting the character.

2. We are not tied down to a particular parameter or feature set. All the necessary information for spotting the characters can be obtained directly from the speech signal.

3. We are also avoiding creation of complex pronunciation dictionary for converting words into phoneme symbol sequence.

4. The overall complexity of the system does not grow with the size of the vocabulary or task.

The only disadvantage, as we see, is the large number of (typically 5000) characters to be spotted for a given Indian language. But, since each character expert can be implemented independently, we propose to exploit the inherent parallei implementation feature in the final design of our system to overcome this disadvantage.

Each character expert uses knowledge relevant to spot that character. While primarily acoustic-phonetic knowledge is used, the rules may also incorporate either directly or indirectly other knowledge sources such as lexicai, syntactic, semantic, etc.. The knowledge is incorporated as a set of rules, and these rules dictate the parameters and features to be extracted from the speech signal.

The rules for each character expert are organised under the following four broad

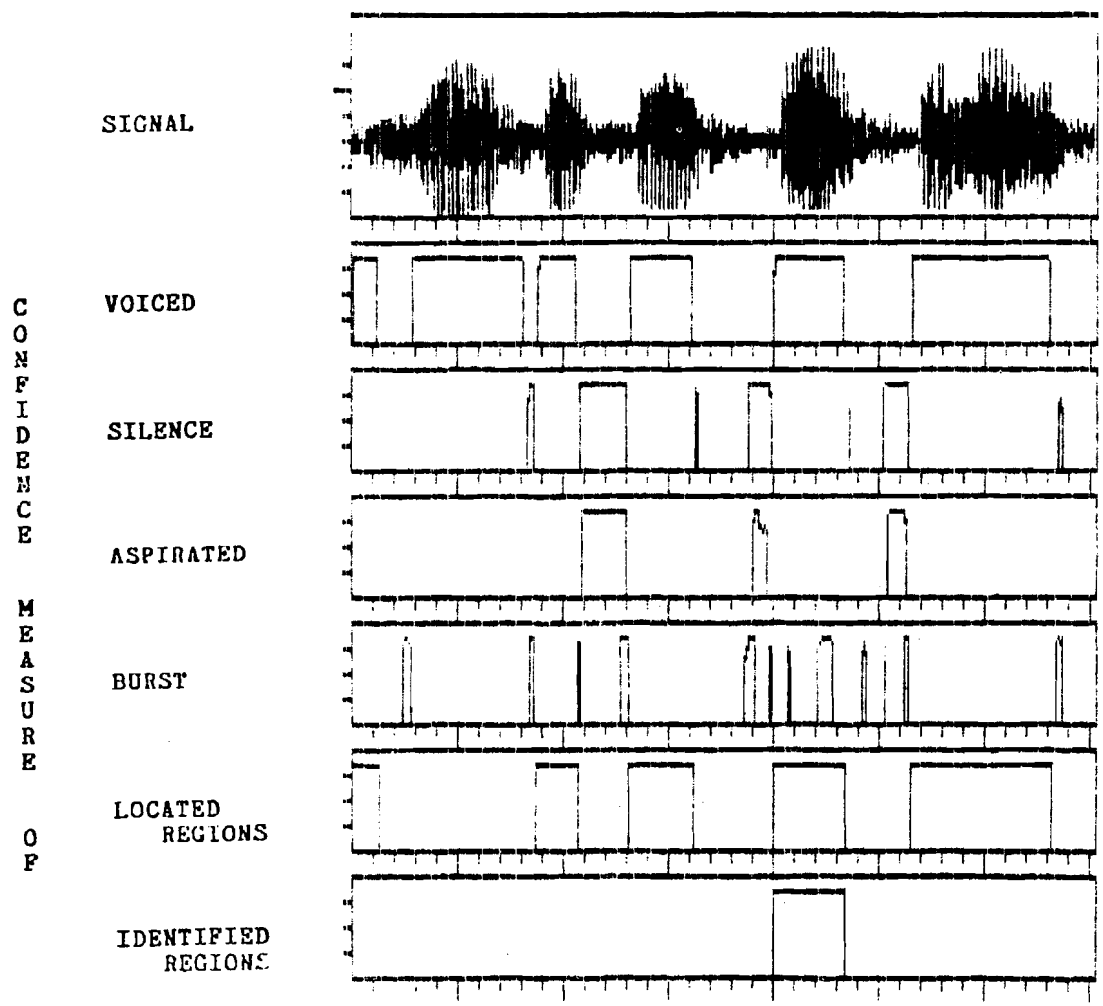


Fig.6 Illustration of क (kaa) Expert

categories :

1. Rules to locate the possible presence of the character in the input speech.

2. Intrinsic cues to recognise the speech segments of the character through a description in terms of articulatory and acoustic features obtained from an acoustic-phonetlc expert.

3. Rules that capture variations in the acoustic correlates of speech segments due to the influence of other segments within a character.

4. Rules that describe the variation that the acoustic features of a character undergo in different character contexts.

It is not possible to spot uniquely a character in an utterance because of the ambiguous nature of the speech signal and also the availability of the partial knowledge to process the speech signal. Fuzzy logic is used to give confidence measures for the conclusions arrived at each stage of the character expert. The outputs of all character experts are combined by second level expert which uses some additional language-specific constraints to get a unique character sequence.

Fig. 6 illustrates the performance of the character का (kaa) in an utterance for the sentence ज्योतिषी का चुनाव (dzjyothishi kaa chunav). Based on the features corresponding to voiced, aspirated, silence and burst, four regions were identified as possible locations for the sound for का (kaa). The first region was eliminated based on the condition that the formant (vocal tract resonance) structure does not correspond to that of the utterance for का (kaa). The second region was eliminated because the burst spectrum peak was very high. The third region was identified as का (kaa) based on the condition that the burst spectrum peak was in the required range and the formant structure also was that for का (kaa). The fourth region was eliminated again because of the mismatch of the formant structure. Thus the का (kaa) expert correctly spotted the region in the given utterance.

5. CONCLUSION

In this paper we presented an overview of the problems in speech-to-text and text-to-speech conversion systems. In particular, we have discussed the state of art in the development of voice Input/Output systems to a computer. We feel that systems for English-like languages may not easily be extendable for use with Indian languages. We have proposed a different approach to the design of Voice Input/Output systems for Indian Languages, which can exploit

the phonetic nature of the language and its orthography. We have discussed the progress made in our laboratory in the development of isolated word speech recognition for Hindi alphabet and also for spotting Hindi characters from continuous speech.

Currently we are developing expert systems to spot all possible characters of Hindi from continuous speech. We are also working on the modules for representation and activation of linguistic knowledge sources to remove the ambiguities in the character string generated by the signal-to-symbol transformation stage. Simultaneously, we are also developing a text-to-speech system for an Indian language, by exploiting the linguistic and phonetic features of the language. We expect some simple versions of these systems to be operational within the next three years.

Acknowledgments

The author would like to gratefully acknowledge the many significant contributions of his colleagues and students of the speech group in this effort.

References

1. Special Issue on "Man-Machine Speech Communication", Proceedings of IEEE, November 1985.

2. W. A. Lea, Trends in Speech Recognition, Prentice-Hall Inc., 1980.

3. Proceedings of SPEECH TECH-87 - Voice Input/Output Applications Show and Conference, April 28-30 1987, New York, Published by Media Dimensions Inc., New York, NY 10010.

4. Proceedings of European Conference on speech technology, Edinburg, UK, September 2-4, 1987.

5. W. Sidney Allen, Phonetics in Ancient India, London, Oxford university Press, 1953.

6. Technical reports by speech Research Croup in the department of Computer Science and Engineering, Indian Institute of Technology, Madras 600 036.

7. B.Yegnanarayana, S.Raman and R.Sundar, "Signal-dependent analysis for speech recognition", Proceedings of International Conference on Speech Input/Output Techniques and Applications, London, March 24-26, 1986.

8. P. Eswar, S. K. Gupta, C. Chandrasekar , B. Yegnanarayana and K. Nagamma Reddy, "An acoustic-phonetic expert for analysis and processing of continuous speech in Hindi", Proceedings of European Conference on Speech Technology, Edinburgh, September 2-4 1987.