Tata Institute of Fundamental Research (TIFR)

1. Name of the Technology

Hindi Speech Data base.

2. Nature of Technology

This is related to the speech technology. The data base is meant to be supportive for developing Automatic Speech Recognition (ASR) systems in Hindi.

3. Level: (Product / Technology / Sub-system) : Product

4. Technical Description of the Technology / Product including Basic block

diagram, Algorithm used, O/S used, Front-end / user interface, and

Specification of the Technology / Product:

An extensive corpus of segmented and labeled speech data is required for adequate training of speech recognition systems. Since existing databases in foreign languages cannot be used 'as is' for Indian languages (owing to language specific effects), there is a need for developing anew. As part of VOICE-II project, a labeled and time-aligned database of spoken Hindi sentences has been generated with the collaborative efforts of the speech groups at TIFR, Mumbai and CEERI, New Delhi. Although the intended principal end-use of this multi-speaker, continuous speech database is in training of a speech recognition system for Hindi, the design specifications of thedatabase are so general that it aids tasks such as speaker recognition, study of acoustic-phonetic correlates of the language. It is comprehensive enough to effectively capture phonetic, acoustic,intra-speaker and inter-speaker variabilities in Hindi speech.

A general purpose speech database should ideally have sentences containing all phonemes of the language in various contexts. This, in turn, requires a phonetically rich corpus of sentences which is not available for Hindi. The manual generation of these sentences is

laborious and difficult. A semi-automatic approach was followed to arrive at an appropriate set of sentences. This involved choosing, based on an objective criterion, sets of phonetically rich sentences out of a large corpus of online Hindi text. Hindi sentences, entered into computer at IIT, Kanpur and CEERI, New Delhi were collected and pooled. The Text-to-phoneme module for Hindi developed at Deccan College, Pune was used to generate expected phonemic transcriptions of all these sentences. Several criteria for deriving optimal sets of sentences were explored. GIST software was used to obtain hardcopy of these sentences.

As a result of this effort, we now have two sets of sentences:

Two dialect sentences, containing all phonemes of Hindi. 100 sets of eight phonetically rich sentences. These sentences cove all contexts between acoustically meaningful classes of phonemes.

100 sets of sentences to be spoken by 100 speakers were converted to a form acceptable by a Hindi word processor on unix. The first set of 2 sentences, read by each speaker, were designed such that they contain almost all the phonemes of Hindi. The second sets of 8 sentences were distinct for each speaker. The sentences in the second set were chosen so that they collectively cover most phonemic contexts. Specifically, the design of sentences aimed at covering all the left and right contexts of Broad Acoustic Classes of phones. Also, due emphasis was given to special phonemic characteristics of sounds of Indian languages such as aspiration and retroflexion. While the first set of 2 sentences can be used for the study of acoustic and phonological variations with respect to change of speaker and dialect, the wide variety of phonetic context variations provided by the second set of sentences can be used for training of speaker independent speech recognition systems. Desktop speech as well as high quality speech was simultaneously recorded by using a close-talking microphone as well as a desk mounted microphone. This has the advantage that segmentation and labeling of a sentence of high quality speech can be used to segment the desktop speech by simply time shifting the signals appropriately.

500 Hindi sentences spoken by 50 speakers, and recorded using close microphone, were hand segmented in terms of phone-like labels by trained personnel using visual displays of speech such as waveform, spectrogram. The units of representation (labels) were so chosen that each unit is acoustically homogeneous and the labels are acoustically fairly

distinct from each other. Most labels represent phonemes while some represent sub-phonetic units. This corpus can be used by researchers for studying acoustic-phonetic aspects of Hindi as well as by technology developers for developing Hindi Speech Recognition systems.

For further details, the interested readers may refer to the following

publications:

(i) Samudravijaya K, K.D.Rawat, and P.V.S.Rao,

"Design of Phonetically Rich Sentences for Hindi Speech Database",

J. Ac. Soc. Ind. vol. XXVI, December 1998, pp. 466-471.

(ii) Samudravijaya K<, "Knowledge Based Spectral Subtraction",

Proc. of the Int. Conf. on Knowledge Based Computer Systems, eds.

M.Sasikumar et al., pp. 237-246, Mumbai, 1998.

(iii) Samudravijaya K, P.V.S.Rao, and S.Agrawal,

"Hindi Speech Database", to be presented at International Conference

on Spoken Language Processing, October 16-20, 2000, China.

5. Representative Snapshot / screenshot of the Technology / Product: Not Available

6. Scalability / Portability / Expandability:

This can be used in any Window or Unix system.

7. Readiness of Transfer of Technology (ToT):

The CD of the data base are available. This was handed over to the Department of Information Technology.

8. Availability of documentation:

Brief description and user guide is available.

9. Testing of the Product / Technology:

The product can be tested only in conjunction with the speech recognition systems. As such systems are now evolving, it will be possible to check the efficacy of the data base in near future.

10. IPR / Open-source:

As the CD has been handed over to the MICT, it depends on them. It is however suggested that the data base be available for R & D personnel of speech technology in Indian languages either free or at a nominal price.

11. Potential beneficiaries:

Direct beneficiaries will be the developers of ASR in Hindi. However, R & D efforts in any area of Hindi speech technology may be indirectly benefited from this. For example, speech data can be used for analysis, synthesis or other purposes.

12. User-agency tie-up: Nil

13. Name and address of the Resource Person:

Dr. K. Samudravijaya,

School of Technology and computer Science,

Tata Institute of Fundamental Research,

Homi Bhabha Road, Mumbai 400 005