Designing and Development of Speech Database The Technical Review

Conference Paper · January 2011

CITATIONS
0

READS
54

4 authors:

Aaron M. Oirere
Dr. Babasaheb Ambedkar Marathwada Unive…
4 PUBLICATIONS **3** CITATIONS
SEE PROFILE

Dr. Vishal Waghmare
Shri Swami Vivekanand Shikshan Sanstha, K…
24 PUBLICATIONS **59** CITATIONS
SEE PROFILE

Ganesh Bapurao Janvale
MGM, Institute of Biosciences and Technology
23 PUBLICATIONS **29** CITATIONS
SEE PROFILE

Ratnadeep R. Deshmukh
Dr. Babasaheb Ambedkar Marathwada Unive…
136 PUBLICATIONS **195** CITATIONS
SEE PROFILE

CHAPTER 22

Designing and Development of Speech Database – The Technical Review

Aaron M. Oirere, Vishal Waghmare, Ganesh Janvale and Ratnadeep Deshmukh
Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad-431004 (MS) India

ABSTRACT

The speech database is one of the crucial items used for the automatic speech recognition system (ASR System). The paper reviews numerous techniques to design and development speech databases for different purpose. The paper presents here compression of four different language databases which are Marathi, Telugu, Tamil, Hindi and Spanish. It also gives the proposal to design and development of Swahili language database for automatic recognition system.

1. INTRODUCTION

Language technologies can provide solutions in the form of natural interfaces so that digital content can reach to the masses and facilitate the exchange of information across different people speaking different languages. Our overall goal is to develop speech recognition and speech recognition systems for most of Swahili language. Swahili is a Bantu language that serves as a second language to various groups traditionally inhabiting parts of the East African coast. About 35% of the Swahili vocabulary derives from the Arabic language, gained through more than twelve centuries of contact with Arabic-speaking inhabitants of the coast of Zanj. It also has incorporated Persian, German, Portuguese, English and French words into its vocabulary through contact during the last five centuries. Swahili has become a second language spoken by tens of millions in three countries, Tanzania, Kenya, and Congo (DRC), where it is an official or national language.

In this paper, we discuss the comparatives of different language database and Swahili language database for building large vocabulary speech recognition systems. The work could also be of great interest and significance to a large number of speech scientists as languages have certain properties that make them a special case. Most of the Indian languages are phonetic in nature. This is to say that there exists a one-to-one correspondence between the orthography and pronunciation in these languages. There are a number of phones present in the Indian languages, like the retroflexes, aspirated stops etc, which English and most European languages don t possess. At the same time, there are not many differences between the phone sets of individual Indian languages. This paper is organized as follows: Section 2 describes the comparison of techniques to design and development speech databases. Section 3 describes the ideas to design and develop Swahili language database. The concluding remarks and future work are stated in the section 4.

2. DATABASE COLLECTION IN VARIOUS LANGUAGES

2.1 Mandarin – English

2.1.1 Data collection

The data was taken and categorized in three categories: native mandarin corpus, native English corpus and mandarin accented English and were recording through telephone line and digitalized at 8 KHz sampling rate with 18 bits resolutions. Mandarin was taken from the native Chinese, English were recorded in higher rate and were band limited to 4 KHz by down sampling and was collected in the house labeled as GTC (general Telephone conversation corpus) speech data was from 60 males and 60 females speakers each with 200 utterances for every day conversations.. The data was taken in restaurants, streets and other noisy places with different background noises, speaker fluency an accent.

2.1.2 Data base

It was chosen from taking single set of bilingual acoustic set of bilingual acoustic model derived by phone clustering. It was developed instead of using two separate monolingual models for each language. The novel two-pass phone clustering method based on confusion matrix (TCM) was presented and compared with the log-likelihood method and it was found that TCM had better results than log-likelihood.

The data was taken in two approaches that is: pronunciation and acoustic. Acoustic dealt with the utterance; which were collected in a region where the matrix language was spoken. It was categorized in three stages: a single set of acoustic to handle both inter- and intra-sentence language switching, a two- pass cross language phone clustering approach based on confusion matrix, optimal way of merging the phone sets of English and mandarin. By merging and clustering the phone set of these two languages, anew set of mono-phones covering both language are determine. The data consist the names of the singers and titles of the songs, they had total of 10179 utterances: 8183 mono-mandarin, 1650 mono-English and 346 bilingual

2.2 Hindi Database

2.2.1 Data collection

In the sound unit the sentence were represented as sequence of phonemes. The units were acoustically homogeneous and fairly distinct from each other. Acoustic-phonetic features of Hindi language differ from that of European language. Aspiration is a phone in Hindi unlike in English. Rectroflexion features were occurring in prominent places in Hindi. The list all the acoustic –phonemic in complete closure of oral tract i.e. plosives, affricates and nasals labeling was done by special symbols which represented devoicing of the voiced closure, voicing of a unvoiced closure, voiced glottal fricative, tongue, click, glottal stops, voiced segmentations, vowel and silence.

2.2.2 Designing of sentences

Designing of the sentence was done in three format i.e. one contain all the phonemes, another rich in phonemic context and final used the identification of dialect/accent. The sentences were syntactically valid and meaningful. 10 sentences were taken in two parts: 2 sentences containing dialects and 8 sentences containing as much of phonetic context as possible. The sentences were made to be syntactically valid, meaningful, natural, simple and short for easy understanding and feature extraction. Designed that each subject was to speak 10 sentences containing of two parts first consist two dialects which contain all the phonemes of the language and secondly they cover the large set of diverse phoneme context.

2.2.3 Designed strategy

100 sets of Hindi sentences were spoken by 100 speakers and were chosen from a corpus of machine readable Hindi text from diverse source. The short sentences were represented by less than 80 ASCII character were retrained. A text to phoneme was used after taking in account the micro-phonemic rule in Hindi. Dialect sentences were selected which

contained maximum number of dialect Hindi phonemes and speed sounds of manner and place of articulation were minimized.

2.2.4 Data collection

The data was collected in two ways; in the room and the field.

The speaker was 100 and each of them read 10 sets phonetically rich sentences which were digitally recorded. In the room there was a close talk of a distance of 5 cm from the mouth and another which was in the field mounted in the desk at a distance of 1 meter and it was collected in the simple shift of the time indices. Then the data was segmentation and labeling. The sampling was done at 16 KHz and digitalized with 16 bits. Vowel duration was also done for it was useful for understanding and capturing systematic variations occurring in natural speech. Occurrences of vowels were more than consonants. The durations were differing from one to another. This was done by the mean and standard deviation. The durations were checked on native Hindi speakers and non-native speakers and it was found that systematic behaviors were distinct.

2.2.5 Database

The database was developed well to capture phonetic, acoustic, intra-speaker and inter-speaker variability in Hindi. The database contains 500 sentences taken from 50 speakers. Vowels duration was done synthesis in high quality speech which was done by the help of mean and standard variation of the vowels from the native Hindi and non-native. The database was well done to capture phonetic, acoustic, intra-speaker and intra-speaker.

3. PROPOSAL FOR DEVELOPMENT OF SWAHILI LANGUAGE DATABASE

In this section, the various steps involved in building the speech corpora are detailed. Firstly, the recording media will be chosen so as to capture the effects due to channel and microphone variations. For the databases that will be built for the Swahili language ASRs, the speech data will be recorded over calculated number of landline and cellular phones using a multi-channel computer telephony interface card.

3.1 Speaker Selection

Speech data is collected from the native speakers of the language who will comfortable in speaking and reading the language. The speakers will be chosen such that all the diversities attributing to the gender, age and dialect are sufficiently captured. The recording is clean and has minimal background disturbance. Any mistakes made while recording will be by re-recording or by making the corresponding changes in the transcription set.

3.2 Data Statistics

Speakers from various parts of the respective states (regions) will carefully recorded in order to cover all possible dialectic variations of the language. Each speaker will be recorded 52 sentences of the optimal text. To capture different microphonic variations, four different cellphones or landline will be used while recording the speakers.

4. CONCLUSION

In this paper, we discussed the optimal design and development of speech databases for the English and Hindi language. We hope the simple methodology of database creation presented will serve as catalyst for the creation of speech databases in all other languages. We also proposed the methodology to design and develop Swahili language speech database for automatic recognition system. We hope the ASRs created will be served as baseline systems for further research on improving the accuracies in each of the languages. Our future work is focused in tuning these models and test them using language models built using a larger corpus and to develop large vocabulary speech database recognition system.

References

1. http://tdil.mit.gov.in/corpora/ach-corpora.htm#tech
2. L.Rabiner, A Tutorial on Hidden Markov models and Selected Applications in Speech Recognition, Proc. Of IEEE, Vol. 77 No. 2, 1989.
3. Samudravijaya K, P.V.S.Rao, and S.S.Agrawal, Hindi Speech Database, Proc. Int. Conf. on Spoken Language processing (ICSLP00), Beijing, China, October 2000, CDROM paper: 00192.
4. Singh, S.P., et al Building Large Vocabulary Speech Recognition Systems for Indian Languages, International Conference on Natural Language Processing, 1:245-254, 2004.
5. Richard Winsky. 1997. "Definition of Corpus, scripts and standards for Fixed Networks", SpeechDat project, doc ref LE2-4001-SD1.1.3, 22.