Volume 3, Issue 10, October 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Development of Multi-Variability Speech Corpus of Adi Language for Speech Recognition Researches

Kapang Legoh*
Department of CSE, North Eastern Regional Institute of Science &Technology (NERIST), India

Utpal Bhattacharjee
Department of CSE, Rajiv Gandhi University (RGU), India

T. Tuithung
Department of CSE, North Eastern Regional Institute of Science & Technology (NERIST), India

Abstract—To develop an Automatic Speech Recognition (ASR) System using a language, standard speech database of that language is required. In this paper, we present a description of Adi phonemes and a methodology and typical experimental set up for development of multi-variability Adi speech corpus which will be useful for development of robust Automatic Speech Recognition System and speech researches in Adi language. The Adi is a major collective tribe living in the Himalayan hills of Arunachal Pradesh, the North Eastern most state of India, having international borders with China, Myanmar and Bhutan and the language spoken by Adi tribe is the Adi language. Session, speaker, dialectical, device and multi-style variabilities are considered while developing the speech database. We have specifically designed and developed Adi Speech databases in two phases recorded in three different sessions for each speaker using four different recording devices. Modes of speech used are connected, read and conversational. The speech data have been collected from 20 male and 20 female speakers in first phase and 20 male and 15 female speakers in second phase for training and testing purposes. Annotated Adi Speech corpus and segmented phoneme database are created for development of robust phone based speech recognition system.

Keywords— Adi Language, Phonemes, Adi Text Corpus, Multi-Variability, Robust ASR, Praat Tools.

I. INTRODUCTION

Speech is the most natural and effective mode of communication between human beings. With the rapid development of the speech technologies, human-machine communication using natural speech has received wide attention from both academic and business communities worldwide since last five decades [1, 2]. It becomes more interesting if one can communicate with machine in his/her native language; and Automatic Speech Recognition (ASR) is the process of automatic extracting and determining linguistics information conveyed by a speech wave using computers [3]. Creation of proper and standard speech database is the basic requirement for development of ASR system in any language. With the recent advances in speech and speaker recognition technologies, sufficient speech corpus is required to properly support the researches and advances in the speech technologies for a language as the quality of ASR system also depends on the quality of speech corpus. Speech corpus has already been developed for many foreign languages [5, 6, 7], Indian languages [8, 9, 10, 11] and some languages of North East India [12, 13] by various researchers.

One of the main challenges before ASR researchers is to unify acoustic processing and to adapt the architecture of the ASR system to cover the broadest range of languages and situations like real-life environments and to build language-independent systems. Different models like ANNs, stochastic models using Hidden Markov models and hybrid models are used for speech recognition by the researchers [4, 14]. However, there are many practical challenges in the development of robust ASR system. The recognizers which are highly efficient in noise free conditions, fails drastically when tested in noisy environments. The environment in which the recognizer is supposed to operate naturally also influences the recognition performance. Adverse conditions like environmental noises, acoustic distortions, recording conditions, gender and age of speakers, mismatches in microphones, channels and sessions affect the performance of ASR systems. Speaker variability, speaking rates, vocabulary sizes, linguistic constraints, mode of speech like isolated, connected, continuous and spontaneous speeches and mismatches in training and testing influence the recognition performance of the ASR systems. To construct robust ASR system a standard multi-variability speech database is required to create robust speech features and models to enhance the performance of ASR systems under various adverse environments.

In this work, the Adi language, which is the most widely spoken language of Arunachal Pradesh, the North Eastern most state of India, is considered for the development of multi-variability speech corpus. It will be a framework and can be used by researchers and to explore other similar tribal languages of the state and North East India. The speech corpus contains speeches from 75 speakers; recorded from 40 male and 35 female Adi native speakers in two phases, Phase-I and Phase-II. The three main dialectical variations of Adi from three districts; East Siang, Upper Siang and Lower Dibang Valley districts of the state have been considered in this work. This paper is organized as follows. Section II briefly describes design and development of ideal speech corpus and a very brief survey on standard speech corpus for

ASR. Adi language, Adi phonemes and the syllable structures of Adi language are briefly described in Section III. Section IV describes the design and development of various Adi text corpus/stories to be used for collection of speech data. Section V describes the technical details of creation of multi-variability speech database of Adi. Features of the Adi speech database are described in Section VI and finally conclusion of the paper is described in Section VII.

II. A BRIEF SURVEY ON STANDARD SPEECH CORPORA

Use of standard speech corpus for development and evaluation of speech and speaker recognition systems helps in promoting progresses in the development of speech technologies and researches. For the development of an ideal corpus for speech and speaker verification researches in different languages under various environments, there are lots of important criteria [13, 21, 22, 23] that should be followed and are listed below:

- Enough speech must be collected from enough speakers in order to validate an experiment under study.
- Speech corpus should contain equal numbers of male and female speeches to perform experiments exclusively for male or female or both.
- Speakers of wide ranging age and varying educational backgrounds should be considered in order to study the effect of age on pronunciations and tract speaking rates, intelligence levels etc.
- A good corpus will therefore strive to include as many styles and registers of language as possible, including samples of spoken language.
- Corpus should be designed for closed set and open set ASR experiments.
- Data should be collected with different types of microphones, transmission channels and medium such as mobile phones, laptop, fixed mounted headphone etc. in order to track the effect of device variability on the ASR systems.
- Data should be collected with multiple training and testing sessions in order to keep track of the effect of intersession variability on ASR or Speaker Verification performance.
- Speech data for a particular language should be collected in all possible dialect in order to track all possible variations in pronunciations of that language.
- Text materials given to speaker should be phonetically sound and balanced such that it contains all possible vowels, nasal consonants, fricatives and nasal-to-vowel co-articulations etc.
- Data should be collected in wide range of acoustic environments such as home, office or telephone booth or road or cars or noisy conditions or workstations etc. in order to make systems more realistic.
- On the issue of spontaneous speech, it is decided that the majority of the corpus should be read speech, for economic reasons, whereas the majority of the test data (which comprises a small fraction of the total data) should be spontaneous speech. The reason for these decisions is that the large amount of read speech will provide greater training benefits than smaller amounts of spontaneous speech, while using spontaneous speech for testing would better validate the technology for a relatively small increase in cost.

Standard speech corpora for training and testing of speech and speaker recognition systems have been reported from many researchers and organizations around the world since last 20 years. There are publicly available standard corpora for speech and speaker recognition. Some of the standard organizations like the Linguistic Data Consortium (LDC) [15], the European Language Resources Association (ELRA) [16] and the Oregon Graduate Institute (OGI) [17] are some of the main supplier of the standard corpora. Brief description of the major speech corpora for speech and speaker recognition, like Texas Instruments Massachusetts Institute of Technology TIMIT(LDC) database, the Italian speech corpus for Speaker Identification and Verification Archives SIVA(ELRA), French PolyVar (ELRA), European project POLYCOST(ELRA), the YOYO(LDC) corpus, Switchboard I-II and NIST Evaluation Subsets (LDC), Speaker Recognition Corpus (OGI) from the Center for Spoken Language Understanding, Oregon Graduate Institute of Science & Technology, Australian National Database of Spoken Language ANDOSL database and Digit-SPL speech database developed at Griffith University are described in[13, 18, 19, 20].

However, most of the speech corpora have been developed in American and Western context. These corpora are not suitable for speech recognition researches for tribal languages like Adi and other languages of North East India and Asia. Therefore, development of standard speech corpus of Adi language is required and such speech database will help preserve and develop the language and also useful for development of speech technologies in the tribal language context and researchers to achieve convergence in the speech technologies towards development of universal ASR system. It is required to develop multi-variability speech database to construct and evaluate the systems in more realistic way.

III. BRIEF DESCRIPTION OF ADI LANGUAGE, PHONEMIC AND SYLLABLE STRUCTURES

A. Adi Tribe and the Language

The Adi is a major collective tribe living in the Himalayan hills of Arunachal Pradesh, the North Eastern most state of India, sharing international borders with China in the North, Bhutan in the West and Myanmar in the East and Assam and Nagaland states of India in the South. The Adis are found in the temperate and sub-tropical regions within the districts of East Siang, Upper Siang, West Siang, Lower Dibang Valley and Lohit districts of the state. Some of them are found in Southern Tibet, around the areas near the Indian border [24]. In British era the Adi was called as *Abor*, which means "uncontrol/independent/savage" [24, 25]. The term *Abor* is a deprecated eponym from Assamese. However, the use of *Abor* name is traced back to the first century AD [25] in which *Pliny the elder* mentioned a great valley in the

Himalayas inhabited by '*Abariman*' a wild tribe and Adi tribe had been equated and identified with *Abariman* by the British. The literal meaning of Adi is "hill" or "mountain". The language spoken by the Adi tribe is called the Adi language and it belongs to a Tibeto-Burman family of languages which is spoken with minor dialectical variations among all the Adi sub-groups in different districts of the state. Other similar languages of the state which belong to same Tibeto-Burman family are Nyishi, Galo, Apatani, Tagin etc. The present work covers the three main dialectical variations of Adi language in three districts of the state; East Siang, Upper Siang and Lower Dibang Valley.

Adi language is rich in oral literature. It has several forms like Bari, Aabang, Pénge and Abee which are used for different occasions. The Bari and Aabangs are used in festive, religious and cultural ceremonies. The Pénge is used in funeral hymns and Abee in political speeches. The Aabang of Adi has been compared and equated with the Puranas of Sanskrit literature [25]. Despite its richness in oral literature the Adi lacks script of its own and therefore Roman script is used in Adi literature with some diacritic symbols to represent unique phonemes of the Adi. For the first time in the history, the Roman script was used for Adi by J. Herbert Lorrain in 1900 [27]. Although Adi lacks script of its own, all the forms of Adi language have been preserved by memorising and orally passing down to the next generation by the elders in the form of rhapsodies, ballads, lullabies, poems/verses, etc. since time immemorial.

The orthographic convention used in this paper for Adi phonemes is the one used by Adi Agom Kebang (AAK) [24]. The AAK is the Apex/Supreme Council of the Adi tribe that works for the preservation and development of Adi language.

B. Adi Phonemes

In this study, 50 Adi phonemes have been considered in which there are 7 short vowels, 7 long vowels, 16 consonants, 19 diphthongs and 1 triphthong. In Adi language, the length/duration of vowels is of high significance since the meaning of same word is decided by the duration (short or long) of vowel phonemes in that word, like [atɔ] and [aːtɔ] which mean *father-in-law* and *come here* respectively. Therefore, for each short vowel there is an equivalent long vowel which gives different meaning to the word. The Adi phonemes have been described using phonetic properties like place of articulation (bilabial, alveolar, palatal, velar and glottal), manner of articulation (stops, nasals, liquids, glides, fricatives and affricates) and type of excitation (voiced or unvoiced) for consonants; and position, height and motion of tongue and shape of lips for vowel phonemes. We have represented the Adi phonemes in / / and corresponding International Phonetic Alphabet (IPA) representation of the phonemes in [ ] in Table I and Table II for consonant and vowel phonemes respectively. In Table II both short and long vowels are shown. The long vowels are represented with two consecutive characters. Diphthongs and Triphthong are shown in Table III and Table IV respectively. The Adi language does not have aspirated, retroflex, dental and labio-dental fricative sounds. The fricatives like [s] and [h] are interchangeably used in Adi speech without any alteration to the meaning of the word like /*hidum*/ or /*sidum*/ in which both have same meaning, which means deer. The following tables give details of Adi phonemes.

Table I. Consonant phoneme characteristics of Adi.

	Bilabial	**Alveolar**	**Palatal**	**Velar**	**Glottal**
Stops (v) (uv)	/p/ [p] /b/ [b]	/t/ [t] /d/ [d]		/k/ [k] /g/ [g]	
Nasals (v)	/m/ [m]	/n/ [n]	/ny/ [ɲ]	/ng/ [ŋ]	
Roll/ Liquid (v)		/r/ [r]			
Fricatives(uv)		/s/ [s]			/h/ [h]
Affricates (v)		/j/ [dʒ]			
Approximant/glide (v)			/y/ [j]		
Lateral Approxmant/ Liquid (v)		/l/ [l]			

(v = *voiced*, uv = *unvoiced*)

Table II: Vowel phoneme characteristics of Adi.

	Front (Unrounded)	**Central (Unrounded)**	**Back (Rounded)**
Close (Short) (Long)	/i/ [i] /ii/ [iː]	/í/ [ɨ] /íí/ [ɨː]	/u/ [u] /uu/ [uː]
Close-mid(Short) (Long)	/e/ [e] /ee/ [eː]		
Mid (Short)		/é/ [ə] /éé/ [əː]	

(Long)			
Open-mid(Short) (Long)			/o/ [ɔ] /oo/ [ɔː]
Open (Short) (Long)		/a/ [a] /aa/ [aː]	

Table III: Adi diphthongs.

Sl. No.	Diphthongs	Example in Adi	Sl. No.	Diphthongs	Example in Adi
1	/aé/ [aə]	ng**aé**t	11	/ía/ [ɨa]	**ía**m
2	/ai/ [ai]	p**ai**	12	/íé/ [ɨə]	lood**íé**m
3	/ao/ [aɔ]	y**ao**	13	/ío/ [ɨɔ]	**ío**ng
4	/au/ [au]	t**au**k	14	/oa/ [ɔa]	d**oa**to
5	/éa/ [əa]	g**éa**to	15	/oé/ [ɔə]	r**oé**
6	/éi/ [əi]	**éi**yot	16	/oi/ [ɔi]	**oi**to
7	/ia/ [ia]	**ia**k	17	/ua/ [ua]	p**ua**kto
8	/ié/ [iə]	**ié**t	18	/ué/ [uə]	l**ué**lule
9	/ie/ [ie]	**ie**sile	19	/ui/ [ui]	p**ui**to
10	/io/ [iɔ]	y**io**			

Table IV: Adi triphthong.

Triphthong	Example in Adi
/uai/ [uai]	Tatt**uai**, Tit**uai**

C. Syllable Structures of Adi Words

The Adi language has four kinds of syllable structures. They are [V], [VC], [CV] and [CVC]. All the syllable patterns occur in all the places in a word; in the initial position, in the medial position and in the final position. It has monosyllabic and disyllabic words and it permits up to trisyllabic words. The quadrisyllabic word is generally formed by combining two independent words, like a compound words, and then used by the Adi speakers. The Adi language has right-to-left *directionality* for syllabification rule and it does not permit consonant clusters. In a word, the moment it gets one consonant as onset, it becomes satisfied and the remaining consonant becomes coda of the second syllable from the right. So, if a pattern like [CVCCV] occurs in a word, then the syllabification will always be done as [CVC.CV] in Adi. In the present work, we have considered all four syllable structures of Adi by preparing text corpus for read speech in such a way that all types of syllable structures are covered with each and every phoneme of Adi in the form of prevocalic, intervocalic and postvocalic. Detail is described in section 4.

IV. ADI TEXT CORPUS SELECTION

We have developed phonetically rich and balanced text corpus/stories for read speech recording to ensure phonemic balance and coverage of all phonetics information of the Adi language. All types of syllable structures are used to cover allophones of each phoneme. The Adi Text corpus for different modes of speech recording has been created in the following way.

A. Isolated Words/Phonemes

List of isolated words containing at least five words each for every vowel (long and short) and consonant has been developed. All types of syllables mentioned in section III(C) are considered for every vowel and consonant while preparing the word list. Prevocalic, intervocalic and postvocalic forms of all consonants with each and every vowel and vice-versa are considered to cover acoustic studies for nasalized vowels, liquids and glides, fricatives etc. This word list contains 421 words. Similarly, a list of words containing all diphthongs and triphthong has been prepared in which occurrence of words containing each diphthong and triphthong is at least three. Also to ensure that the entire Adi phonemes (vowels and stops, nasals, fricatives, affricates, approximant consonants) are covered in the speech corpus, a chart of isolated phonemes (14 vowels and 16 consonants) has been developed. All isolated consonants have been pronounced with suffix /o/ during recording. All these text lists have been recorded in connected mode.

B. Phonemically Balanced Paragraph

A contextual Adi text corpus/story has been developed in which a description of day today activity of Adi people is mentioned. This corpus contains 9 sentences, 198 words and 823 phonemes. This paragraph is phonemically rich and balanced as it contains all 50 Adi phonemes under the study and occurrence of each phoneme varies from 4 to 104. This text has been recorded in read mode.

C. Open Text Corpus

To increase the amount of continuous speech database, three passages of Adi text have been created from the contents of local News paper. Speech of a political leader on Adi tribe, its language, culture, tradition and contribution of Adi tribe in building modern Arunachal Pradesh, published in local news paper has been selected and converted to phonemically rich Adi text. This contextual text corpus contains 400 words which were also recorded in read mode.

D. Spontaneous Text Corpus

To develop spontaneous speech database, question answer sessions were held to record spontaneous response from the speakers. A series of questions was prepared which were asked to the speakers. It contains questions related to speaker's biodata (name, addresses, qualifications, professions, etc.), information on family members, day today activities of speakers, current affairs, and hobbies. The recording duration for each conversation was around 120 seconds to 180 seconds.

V. MULTI-VARIABILITY ADI SPEECH DATABASE FOR SPEECH RECOGNITION

In this section, we describe the technical details of speech collection for development of multi-variability speech database for speech technology researches in Adi language. To develop multi-variability Adi speech database, we collected the speech data in two phases, phase-I and phase-II. In each phase we collected speech data using multiple recording devices simultaneously/in parallel, in multi-sessions, in multi-style speech modes and under different background environments. Almost equal numbers of male and female speakers, ranging from age group of 17 to 55 years, were selected in order to include the variabilities like age, gender, session, microphone, dialect etc. Radio news readers in Adi, native public orators and Adi singers also contributed the speech data. 40 speakers (20 male and 20 female) were selected in first phase and 35 speakers (20 male and 15 female) were selected in second phase for recording. The phase-I data set were collected in quite room and phase-II data set were collected by visiting speaker's residence and in guest houses in uncontrolled environment. All other variabilities are same for both data sets. Initially the speakers were provided with the text corpus and they were trained for 20 to 30 minutes prior to first session recording.

A. Phase-I Speech Data collection

The variabilities considered while collecting the Adi speech data in phase-I are:

- Multi-sessions: All modes of speech described in section IV were recorded in three sessions for each speaker.
- Multiple microphones: Four recording devices were used in parallel to collect the speech data.
- Multi-style speech modes like connected, read and conversational speeches were used.
- 20 male and 20 female speakers ranging from age group of 17 to 55 years were selected from three districts.

In phase-I, a headset microphone was connected to a Dell laptop for recording; two different mobile phones and one digital voice recorder were used. These are the commonly used portable devices used for voice recording available to people. The head set microphone was set parallel to the wave front and other three devices were kept at a distance of around 45 to 50 cm from the speaker's mouth. Table V describes the details of devices and formats used for collecting speech data, Table VI describe the occurrences of dialect variations of three districts in each phase of speech data collection and Table VII describes the detail specification of the multi-variability speech database.

B. Phase-II Speech Data Collection

The variabilties and hardware set up considered in phase-II are same as those of phase-I except the recording environments which were uncontrolled in case of phase-II, with fan and Air Conditioner switched on, and the head set microphone was connected to Sony Vaio ultra laptop. In this phase 20 male and 15 female speakers were selected for recording from the three districts.

Table V: Details of devices used for collecting speech data.

Devices	**Make/Model**	**Sampling Rate**	**Recording format**
Headset Microphone	HP, BHP-95NCV	16 KHz	wav
Digital Voice Recorder (DVR)	Sony ICD-UX523F	44.1 KHz	mp3
Mobile Handset-1	Samsung GT - S3653	8 KHz	amr
Mobile Handset-2	Nokia – E71	8 KHz	wav

Table VI: List of occurrences of dialectical variations in phase-I and phase-II

Dialect Variation/District	**Phase-I**		**Phase-II**	
	Male	**Female**	**Male**	**Female**
Lower Dibang Valley	10	10	5	5
East Siang	5	5	8	5
Upper Siang	5	5	7	5

Table VII: Detail specification of the Adi speech corpus.

Language	Adi (Three dialectical variations of Adi)
Speakers	Phase-I: 20 male, 20 female & Phase-II: 20 male, 15 female
Number of sessions	3
Intersession interval	1 to 3 week
Data type	Speech
Type of speech	Connected, Read & Spontaneous
Sampling rate	16KHz
Sampling format	PCM mono channel, 16 bit resolution
Acoustic environment	Close room (phase-I) and uncontrolled (phase-II)
Application	Robust Speech Recognition for Adi language, Researches in mismatched environments.
Speech duration	11 to 12 minutes a session (3 sessions, 4 devices & 75 speakers = 165 hrs. of speech approx.)
Training session	Phase-I speech data to be used
Testing session	Phase-II speech data to be used
Phoneme database	Manually segmented using Praat tools

VI. DESCRIPTION OF SPEECH CORPUS AND THE PHONEME DATABASE

In this paper, we have developed a multi-variability speech database of Adi language of Arunachal Pradesh for development and evaluation of robust ASR systems and for aiding researches on the unexplored tribal languages of the state. Some important features of the Adi speech corpus developed are mentioned below:

- The speech data were collected in quite room with fan and air conditioner switched off for phase-I data set and speakers were invited in the office chamber and at home for recording.
- For phase-II data set speech data were collected by visiting the speaker's residences and some were recorded in guest houses with uncontrolled environment i.e., fan and air conditioner was switched on.
- Four devices, mentioned at Table V, were used to collect speech data and the speech was recorded simultaneously with the four devices in each session. These devices are used as they are highly portable and most commonly used by people. Different sampling rate, file format and resolution have been used while recording. This gives lots of variabilities (device, sampling, format and environment) in the collected data.
- The speech data were recorded in three different sessions having intervals of one to three weeks in order to have session variability in the collected data for each data set.
- The text corpus described in section IV (A) was recorded in connected mode, section 4 (B) and (C) in read mode and section 4 (D) in conversational mode.
- Speakers ranging from 17 to 55 years of age group were selected with different educational background ranging from metric to post graduate. Almost equal number of male and female speakers was selected from three districts to maximize the coverage of speaker variability and dialectical variations.
- Combining speeches recorded from four devices in three sessions, each speaker contributed approximately two and half hour speech.
- All speech data collected were converted to 16 KHz sampling rate, wav and PCM mono format with 16 bit resolution and amplified before phoneme segmentation. Phoneme segmentation was done manually using the Praat tools [26] for construction of phone database from the speech corpus which will be used for phone based speech recognition researchers.

VII. CONCLUSION

This paper presents the optimal methodology for design and development of a multi-variability speech corpus containing around 165 hours of Adi speech recorded using four different devices in three different sessions from 75 speakers selected across three districts of the state of Arunachal Pradesh and covers three main dialectical variations of Adi language. We hope that this work will serve as a catalyst for development of speech databases of all other languages of the state and the region and promote further scientific researches on the still unexplored tribal languages of the state. Future work would include collection of more speech data from three districts. A phone based ASR system is being developed on the speech corpus by the authors.

ACKNOWLEDGMENT

The authors sincerely thank Mr. Bolin Pao, Deputy Manager (IT), NEEPCO for his efforts and help towards the collection of phase-II speech data set.

REFERENCES

[1] B. H. Juang and L. R. Rabiner, "Automatic speech recognition – A brief history of the technology development," K. Brown (Ed) *Encyclopedia of Language and Linguistics*, Elsevier, 2005.

[2] S. Furui, "50 years of progress in speech and speaker recognition," *ECTI Transactions on Computer and Information Technology*, Vol. 1, No.2, November 2005.

[3] S. Furui, *Digital Speech Processing, Synthesis and Recognition,* Marcel Dekker, Inc., 2001.

[4] A. Amrouche, Md. Debyeche, A. Taleb-Ahemad, J. M. Rouvaen and M. C. E. Yagoub, "An efficient speech recognition system in adverse conditions using the non parametric regression,"*Engineering Applications of Artificial Intelligence,* Elsevier, 2010, 23, pp. 85-94.

[5] R. Carre, R. Descout, M. Eskenazi, J. Mariani, and M. Rossi, "The French language database: defining, planning, and recording a large database," *IEEE International Conf. on ICASSP'84,* vol. 9, pp. 324-327.

[6] T. K. sagisaka, Y. Katagiri, S. Morikawa, S. Watanabe, T. Kuwabara, H., "Construction of a large-scale Japanese speech database and its management system," *IEEE International Conf. on ICASSP'89,* vol. 1, pp. 560-563.

[7] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, and N. Dahlgren, "DARPA, TIMIT Acoustic-phonetic continuous speech corpus CD-ROM," NIST, 1990.

[8] B. Das, S Mandal and P. Mitra, "Bengali speech corpus for continuous automatic speech recognition system," *International Conference on Speech Database and assessments (Oriental COCOSDA),* October 2011, pp. 51-55

[9] G. Anumanchipalli, Rahul Chitturi, Sachin Joshi, Rohit Kumar, Satinder Singh, R.N.V Sitaram and S.P. Kishore, "Development of Indian Language Speech Databases for Large Vocabulary Speech Recognition Systems", *Proceedings of International Conference on Speech and Computer (SPECOM),* Patras, Greece, Oct 2005.

[10] Technology Development for Indian Languages, URL:http://tdil.mit.gov.in.

[11] Haris B. C., G. Pradhan, A. Mishra, S. Shukla, R. Sinha and S. R. M. Prasanna, "Multi-variability speech database for robust speaker recognition," *National Conference on Communication (NCC),* Bangalore, Jan 2011.

[12] Resource Centre for Indian Language Technology Solutions, IIT Guwahati, URL:http://www.iitg.ernet.in/rcilts/.

[13] U. Bhattacharjee, Kshirod Sarmah, "Development of a speech corpus for speaker verification research in multilingual environment," *IJSCE,* 2013, vol. 2(6), pp. 443-446.

[14] S. K. Riis and A. Krogh, "Hidden neural networks: A framework for HMM/NN hybrids," *Proc. ICASSP-97,* Munich, Germany.

[15] Linguistic Data Consortium. URL: http://www.ldc.upenn.edu/.

[16] European Language Resources Association. URL: http://www.icp.grenet.fr/ELRA/

[17] Oregon Graduate Institute. URL: http://cslu.cse.ogi.edu/

[18] B. R. Wildermoth and K. K. Paliwal, "GMM based speaker recognition on readily available databases," *Proc. Microelectronic Engineering Research Conf., 2003.*

[19] J. P. Campbell, Jr. and D. A. Reynolds, "Corpora for the evaluation of speaker recognition system," *Proceedings of ICASSP,* 1999, vol. 2, pp. 829-832.

[20] J. B. Millar, P. Dermody, J. M. Harrington and J.Vonwiller,, "The Australian national database of spoken language," *Proceedings of ICASSP,* 1994, vol. 1, pp. 97-101.

[21] G. R. Doddington, " CSR corpus development," *HLT'91 Proceedings of the Workshop on Speech and Natural Language,* pp. 363-366.

[22] G. R. Doddington, M. A. Przybocki, A. F. Martin and D. A. Reynolds, "The NIST speaker recognition evaluation overview: Methodology systems, results, perspective," *Speech Communication,* 2000, 31, pp. 225-254.

[23] H.A. Patil and T. K. Basu, "Development of speech copora for speaker recognition research and evaluation in Indian languages," *Int J Speech Techno*logy, 2008, Vol. 11, pp. 17-32.

[24] Adi Agom Kebang (AAK), URL: http://www.adiagomkebang.org/

[25] Sachin Roy, "*Aspects of Padam Minyong Culture*," North East Frontier Agency, 1960.

[26] Boersma, P., Weenink, D., "Praat: doing phonetics by computer", URL: http://www.praat.org, accessed in October 2012.

[27] J. Herbert Lorain, *"A Dictionary of the Abor-Miri Language,"* Mittal Publication, 1910.

