IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 2, Ver. V (Mar.-Apr. 2017), PP 84-89
www.iosrjournals.org

Automatic Identification of Spoken Language

Tanusree Sadhukhan[1] ,Shweta Bansal[2] ,Atul Kumar[3]
[1,2,3] *KIIT College of Engineering,Gurugram*

Abstract: *Spoken language identification is the process of mapping continuous speech to the language itbelongs to. Applications of spoken language identification include front-ends for multilingual speech recognition systems, retrieval of web data, automatic customer routing in call centers.This paper present the development of spoken language identification system for under resource Indian languages i.e Hindi and Manipuri.The system which has been designed is able to identify the languages with a great accuracy.*

I. Introduction

From the last two decades speech recognition has been conducted based on Hidden Markov Models (HMMs) and n-gram language models[3].Though they have limitations but also have given a successful result.In our day to day life spoken language identification system has a great importance.For our country where more than 200 languages are currently being spoken,the automatic language identification systems have great importance. The task of the system is to quickly and accurately identify the spoken language Hindi and Manipuri.Till now we have designed the system only for Manipuri and Hindi,but we are trying to incorpoprate the many other languages like Urdu,Punjabietc.We have found many differences in the features of Hindi and Manipuri languages.Hindi and Manipuri have different phone set. Some times they share common phones. But phoneme frequencies are different in both languages. Phonotactics are also different. Both the languages have its own vocabulary set. So The rules of the formation of words are also different.For Hindi,Khariboli dialect is considered here [1]and Bishnupriya dialect is chosen for Manipuri.In this paper ,The methodology of the system is discussed mainly in the following section .

DatabasePreperation,which includes Data Collection ,Speaker Selection and Data Recording.

Data Processingdescribes DataAnalysis and Model Building Processes.This language-ID system operates in two phases: training and recognition. During the training phase, the typical system is presented with examples of speech from a variety of languages.**Classification** performs the actual identification procedure.

II. Methodology

2.1 Database Preperation

It is the first step which we follow during language identification. It includes the following steps.

Data collection : We have taken 300 sentences for Hindi language and 300 sentences for Manipuri languages.These sentences are recorded by 20 various speakers.We have extracted 1000 unique words from hindi sentences and 1200 unique words from Manipuri sentences.

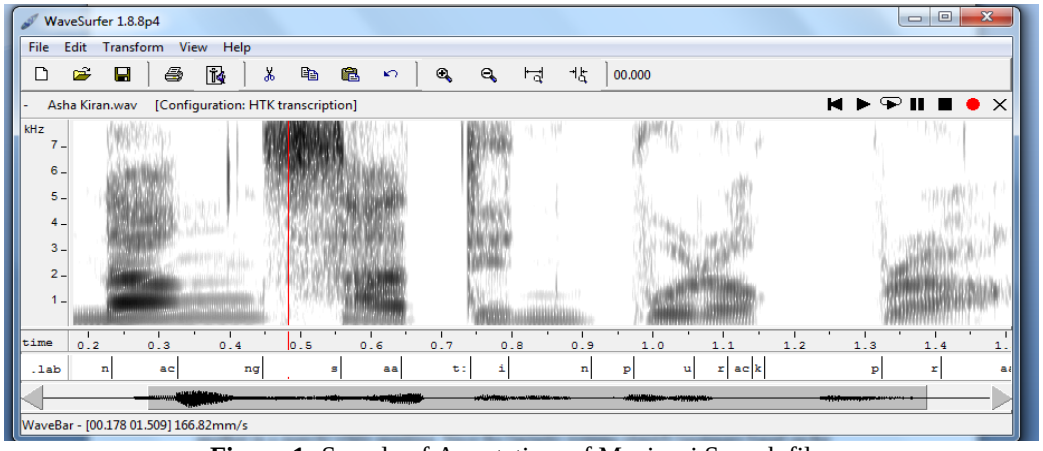

Figure 1: Sample of Annotation of Manipuri Speech file

Speaker Selection:20 Native speakers are selected for both the languages.Out of 20 speakers 10 were female and 10 were male.The age of the speakers are between 18 to 40.They didn't have any articulatory disorder.All of them have basic qualification atleast 10+2 level.

Data Recording: The sentences are recordedwith the help of unidirectional microphone and the distance between mouth and microphone was minimized (3-4 cm) during recording.Recording was done in studio environment and each word was recorded 10 times in separate file with .wav extension.All recording is done by SHRUE unidirectional mike.Goldwave is used for recording the speech sounds.Labeling is done with the help of wave surfer.

Annotation : Sentence are segmented in phoneme level.The above figure(Figure 2) represents the annotation of the Manipuri speech file :Nang Satin Purakpra ,i.e ***Nacng saat:in purackpraa.***

2.2Data Processing

Analog signal is converted into digital signal and then the digital data is processed to prepare dictionary and models.Here,various levels of speech features are extracted from the raw. The acoustic speech feature is a simple compact representation of the raw speech sound and can be modeled by the cepstral features such as Mel Frequency Cepstral Coefficient (MFCC) or Perceptual Linear Prediction (PLP) coefficient. The phonotactics refers the study of the permissible set of allowed sequences of speech sounds in a given language. The N-gram language model (LM) can be used to model the phonotactic features. These data are used for language identification.The flow of the process is descried below(Figure 2).[11][6]

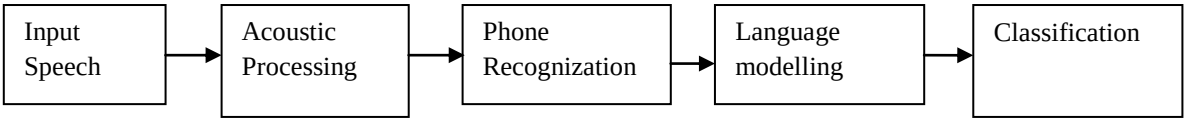

Figure 3: Language Identification Process

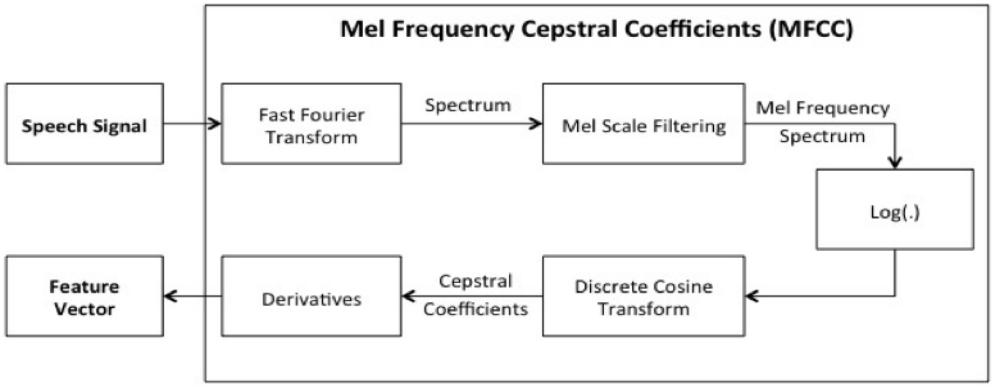

Figure 4:MFCC Process

Pre-processing

The speech signal is an analog signal; to process digitally, we have to convert it into digital form. Analogue to digital convertor is used for this function. After this, the signal has to be preprocessed. Signal preprocessing involves some crucial steps like Background Noise Elimination, Pre-Emphasis Filtering, Framing and Windowing.[7]

Feature Extraction

Feature extraction is the method of extracting the limited amount of useful information from high dimensional data.The foremost measure in any Automatic LanguageRecognition system is to extract features, i.e. identify the parts of the audio signal that are good for identifying the linguistic content and discarding all the other stuff which takes information like background noise, emotion and so on.The most commonly used feature extraction methods in automatic address recognition is Mel-Frequency Cepstral Coefficients (MFCC).Prior to the introduction of MFCCs, Linear Prediction Coefficients (LPCs) and Linear Prediction Cepstral Coefficients (LPCCs) were popular methods. [5] [6]

Feature extraction reduces the bandwidth from 16,000 samples per second (speech sampled at 16 kHz) to about 3,900 features per second (39 features per frame × 100 frames per second with 25 ms window size). Clearly, this step is crucial to the system, as any loss of useful information cannot be covered in later processing.The Input for the computation of the MFFCs is a speech signal in time domain.The process is described above(figure 3).

Pronunciation Dictationary

During identification, the succession of symbols generated by the acoustic element is compared with the set of words present in the lexicon to produce an optimal sequence of words that compose the system's final output.This lexicon (or dictionary) is used to provide the mapping between words and phones (or sub-word units). It holds information about which words are known to the system and also how these words are pronounced, i.e., what their phonetic representations look like. Here We have considered the Bengali script of Manipuri language.Some parts of Hindi and Manipuri dictionary are given below(Table1).As an example, Transcription of the word SCHOOL is same but the word BISCUIT is different in both the languages.

Table 1: Sample of Hindi and Manipuri Pronounciation Dictionary

Hindi	Transcription	**Manipuri**	Transcription
अगर	ac g ac r	Achaba	ac cp aa b aa
उसको	uc s k o	amu	aa m u
बिस्कुट	b ic s k uc t:	biscuit	b i s k u t:
घूमने	gh u m n e	damak	d: aa m ac k
जिसमें	j ic s m e	lampak	l ac m p aa k
हमले	h ac m l e	Phakse	ph ac k s e
स्कूल	s k u l	School	s k u l

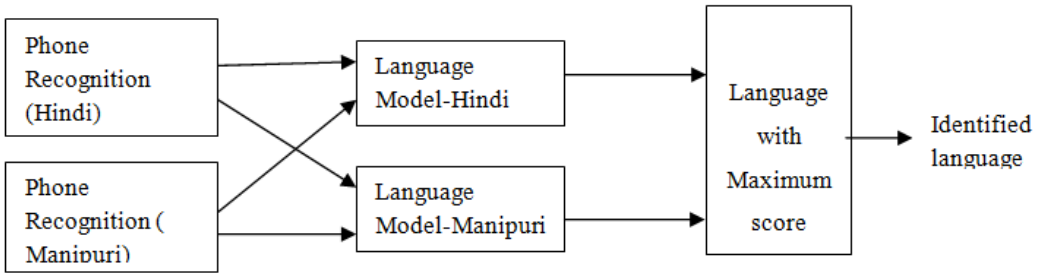

Figure 5 : Block diagram of the system(using PPRLM)

Language Model

Here,We have used PPRLM architecture (Parallel Phone Recognition followed by Language Model). In PPRLM system two parallel subsystems are made(one for Hindi and other for Manipuri language), where each subsystem consists of a phone recognizer with a different phone set for the particular language. The phone recognizer extracts phonotactic attributes from the speech input to characterize a language. These two parallel subsystems are made to capture the phonetic diversification available in the speech input. [8][12]

We have used two front-end phone recognisers instead of a single recogniser. Input speech is sent through both the phone recognizers where one is trained on Hindi language and another is trained on the Manipuri. For each recogniser, the phone strings are then scored with the help of Language models. The process is shown above(Figure 4).

The language model is combined with an acoustic model that models the pronunciation of different words.The acoustic model generates a large number of candidate sentences, together with probabilities; the language model is then used to reorder these possibilities based on how likely they are to be a sentence in the language.The systems use N -gram language models to lead the search for correct word sequence. Common feasible N-gram models are tri –gram.Hindi has almost 43 phonemes where Manipuri has less no of phoneme because mulitple phonemes (**like ছ শ স ষ is pronounciated as S only**) are represented as a single sound.Hindi and Manipuri phone sets are given below(Table 2).

Table 2: Common phone set for Hindi and Manipuri

Hindi Phoneme	**Common Phone Label**	**Hindi Phoneme**	**Common Phone Label**	**Manipuri Phoneme**	**Common Phone label**	**Manipuri Phoneme**	**Common Phone label**
अ	**ac**	**ढ**	**d:h**	**অ**	**ac**	**ণন**	**n**
आ	**aq**	**ण**	**n:**	**আ**	**aa**	**প**	**p**
औ	**a**	**त**	**t**	**ইঈ**	**i**	**ফ**	**ph**
इ	**i**	**थ**	**th**	**উঊ**	**u**	**ব**	**b**
ई	**ic**	**द**	**d**	**এ**	**e**	**ভ**	**bh**

উ	u	ধ	dh	ঐ	ae	ম	m
ঊ	uc	ন	n	ও	o	য়	y
এ	e	প	p	ঔ	ou	ড়়	r
ঐ	ae	ফ	ph			ল	l
ও	o	ব	b			হ	h
ক	k	ভ	bh	ক	k		
খ	kh	ম	m	খ	kh		
গ	g	য	y	গ	g		
ঘ	gh	র	r	ঘ	gh		
ঙ	ng	ল	l	ঙ	ng		
চ	c	ব	v	চ	cp		
ছ	ch	শ	sh	ছশসষ	s		
জ	j	স	s	জ	jp		
ঝ	jh	হ	h	ঝ	jph		
ঞ	nj	ড়়	z	টত	t:		
ট	t:			ঠথ	t:h		
ঠ	t:h			ডদ	d:		
ড	d:			ঢধ	d:h		

Acoustic Modelling

The systems perform phonetic tokenization followed by phonotactic analysis.

This process is divided into two parts – front end process and back end process.Phone Recognization is the front end process which is used to implemented using ***HMM****(Hidden Markov Model)* or ***GMM****(Gaussian Markov Model)*model.In back end process N-***gram language model*** is prepared, one for each language.[1][8]

Acoustic models are used to link the observed features of the speech signal with the expected phonetics of hypothesis sentence. Being the primary component of this system, acoustic model accounts for most of the computational load and performance of the organization.The most typical implementation of this process is probabilistic, making usage of Hidden Markov Models.To generate mapping between the basic speech units (phones, syllables) and the acoustic observations, a rigorous training procedure is adopted.A phonetically rich and balanced database is needed to develop the acoustic models.

2.3Classification

A Classifier is a portion of the speech recognizer which performs actual identification with the aid of trained acoustic and language models.The determination on the final transcription (recognized words) must be removed by combining and optimizing the information of acoustic and language models.We have taken 70% of the test data for for training and rest 30% for testing.

Step 1: Building Model on Trained Data

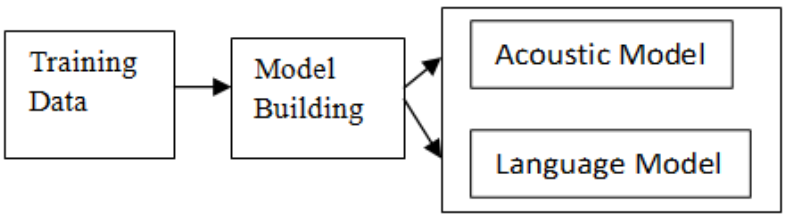

Figure 6: Block diagram of Training Process

Step2: Identification of spoken language using the models built on step1

Acoustic and Language model from

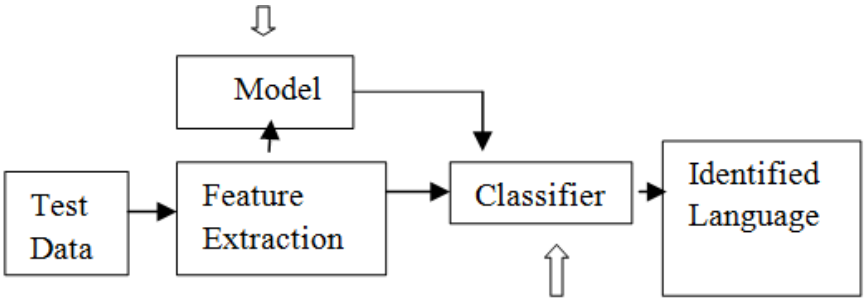

Figure 7: Block diagram of Testing Process

III. Result

The system has been tested using 30% of Hindi and Manipuri speech files.
We have got the result of 96.34% with great accuracy.

IV. Conclusion

In this report, we have shown the operation of integrated speech recognition and language recognition systems for the, Hindi and Manipuri.Furthermore, we have demonstrated that language identification performance can be improved by means of large training data set. Indian speakers speak the same native accents to speak any language. For example, even if a Manipuri speaker speaks English, there is a chance of falsely identifying it as a Manipuri due to his accents. Acoustic feature systems differentiate each language based on the physical sound patterns used to speak languages. Also, there is a chance of occurring code mixing (mixing of two or more languages in a speech) and code switching (moving from one language to another in a speech) while speaking. Since the fantastic systems identify languages based on the frequency of occurrence of phone sequence or a subset of phone sequence, it results better performance than the acoustic systems. We have mainlychosen native speakers for the recording of Hindi and Manipuri sentences. A collection of the robust database is the main supporting aid in designing a good LID systemThis system will be used to be demonstrated to give a superior performance [9].The focus during experimental test runs and evaluation is on the suitability for use in real-time identification scenarios.

Acknowledgments

I would like to thanksDr. S.S. Aggarwal, Dr. Sudhir KumarSharma for their kind support and valuable suggestions.

References

[1]. Comparison of :Four Approaches to Automatic Language Identification of Telephone Speech,IMarc A. Zissman, Member, IEEE.[1, JANUARY 1996]
[2]. AUTOMATIC SPOKEN LANGUAGE IDENTIFICATION, Liang Wang [2008]
[3]. Automatic Language dentification for Natural Speech Processing Systems Student Research Paper of Michael Heck At the Department of Informatics Institute of Anthropomatics (IFA) Interactive Systems Laboratories (ISL) Supervisors: Prof. Dr. Alex WaibelDr. Sebastian Stüker Duration: 01. June 2011 – 01. September 2011.
[4]. A Vector Space Modeling Approach to Spoken Language Identification Haizhou Li, Senior Member, IEEE, Bin Ma, Senior Member, IEEE, and Chin-Hui Lee, Fellow, IEEE
[5]. http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/
[6]. https://www.ce.yildiz.edu.tr/personal/fkarabiber/file/19791/BLM5122_LN6_MFCC.pdf
[7]. Continuous Hindi Speech Recognition using Monophone based Acoustic Modeling Ankit Kumar, Department of Computer Engg National Institute of Tech Kurukshetra, India ,MohitDua, Department of Computer Engg National Institute of Tech Kurukshetra, India ,TriptiChoudhary, Department of Elect &CommVishveshwarya Inst. of Tech Greater Noida, India-2014
[8]. Using Phone Recognition and Language Modelling (PRLM) for Automatic Language Identification by UrosRapajic Supervisor: Dr. P. Naylor
[9]. Spoken Language Recognition StanisławKacprzak 27.03.2014, Kraków, Seminarium DSP Based on, Spoken Language Recognition: From Fundamentals to Practice" Haizhou Li; Bin Ma; Kong Aik Le.
[10]. Comparison of four approaches to automatic language identification of telephone speech. IEEE Transactions on Speech and Audio Processing, 4(1), 31-34 Article in IEEE Transactions on Speech and Audio Processing · February 1996 DOI: 10.1109/TSA.1996.481450 · Source: IEEE Xplore CITATIONS 425
[11]. Automatic Spoken Language Identification Utilizing Acoustic and Phonetic Speech Information by Kim-Yung Eddie Wong, BEng(Hons), BIT, Speech and Audio Research Laboratory School of Electrical & Electronic Systems Engineering June 2004
[12]. Automatic speech recognition for under-resourced languages: A survey by Laurent Besaciera, Etienne Barnardb, Alexey Karpovc, Tanja Schultzd, January 2014.
[13]. Marc A. Zissman, Kay M. Berkling," AUTOMATIC LANGUAGE IDENTIFICATION",Speech Communication,35(1-2):115-124,Aug 2001.
[14]. K. Kirchoff,S.Parandekar,and J.Bilmea,"Mixed Memory Markov Models for Automatic Language Identification",ICASSP 2002,Vol.1,pp.761-764,2002.
[15]. Hossan M.A., A novel approach for MFCC feature extraction,IEEE Xplore,13-15 Dec,2010.

[16]. Shikha Gupta 1,Jafreezal Jaafar 2,Wan Fatimah wan Ahmad 3 and Arpit Bansal 4,"FEATURE EXTRACTION USING MFCC",Signal & Image Processing:An International Journal(SIPIJ) Vol.4,No 4,August 2013.
[17]. H. Li,B.Ma, and C.H Lee,"A Vector Space Modeling Approach to Spoken Language Identification",IEEE transaction on Audio,Speech and Language Processing,Vol. 15. No. 1,pp.271-284,2007.
[18]. Y. K. Muthusamy, E. Barnard and R. A. Cole, "Reviewing automatic language identification," in IEEE Signal Processing Magazine, vol. 11, no. 4, pp. 33-41, Oct. 1994.
[19]. L. R. Bahl, F. Jelinek, R. L. Mercer, "A maximum likelihood approach to continuous speech recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 5, pp. 179-190, 1983.
[20]. T. J. Hazen, V. W. Zuc, "Automatic language identification using a segment-based approach", In Proceedings 3rd European Conference on Speech Communication and Technology (Eurospeech 93), 1993-September.
[21]. T. J. Hazen, V. W. Zuc, "Recent improvements in an approach to segment-based automatic language identification", Proceedings International Conference on Spoken Language Processing 94, 1994-September.
[22]. M. A. Zissman, "Automatic language identification using gaussian mixture and hidden markov models", Proceedings IEEE International Conference on Acoustics Speech and Signal Processing 93, 1993-April.
[23]. M. A. Zissman, E. Singer, "Language identification using phonetic class recognition and N-gram analysis", Speech Research Symposium XIII, pp. 400-409, 1993-June.
[24]. Y. K. Muthusamy, K. M. Berkling, T. Arai, R. A. Cole, E. Barnard, "A comparison of approaches to automatic language identification using telephone speech", Proceedings 3rd European Conference on Speech Communication and Technology (Eurospeech 93), 1993-September.
[25]. M. A. Zissman, E. Singer, "Automatic language identification of telephone speech messages using phoneme recognition and N-gram modeling", Proceedings IEEE International Conference on Acoustics Speech and Signal Processing 94, pp. I-305-I-308, 1994-April.
[26]. Y. K. Muthusamy, R. A. Cole, "Automatic segmentation and identification of ten languages using telephone speech", Proceedings International Conference on Spoken Language Processing 92, 1992-October.
[27]. L. F. Lamel, J-L. S. Gauvain, "Language identification using phone-based acoustic likelihoods", Proceedings IEEE International Conference on Acoustics Speech and Signal Processing 94, pp. I-293-I-296, 1994-April.