International Journal of Innovative Research in Engineering & Management (IJIREM)
ISSN: 2350-0557, Volume-8, Issue-6, November 2021
https://doi.org/10.55524/ijirem.2021.8.6.109
www.ijirem.org

Automatic Language Identification of Spoken Hindi and Manipuri

Dr. Bhupesh Ghai
RIMT University, Mandi Gobindgarh, Punjab, India
Email Id- bhupeshghai@rimt.ac.in

ABSTRACT

The technique of mapping continuous speech to the language it belongs to is known as spoken language identification. Front-ends for bilingual voice recognition frameworks, online data retrieval, and automated customer handling in customer support are all examples of spoken language identification applications. The creation of a spoken language recognition system for under-resourced languages of India such as Hindi and Manipuri is presented in this research work. The technology that has been developed is capable of accurately identifying languages. This technology will be used to demonstrate that it is capable of producing superior outcomes. Experimental test runs and evaluation are used to determine if the system is suitable for use in real-time identification scenarios.

Keywords

Automatic Language Identification, Feature Extraction, MFCC, PPRLM, Spoken Language Recognition.

1. INTRODUCTION

Hidden Markov models (HMMs) and N-gram language models have been used in voice recognition over the last two decades [1,2]. They have limits, yet they have also produced a good outcome. The significance of a spoken language detection algorithm in our daily lives cannot be overstated. Automated recognition systems are critical in our nation, since more than 190 languages are spoken at any one time. AISL's job is to rapidly and correctly identify Hindi and Manipuri as spoken languages. We have only developed the system for Hindi and Manipuri so far, but we are working on adding Urdu, Punjabi, and other languages. Many distinctions between the Hindi and Manipuri languages have been discovered. The phonetic alphabets in Hindi and Manipuri are distinct. They may use the same phone. However, the frequency of phonemes in each languages are different (3,4). Phonotactics differs as well. Each language has its unique collection of words. As a result, the rules for word construction are likewise varied. The dialect of For Hindi, Khariboli, is taken into account here. The technique of the method is mostly covered in the next section of this research paper [5–9].

- *Preparing the Database:*
  Collection of data, selection of speaker, and data recording are all included in data preparation.
- *Processing of Information:*
  It covers procedures such as data analysis and model development. The training and recognition stages of these language-ID systems work in tandem. The usual system is provided with samples of speech from a range of languages throughout the training phase.
- *Classification:*
  It is in charge of the actual identifying process.

2. METHODOLOGY

2.1. Database Preperation

It is the initial stage in the language identification process. It consists of the stages listed below.

2.1.1. Data Assortment

We took 200 sentences for the Hindi language and 200 sentences for the Manipuri language. The sentences were then recorded by 50 different people. From Hindi sentences, we selected 1000 unique words, and from Manipuri phrases, we retrieved 1100 unique words.

2.1.2. Speaker Selection

For both languages, fifty native speakers were chosen, twenty of whom were female while 30 of them were male. The speakers ranged in age from 18 through 40, and none of them had any articulatory issues. Most of the individuals have had at least a 10+2 level of education.

2.1.3. Audio Recording

The phrases were captured using a unidirectional mic with the space between lips and mic kept to a minimum [3-4]. The recording took place in a studio setting, and each syllable was recorded ten times in a separate.wav file. The SHRUE unidirectional microphone was used to record everything. With the assistance of a Wavesurfer, the labeling is completed and the voice samples are recorded using Goldwave [10–14].

2.1.4. Annotation

Phoneme-level segmentation is used to segment sentences. The annotations of the Manipuri audio file is shown in the diagram below in Fig 1.

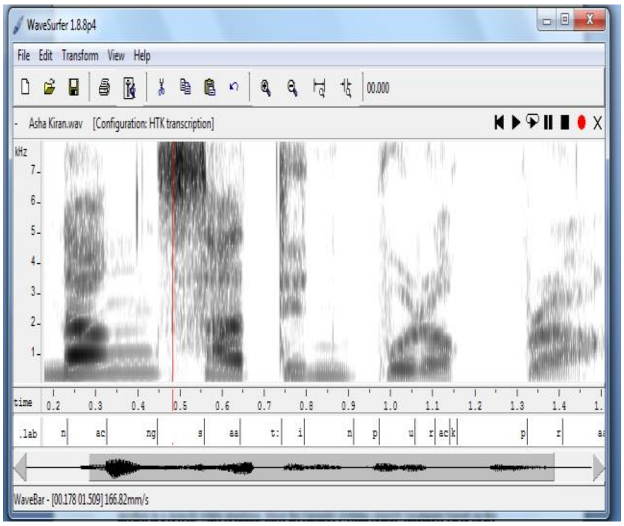

WaveSurfer 1.8.8p4
File Edit Transform View Help
00.000
Asha Kiran.wav [Configuration: HTK transcription]
kHz
7
6
5
4
3
2
1
time 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.
.lab n ac ng s aa t: i n p u r ac k p r aa
WaveBar - [00.178 01.509] 166.82mm/s

Figure 1: Sample of Manipuri Speech Annotation using Wavesurfer v1.8

2.2.Data Processing

The analog signal is transformed to digital data, which is subsequently processed to create dictionaries and models. From the raw, different layers of speech characteristics are retrieved. The acoustic speech feature is a concise description of the actual spoken voice that may be modeled using cepstral features like the Perceptual Linear Prediction (PLP) or the Mel Frequency Cepstral Coefficient (MFCC). The study of the acceptable set of authorized sequences of spoken sounds in a particular language is referred to as phonotactics. The phonotactic characteristics may be modeled using the N-gram language model (LM). These details are used to identify languages [15–19]. The process's flow is shown above in Fig. 2.

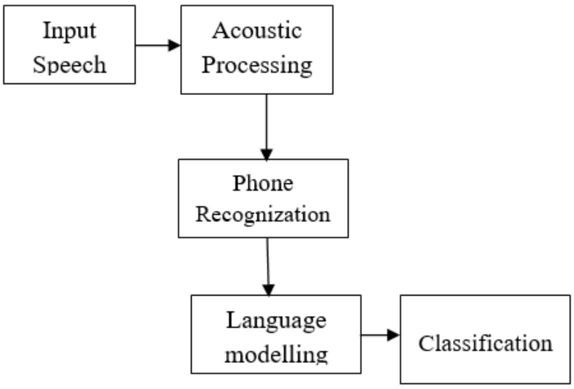


Figure 2: The above figure discloses a general Language Identification Process

2.2.1. Pre-processing

We must transform the analog voice stream into digital form in order to process it digitally. This function is performed using an analog to digital converter. Following that, the signal must be pre-processed. Ambient sound reduction, pre-emphasis filtration, framing, and windowing are all important stages in signal pre-processing.

2.2.2. Feature Identification & Retrieval

Feature extraction is the process of deriving a small percentage of usable data from a large quantity of information collected. It is the first step in any automatic language identification system to harvest features, which means identifying the sections of the voice signal that really are helpful for assessing linguistic content and jettisoning everything else, such as background noise, emoticons, and so on. The MFCC technique as depicted in Fig. 3, is one of the most widely utilized feature extraction algorithms in automated language recognition [20,21]. LPCs and LPCCs were prominent techniques prior to the advent of MFCCs.

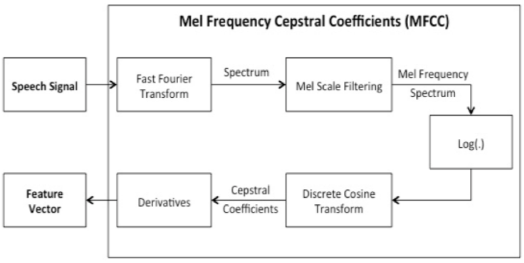


Figure 3: MFCC Process is shown in this figure. Here speech signal is feed initially to the system which generates feature vectors as output

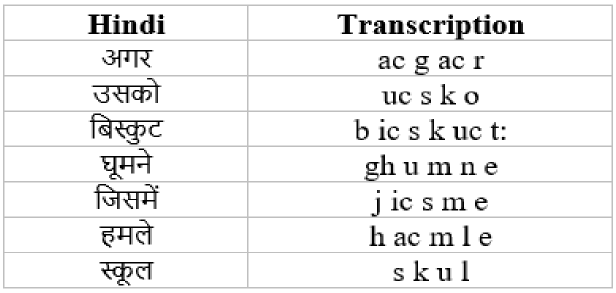
Feature extraction lowers the bandwidth from 16,000 samples per second, or 16 kHz speech, to approximately 3,900 features per second, or 39 features per second and 100 frames per second with a 25ms sliding window. This phase is clearly critical to the algorithm, since any loss of valuable information cannot be compensated for later. The MFFCs are computed using a time-domain speech stream as input.

3. RESULTS AND DISCUSSION

3.1. Pronunciation Dictationary

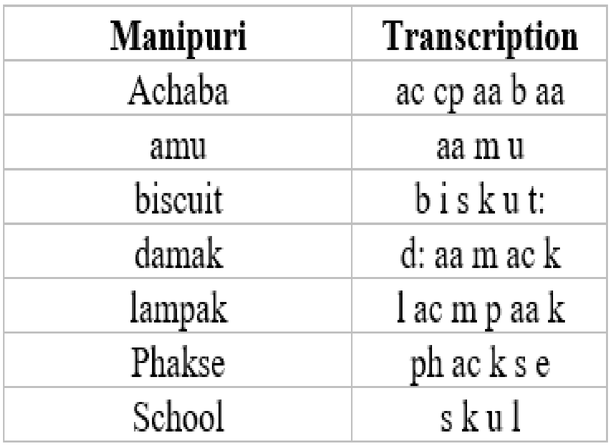
Throughout recognition, the acoustic element's sequence of symbols is analyzed to the bunch of words in the lexicon to create an optimum sequencing of words that make up the system's final outcome[22]. The linkage among words and phones or sub-word components is provided by this dictionary or lexicon as clearly can be seen in Table 1 and Table 2. It stores information as to which words the algorithm recognizes as well as how those words are spoken, or even what respective phonetic equivalents seem like. The meitei script of the Manipuri language has been examined here. Below are excerpts from the Hindi and Manipuri dictionaries. The term "Ball", for example, has the same transcription in both languages, while the word "Snack" has a distinct one.

Table 1: Sample of Hindi Pronounciation Dictionary



Hindi	Transcription
अगर	ac g ac r
उसको	uc s k o
बिस्कुट	b ic s k uc t:
घूमने	gh u m n e
जिसमें	j ic s m e
हमले	h ac m l e
स्कूल	s k u l

Table 2: Sample of Manipuri Pronounciation Dictionary



Manipuri	Transcription
Achaba	ac cp aa b aa
amu	aa m u
biscuit	b i s k u t:
damak	d: aa m ac k
lampak	l ac m p aa k
Phakse	ph ac k s e
School	s k u l

3.2. Language Model

We've utilized the PPRLM framework in this case, which stands for Parallel Phone Recognition followed by Language Model and is illustrated in Fig. 4[23]. Two parallel sub-systems one for each Hindi and Manipuri language are created in the PPRLM system, each consisting of a phone recognition system with a distinct phone set for the respective language. To describe a language, the phone recognizer collects phonotactic characteristics from the voice input. The purpose of these two simultaneous subsystems is to capture the phonetic diversity present in the voice input.

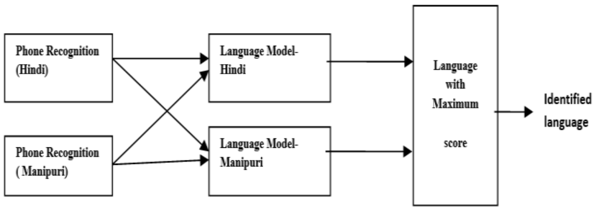

Phone Recognition (Hindi)
Phone Recognition (Manipuri)
Language Model-Hindi
Language Model-Manipuri
Language with Maximum score
Identified language

Figure 4: Block diagram of the system (using PPRLM)

Table 3: Common phone set for Hindi and Manipuri

Hindi Phoneme	Common Phone	Hindi Phoneme	Common Phone	Manipuri Phoneme	Common Phone	Manipuri Phoneme	Common Phone
अ	ac	ढ	d:h	অ	ac	ণ ন	n
आ	aq	ण	n:	আ	aa	প	p
औ	a	त	t	ই ঈ	i	ফ	ph
इ	i	थ	th	উ ঊ	u	ব	b
ई	ic	द	d	এ	e	ভ	bh
उ	u	ध	dh	ঐ	ae	ম	m
ऊ	uc	न	n	ও	o	য়	y
ए	e	प	p	ঔ	ou	র ড়	r
ऐ	ae	फ	ph			ল	l
ओ	o	ब	b			হ	h
क	k	भ	bh	ক	k		
ख	kh	म	m	খ	kh		
ग	g	य	y	গ	g		
घ	gh	र	r	ঘ	gh		
ङ	ng	ल	l	ঙ	ng		
च	c	व	v	চ	cp		
छ	ch	श	sh	ছ শ স ষ	s		
ज	j	स	s	জ	jp		
झ	jh	ह	h	ঝ	jph		
ञ	nj	ज़	z	ট ত	t:		
ट	t:			ঠ থ	t:h		
ठ	t:h			ড দ	d:		
ड	d:			ঢ ধ	d:h		

Instead of using only one front-end phone recognizer, we utilized two. The input voice is transmitted to both phone recognizers, one of which is based on the Hindi language while the other on the Meitei language. The phone sequences are then rated using Language models for each recognizer. The linguistic model is coupled with an audio model that simulates various word pronunciations. The acoustic model produces a huge set of possible utterances with probabilities, which are then reordered by the language model depending on how probable those were to become a statement in the language. To guide the search for the right word sequence, the systems utilize N-gram language models. The tri-gram model is a common N-gram model that is practical. Because several phonemes such as ছ, শ, স & ষ , which is only pronounciated as S, are expressed as a single voice in Hindi, Manipuri has fewer phonemes. Above in Table 3 are phone sets in Hindi and Manipuri.

3.3. Acoustic Modelling

Phonetic tokenization is proceeded by phonetic evaluation in the approache. This procedure is split into two parts: the front-end and the back-end. Phone recognition is a front-end procedure that uses the Hidden Markov Model (HMM) or Gaussian Markov Model (GMM) to implement modelling [24]. A N-gram linguistic model is created in the back-end process, one for each language.

The measured characteristics of the speech waveform are linked to the anticipated pronunciation of the hypothesis phrase using acoustic modelling. The acoustic model, as the system's main component, is responsible for the majority of the foundation's computation complexity and effectiveness. The stochastic approach of this procedure, which employs HMM, is the most common. A intensive training process is used to create mapping between both the fundamental speech entities i.e., phones or syllables and the acoustic measurements. To build the acoustic simulations, a phonetically diverse and neutral database is required.

3.4. Classification

A Classifier is a component of a speech recognizer that uses learned acoustic and linguistic models to accomplish real identification[25]. By integrating and improving the information from acoustic and language models, the final transcription i.e., identified words, must be eliminated. Seventy percent of total of the dataset was used for training, while the remaining thirty percent was used for testing. Fig. 5 and Fig. 6 shows the process of training and testing respectively.

3.5. Building Model on Trained Data

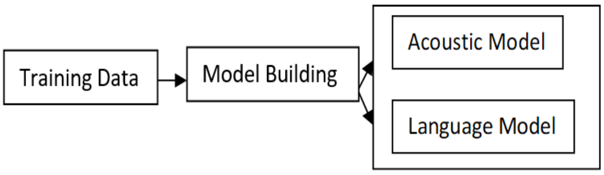


Figure 5: Illustrates the training process flow. Here, Language and Acoustic models are generated from the training datasets

3.6. Identification of Spoken Language Using the Models

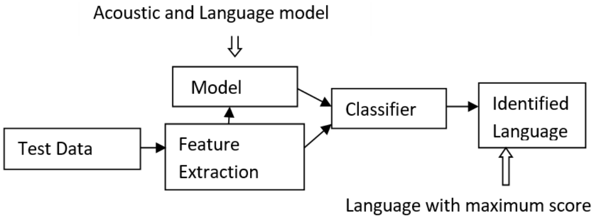


Figure 6: Illustrates the testing process. Here, the test data is compared with the previous generated model and the language is identified based on the methodology discussed in previous section

Thirty percent of Hindi and Manipuri voice files were used to test the system and with an excellent precision, we were able to get a result of 95 percent.

4. CONCLUSION

We showed the functioning of combined voice recognition and language identification techniques for the Hindi and Manipuri languages in this paper, as well as how huge training datasets may enhance language identification effectiveness is also discussed. In most instances, Indian speakers have the same indigenous accents while speaking any language. For illustration, even though a Manipuri person talks English, owing to his accents, it is possible to misidentify him as a Manipuri. Acoustic feature algorithms distinguish each dialect depending on the actual sound patterns that are utilized to communicate. Additionally, code mixing i.e., the blending of two different languages in a speech, and code switching i.e., the change of one language to the other in an utterance during speaking are possible. This system outperform acoustic systems because they detect languages depending on the rate of occurrence of phone sequences or a selection of phone sequences. This technology will be utilized to show that it can provide better results. The appropriateness for application in real-time identification situations is the subject of experimental test runs and assessment.

REFERENCES

[1] Karan B, Sahoo J, Sahu PK. Automatic speech recognition based Odia system. In: 2015 International Conference on Microwave, Optical and Communication Engineering, ICMOCE 2015. 2016.

[2] Upadhyaya P, Farooq O, Abidi MR, Varshney YV. Continuous Hindi speech recognition model based on Kaldi ASR toolkit. In: Proceedings of the 2017 International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2017. 2018.

[3] Bansal S, Sharan S, Agrawal SS. Corpus design and development of an annotated speech database for Punjabi. In: 2015 International Conference Oriental COCOSDA held jointly with 2015 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE). IEEE; 2015. p. 32–7.

[4] Sinha S, Sharan S, Agrawal SS. O-MARC: A multilingual online speech data acquisition for Indian languages. In: 2017 20th Conference of the Oriental Chapter of the International Coordinating Committee on Speech Databases and Speech I/O Systems and Assessment (O-COCOSDA). IEEE; 2017. p. 1–6.

[5] Bhatia R, Wadhawa D, Gurtu G, Gaur J, Gupta D. Methodologies for the synthesis of pentacene and its derivatives. Journal of Saudi Chemical Society. 2019.

[6] Shabbir M. Textiles and clothing: Environmental concerns and solutions. Textiles and Clothing: Environmental Concerns and Solutions. 2019.

[7] Shabbir M, Naim M. Introduction to textiles and the environment. Textiles and Clothing: Environmental Concerns and Solutions. 2019.

[8] Sharma S, Hussain MS, Agarwal NB, Bhurani D, Khan MA, Ahmad Ansari MA. Efficacy of sirolimus for treatment of autoimmune lymphoproliferative syndrome: a systematic review of open label clinical studies. Expert Opinion on Orphan Drugs. 2021.

[9] Hussain S, Singh A, Zameer S, Jamali MC, Baxi H, Rahman SO, et al. No association between proton pump inhibitor use and risk of dementia: Evidence from a meta-analysis. J Gastroenterol Hepatol. 2020;

[10] Hussain S, Singh A, Habib A, Hussain MS, Najmi AK. Comment on: "Cost Effectiveness of Dialysis Modalities: A Systematic Review of Economic Evaluations." Applied Health Economics and Health Policy. 2019.

[11] Kumar N, Singh A, Sharma DK, Kishore K. Novel Target Sites for Drug Screening: A Special Reference to Cancer, Rheumatoid Arthritis and Parkinson's Disease. Curr Signal Transduct Ther. 2018;

[12] Goswami G, Goswami PK. Artificial Intelligence based PV-Fed Shunt Active Power Filter for IOT Applications. In: Proceedings of the 2020 9th International Conference on System Modeling and Advancement in Research Trends, SMART 2020. 2020.

[13] Yadav CS, Yadav M, Yadav PSS, Kumar R, Yadav S, Yadav KS. Effect of Normalisation for Gender Identification. In: Lecture Notes in Electrical Engineering. 2021.

[14] Thappa S, Chauhan A, Anand Y, Anand S. Thermal and geometrical assessment of parabolic trough collector-mounted double-evacuated receiver tube system. Clean Technol Environ Policy. 2021;

[15] Solanki MS, Goswami L, Sharma KP, Sikka R. Automatic Detection of Temples in consumer Images using histogram of Gradient. In: Proceedings of 2019 International Conference on Computational Intelligence and Knowledge Economy, ICCIKE 2019. 2019.

[16] Anand V. Photovoltaic actuated induction motor for driving electric vehicle. Int J Eng Adv Technol. 2019;

[17] Singh D. Robust controlling of thermal mixing procedure by means of sliding type controlling. Int J Eng Adv Technol. 2019;

[18] Pandey B, Sharma KP. Radar Transmogrification Technology: Support for Unmanned System. In: Proceedings - 2019 Amity International Conference on Artificial Intelligence, AICAI 2019. 2019.

[19] Chauhan A, Tyagi V V., Sawhney A, Anand S. Comparative enviro-economic assessment and thermal optimization of two distinctly designed and experimentally validated PV/T collectors. J Therm Anal Calorim. 2021;

[20] Kuamr A, Dua M, Choudhary T. Continuous Hindi speech recognition using Gaussian mixture HMM. In: 2014 IEEE Students' Conference on Electrical, Electronics and Computer Science. IEEE; 2014. p. 1–5.

[21] Sharan S, Bansal S, Agrawal SS. Speaker-Independent Recognition System for Continuous Hindi Speech Using Probabilistic Model. In: Advances in Intelligent Systems and Computing. 2018. p. 91–7.

[22] Le VB, Besacier L. Comparison of acoustic modeling techniques for Vietnamese and Khmer ASR. Proc Annu Conf Int Speech Commun Assoc INTERSPEECH. 2006;1(June 2014):129–32.

[23] Sarmah K, Bhattacharjee U. GMM based Language Identification using MFCC and SDC Features. Int J Comput Appl. 2014;

[24] Kumar K, Aggarwal RK, Jain A. A Hindi speech recognition system for connected words using HTK. Int J Comput Syst Eng. 2012;

[25] Córdoba R, San-Segundo R, Macías J, Montero JM, Barra R, D'Haro LF, et al. Integration of acoustic information and PPRLM scores in a multiple-Gaussian classifier for Language Identification. In: IEEE Odyssey 2006: Workshop on Speaker and Language Recognition. 2006.