International Journal on Advances in Engineering Technology and Science
Volume: 1, Issue: 2, December 2015, e-ISSN: 2455-3131

Speech Conversion to Devanagari Script

Sabu Kamini

Electronics and Telecommunication Department
Govt. College of Engg. Jalgaon
kaminimsabu@gmail.com

***Abstract* — Speech to text conversion is a kind of speech recognition, where in the speech sound is written down in text form. This paper deals with conversion of speech into devanagari script. Speech to text conversion will help easy and faster recording of speech signal. It will also help communication with deaf people. This paper proposes use of phonetic model to convert speech to devanagari script. MFCC features are used for feature extraction.**

***Keywords* — speech recognition, speech-to-text conversion, devanagari script**

I. Introduction

Speech is commonly used for communication between two persons. Sometimes it is necessary to convert the speech into text for recording purposes. Speech to text conversion is also helpful for communication with deaf people who can't hear but can read the spoken text.

Since the accent of every person is different and dependent on his surroundings, it is many times difficult to understand what is being spoken. Many videos are thus attached with subtitles.

Speech to text conversion is the process of converting speech into text using some computer program. It is also termed as speech recognition though later has some broader sense. Speech recognition refers to understanding the speech signal while speech to text conversion has nothing to do with speech understanding.

Speech to text conversion is very much useful for recording what is being spoken. Human writing speed is very slow and hence if computer can write what is being spoken, the speed of human work will be more.

Again speech to text conversion may be used further for speech recognition, where computer can actually decode the meaning and perform the desired task based on the spoken instructions.

There are a lot of systems in use for speech to text conversion for English language. This paper deals with the speech conversion to devanagari text.

Section 2 introduces devanagari script and its corresponding phonemes. Section 3 describes the basic architecture of speech recognition system. Section 4 gives concluding remarks.

II. Devanagari Script Pronunciation Fundamentals

According to [6], Devnagari script is a script of phonemes arranged in a well structured scientific manner showing unambiguous classification and grouping of phonemes according to the organs used in producing that sound.

In devanagari, the letter order is based on phones. The manner and position of occurrences of combination of consonants and vowels defines its phone. Each letter has its unique pronunciation which can't be imitated by any other choice of letters.

The relation between a devanagari character and its phones is shown in Fig.1.

Devanagari script thus gives a unique representation for each and every word spoken by human irrespective of who is speaking and the context of speech.

This feature may help speech-to-text conversion while dealing with Indian languages. This feature will save us from the need of large vocabulary storage in the system.

अ	आ	इ	ई	उ	ऊ	ए	ऐ	ओ	औ
a	$aː$	i	$iː$	u	$uː$	e	$eː$	o	$oː$
a	A	i	I	u	U	e	E	o	O

क	ख	ग	घ	ङ
k	k^h	g	g^h	$ŋ$
k	kh	g	gh	gñ
च	छ	ज	झ	ञ
$tʃ$	$tʃ^h$	$dʒ$	$dʒ^h$	$ɲ$
c	ch	j	jh	jñ
ट	ठ	ड	ढ	ण
$ʈ$	$ʈ^h$	$ɖ$	$ɖ^h$	$ɳ$
T	Th	D	Dh	N
त	थ	द	ध	न
t	t^h	d	d^h	n
t	th	d	dh	n
प	फ	ब	भ	म
p	p^h	b	b^h	m
p	ph	b	bh	m

य	र	ल	व	श	ष	स	ह
j	r	l	$ʋ$	$ʃ$	$ʂ$	s	h
y	r	l	w	s̃	S	s	h

Fig.1 Hindi Alphabet Phoneme Relation as in [2]

In [2], different phonetic and acoustic features of devanagari and its pronunciation are described.

Ref. [3] summarizes various research works on speech recognition in Indian languages.

III. Speech Recognition Architecture

Speech recognition is a type of pattern recognition. Therefore, the overall system consists of two phases – training and testing.

According to [1], speech recognition process requires following steps – voice recording, word boundary detection, feature extraction and recognition with the help of models.

Initially, voice recording is performed for different speakers. Voice recording involves recording different sound (speech) signals for making database/knowledge models.

© 2015, IJAETS

Word boundary detection involves detecting the start and end of spoken word. Normally, duration of pause helps identify the start and end of word.

Feature extraction involves identifying various characteristics of the sound signal like amplitude and energy. These features form the training parameters and are used for class formation during training. Testing phase involves matching the input signal parameters with the class model parameters.

Features of the input sound signal are compared with the features of database in knowledge models and recognition is performed. Three important types of models are used in speech recognition.

A. Language Models

Language Models use the context of word usage for modelling. They also take into account the probability of occurrence of a specific word.

B. Acoustic Models

Phonetic, acoustic or language models are prepared and used as knowledge models. Acoustic model is of two types – word model or phone model.

Word model involves modelling each word as a whole. Properties of each word form a new class. Each word is modelled separately during training; during recognition, on the other hand, the matching is also performed based on word-by-word basis.

In phone model, instead of modelling each word separately, part of word is modelled based on phones. Word is recognised as a sequence of phones.

Hidden Markov Model is the most popular acoustic model.

Fig. 2 indicates typical system architecture for speech recognition system. Feature extraction is initially performed on number of voice recordings to yield various knowledge models during training phase. During testing phase, heard sound signal is given to the system as input. The same feature extraction procedure is carried out and the features are matched. Recognised speech is obtained by arranging the estimated words in sequence.

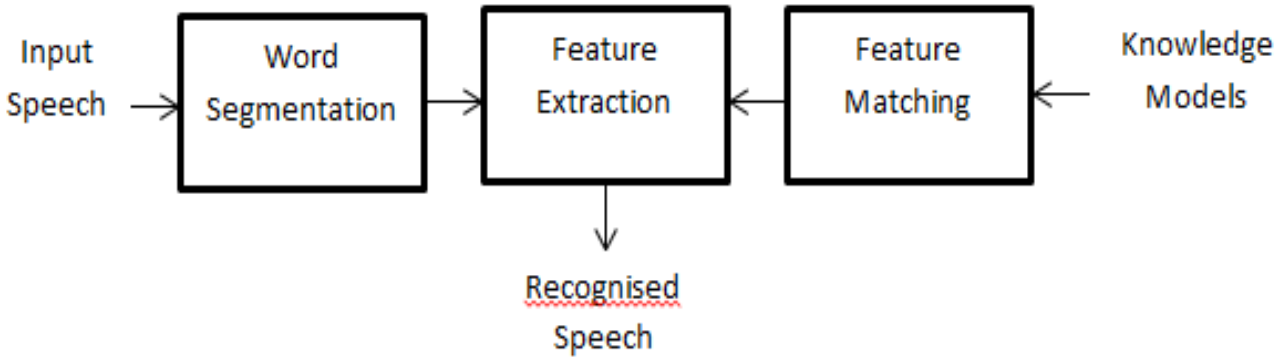

Fig. 2 A Simple Speech Recognition System Architecture

Ref. [5] enlists number of features used for feature extraction. Two approaches are preferred for feature extraction viz. Mel Frequency Cepstral Coefficient (MFCC) and Linear Predictive Coding (LPC).

A. Mel Frequency Cepstral Coefficient (MFCC)

MFCCs are commonly used as features for speech recognition. These coefficients are obtained from a cepstrum where frequency bands are spaced on mel scale – the scale of human perception for voice.

B. LPC Cepstral Coefficients

Linear Predictive Coding is a method normally used for predicting future samples. LPC Cepstral Coefficients are obtained from LPC parameters which are used as features for speech recognition.

IV. CONCLUSIONS

Normally, Indian languages are phonetic languages. Their pronunciation clearly indicates how they are spelt. Therefore, phonetic models form the best modeling for speech to text conversion as far as languages with devanagari script are considered.

The proposed system for speech to devanagari text conversion is based on phonetic modeling. Feature extraction uses Mel Frequency Cepstral Coefficients.

REFERENCES

[1] Neema Mishra, Urmila Shrawankar, Dr. V. M Thakare, “An Overview of Hindi Speech Recognition”, in Proceedings of the International Conference Computational Systems and Communication Technology, 2010, paper.

[2] Rohini Shinde, V.P. Pavar, “A Review on Acoustic Phonetic Approach for Marathi Speech Recognition”, International Journal of Computer Applications, Vol. 59(2), pp. 40-44, Dec.2012

[3] Cini Kurian, “A Survey on Speech Recogntion in Indian Languages”, International Journal of Computer Science and Information Technologies, Vol. 5 (5) , pp.6169-6175, 2014

[4] Susane Wagner, “Intralingual speech-to-text-conversion in real-time: Challenges and Opportunities”, MuTra 2005 – Challenges of Multidimensional Translation: Conference Proceedings, pp. 1-3,2005

[5] Urmila Shrawankar, Dr. Vilas Thakare, Techniques for Feature Extraction In Speech Recognition System : A Comparative Study”

[6] D. S. Shete, S.B. Patil, “Devnagari Phonetic Speech Analysis”, IOSR Journal of VLSI and Signal Processing (IOSR-JVSP),Vol. 3(2), pp. 62-66, Sept-Oct,2013.