Hindi Speaking Person Identification using Zero Crossing Rate

Arif Ullah Khan, L.P.Bhaiya Sumit Kumar Banchhor

***Abstract*: Information from speech recognition can be used in various ways in state of the art speaker recognition systems. This includes the obvious use of recognized words to enable the use of text dependent speaker modeling techniques. In this paper text dependent speaker identification method is used. This system contains training phase, the testing phase and recognition phase. In the training phase, the feature word is extracted. During the testing phase, feature matching takes place. The feature that extracted is stored in the data base. During the recognition phase, the features are extracted by same techniques and are compared with the template in the database.**

Differences of physiological properties of the glottis and vocal tracts are partly due to age, gender and/or person differences. Since these differences are related in the speech signal, acoustic measures related to those properties can be helpful for speaker identification. Acoustic measure of voice sources were extracted from 3 utterances spoken by 10 peoples including 5 male and 5 female talkers (aged 19 to 25 years old). In this paper, the eature of the extraction takes place by Zero Crossing Rate (ZCR).

Index Terms: Speech recognition, feature extraction, zero-crossing rate.

I. INTRODUCTION

Speech is natural mode of communication for people in our lives. In a fundamental aspect, speaker recognition and speech recognition are dual problem. In speaker recognition the goal is to identify the speaker irrespective of what is being said, in speech recognition the goal is to recognize what is being said irrespective of who is speaking. Thus in speaker recognition, one of the fundamental problem is to normalize for variability due to speaker's choice of phones, words, and so on; conversely, in speech recognition, a basic challenge is to normalize out speaker differences. Automatic speaker recognition technology is becoming increasingly widespread in many applications.

Speaker recognition is an example of biometric personal identification. This term is used to differentiate techniques that base identification on certain intrinsic characteristic of the person (such as voice, finger-prints, retinal patterns, or genetic structure) from those that use artifacts for identification (such as keys, badges, magnetic cards, or memorized passwords). Thus a prime motivation for studying speaker recognition is to achieve more reliable personal identification. This is particularly true for security applications, such as physical access control, computer data access control. Convenience is another benefit which accrues to a biometric system.

Application also exist which depends uniquely upon the identification of a person by his/her voice.

Recently cellular phones and the internet have made it possible to convey audio in digital form. It also enables the control access to service such as database access related services, information services, security control for confidential information area and forensic application. Human uses voice recognition everyday to distinguish between speakers and genders. Once correctly setup, the system should recognize over 95% of who said if you speak clearly. Voice recognition is software technique that identifies the closest results in pre-entered recognition data by extracting and analyzing the voice feature of people delivered on computer through microphones. Speech recognition sets its goal at recognizing the spoken word in speech; the aim of speaker recognition is to identify the speaker by extraction, characterization and recognition of the information contained in the speech signal.

II. LITERATURE REVIEW

Zero-crossing rate is proposed for sex identification and result of about 97% for gender classification is obtained [6]. Many attempts in speaker recognition have taken place in last thirty years. Major efforts have been made to develop methods for extracting information from audio-visual media, in order that they may be stored and retrieved in database automatically. Recent work has been done on segmentation of the audio signal and then classification into one of two main categories: speech or music [2, 3]. Zero-crossing rate is proposed for musical instrument identification and result reflects more effectively the difference in musical instrument [4]. An approach for separating the voiced/unvoiced part of the speech in a simple and efficient way, the algorithm shows good result in classifying the speech as the segmented speech into many frames [5]. However the issue is yet far from being solved. The work on recognition from is still remains crucial.

The performance achieved by listening, visual examination of spectrograms, and automatic computer techniques, attempt to provide a perspective with which to

Manuscript received on July, 2012

Arif Ullah Khan **,**Electronics& telecomm,CSVTU/RSR-RCET, Bhilai, India.

Lalit P Bhaiya, Electronics& telecomm, CSVTU / RCET, Bhilai, India.

Sumit Banchhor, Electronics & Telecomm, CSVTU/ GDR RCET, Bhilai, India.

evaluate the potential of speaker recognition and productive direction for research into and application of speaker recognition technology [1].

III. BASIC STRUCTURE OF SPEAKER RECOGNITION SYSTEM

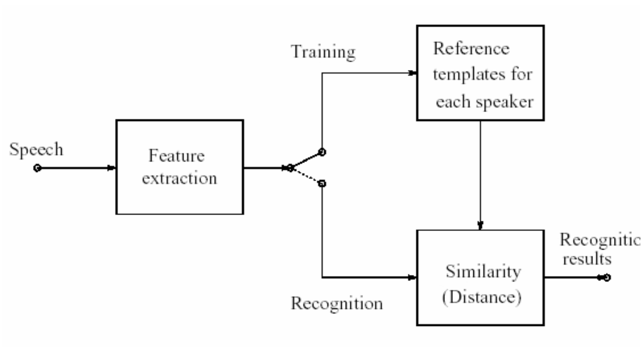

Fig-1 Structure of speech recognition system

The process of speaker recognition consists of the training phase and the recognition phase. In the training phase, the feature of speaker's speech signal is stored as reference features. The feature vectors of speech are used to create a speaker's model. The numbers of reference templates that are required for efficient speaker recognition depend upon the kind of features or techniques used for recognizing the speaker.

In the recognition phase, features similar to the ones that are used in the reference template are extracted from an input utterance of the speaker whose identity is required to be determined. The recognition decision depends upon the computed distance between reference template and template devised from the input utterance. The template of the registered user, whose distance with the input utterance template is the smallest, is finally selected as the speakers input utterance.

3.1 TEXT-DEPENDENT SPEKAER RECOGNITION METHODS

Speaker is required to utter a predetermined set of words or sentences. Features of voice are extracted from the same utterance. These systems can be deceived because someone who plays back a recorded voice of a registered speaker saying that the key words or sentences can be accepted as a registered speaker. The text dependent speaker recognition is most commercially viable and useful technology. Speaker identification can be thought of, as the task of determining who is talking from a set of known voices of speakers.

IV. METHODOLOGY

The target sample was manually segregated using GOLDWAVE software and stored with .wav extension.

V. ZERO-CROSSING RATE FOR VOICED/UNVOICES SPEECH

Zero-crossing rate is a measure of the number of times in a given time interval that the amplitude of the speech signals passes through a value of zero. Because of its random nature, zero-crossing rate for unvoiced speech is greater than that of voiced speech. Zero-crossing rate is an important parameter for voiced/unvoiced classification and for endpoint detection. Detecting when a speech utterance begins and ends is a basic problem in speech processing. This is often referred to as endpoint detection. End point detection is difficult if the speech is uttered in a noisy environment.

Many pitch detection algorithms are based on measurement of short term energy signal and zero-crossing rate and attempt to defect as accurately as possible the changes that there quantities undergo at the beginning and end of an utterance. The basic operation of algorithm is as follows. A small sample of background noise is taken during a 'silence' interval just prior to commencement of the speech is signal. The short-time energy function of the entire utterance is then computed. A speech threshold is determined which takes into account silence energy and peak energy. Initially, endpoints are assumed to occur where the signal energy crosses this threshold. Correction to these initial estimate, are then made by computing zero-crossing rate in the vicinity of endpoints and by comparing it with that of silence. If detectable changes in zero-crossing rate occur outside the initial thresholds, endpoints are re-designated to points at which the changes take place.

5.1 PERSON IDENTIFICATION USING ZERO-CROSSINGS

The notion of zero crossing is defined to be "The *number of times in a sound sample that the amplitude of the sign wave changes sign"*

For a10ms sample of clean speech, the zero-crossing rate is approximately 12 for voiced speech and 50 for unvoiced speech. For clean speech the zero-crossing rate should be useful for detecting region of silence, as the zero-crossing rate should be zero.

Unfortunately, very few sound samples are recordings of perfect clean speech. This means that there is some level of background noise, that interferes with the speech, meaning that the silent region actually have quite a high zero-crossing rate as the signal changes from just one side of zero amplitude to the other and back again. For this reason a tolerance threshold is included in the function that calculates zero-crossing to try and alleviate this problem. The thresholds work by removing any zero-crossings, which do not start and end a certain amount from the zero value.

In this study we have used a threshold of 0.001. This result states that any zero-crossings that start and end in the range of 'x', where $-0.001 < x < 0.001$, are not included in the total number of zero-crossing for that window. This enables us to filter out most of the zero-crossings that occur during silent region of the sample due only to background noise.

VI. RESULTS

It indicates the frequency of signal amplitude sign changes. To some extent, it indicates the average signal frequency as:

$$ZCR = \frac{\sum_{n=1}^{N} \left|\operatorname{sgn} x(n) - \operatorname{sgn} x(n-1)\right|}{2N}$$

Where *sgn[]* is a signum function and *x(m)* is discrete audio signal.

In mathematical terms, a “zero-crossing” is a point where the sign changes (e.g. from positive to negative), represented by a crossing of the axis (zero value) in the graph of the function. The zero-crossing is important for systems which send digital data over AC circuits, such as modem, X10 home automation control systems for Lionel and other AC model trains. Counting zero-crossing is also a method used in speech processing to estimate the fundamental frequency of the speech.

Zero-crossing rate is important because they abstract valuable information about the speech and they are simple to compute.

.Cha:-

S.No	Person	Frames					
		0	5	10	15	20	25
1	F1	39	56	39	39	40	38
2	F2	64	66	43	48	44	41
3	F3	46	4	55	47	53	49
4	F4	38	48	46	48	56	43
5	F5	28	44	40	39	48	55

TTa:-

S.No	Person	Frames					
		0	5	10	15	20	25
1	F1	36	50	42	37	37	45
2	F2	48	48	44	36	46	42
3	F3	51	52	52	51	52	49
4	F4	49	56	57	53	62	55
5	F5	52	41	46	43	40	48

Ka:-

S.No.	Person	Frames					
		0	5	10	15	20	25
1	F1	39	56	39	39	40	38
2	F2	64	66	43	48	44	41
3	F3	46	4	55	47	53	49
4	F4	38	48	46	48	56	43
5	F5	28	44	40	39	48	55

Cha:-

S.No	Person	Frames					
		0	5	10	15	20	25
1	M1	40	154	55	38	38	38
2	M2	51	194	46	32	51	34
3	M3	39	143	42	112	34	45
4	M4	40	158	31	35	40	41
5	M5	48	157	60	60	46	44

Tta:-

S.No	Person	Frames					
		0	5	10	15	20	25
1	M1	49	52	41	41	46	42
2	M2	57	59	59	51	51	37
3	M3	43	48	36	45	44	45
4	M4	39	41	63	38	45	45
5	M5	43	51	44	58	59	27

Ka:-

S.No	Person	Frames					
		0	5	10	15	20	25
1	M1	41	45	37	37	52	52
2	M2	46	75	44	37	40	43
3	M3	47	49	49	35	55	46
4	M4	47	75	36	36	35	30
5	M5	48	44	57	59	45	45

VII. FURURE PLANS

Speech recognition is a difficult task. A speaker recognition works based on the premise that a person’s speech exhibits characteristics that are unique to the speaker. However this task has been challenged by the input speech signal as many as it can. Speaker recognition technology is very interesting so we have been doing research in this field. In addition, the future investigation in the zero-crossing peak amplitude (ZCPA) for ASR will be undertaken to evaluate the adaptation model in other types of noise such as convolute and other type of noise.

VIII. CONCLUSION

In this paper we present a method for isolated hindi word recognition based on zero-crossing feature. The isolated word recognition method consist feature extraction phase. In feature extraction, end points are detected and noise is removed using end point detection algorithm. The role of speaker recognition can range from simple transcription to enable text-dependent modelling to the extraction of novel speaker features that characterize the behaviour of the speaker recognizer.

Results from a selection of these techniques show that such features have much potential.

Result shows that the estimation of zero crossing rate reflects more effectively the difference in different people speaking in Hindi.

REFERENCES

1. Yiu - Kei Lau and Chok- Ki Chan,"Speech recognition based on zero-crossing rate", IEEE Transactions on acoustics, speech and signal processing, Vol.ASSP-33, No.1.
2. Costas panagiotakis and George tziritas, "A speech/music discriminator based on RMS and zero-crossings", IEEE transactions on multimedia.vol.7, no.1, February 2005.
3. Sumit Kumar Banchhor, Om Prakash Sahu, Prabhakar, "A Speech/Music Discriminator based on Frequency energy, Spectrogram and Autocorrelation", IJSCE, Volume-1, Issue-6, January 2012
4. Sumit kumar Banchhor and Arif Khan, "Musical Instrument Recognition using Zero Crossing Rate and Short-time Energy", Volume 1– No.3, February 2012.
5. Bachu R.G, Kopparthi S, Adapa B, Barkana B.D, "Separation of voiced and unvoiced using zero crossing rate and energy of the signal".
6. Sumit kumar Banchhor and S. K. Dekate, "Text-dependent Method for Gender Identification through synthesis of voiced segments",IJEST, Volume- 3, No. 6, June 2011.