2011 International Conference on Signal, Image Processing and Applications
With workshop of ICEEA 2011
IPCSIT vol.21 (2011) © (2011) IACSIT Press, Singapore

Mathematical Model for Detection of Dissimilar Patterns of Speech Signal of Chhattisgarhi Dialects using Wavelet Transformation

Madhuri Gupta , and Akhilesh Tiwari

[1] S.S.C.E.T ,Bhilai(Chhattisgarh)

Abstract. This paper will presents a mathematical model of finding a relations between dissimilar speeches signals which are speaker independent taken from common set of samples of Chhattisgarhi language and dialects commonly spoken at different regions of Chhattisgarh. The objective of this paper is to detect similar and dissimilar patterns of different Chhattisgarhi dialects through wavelet analysis methods which are useful for complex speech signal analysis.

Keywords: Speech Recognition, Speech parameterization, Wavelet transformation.

1. Introduction

Speech is the primary means of communication between people. For reasons ranging from technological curiosity about the mechanisms for mechanical realization of human speech capabilities, to the desire to automate simple tasks inherently requiring human-machine interactions, research in automatic speech recognition (and speech synthesis) by machine has attracted a great deal of attention over the past five decades. The speech signal conveys many levels of information to the listener. At the primary level, speech conveys a message via words. But at other levels speech conveys information about the language being spoken and the emotion, gender and, generally, the identity of the speaker. While speech recognition aims at recognizing the word spoken in speech, the goal of automatic speaker recognition systems is to extract, characterize and recognize the information in the speech signal conveying speaker identity.

Chhattisgarhi is a dialect of Hindi Language or language of its own right and it is spoken and understood by the majority of people in Chhattisgarh. Chhattisgarhi was also known as Khaltahi to surrounding people and as Laria to Oriya speakers. Chhattisgarhi has several identified dialects of its own, in addition to Chhattisgarhi proper. Overall Chhattisgarhi can be divided into 26 different types of dialects.

In this paper the system will be focus on 3 types of dialects that are Baheliya, Bhunjwari and Bilaspuri. The system works on speech parameterization using *Filter bank-based cepstral parameters.* Based on different filters, histogram, standard deviation, mean and mode the system will be able to analysis difference among different dialects and also able to conclude the similarities.

2. Analysis of Chhattisgarhi dialects using wavelet transformation

The fundamental idea behind wavelets is to analyze according to the scale. Indeed, some researchers in the wavelet field feel that, by using wavelets, one is adopting a whole new mindset or perspective in processing data. Wavelets are functions which satisfy certain mathematical requirements and are used in representing data or other functions. Approximation using superposition of functions has existed since the early 1800's, when Joseph Fourier discovered that he could superpose sins and cosines to represent other functions. However, in wavelet analysis, the scale that one uses for looking at data plays a vital role.

2.1. Wavelet Algorithm

Wavelet algorithms process data at different scales or resolutions. If we look at a signal with a large “window," we would notice gross features. Similarly, if we look at a signal with a small "window," we would notice small discontinuities.

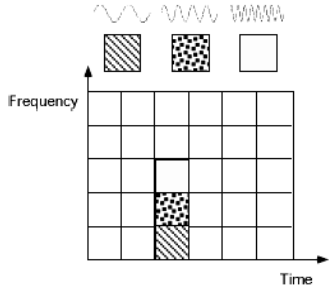

Fig. 2.1.Fourier basis functions, time-frequency tiles and Coverage of the Time-frequency plane

Fig. 2.1 shows a windowed Fourier transform, where the window is simply a square wave [10-11]. The square wave window truncates the sine or cosine function to fit a window of a particular width. Because a single window is used for all frequencies in the WFT, the resolution of the analysis is same at all locations in the time-frequency plane.

For the analysis of this model the system load the signal from wavelet menu and then get the sample analyzed waveform for baheliya dialect for sentence “Main Jao Chu” as depicted in figure 2

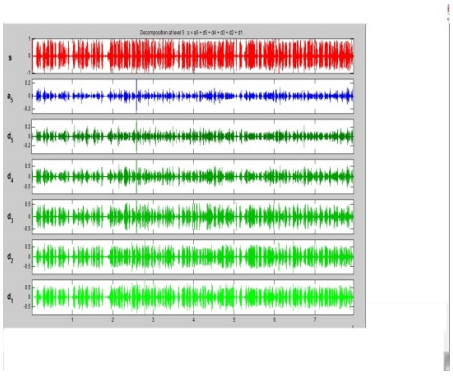

Figure 2.2: Analyzed acoustic wave signal of Sentence “Main Jao Chu”.

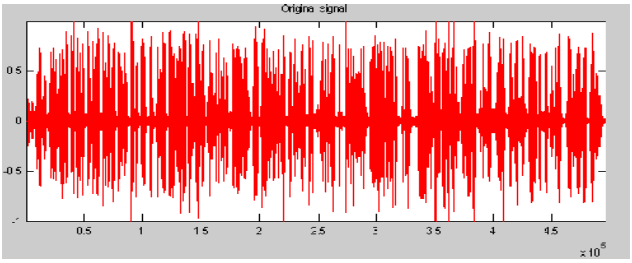

Figure 2.3: Original Wave signal of Bhunjwari Dialect

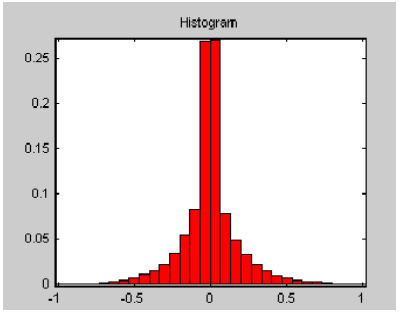

Fig. 2.4 Calculated Histogram of Bhunjwari Dialect

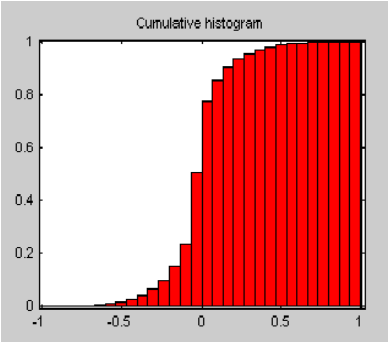

Fig. 2.5: Sample Calculated Cumulative Histogram of Bhunjwari Dialect

To calculate the statistical distribution of wavelet coefficients, the standard deviation is considered in terms of second momentum coefficient. It is noted that the mean of speech signal is zero. If we assume that the wavelet coefficients distribution is modeled as Laplace density function, then distribution of these coefficients is related directly to its standard deviation. In this way, assume that the output signal, y, is achieved summation of input, w, and noise, n. Then the variance of output is as follows (Eq. 1).

$$VAR[\mathrm{y}] = VAR[w] + VAR[n] \tag{1}$$

The mean of speech and noise signals is zero, so we have:

$$VAR[\mathrm{y}] = MEAN\ [\mathrm{y2}] \tag{2}$$

The standard deviation is calculated as follows:

$$\hat{\sigma} == \sqrt{mean\ [y2]} - \sigma_n^2 \tag{3}$$

The original signal is shown in figure 2.3; on the basis of original signal the system will generate a histogram and cumulative histogram. Histogram and Cumulative histogram of calculated standard deviations is shown in figure 2.4 and 2.5.

3. Speech Parameterization

Speech parameterization consists in transforming the speech signal to a set of features vectors. The aim of this transformation is to obtain a new representation which is more suitable for statistical modeling and the calculation of a distance or any other kind of score. Most of the speech paramerization used in speaker verification systems relies on a cepstral representation of speech.

3.1. Filter bank-based cepstral parameters

The speech signal is first pre-emphasized, that is, a filter is applied to it. The goal of this filter is to enhance the high frequencies of the spectrum, which are generally reduced by the speech production process. The pre-emphasized signal is obtained by applying the following filter:

$$xp(t) = x(t) - a \cdot x(t-1). \tag{4}$$

Values of a are generally taken in the interval [0.95, 0.98]. This filter is not always applied, and some people prefer not to pre-emphasize the signal before processing it. The analysis of the speech signal is done locally by the application with a window whose duration in time is shorter than the whole signal. This window is first applied to the beginning of the signal, and then moved further until the end of the signal is reached. Each application of the window to a portion of the speech signal provides a spectral vector. For the length of the window, two values are most often used: 20 milliseconds and 30 milliseconds.

4. Detection of similar pattern via likelihood detection

Given a segment of speech *Y* and a hypothesized speaker *S*, the task of speaker verification, also referred to as detection, is to determine if *Y* was spoken by *S*. An implicit assumption often used is that *Y* contains speech from only one speaker. Thus, the task is better termed single speaker verification. If there is no prior information that *Y* contains speech from a single speaker, the task becomes multispeaker detection.

The system essentially implements a likelihood ratio test to distinguish between two hypotheses: the test speech comes from the claimed speaker or from an imposter. Features extracted from the speech signal in front-end processing are compared to a model representing the claimed speaker, obtained from a previous enrolment, and to some model(s) representing potential imposter speakers (i.e., those *not* the claimed speaker). The ratio (or difference in the log domain) of speaker and imposter match scores is the likelihood ratio statistic ( ), which is then compared to a threshold ( ) to decide whether to accept or reject the speaker. The general techniques used for the three main components, front-end processing, speaker models, and imposter models, are briefly described next.

The optimum test to decide between these two hypotheses is a likelihood ratio (LR) test given by

$$\frac{p(Y|H0)}{p(Y|H1)} \begin{cases} > \theta, accept\ H0, \\ < \theta, accept\ H1, \end{cases} \quad (5)$$

Where $p(Y|H0)$ is the probability density function for the hypothesis H0 evaluated for the observed speech segment *Y*, also referred to as the "likelihood" of the hypothesisH0 given the speech segment.2 The likelihood function for H1 is likewise $p(Y|H1)$. The decision threshold for accepting or rejecting H0 is θ. One main goal in designing a speaker detection system is to determine techniques to compute values for the two likelihoods $p(Y|H0)$ and $p(Y|H1)$.

5. 5. Result

On the basis of Mean, Median, Mode and SD the system will be able to differentiate among three dialects i.e. Baheliya, Bhunjwari and Bilaspuri.

Table1: Result for the different dialects

Criteria	Baheliya	Bhnjwari	Bilaspuri
Mean	-0.0001097	-0.0001093	-0.0002385
Median	-0.0007324	0.0002441	-0.0004883
Mode	-0.03345	0.0188	-0.3345
SD	0.2139	0.5367	0.2001

6. Conclusion

On the basis of different criteria the system will able to conclude the dissimilar patterns of Chhattisgarhi dialects. The result can be useful for the creation of database for the particular state as well as for the security purpose. The future enhancement for the error fining and noise reduction can be possible for the new modeling of the system.

7. Acknowledgment

The real spirit of achieving a goal is through the way of excellence and asteroids discipline. I would have never succeeded in completing my task without the cooperation, encouragement and help provided to me by

various personalities. I want to thank SSCET, Bhilai for providing me the necessary software, tools and other resources to deliver my research work. With deep sense of gratitude I express my sincere thanks to my esteemed and worthy Guide Prof. Akhilesh Tiwari, SSCEC, Bhilai for their valuable guidance in carrying out this work under their effective supervision. Encouragement, enlighten and cooperation.

References

[1] Madhuri Gupta & Akhilesh Tiwari .―Speech Analysis of Chhattisgarhi (dialect) Speech signal of different regions of Chhattisgarh. International Conference on emerging trends in soft computing (SCICT), March 2011,Bilaspur,India.

[2] Leh Luoh,Yu-Zhe Su and Chih-Fan Hsu. ― Speech signal processing based emotion recognition ‖ .2010 International Conferences on System Science and Engineering. 978-1-4244-6474-61101$26.00 © 2010 IEEE.

[3] Shiv Kumar, Member, IACSIT, IAENG Aditya Shastri and R.K. Singh- "An Approach for Automatic Voice Signal Detection (AVSD) Using Matlab" International Journal of Computer Theory and Engineering, Vol. 3, No. 2, April 2011 ISSN: 1793-8201.

[4] Mansour Sheikhan1, Mohammad Khadem Safdarkhani2, Davood Gharavian3. Presenting and Classification Based on Three Basic Speech Properties, Using Haar Wavelet Analyzing ‖ . 22nd International Conference on Signal Processing Systems (ICSPS 2010).

[5] Douglas A. Reynolds "Automatic Speaker Recognition: Current Approaches and Future Trends" ICASSP 2001.

[6] Ilyas Potamitis and George Kokkinakis, _Speech Separation of Multiple Moving Speakers Using Multisensory

[7] Multistage Techniques ‖ , IEEE TRANSACTION ON SYSTEM, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, VOL. 37, NO. 1, JANUARY 2007.

[8] G. R. Doddington, _Speaker Recognition based on Idiolectal Differences between Speakers, ‖ Eurospeech 2001.

[9] A. D. Andrews, M. A. Kohler and J. P. Campbell, Phonetic Speaker Recognition, ‖ Eurospeech 2001.