INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY

WINGS TO YOUR THOUGHTS.....

Classification of Speech Emotion using MFCC

Arpit Parikh[1], Sameena Zafar[2], Mukesh Saini[3]

[1]PG Research Scholar, Department of EC,
Patel Institute of Engineering & Science,RGPV University, Bhopal, India
arpit11parikh@gmail.com
[2]Assistant Professor, Head of Department, Department of EC,
Patel College of Science & Technology,RGPV University, Bhopal, India
sameena_zafar82@yahoo.com
[3]Assistant Professor, Department of EC,
Patel College of Science & Technology, RGPV University, Bhopal, India
sainimukesh16@gmail.com

Abstract:*Now a day, the problem of detection of depressed person from the speech signal becomes initial matter. We know depression is a common and serious mental disorder. Now a day's lots of People in India are identified as suffering from depression. It is a trend of increase the prevalence of depression. In this study, the speech signal is identified of depressed and normal person. Analysis of voice for depressed and normal person is done using Mel Frequency Cepstral Coefficients (MFCC) Extract feature using this method and then store this feature.After that, compare the feature for detection.Analysis of voice for depressed and normal person is done by Mel Frequency Cepstral Coefficients (MFCC).Various parameters are used for the analysis of depressed speech and normal speech, but MFCCs based parameter provide the most effective information then other parameter because depressive speech can contain more information in the higher energy bands when compared with normal speech*

Keywords: *Mel Frequency Cepstral Coefficients (MFCC), discrete Cosine Transform (DCT), Speech Recognition, Depression.*

1. INTRODUCTION

There are many types of pressure living in present. Because the pressure becomes heavy, some people are living with depression. There are 15% patients suffering from depression will suicide themselves and there are 87% suicides diagnosed with depression during their lifetime. Suicide is the second reason of death between the age of 15 and 24 years, and the third reason of death between age of 25 and 44 years. Moreover, the percentage of depressed person/patients in treatment was low.Even if the depressed patients were in treatment, doctors usually relied on their medical history and clinical observation [1].

The electrocardiogram, blood volume pulse, and Galvanic skin response are as measures depression for clinical application [2]. In sound processing, the Mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear Mel scale of frequency. It was originally developed for speech recognition but is now one of the most common feature extraction techniques in speaker verification today. The idea behind using MFCC is the assumption that the human hearing is an optimal speaker recognizer, although this has not yet been confirmed by studies.

Mel-frequency Cepstral coefficients (MFCCs) are coefficients that collectively make up an MFC. They are derived from a type of Cepstral representation of the audio clip (a nonlinear "spectrum-of-a-spectrum"). The difference between the cepstrum and the Mel-frequency cepstrum is that in the MFC, the frequency bands are equally spaced on the Mel scale, which approximates the human auditory system's response more closely than the linearly-spaced frequency bands used in the normal cepstrum. This frequency warping can allow for better representation of sound, for example, in audio compression.

2. LITERATURE

The idea of recognizing distinctive patterns and tone of voice in patients with high risk suicide was introduced by two clinical psychologists, Drs. Stephen and Marilyn Silverman. Both had experience in treating patients with near term suicidal risk. They began research in the 1980s by collecting and analyzing suicidal tape recordings obtained through therapy sessions in an uncontrolled environment, and notes and interviews made shortly before suicide attempts. They describe the similarity of vocal speech between depressed and suicidal patient but notice changes occur considerably in the tonal quality and acoustical characteristics when the patient enters the suicidal state [3]. Several other researchers continue to study the relation of vocal tract characteristic to depression and suicidal risk.

France [4] began the research by extracting and analyzing fundamental frequency (Fo), amplitude

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS.....

modulation (AM), the formants and power distribution (PSD) on speech samples. Among these perceptual qualities, formant and PSD features appeared to be distinguishing vocal features when discriminating between suicidal and major depressed patients compared to the ones collected from control groups. Linear Predictive Coding (LPC) was preferred over Long-term-average (LTAS) to calculate the formant frequencies and bandwidths due to the volume of speech analyzed which made it computationally expensive even though the LTAS approach provides a more accurate representation of the formant properties. The classical Welch method was used when extracting the PSD. The energy spectrum was investigated on the percentages of total energy in frequency sub-bands with a bandwidth of 500 Hz over the frequency range of 0-2000 Hz. It was reported that most energy are distributed in the range of 0-2000 Hz. Features were integrated when performing classification in order to obtain the best parameter combination in distinguishing between the suicidal and non-suicidal groups. Multi-parameter classification was shown to be more effective than classification of single parameters.

Ozdas [5] studied the discriminating power of lower order mel-cepstral coefficients (MFCC) among suicidal, major depressed and non-suicidal patients. Vocal tract characteristic using a non-model based approach for near term suicidal risk assessment was the focus of this study. The effects of source (excitation) and filter (vocal tract) on suicidal state were the two domains examined. Vocal jitter and slope of the glottal flow spectrum were two other excitation features that were further investigated related to the excitation signal, whereas in the filter domain, speech features are investigated through Cepstral analysis. There were variations among different psychological states with the use of mel-cepstral filter bank coefficients and the results suggested that the use of MFCC features could provide useful measurements for identification of a possible suicidal state. The use of Gaussian mixture models yielded better class approximation when performing classification for individual diagnostic groups.

3. METHODOLOGY

MFCCs are commonly derived as follows [6]:

1. Framing and windowing of signal
2. Take the Fourier transform of (a windowed excerpt of) a signal.
3. Map the powers of the spectrum obtained above onto the Mel scale, using triangular overlapping windows.
4. Take the logs of the powers at each of the Mel frequencies.
5. Take the discrete cosine transform of the list of Mel log powers, as if it were a signal.
6. The MFCCs are the amplitudes of the resulting spectrum.

All steps for Mel frequency cepstral coefficients is combined in block diagram and this block diagram is shown as below.

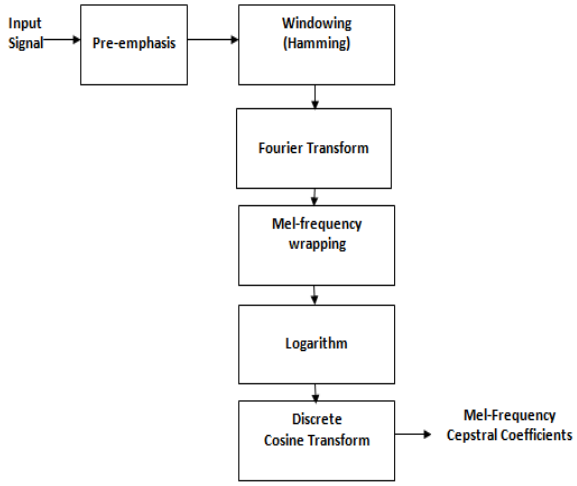

Figure 1: Block diagram of MFCCs

The goal of pre-emphasis is to compensate the high-frequency part that was suppressed during the sound production mechanism of humans. Moreover, it can also amplify the importance of high-frequency formants.

The speech signal s(n) is sent to a high-pass filter:

$$s_2(n) = s(n) - a*s(n-1) \quad (1)$$

The Hamming window is defined as:

$$h(n) = \begin{cases} (1-\alpha) - \alpha\cos\left(\frac{2\pi n}{N-1}\right), 0 \le n \le N-1 \\ 0, \quad otherwise \end{cases}$$

We multiple the magnitude frequency response by a set of 20 triangular band pass filters to get the log energy of each triangular band pass filter. The positions of these filters are equally spaced along the Mel frequency, which is related to the common linear frequency f by the following equation:

$$M=2595\log_{10}\left(\frac{f}{700}+1\right) \quad (2)$$

Where f is the actual frequency and mel(f) is the perceived one.

The mel scale is based on how the human hearing perceives frequencies. It was defined by setting 1000 mels equal to 1000 Hz as a reference point. Mel-frequency is proportional to the logarithm of the linear frequency, reflecting similar effects in the human's subjective aural perception. Then listeners were asked to adjust the physical pitch until they perceived it as two-fold ten-fold and half, and those frequencies were then labelled as 2000 mel, 10000 mel and 500 mel respectively. The resulting scale

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS.....

was called the mel scale and is approximately linear below frequencies of 1000hz and logarithmic above. When we run the program, three messages are displayed in GUI. 'Start record', 'Browse Wav file' & 'Recognize 'is displayed as shown in below figure.

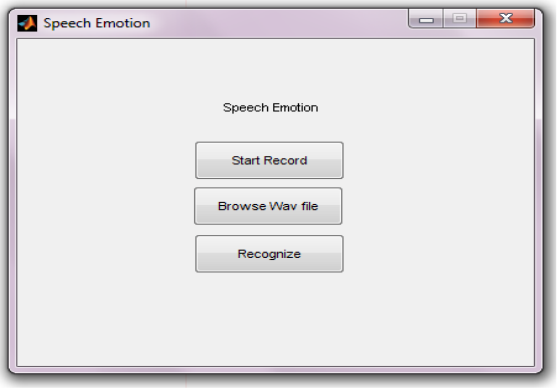

Figure 2: Graphical user interface

When we push the button 'Start Record', program asks to record of audio signal or voicefor 1 second and when we push the button 'Browse Wav file', program asks to browse wave file from the system and to record of audio signal or voicefor 1 second. After Recording When we push the button 'Recognize', of the Audio signal, , if feature of voice matched with depressed data, then it display message 'Depressed' as shown below:

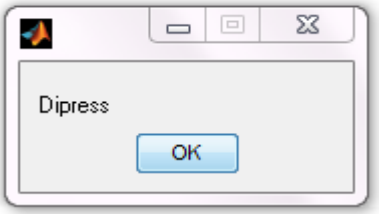

Figure 3: Display message for depress person

After recording, if feature of voice is not matched with depressed data, then it displays message 'Normal' as shown below.

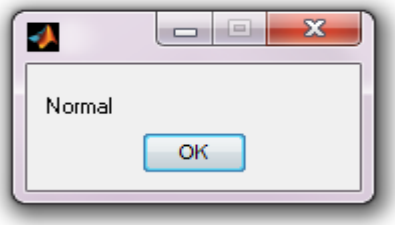

Figure 4: Display message for depress person

4. RESULT

Here are some images of the result obtained during recognition of speech. It seems to be very efficient and robust to implement in real time working environment.

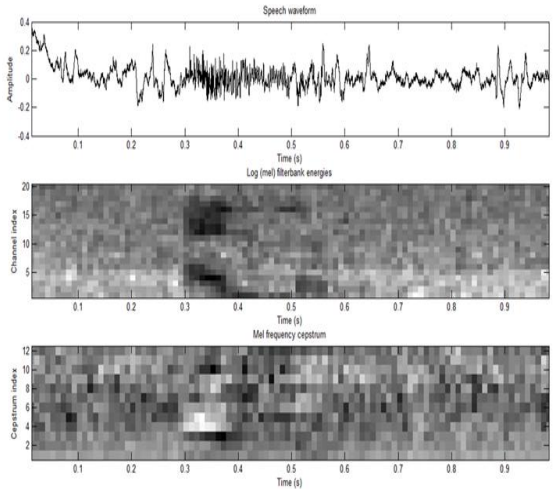

Figure 5: Output wave form of MFCCs

5. CONCLUSION

After evaluating the speech sample, conclude that vocal properties represented by Mel-Frequency Cepstral Coefficients (MFCC). This is providing most effective discriminate between depressed/normal people. Several insights have been gained from this study of depressed/neutral classification. Voiced speech segments appear to be mildly preferable for the purpose; however segment selection is not critical. Feature warping needs to be used with care; however its behavior in this investigation is similar to that for emotion recognition, despite the differences in the structure of the respective recognition problems.

REFERENCES

[1] M. W. Huang, K. S. Cheng, "The Evaluation of Cognitive Function with Visual Long-term Evoked Potentials in Schizophrenic Patients", Institute of Biomedical Engineering National Cheng Kung University, 2002.

[2] Yen-Ting Chen, I-Chung Hung, Chun-JuHou, "Physiological Signal Analysis for Patients with Depression", Department of Electrical Engineering, Southern Taiwan University, Tainan, Taiwan.

[3] Thaweesak Yingthawornsuk. Ph.D. Thesis. "Acoustic Analysis of Vocal Output Characteristics for Suicidal Risk Assessment "Graduate School of Vanderbilt University, December, 2007.

[4] France, D.J., PhD, Thesis "Acoustical properties of speech as indicators of depression and suicidal risk", Vanderbilt University, August, 1997.

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY

WINGS TO YOUR THOUGHTS.....

[5] Asli Ozdas, Richard G. Shiavi, Senior Member, IEEE, Stephen E. Silverman, Marilyn K. Silverman, and D. Mitchell Wilkes, Member, IEEE. "Investigation of Vocal Jitter and Glottal Flow Spectrum as Possible Cues for Depression and Near-Term Suicidal Risk", IEEE transactions on Engineering, Vol. 51, No. 9, September 2004.

[6] Mel-frequency cepstrum coefficients (MFCCs),http://mirlab.org/jang/books/audiosignalprocessing/speechFeatureMfcc.asp?title=12-2%20MFCC.