Speech Emotion Classification Using DWT and Neural Network

Ms.Abinaya.B[1], C.Abishai[2], S.Ashika [3]
Student[1,2,3], Dept. of ECE[1,2,3], UG Scholar, Dept. of ECE[1,2,3]
Hindusthan College of Engineering and Technology,
Coimbatore, Tamilnadu, India.

Abstract - This paper provides the various sorts of the emotion classification of the person. That format is going to be identified using neural network. Translating of spoken words into text remains a challenging task thanks to the high viability in speech signals. Neural Network is becoming a mainstream technology for speech recognition, taking the sides of the signal based information using DWT(discrete wavelet transformation)and are using in GLCM (gray level co-occurrence matrix) features for speech emotion recognition and have coding at an increasingly larger scale. Including methods to enforce temporal smoothing and a way to include a previous distribution to constrain the extracted parameters. Basically these are used at the social media applications.

Keywords:-input signal, DWT, KLT, Neural Network.

I. INTRODUCTION

The fundamental purpose of speech is communication, i.e., the transmission of messages. A new message represented as a sequence of discrete symbols is often quantized by its information content in bits, and therefore the rate of transmission of data Is measured in bits/second (bps)will be shown. In speaking, also as in many human-engineered transmission systems, the Knowledge to be transmitted is encoded within the sort of a continuously varying (analog) waveform which will be transmitted, recorded, manipulated, and ultimately decided by a person's listener. Within the case of speech, the elemental analog sort of the message is an acoustic waveform, which we call the speech signal. Speech signals are often converted to an electrical waveform by a microphone, further manipulated by both analog and digital signal processing, then converted back to acoustic form by a loudspeaker, a telephone handset or headphone, as desired in this project. Signals are usually corrupted by noise within the world. To scale back the influence of noise, two research topics are the speech enhancement and speech recognition in noisy environments have arose in this. The main application of the project is detection of emotion voice classification.

II. RELATED WORK

Recognition of emotion in speech allows to spot the spirit of humans. Sets of features including acoustics, spectral, and non-linear dynamics measures are utilized in the literature for speech emotion recognition. Set of features supported the Discrete Wavelet Transform (DWT) to classify differing types of emotions like anger, happy, disgust, normal and sad , in scenarios where the speech signal is contaminated with noise or is coded by telephone channels. In this project we are taking classification of the emotion classification using as of neural network. in that we have a layers based on that only we are getting the result.

III. EXISTING SYSTEM

Sampling frequency means pc collects what percentage speech samples per second, which not only is describing voice files‟ pitch, timbre but can also measure the sound card, sound file's quality. The frequency is higher while the interval time of sampling is shorter, which suggests computer gets more sound sample data during a unit time, therefore the shape of acoustic wave forms are going to be more precise. Preprocessing speech signal commonly involves removing low-frequency ground noise, normalizing the intensity enhancing data.

Peak Value Detection:-Maximum Stem on Input Voice Signal (Low Pass Signal) using “max” command process on “Feature_Peak=max(low_pass_signal)”

Minimum Value Detection:-Minimum Stem on Input Voice Signal (Low Pass Signal) using “min” command process on “Feature_Minimum=min(low_pass_signal)”

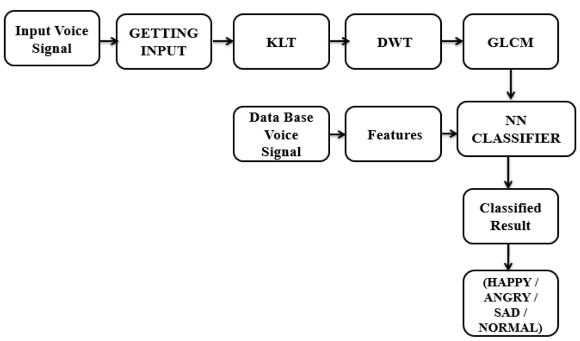
Mean (Average) Value Detection:- Minimum Stem on Input Voice Signal (Low Pass Signal) using “min” command process on”Feature_Average=mean(low_pass_signal)”

A Fast Fourier transform is an algorithm to compute the Discrete Fourier transform (DFT) and it’s inverse. Fourier analysis converts time (or space) to frequency (or wave number) and the other way around . An FFT rapidly computes such transformations by factorizing the DFT matrix into a product of sparse (mostly zero) factors. SVM may be a simple and efficient computation of machine learning algorithms, and is widely used for pattern recognition and classification problems, and under the conditions of limited training data, it can have a really good classification performance compared to other classifiers. SVM may be a binary classifier to research the info and recognize the patterns for classification and multivariate analysis and Adaboost is an iterative algorithm that focuses on both continuous valued input and textual input by text categorization. It combines many simple and moderately accurate rules into one highly accurate rule.

IV. PROPOSED SYSTEM

In this proposed method we are taking the speech classification and after that detecting the speech emotion.the below block diagram explanation will be there.

Block diagram



Input Voice signal:-

Sampling frequency means pc collects what percentage speech samples per second, which not only is describing voice files" pitch, timbre but also can measure the sound card, sound file's quality. The frequency is higher while the interval time of sampling is shorter, which suggests computer gets more sound sample data during a unit time, therefore the form of sound wave forms are getting to be more precise.

Preprocessing:- It is the second step in speech signal commonly involves removing low-frequency background noise, normalizing the intensity enhancing data.

KLT:-
To obtain these silence intervals, we proposed an efficient voice activity detector supported outputs of principle component Eigen filter; the best eigenvalue of speech KLT. Enhancement is performed by the technique modifying each KLT component thanks to its noise and clean speech energies. the target is to attenuate the produced distortion when residual noise power is restricted to a selected level. At the top , inverse KLT is performed and an estimation of the clean signal is synthesized.

DWT:-

The Discrete Wavelet Decomposition (DWD) may be a popular method used for transform features extraction signal processing.The DWT allows the level based decomposition of an signal with a wavelet basis consistent with a coffee pass and high pass filter signal. The resolution is reduced by one-half at each level by sub sampling data (signal) by two.

GLCM:-

ENERGY:- Energy is a simple short-time speech measurement. It is defined as

$$S(k) = \frac{\sum_{l=1}^{N_{ki}} [d_{ki,l}]}{N_{ki}}$$

The normalized energies S(k) are computed from the wavelet coefficients dki,l for fame.

TEAGER ENERGY OPERATOR

TEO of the sequence x[n] is defined as:

$$TEO\{x(n)\} = | x_n . x_n^* - x_{n+1}.x_{n-1}^* |$$

Where xn, xn-1, and xn+1, are the actual, past and future samples of x[n], respectively.

$$S_{tf}(ki) = \frac{\sum_{l=1}^{N_{ki}} | d_{ki,l}.d_{ki,l}^* - d_{ki,l+1}.d_{ki,l-1}^* |}{N_{ki} - 2}$$

Entropy:-The entropy S is proportional to the logarithm of the number of possible microscopic conficonfigurations W, under the assumption that all possible micro states of the system are equally likely and that constraints on energy mentioned.

$$H1(ki) = -\sum_{l=1}^{N_{ki}} [|d(ki,l)|^2 \cdot log\, |d(ki,l)^2|]$$

$$H2(ki) = -\sum_{l=1}^{N_{ki}} log\, |d(ki,l)^2|$$

Neural network:-

Neural network may be a machine learning algorithm inspired from the working of human brain which enable a system to find out from some observational data. an easy neural network contains an input layer, one hidden layer and an output layer.

V. RESULTS

Take an input voice and apply the filtering from pre-process.

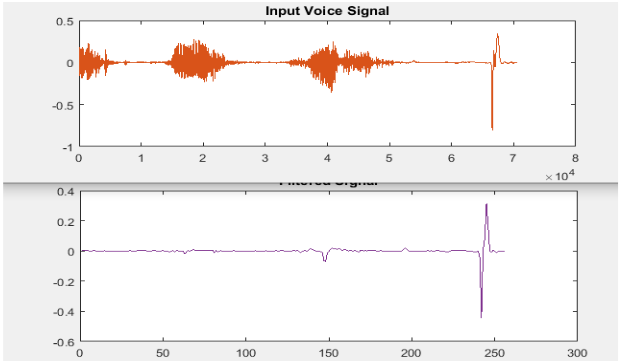


Figure1.0 input voice and filtering

After that apply to the DWT technique.

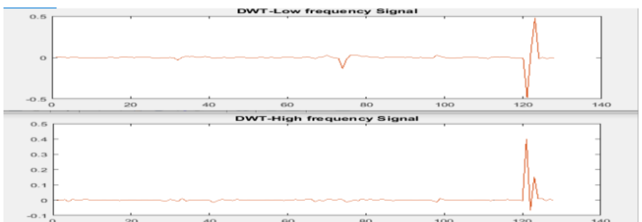


Figure1.2. applied DWT for input voice.

Finally, we can get the result weather that are in which emotion using Neural Network will be shown below.

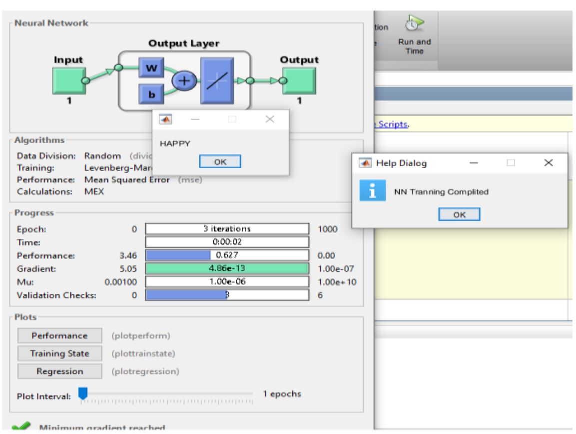


Figure1.3. getting output using Neural Network

VI. CONCLUSION AND FUTURESCOPE

Automatic classification of emotional speech in non-controlled conditions is being addressed. Noise-free conditions. With additive white Gaussian noise. The main theme of the project is detecting the different emotion classification using best algorithm of neural network. Basically there are used and detect the features using feature parameters from the GLCM. Based on this we can go with the neural network to get the emotion classification.for the future purpose we can go with the we can give the different voice in google assistant.

VII. REFERENCES

[1]. J.C. Vasquez-Correa, N. Garcia, J.F. Vargas-Bonilla, J.R.Orozco-Arroyave, J.D.Arias-Londono, O.Licia Quintero M. “Evaluation of Wavelet Measures on Automatic Detection of Emotion in Noisy and Telephony Speech Signals” 2014 IEEE 978-1-4799-3532-1/14/$31.00 .

[2]. P. Gupta and N. Rajput, “Two-stream emotion recognition for call center monitoring,” in INTERSPEECH, 2007, pp. 2241–2244.

[3]. J. Pohjalainen and P. Alku, “Automatic detection of anger in telephone speech with robust autoregressive modulation filtering,” in Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, May 2013, pp. 7537–7541.

[4]. F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, and W. B., “A database of German emotional speech,” Proc of the INTERSPEECH 2005, pp. 1517–1520, 2005.

[5]. C. Clavel, I. Vasilescu, L. Devillers, G. Richard, and T. Ehrette, “Fear- type emotion recognition for future audio-based surveillance systems,” Speech Communication, vol. 50, no. 6, pp. 487–503, Jun. 2008.

[6]. R. M. Nickel and R. Martin, “Memory and complexity reduction for inventory-style speech enhancement systems,” in Proc. EUSIPCO, Barcelona, Spain, Sep. 2011, pp. 196–200.

[7]. D. Kolossa, A. Klimas, and R. Orglmeister, “Separation and robust recognition of noisy, convolutive speech mixtures using time-frequency masking and missing data techniques,” in Proc. Workshop Applicat. Signal Process. Audio Acoust. (WASPAA), Oct. 2005, pp. 82–85.

[8]. C. Breithaupt, T. Gerkmann, and R. Martin, “Cepstral smoothing of spectral filter gains for speech enhancement without musical noise,” IEEE Signal Process. Lett., vol. 14, no. 12, pp. 1036–1039, Dec. 2007.

[9]. R. Nickel, R. F. Astudillo, D. Kolossa, S. Zeiler, and R. Martin, “Inventory- style speech enhancement with uncertainty-of-observation techniques,” in Proc. ICASSP, Kyoto, Japan, Mar. 2012, pp. 4645–4648.

[10]. I. Cohen, “Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging,” IEEE Trans. Speech Audio Process., vol. 11, no. 5, pp. 466–475, Sep. 2003.

[11]. P. Vary and R. Martin, Digital Speech Transmission: Enhancement, Coding and Error Concealment. Chichester, U.K.: Wiley, 2006.

[12]. P. C. Loizou, Speech Enhancement - Theory and Practice. Boca Raton, FL, USA: CRC, Taylor and Francis, 2007.

[13]. X. Xiao and R. M. Nickel, “Speech enhancement with inventory style speech resynthesis,” IEEE Trans. Audio, Speech, Lang. Process., vol. 18, no. 6, pp. 1243–1257, Aug. 2010.

[14].J. Ming, R. Srinivasan, and D. Crookes, “A corpus-based approach to speech enhancement from non stationary noise,” IEEE Trans. Audio, Speech, Lang. Process., vol. 19, no. 4, pp. 822–836, May 2011.
[15].J. Ming, R. Srinivasan, and D. Crookes, “A corpus-based approach to speech enhancement from non stationary noise,” in Proc. INTERSPEECH, Makuhari, Japan, Sep. 2010, pp. 1097–1100.