Voice Stress Analysis using Linear Predictive Coding in MATLAB

Dinesh Kumar
Dept. Of Electronics and Communication, Maharaja Surajmal Institute of Technology, New Delhi, India.

Upasana Sharma
Dept. Of Electronics and Communication, Maharaja Surajmal Institute of Technology, New Delhi, India.

Shweta Solanki
Dept. Of Electronics and Communication, Maharaja Surajmal Institute of Technology, New Delhi, India.

Abstract – Speech in humans is generated by a series of formants generated by the resonances in the nasal, pharyngeal and vocal tracts. Linear prediction of speech is one of the most widely used techniques for analysis of speech samples. Voice stress analysis as a field of study holds immense importance to human life. Stress analysis has been used in critical situation endangering life and in several medical scenarios as well. This paper uses Linear Predictive Coding (LPC) to analyse the stress levels in the voice samples by comparing them to existing samples and outputting a human emotion.

Index Terms – LPC, MATLAB, stress analysis, emotion recognition.

1. INTRODUCTION

Linear Predictive Coding (LPC) was chiefly created for audio and speech signal processing. LPC processes the speech signal and compresses the spectral envelope of the waveform making it easier to analyse and compare particular sections of a waveform. Size and quality are important metrics when aiming to process speech via a wireless medium (mostly used in practical applications) and thus LPC is used to encode a speech signal at low bit rate but without compromising quality. This process flow contributes to proper comparison and accuracy of results. LPC was conceptualized by S. Saito and F. Itakura of NTT based on automatic phonetic discrimination using maximum likelihood approach. Speech analysis using LPC is devoid of transmission errors since it transcodes and transmits spectral data. This paper aims to analyse a speech signal by comparing it to a standard voice sample.

2. RELATED WORK

Various samples of speech were taken to detect the emotion. There are various steps involved in the process of detecting the emotion using speech. These steps have been discussed below:

- Feature extraction
- Feature selection
- Simulation results in MATLAB

2.1. Feature extraction

2.1.1 Features for Emotion Recognition System

One of the important aspect of design of a emotion recognition system is the choice of features that are best suited to characterize the different emotions efficiently. Feature extraction aims at calculating relevant features of speech signal this helps in removing the redundancy by extracting useful information. It transforms initial sets of measured data into derived features also called as feature vectors.

There are two ways either the speech signal There are some issues which must be considered in feature extraction.The first issue deals with region of analysis used to extract features. can be divided into intervals called frames from each a local feature vector is extracted or global statics can be extracted from whole speech utterance. Second issue is to decide best feature type e.g. pitch, energy etc.

Features like mean of the pitch or energy contours are dependent on signal duration. [5] Furthermore, speech rate, which is known to be useful for recognition of highly active emotions, is calculated as density of pitch or energy peaks over time. Therefore, a fragment of speech should be long enough to make computation of these features reliable.

Feature extraction is done by tracking summary statistics from different acoustic parameters of the speech signal. In feature extraction we exclude the linguistic features of speech but focus on acoustic and prosodic ones including pitch , power ,formant frequencies and their bandwidths ,linear prediction coefficients (LPC) [3] .Feature extraction process involves two steps: Spectral feature extraction and Prosodic feature extraction.

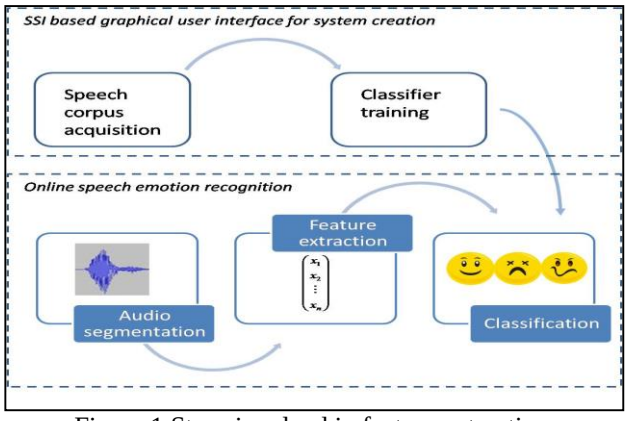

Figure 1 Steps involved in feature extraction

2.1.2 Spectral feature extraction using LPC

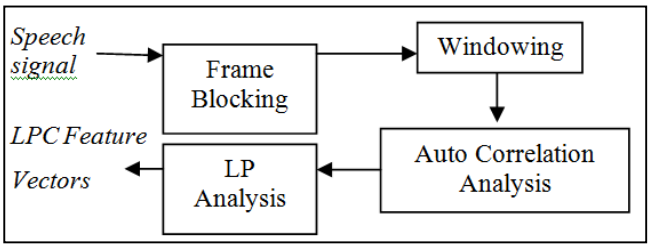

Figure 2 Steps involved in LPC

LPC is a tool used for representing spectral envelope of speech signal in compressed form. It is mostly used in audio signal processing and speech processing. LPC gives an accurate estimate of speech parameter. It approximates speech samples as a linear combination of past speech samples.

The predictor coefficients, a unique set of parameters are determined through minimizing the amount of squared differences between the actual parameters and predicted ones[2].LPC also helps in saving bits.

The basic steps of LPC processor includes:

- Pre emphasis: speech signal experiences some roll off. As a result, major part of the spectral density resides in low frequency component but the information in high frequency component is as important as low frequency so we provide boosting to high frequency components this is called pre emphasis.
- Frame blocking: the output of pre emphasis is blocked into frames of N samples with each frame separated by adjacent M samples.
- Windowing: data windowing lowers the variance of autocorrelation matrix estimate.
- Autocorrelation analysis: It determines the prediction error in terms of polynomial coefficients.
- LPC analysis: It converts each frame into LPC parameter set by using Durbin's method [8].

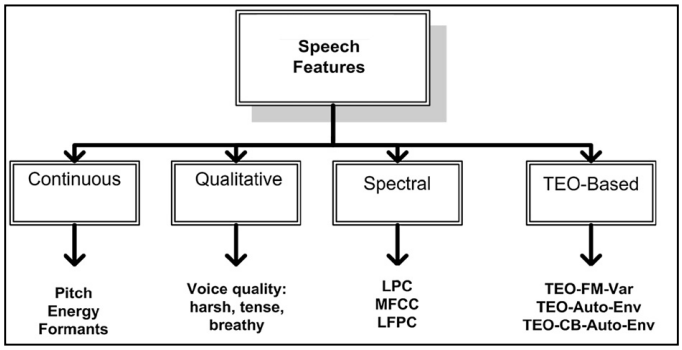

Figure 3 Categories of speech features

2.1.3 Prosodic feature extraction

In prosodic feature extraction, pitch loudness and formants of the speech signal are extracted and then minimum value, maximum value and statistical moments – mean, variance are calculated to obtain the feature vectors.

Formants are the meaningful frequency components of a speech signal.

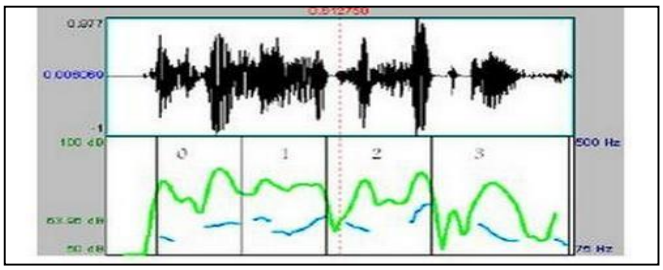

Figure 4 Speech signal, its energy contour and pitch contours

2.2 Feature selection

To identify the features which contribute more to in the classification out of the derived features which are input to classifiers. This determines those properties of speech or features are helpful in discriminating the emotions and improves classification accuracy[4]. The process of searching for subset of features that give best classification is time consuming .So forward selection and backward elimination is used to find the features that provide best classification. Forward selection sequentially add one feature at a time that most increases or least decreases classification accuracy. Backward features starts with all features and sequentially deletes the next feature that most decreases or least increases classification accuracy[6].

2.3 Simulation results in MATLAB

MATLAB R2008b is used for feature extraction and classification. Features are extracted from statistical moments

of the sequence. To select informative features forward feature selection algorithm is used.

Accuracy is calculated as the ratio of correctly labeled samples to the total samples.[4]

2.3.1 Preparation of raw database

Four Audio Samples corresponding to each Emotion are recorded. These samples are then read and sampled using the wavread() function in MATLAB. A unique vector is defined to store the sampled coefficients corresponding to each audio sample. Thus we obtain a total of 16 vectors at this stage i.e. four corresponding to each emotion.

2.3.2 Achieving uniformity of size of the Sampled Audio Sequences in the Raw Data Base

Each of the 16 vectors obtained in the previous stage have different dimensions i.e. they may or may not be row vectors. In order to convert them into row vectors we employ reshape() function.

2.3.3 Conditioning of the Raw Audio Samples

Conditioning implies re-moving Noise and regions of silence from these audio vectors.

- Removing the DC Component by subtracting the mean of each vector from that vector itself.
- Removing Regions of silence by using the myVAD() function in MATLAB.

2.3.4 Application of LPC to obtain Feature vectors

The audio samples obtained from the previous stage are free from noise and regions of silence. The next logical step is to remove redundancy from these audio vectors which will ensure robust extraction of features from the same. The LPC Feature Extraction technique is used to achieve this step. The LPC Feature Extraction Technique is readily available in the form of the predefined lpc() function in MATLAB. The resultant vectors obtained at this stage contain the features unique to the corresponding Emotion. Hence these vectors are known as feature vectors.

2.3.5 Combining all the obtained feature vectors to build a reference data base

In the previous stage we obtain 4 feature vectors corresponding to every emotion. The mean of the 4 Feature Vectors corresponding to each Emotion is computed to yield a unique Feature Vector for each emotion. Thus at this stage we are left with only 4 Feature Vectors. The Feature Vector corresponding to each emotion contains features that are unique to that emotion only. These 4 Feature vectors are now assembled to obtain the desired Reference Database matrix.

2.3.6 Use of Minimum Euclidean Distance Algorithm to predict the emotion contained in the input audio sequence

The reference database matrix obtained in the previous stage l now serves as a threshold for determining the emotion contained in the input audio sequence. The input audio Sequence is compared with the reference database using the Minimum Euclidean Distance Algorithm. The detected emotion is the one corresponding to that row of the Reference Database Matrix with which the input audio sequence has maximum correlation.

3. RESULTS AND DISCUSSIONS

The folllowing results were obtained :

Accuracies of Sadness, Anger, Happiness and Sadness were recorded and they vary from each other in the following manner:

	Frame Based		Voiced Segments	
Emotion	#Frames	Acc*(%)	# Segments	Acc(%)
Anger	180	0.65	10	0.70
Happiness	200	0.55	14	0.55
Sadness	100	0.75	11	0.82
Excitement	220	0.72	18	0.77

* Acc: Accuracy

Figure 5 Highest Accuracies resulted from each approach. The results are estimated from Trend lines, instead of original curves

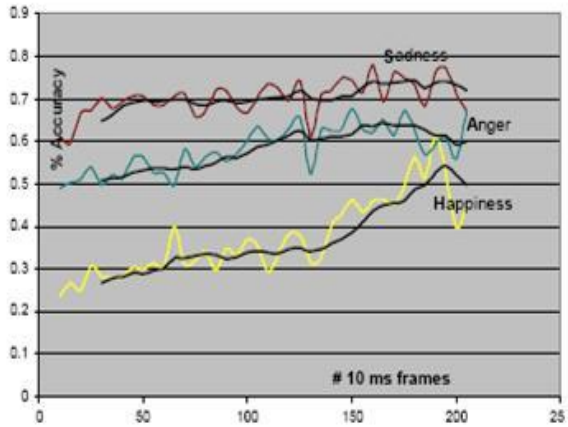

Figure 6 Accuracies of Sadness, Anger and Happiness using different number of frames per sample. over each curve, the moving-average trend line width period of 5 samples is also plotted.

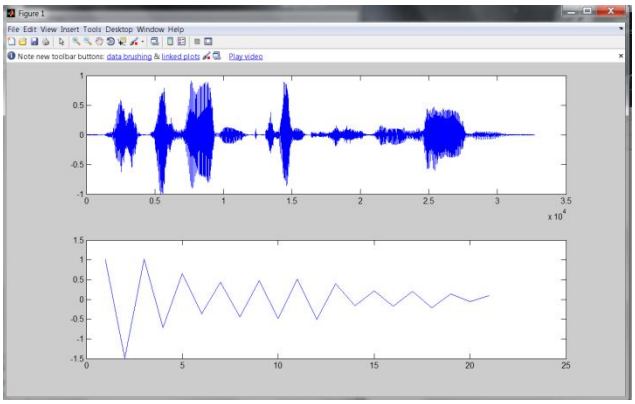

Figure 7 Applying LPC function on voice sample.

4. CONCLUSION

The various voice samples are compared with the Berlin database in order to understand and analyze their stress levels leading to identification of the emotion being expressed. Empirically, we have found that there is a directly proportional relationship between accuracy of the result and number of frames per sample.

The peak accuracy is maximum for sadness hovering around the 70% marks while the least peak accuracy is happiness which tops up at 55%. Anger has a peak accuracy between the two at 65%.

This paper showcases the practical usage of Linear Predictive Coding for voice stress analysis and proves that such a method can be used for industry relevant purposes

REFERENCES

[1] 'Survey on speech emotion recognition: Features, classification schemes and databases': Moataz El Ayadi , Mohamed S. Kamel , Fakhri Karray.

[2] 'A Study of Methods Involved In Voice Emotion Recognition': P. Bhardwaj, S. Debbarma

[3] Williams, U.; Stevens K. N., (1972). Emotion and Speech: some acoustical correlates,

[4] Murray, I. and Arnott, J. L., (2000).Towards the Simulation of Emotion in Synthetic Speech: A Review of the Literature on Human Vocal Emotion, in Journal of the Acoustic Society of America, pp.1097-1108, (1993).

[5] Petrushin, V. A., Emotion Recognition in Speech Signal: Experimental Study, Development and Application, ICSLP 2000, Beijing,

[6] 'Emotion Recognition by Speech Signals':*Oh-Wook Kwon, Kwokleung Chan, Jiucang Hao, Te-Won Lee*

[7] 'EmoVoice | A framework for online recognition of emotions from voice': Thurid Vogt, Elisabeth Andr_e, Nikolaus Bee

[8] 'Emotion Recognition with Speech for Call Centres using LPC and Spectral Analysis ':Rashmirekha Ram, Hemanta Kumar Palo , Mihir Narayan Mohanty

[9] 'Speech Emotion Recognition: Comparison of Speech Segmentation Approaches': Muharram Mansoorizadeh, Nasrollah.M. Charkari

[10] McGilloway. S. Cowie, R.; Doulas-Cowie, E.; Gielen, S.; Westerdijk, M.; Stroeve S. (2000). Approaching Automatic Recognition of Emotion from Voice: A Rough Benchmark,