International Journal of Engineering & Technology, 7 (3) (2018) 1531-1534

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i3.13285
Research paper

Word recognition from speech signal using linear predictive coding and spectrum analysis

Mandeep Singh [1] *, Gurpreet Singh [2]

[1] Assistant Professor, EIED, Thapar Institute of Engineering & Technology, Patiala, India
[2] Research Scholar, EIED, Thapar Institute of Engineering & Technology, Patiala, India
**Corresponding author E-mail:mdsingh@thapar.edu*

Abstract

This paper presents a technique for isolated word recognition from speech signal using Spectrum Analysis and Linear Predictive Coding (LPC). In the present study, only those words have been analyzed which are commonly used during a telephonic conversations by criminals. Since each word is characterized by unique frequency spectrum signature, thus, spectrum analysis of a speech signal has been done using certain statistical parameters. These parameters help in recognizing a particular word from a speech signal, as there is a unique value of a feature for each word, which helps in distinguishing one word from the other. Second method used is based on LPC coefficients. Analysis of features extracted using LPC coefficients help in identification of a specific word from the input speech signal. Finally, a combination of best features from these two methods has been used and a hybrid technique is proposed. An accuracy of 94% has been achieved for sample size of 400 speech words.

Keywords:*Speech Signal Processing; Isolated Word Recognition; Spectrum Analysis; Criminals.*

1. Introduction

In today's era of information technology, mobile communication is the fastest growing field and most popular medium of communication. It serves the humanity not only in formal interaction, but also to help business community, natural disaster victims, people in remote areas and other such emergency conditions. However, the same media is also being misused for enemies of the society like terrorists and rioters. These anti-social people use some typical words or phrases during their communication on mobile phones. If these interactions can intercept timely, then lot of lives can be saved from those anti social elements. Online monitoring of millions of calls is a huge task and very costly. However, if some selected calls (interaction over phone) , which includes some typical keywords like 'bomb', 'kill', 'attack', 'target' etc. can be identified and monitored for further investigation, then criminals can be intercepted. In this paper, a hybrid technique is proposed, which is able to recognize pre-determined isolated words from speech signals.

Every person has a different voice and different intonation. However, in case of similar pronouncing words, the automated word recognition system may produce errors. The intonation between the letters that form the word "KILL" differ from the ones and that form the word "FILL" (similarlyBUT and SHUT). Therefore, digital signal processing and analysis could reveal some features, which can be used to detect such words from a speech signal. As per the US department of homeland security, the list of suspicious words includes words such as 'attack', 'kill', 'terrorism' and 'dirty bomb' alongside dozens of seemingly innocent words like 'pork', 'cloud', 'team' and 'box' [1]. In the present study, speech signals of different 400 words have been collected from each young volunteer (age 19-21 years). All the 20 volunteers were of north Indian English accent. In total 400 words were collected in the form of digital speech signal using a SONY sound recorder with no background noise. By considering similar words (e.g. Fill and Kill) and analyzing them, we got some parameters that are unique for a particular word to identify. J.L. Plaza et al. found maximum frequency which recognized the vowel for speech impaired peoples [2]. A.K. Paul et al. presented time domain analysis for energy and zero crossing rate (ZCR), cepstral analysis for fundamental frequency,Linear Predictive coding (LPC) for formants and also synthesized it using MATLAB [3]. Nica et al. designed Bangla speech recognition system using LPC and Artificial Neural Network [4]. Ning used inverse of spectral features called cepstral for word recognition [5]. Murakami et al. estimated fundamental frequency from noise spectrum [6]. Daqrouq et al. achieved Speaker identification using power spectrum density [7]. Z. Weng et al. obtained Mel-frequency cepstral coefficients using Gaussian mixture density [8]. In most of the existing methods [9-10] to recognize an isolated word, frequency based analysis has been used. However, none of the researcher tried a hybrid technique. In the present research, a method is proposed by considering most discriminating features from Spectrum analysis and LPC analysis.

2. Methodology

Computer based decision making is the need of the hour in online monitoring systems, because it should be fast enough al time to analyze and take the decision in real time conditions. In this study, high quality speech signal has been recorded in 'wav' format and the same is captured in MATLAB for feature extraction. As shown in figure 1, two approaches Spectrum analysis and LPC analysis have been used to extract the features.

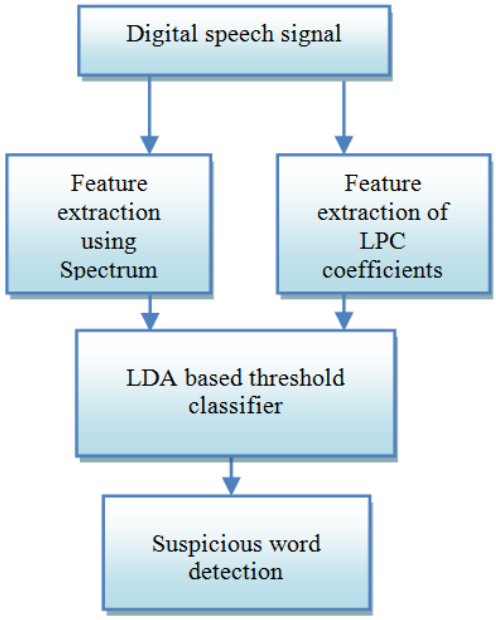

Fig. 1: The Block Diagram of Proposed Methodology

Linear discriminant analysis (LDA) has been used to compare the values of the most discriminating features. The threshold value has been achieved from the training set of 50% samples. Based on the training result threshold values of selected features have been passed to the LDA classifier. In the testing phase the remaining 50% speech signals have been evaluated to classify the word as 'suspicious' or 'normal'. In the end, the name of speech signal file is highlighted, which contains any suspicious word.

2.1. Spectrum analysis

The speech signals were analyzed using digital signal processing techniques. To convert our samples to a series of data that could be analyzed the sound samples were recorded in an archive with ".wav" extension, and was retrieved using "wavread" instruction, in MATLAB. Fast Fourier Transform (FFT) of the signal was obtained and its absolute value was plotted as shown in figures 2a.The spectrum analysis has been done as shown in figure 2b and the significant parameters are represented in the Table 1.

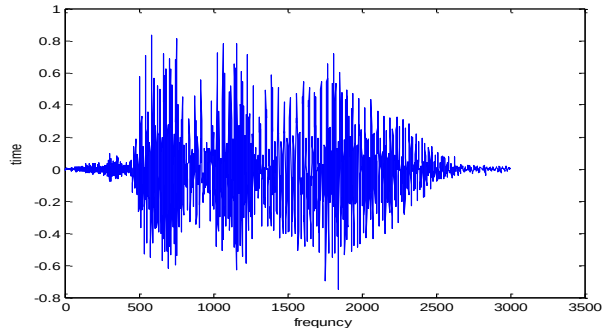

Fig. 2: A) Signal of the Word.

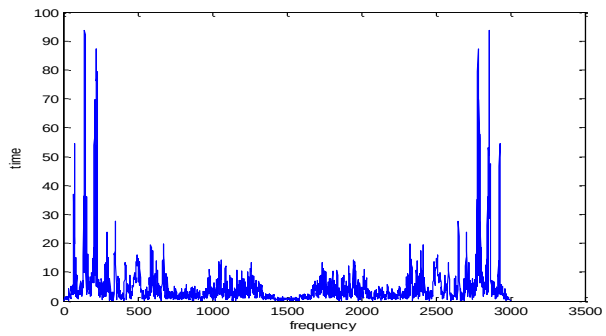

Fig. 2: B) Spectrum of Signal.

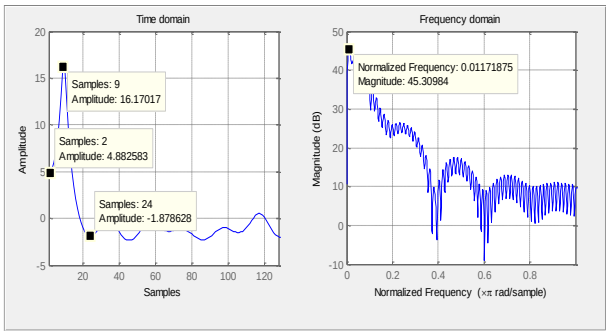

Fig. 3: Time–Frequency Domain Spectrum.

Table 1: Features Selected from Spectrum Analysis

S. No.	Feature
1	Maximum peak (MAX)
2	Minimum peak (MIN)
3	Slope of maximum peak (SLOP)
4	Width of maximum slope peak (WIDTH)
5	Root Mean Square Value (RMS)
6	Mean (MEAN)
7	Median (MEDIAN)
8	Standard Deviation (STD)
9	Total Harmonic Distortion (THD)
10	Inter Modulation Distortion (IMD)
11	Signal to Noise Ratio (SNR)
12	Peak Frequency (PFR)
13	Peak Amplitude (PAM)
14	Total Power (TP)

2.2. LPC analysis

Speech signals were recorded using an audio wave recorder in a room environment. The silence and pause portions were removed from the speech signal. The resulting signal was then filtered for the removal of unwanted background noise. The filtered signal was then windowed followed by LPC coefficient calculation. The feature extractor was based on a standard LPC Cepstrum coder, which converts the incoming speech signal into LPC Cepstrum coefficients [2]. The features extracted from LPC analysis are listed in Table 2. After the above mentioned steps, a new spectrum has been plotted in MATLAB, this is a simultaneous Time and Frequency domain spectrum of a signal and is shown in figure 3.

Table 2: Features Selected from LPC Analysis

S. No.	Feature
1	Maximum peak of LPC spectrum (MxP)
2	Minimum peak of LPC spectrum (MnP)
3	LPC1
4	LPC2
5	LPC 3

2.3. Fisher's linear discriminant analysis (LDA)

Fisher's LDA technique is very useful where there is two-class analysis. This technique is based on the statistical values of a feature.

$$\text{LDA Score} = \left|\frac{\mu_1-\mu_2}{\sqrt{(\sigma_1^2+\sigma_2^2)}}\right|$$

Here μ_1, μ_2 are the mean of same feature for two similar words (like Fill and Kill), σ_1, σ_2 are the standard deviation of same features of two similar words. This technique has been used to select most discriminating features from the available feature set which can be further used to differentiate among two classes. The feature which has higher value of LDA score, that feature has better ability to differentiate between two classes.

3. Results and discussion

Database of 400 sample words recorded by 20 volunteers, have been split into two equal parts. 50% of the data (400 samples by 10 random volunteers) has been used as training data to evaluate the most discriminating features and to formulate their ranges for threshold and further classification. The remaining 50% dataset has been used for testing the proposed method. Table 4 represents the features from spectrum analysis for the word 'Kill', and "Fill'. The word 'Fill' has been analyzed along with 'Kill' to select the most appropriate features for detecting the isolated word 'Kill'. In the same way, all the 400 sample words have been analyzed with respect to a similar sounding word. Table 4 shows the mean values with standard deviation for all the features for the given word. These values later used for calculation of LDA score.

Table 4: Mean and Std. Deviation of All the Features of the Word 'KILL' and 'FILL'

	For the word 'Kill'		For the word 'Fill'	
Feature	Mean (μ_1)	Standard deviation ± (σ_1)	Mean (μ_1)	Standard deviation ± (σ_1)
MEAN	4.51	0.19	4.53	0.63
MEDIAN	2.64	0.23	2.73	0.32
STD	6.49	0.29	6.53	0.85
RMS	7.94	0.15	7.46	0.68
MAX	68.24	3.89	96.66	3.59
MIN	0.004	0.002	0.009	0.004
SLOP	4.19	0.43	6.75	0.55
WIDTH	31.75	3.59	28.00	5.94
PFR(Hz)	336.43	20.45	333.98	11.16
PAM (dB)	-28.73	2.89	-29.70	6.36
TP(dB)	-17.20	1.21	-17.57	3.96
THD (%)	24.12	3.07	34.47	2.65
THD+N (%)	107.99	12.07	135.51	3.73
IMD (%)	35.37	5.50	15.54	2.15
SNR(dB)	-2.15	3.70	-3.13	0.09
MxP	6.47	0.99	13.24	3.36
MnP	-6.46	1.90	-7.60	1.55
LPC1	3.55	0.76	5.06	2.22
LPC2	0.53	0.07	1.56	0.14
LPC3	0.014	0.005	0.020	0.003

Based on the observations and results shown in Table 4, all the features have been evaluated through Linear Discriminant Analysis results [11] of the same parameters of different words. The top seven features having LDA score more than 1.00 are shown in Table 5. Since these features have high score of LDA, therefore, these [7] features have been selected for detection of the isolated word.

Table 5: LDA of Parameters for Words FILL-KILL

Parameters	LDA Score
$LPC2_{(FILL-KILL)}$	6.58
$Max_{(FILL-KILL)}$	5.36
$Slop_{(FILL-KILL)}$	3.66
$IMD_{(FILL-KILL)}$	3.35
$THD_{(FILL-KILL)}$	2.55
$THD+N_{(FILL-KILL)}$	2.17
$Mxp_{(FILL-KILL)}$	1.93

Higher the LDA values, more is the difference between the parameters. We show only upper range values (more than 1) as the lower range values are insignificant for differentiating a correct word. The parameter having higher LDA value can be differentiated better than the parameter having lower LDA value. The Figure 5 shows that Maximum Peak and Slope of Peak values have very large difference for the two words which helps in differentiating the one word from the other. Box plots of these two features also support the significance of their differentiating power.

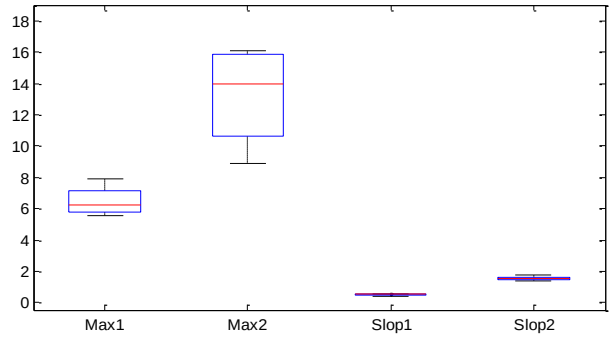

Fig. 5: Box Plots of Features for Words 'FILL and KILL' By LPC Spectrum Analysis.

Similarly, all the sample words have been analyzed and their best features have been identified. In the last step, these features have been passed to information fusion based binary classifier [12] and the required threshold value has been calculated for each word. The accuracy of the proposed features and the method has been evaluated by using the 'testing' dataset (the remaining 50% data set). Table 6 shows the accuracy of three different options;all features of Spectrum analysis, all features of LPC method and the third one (proposed) is considering the most significant features from both domains.

Table 6: Average Accuracy of Word Recognition by Different Methods

Method	Accuracy
Accuracy of Spectrum Analysis	90%
Accuracy of LPC method	92%
Accuracy of Hybrid technique (Proposed)	94%
Ning method [5] on same dataset	92%
Heidari and S. Gobee [9]	72%

4. Conclusion

In this paper, frequency spectrum analysis and LPC analysis methodologies have been used for isolated word recognition. The results show that every word has typical feature values, which can be used to detect that word in a speech signal. In this work, even similar-sound words have also been analyzed to increase the efficacy of the proposed method. Words with similar sound, still have some different features, which help in identify those wordse.g. words FILL and KILL sound similar, however features like Maximum Peak, THD, THD+N, IMD are different which candifferentiate them. In this study, Fisher's LDA and Box Plot methods are used to find most significant features. Finally, those most significant features were used with Information Fusion based binary classifier to recognize the isolated word with an accuracy of 94%, which better than the existing methods.

References

[1] D. Miller, "Revealed: Hundreds of words to avoid using online if you don't want the government spying on you", The Daily Mail UK, 26 May 2012.

[2] J. Leonardo, P. Aguilar, D. Báez-López, Luis Guerrero-Ojeda and J. Rodríguez, "A Voice Recognition System for Speech Impaired People", 14th International Conference on Electronics, Communications and Computers (CONIELECOMP) 2004.

[3] A. Kumar Paul, D. Das, Md. M. Kamal, "Bangla Speech Recognition System using LPC and ANN", 2009 Seventh International Conference on Advances in Pattern Recognition , 2009.

[4] A. Nica, A. Caruntu, G. Toderean, O. Buza, "Analysis and Synthesis of Vowels Using Matlab", IEEE, 2006.https://doi.org/10.1109/AQTR.2006.254662.

[5] D.Ning "Developing an Isolated Word Recognition System in MATLAB". Matlab Digest, 2010.

[6] T. Murakami and Y. Ishida, "Fundamental frequency estimation of speech signals using MUSIC Algorithm", Conference on Acoust. Sci. & Tech., 2001.

[7] K. Daqrouq, W.A. Sawalmeh, “Speaker Identification Wavelet Transform Based Method”, fifth International Multi-Conference on Systems, Signals and Devices 2008.
[8] Z. Weng, L. Li, D. Guo, “Speaker Recognition Using Weighted Dynamic MFCC Based on GMM”, ASID International Conference, 2010.https://doi.org/10.1109/ICASID.2010.5551341.
[9] H. Heidari and S. Gobee, "Isolated Word Command Recognition for Robot Navigation", International Symposium on Robotics and Intelligent Sensors 2012.
[10] E. Chandra, C. Sunitha, “A review on Speech and Speaker Authentication System using Voice Signal feature selection and extraction”, IEEE International Advance Computing Conference, 2009.
[11] R. Fisher, Mclachlan, “The Use of Multiple Measurements in Taxonomic Problems”, 1976.
[12] M. Singh, S. Singh and S. Gupta, “An information fusion based method for liver classification using texture analysis of ultrasound images”, Information Fusion, Vol.19, pp 91-96, 2014.https://doi.org/10.1016/j.inffus.2013.05.007.