The 4th International Conference
"Computational Mechanics
and Virtual Engineering"
COMEC 2011
20-22 OCTOBER 2011, Brasov, Romania

SPEECH DETECTION USING LINEAR PREDICTION CODING BASED ON CORRELATION COEFFICIENT CLASSIFICATION

Tariq Abu Hilal[1], Hasan Abu Hilal[2], Jamal Daqrouq[3]
[1] Dhofar University, Salalah, OMAN, tariq_abuhilal@du.edu.om
[2] Dhofar University, Salalah, OMAN, h_abuhilal@du.edu.om
[3]Gasco Fuel Company, Abu Dhabi, UAE, jdaqrouq@gasco.ae

Abstract: *In this paper, the assumption of Linear Predictive Coding (LPC) as functional to speech is utilized for speaker accurate pattern extraction. Correlation Coefficient as a statistical Measurment is used for features classification, a thresholding factor is adjusted. The system works with the recorded samples, the features tracking capability was excellent with the recorded dataset of National Institute of Standards and Technology (NIST/TIMIT Database); therefore the system can be applied in different speaker-recognition systems for user's authentication and verification. The proposed technique is simulated by MATLAB; the results show sufficient Recognition Rates (RR).*

Keywords: *linear predictive coding, correlation coefficient, Speaker Verification.*

1. INTRODUCTION

Speech detection and recognition by machine has been an aspiration of research for more than six decades and has stimulated such science, in spite of the glamour of designing an intelligent machine that can be able to identify the spoken word and realize its meaning, and regardless of the massive research hard work spent in trying to generate such a machine. The consciousness of that can be divided into two main parts: features extraction, followed by speaker's voices classification, based on the extracted features, still it is hard to achieve the desired ambition of a machine that can recognize spoken dialogue on any issue by all speakers in all environments. One of the majority complex aspects of research in speech recognition is its interdisciplinary temperament, and the affinity of most researchers to apply a monolithic approach to individual problems and disciplines that have been useful to speech recognition problems [1][2]. A basic question in speech recognition is how speech models can be evaluated to carry on their similarity (the expanse between patterns). Depending on the particulars of the recognition systems, pattern association can be done in a broad selection of ways, Concerning the applications of broadly ranges of automation for operator assisted services [3]. Basically the hypothesis is that the speech is comprised of a word or more and to be predictable as a complete unit with no unambiguity in the phonetic contented. Utterances detection algorithms consist of identical components (via time alignment), the calculated succession of spectral vectors of the spoken parts to be created, alongside for each position of the spectral patterns, and picking and building up the voice patterns. One more implied supposition is that every spoken utterance has a plainly distinct beginning and ending point, which could be found using some type of speech endpoint detector. LPC offers a good mold of the speech signal. This is particularly true for the quasi stable state voiced districts of speech in which the all-pole model of LPC offers fine estimates to the vocal zone supernatural envelope the fundamentals of how LPC has been useful in systems. The numerical details and mathematical derivations will be absent here. Prior to relating a general LPC front-end processor for speech recognition, it is valuable to appraise the reasons why LPC has been so extensively used [4, 5].
As significance, pattern matching could be dependably completed, without need to be worried about suspicions in the endpoints of the patterns being compared. This model's function is totally suitable, some applications where the speech to be well-known, those consist of a sequence of words from the recognition vocabulary, such a paradigm is often improper for practical reasons. We present the efficiency and ease of the correlation method, to be applied on the LPC calculated coefficients, to be applied for identical text from different speakers; we investigate the differences of the processed signal. The organization of this paper is as follows. In Part 2, the speech recognition process is summarized. In Section 3, we discuss the proposed system. In Section 4, we illustrate and discuss the results of the proposed method. Finally, conclusions are stated out in Section 5.

2. SPEAKER RECOGNITION PROCESS

The speaker recognition model starts with generating a speech signal by speaking a complete given words, the spoken production is decoded into speech signal as a vector of values, the process of speaker recognition is a combination of the input step, signal processing step, then classification and recognition step. First, a stored data set is used to be processed , afterward the manipulation stage, finally, features of each speech signal are stored as reference features for comparisons, verification is to minimize the error rate and to achieve precise recognition rate [IANG]. The way in which LPC is useful to the analysis of speech signals guides to a logical source-vocal tract disjointing. As an outcome, a frugal depiction of the vocal tract uniqueness which we recognize are in a straight line related to the speech sound being shaped becomes achievable. It is a systematically tractable model. The method of LPC is mathematically accurate and is trouble-free and straightforward to implement in either software or hardware [6].
The feature vectors of speech are used to create a blueprint for each speaker, the number of reference patterns required for successful speech recognition application depends on the sort of features and techniques that the system utilize for identifying any dialogue, those features are extracted from an entered signal of any speaker to be verified, then the acceptance depends on similarity comparison with the stored pattern. As known, the LPC sculpt functions well in speech recognition applications. Incidents have revealed that the presentation of speech recognizers based on LPC front ends, they are more improved than the recognizers those depend on filter-bank front ends.
Correlation coefficient measures differences between the entered and the stored dataset vectors, the difference between an inserted dialogue and all other dataset stored patterns, the same speech vectors will yield high similarity, and then accepted as identical signals. Many methods were proposed to do so in the literature. The similarity in speech recognition is calculated only between the input signal and the stored patterns of the other recorded speakers. If the result is less than an examined threshold, then the word will be known, otherwise will be discarded and rejected.
Speech Recognition is considered as identifying the same speech dialogues or words. A given utterance based on the information restricted in speech signals is the process of deciding identical ones from a known set of others. A fault that may take place in speech recognition is the false rejection; two identical words are rejected, and false acceptance, two different words accepted as the same [7].
Most of speaker recognition systems use a categorization towards stored speaker's processed label and compare with a pre-defined rate, verification and rejection processes depend on a predefined threshold value, if the founded deference is less, the word is acknowledged, if not it will be declined.
The use of recorded dataset, which is subjectively chosen at all times, by electronic recording devices. A text will be provoked as a computer driven and text dependent speaker recognition system. With the unification of speaker and speech recognition systems, over the development in speech recognition precision, the characteristic between text dependent and independent systems will eventually decrease. The text dependent speaker recognition is the most commercially feasible and useful application, even though there is a lot of research which passed through both of them. Due to the promises offered, many trials done to the text dependent methods of speech recognition ignoring their complexity. [1]

3. THE PROPOSED SYSTEM

In this paper Linear Predictive Coding is built and presented, it consists of two main parts; LPC features extracting and similarity comparison statistically, reducing the complexity of using Neural Network. The benefit of the system than the other classification methods is the sufficient features tracking and speedy identification (i.e. the neural networks complication). The transform depends on LPC with a selected coefficients magnitude, to detect the differences of the signal's behavior, to find the exact variation in the occurrence. That what exactly happens in non-stationary signals, such as speech signal. Then the correlation coefficient is used on the extracted coefficients to classify and verify.
The processing of voice as the core and the only way of giving speakers identity has distinct connected troubles. The individuality of a speaker's voice can be degraded by the characteristics of the communication channel or by the surroundings noise. Consequently, speech recognition systems should be able to recognize a wide range of variations of the user's voice. The competence of similar voice characteristics would allow other speakers to be acknowledged by the system. For a high quality speaker recognition, understanding of the human formants speech recognition is very essential and thus speech recognition should depend on a thorough study for components that are used by humans in speech detection. Finding the steady principles of voice, therefore the most significant task for speech recognition. SK [8]
LPC coefficients are just a sampled version of the whole voice signal, and there calculation may devour important amount of time and assets, depending on the resolution requisite. (LPC), which is based on coding, it generates a fast computation of the wave resonance. It is easy to put into practice the totaling time and resources required [Tariq]. Number of coefficients is one of the mainly extensively used signal processing factor, which can be realized by adjustment for the density of the signal that can be measured for the conveyed amount of features information [9].
The speech signal holds many levels of information; above all a message is carried from the beginning to the end of the spoken words. In each level, there exists specific information about the language being spoken, the feeling, gender, and the uniqueness of the speaker. While speech recognition sets its goals at recognizing the spoken words in speech, the automatic identification of speaker and speech recognition are very vigorously associated, the mean of automatic

speaker recognition is to characterize the speaker by extraction, categorization then recognition of the information conveyed in the speech signal [10]
Features are the frequency parts of speech signal those are related to the human distinct vocal tract anatomy form, which is distinguishable for each person's resonance. We use these formants as the basic speaker features carriers [9], which is determined by LPC function equation 1 and the error is shown in equation 2 and 3. which is estimated by using MATLAB ready Library.

LPC determines the coefficients of a forward linear predictor by minimizing the prediction error in the least squares sense. It has applications in filter design and speech coding.

The most common representation is

$$\hat{x}(n) = \sum_{i=1}^{p} a_i x(n-i) \qquad 1$$

Where $\hat{x}(n)$ is the predicted signal value, $x(n-i)$ the previous observed values, and a_i the predictor coefficients. The error generated by this estimate is

$$e(n) = x(n) - \hat{x}(n) \qquad 2$$

$x(n)$ is the true signal value.

These equations are valid for all types of (one-dimensional) linear prediction. The differences are found in the way the parameters a_i are chosen.

For multi-dimensional signals the error metric is often defined as

$$e(n) = \left\| x(n) - \hat{x}(n) \right\| \qquad 3$$

Where $\|.\|$ is a suitable chosen vector norm.

Correlation coefficient is calculated as:

The correlation coefficient matrix represents the normalized measure of the strength of linear relationship between variables as shown in equation 4.

The correlation coefficient $r_{X,Y}$ between two random variables X and Y with expected values μX and μY and standard deviations σX and σY is their covariance normalized by their standard deviations, as follows

$$r_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_x \sigma_Y} = \frac{E\left((X - \upsilon X)(Y - \upsilon Y)\right)}{\sigma_x \sigma_Y} \qquad 4$$

Where E is the expected value operator and cov means covariance. Since μX = E(X), σX2 = E(X2) − E2(X), and likewise for Y, $r_{X,Y}$

Where ρ is the correlation coefficients and E[.] denotes the expectation of the product of the speech signal model, vector X is about the mean value, and the speech signal vector Y is about the mean value that related to the product of the Standard Deviation of X(σX) and Standard Deviation of Y (σY). ρ is efficient likeness or similarity tool judgment between the two vectors X and Y in terms of unity (out of one hundred percent similarity).

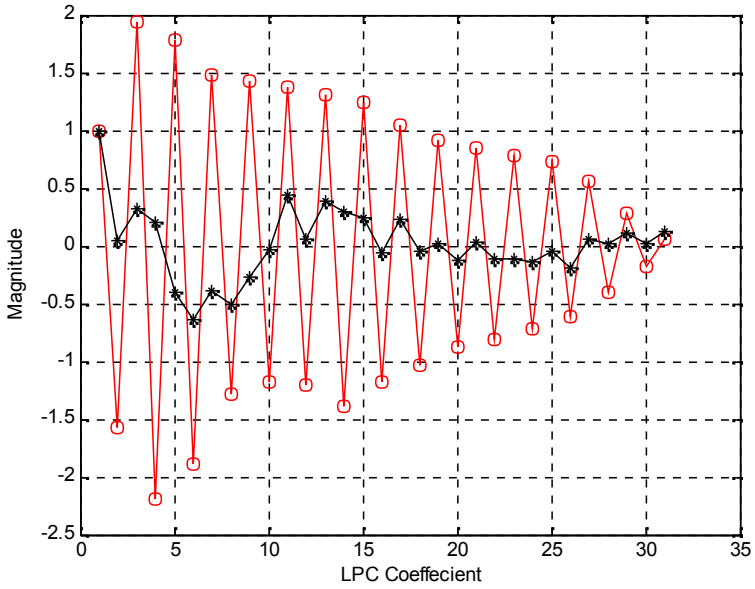

Figure 1: Two Speakers with different dialogue

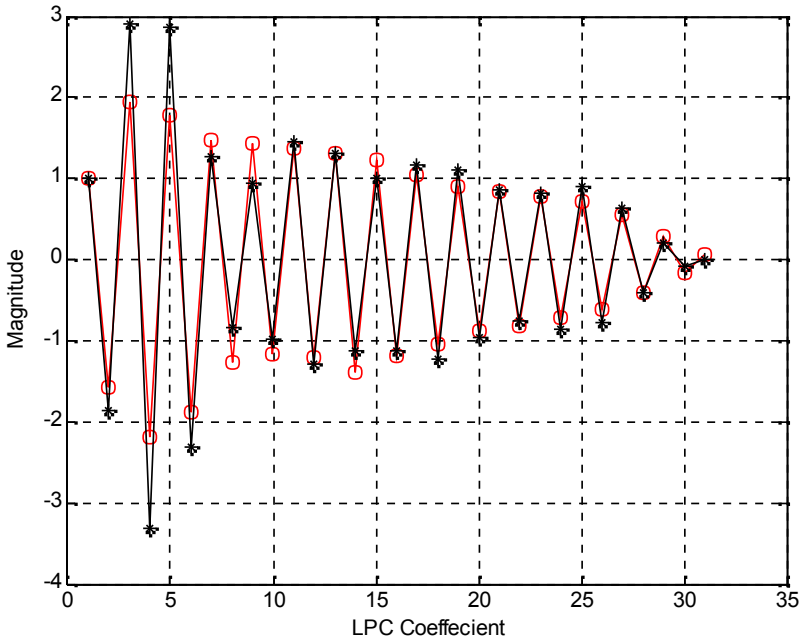

Figure 2: Two speaker with same dialogue

4. SIMULATION RESULTS

Linear Prediction coding (LPC) based speech feature extraction method is investigated by the use of the correlation coefficients. The introduced system depends on two steps: features extraction by LPC over fixed number of coefficients, due to its better capability of formants and features extraction from the signal's frequency, which is more suitable for non stationary signals, this is shown in Figure 1 and 2, and classification based statistical analysis of each dialog vector, in other words, the correlation value of the new input vector is compared with a predefined threshold, the system recognizes and refuses if the result has a low correlation coefficient, and will accept true results based on high values.

The system works with sufficient capability for features tracking and detection, simulation has been conducted on 50 dissimilar speakers. The threshold value between the users is adjusted to minimize the error and maximize the speech

recognition rate. This system reduces the complexity of using other methods like Neural Networks, with sufficient detection results. Text -independant system is used, with MATLAB package simulation tools.
For two different speakers with ten audio recordings, each for the same text uttering, processed and classified; Figure 2 shows the voice signals after processing with LPC, afterward classification results will be done by correlation coefficient.

5. CONCLUSION

Throughout this paper, Linear Predictive Coding (LPC) technique is used to speaker feature extraction and classified by the use of the correlation coefficient. The introduced system depends on two steps features extraction over fixed coffecients number, due to its ability to change signal illustration over the frequency, and classification using the correlation coefficient.
The system works with sufficient results of classification and similarity detection, the system can be applied for some speech classification problems. This Paper presented a description and performance evaluation for new efficient speaker recognition. In these methods, each one utilizes the voice signal, which provides more accurate results. The Process was implemented on the MATLAB simulator, in order to evaluate and compare the efficiency, four scenarios were simulated. The performance of the proposed systems was assessed by calculating the similarity and the correlation coefficient. These parameters estimate the speaker acceptance or rejection according to known threshold. In addition, all trials are evaluated, compared, and analyzed.
To conclude, this speech recognition system is designed to investigate the process of automatic speech recognition. The system is operated in text-independent mode using TIMIT database samples which contains of recorded utterances for different speakers. To sum up, this work can be summarized as follows:
The LPC method can be easily implemented, since it became a ready tool, and it can provide significant performance, it can extract the features in the voice signal is dependent on the coefficient number. The recognition by Correlation coefficient method comes to less cost, where other techniques like Neural Networks may give higher recognition rates. Although many current progress and successes in speech recognition have been attained, there are still many pitfalls for which high-quality solutions remain to be found. Most of these problems occur because of including speaker abnormality, and variations in channel and recording environment. It is essential to examine feature parameters that are stable over time to better detection results, insensitive to the variation of speaking way, depends on vocal resonance, including the speaking level and robust against variations in voice quality due to causes like throat infections and problems. It is also important to develop a way to cope with the problem of distortion due to telephone sets and noises.

REFERENCES

[1] Tariq Abu Hilal, H. Abu Hilal and K. Daqrouq, Speaker Verification System Using Discrete. Wavelet Transform And Formants Extraction. Based On The Correlation Coefficient, IAENG 2011, Hong Kong.
[2] K. Daqrouq, T. Abu Hilal, M. Sherif, S. El-Hajjar, and A. Al-Qawasmi,"Speaker Identification System Using Wavelet Transform and Neural Network," IEEE 2009 International Conference on Advances in Computational Tools for Engineering Applications.-2009 Advances in Computational Tools for Engineering Applications, Lebanon
[3] J.Wu -D., Lin B.-F. , "Speaker identification using discrete wavelet packet transform technique with irregular decomposition Expert Systems with Applications", 36 (2009) 3136-3143.
[4] D. Avci , "An expert system for speaker identification using adaptive wavelet sure entropy", Expert Systems with Applications, 36 (2009) 6295-6300.
[5] Reynolds, D. A., Quatieri, T. F., and Dunn, R. B. (2000). "Speaker verification using adapted gaussian mixture models". Digital Signal Processing," 10(1-3), 19-41.
[6] B. Imperl, "Speaker recognition techniques, Laboratory for Digital Signal Processing", Faculty of Electrical Engineering and Comp. Sci., Smetanova 17, 2000 Maribor, Slovenia.
[7] D. Ranjan Panda and Chittaranjan Nayak, "Eye Detection Using Wavelets And Ann", A Thesis Submitted In Partial Fulfillment Of Department Of Electronics and Instrumentation Engineering National Institute Of Technology Rourkela-769008 -(2007).
[8] N. Rao and A. Govardhan, "Comparative Study of Visible Reversible Watermarking Algorithms" Image Security Paradigm, Engineering College, Hyderabad, A P, India, Vol. 7, No. -177 (2, April 2010).
[9] W. Al-Sawalmeh, Khaled Daqroug and Tareq Abu Hilal, "The use of wavelets in speaker feature tracking identification system using neural network," WSEAS Transactions on Signal Processing archive Volume 5 , Pages: 167-177 Year of Publication: 2009 ISSN:1790-5022 Issue 5, (May 2009).
[10] J. M. Naik, L. P. Netsch, and G. R. Doddington. "Speaker verification over long distance telephone lines." In IEEE Proceedings of the 1989.