Global Journal of Computational Intelligence Research.
ISSN 2249-0000 Volume 1, Number 2 (2011), pp. 159-166

http://www.ripublication.com/gjcir.htm

Recognition of Vernacular Language Speech for Discrete Words using LPC Technique

[1]Omesh Wadhwani and [2]Prof. Amit Kolhe

[1]M.Tech. Student (Digital Electronics)
[1&2]Department of Electronics and Telecommunication
Rungta College of Engineering, Bhilai, Chhatisgarh, India
E-mail: sand_t@rediffmail.com, sand_t@rediffmail.com

Abstract

Vernacular language spoken in various countries creates a limitation on software associated with speech recognition. This paper is an attempt to overcome such problem. The suggested work makes use of Linear Predictive Technique for better interpretation of spoken words. The rule based structure of fuzzy suits very well with closeness of vernacular speech recognition. In this paper we study the feasibility of Speech Recognition with fuzzy neural Networks for discrete Words Different Technical methods are used for speech recognition. Most of these methods are based on transfiguration of the speech signals for phonemes and syllables of the words. We use the expression "word Recognition" (because in our proposed method there is no need to catch the phonemes of words.). In our proposed method, LPC coefficients for discrete spoken words are used for compaction and learning the data and then the output is sent to a fuzzy system and an expert system for classifying the conclusion. The experimental results show good precisions. The recognition precision of our proposed method with fuzzy conclusion is around 90 percent.

Keywords: Vernacular, Words Recognition, Linear Predictive Coding, Feature Extraction, Automatic Speech Recognition, LPC Coefficients, Word error rate

Introduction

The vernacular speech of a particular community is the ordinary speech used by people in a particular community that is noticeably different from the standard form of the language. Especially where Europeans languages were concerned, the linguists of the past normally concentrated on the standard forms of languages. Nonstandard

vernacular forms were silently ignored, excepting only in the study of regional dialects, for which the speech of elderly rural speakers was considered most appropriate; at the same time, the speech of younger speakers or of urban speakers was similarly ignored. Interest in vernacular forms developed only slowly during the twentieth century [1], but it became increasingly prominent with the rise of sociolinguistics in the 1960s. Today, there is intense interest in vernacular forms of the speech.

Automatic Speech Recognition

The goal of automatic speech recognition (ASR) is to take a word from microphone as input, and produce the text of the words spoken. The need for highly reliable ASR lies at the core of many rapidly growing application areas such as speech interfaces (increasingly on mobile devices) and indexing of audio/video databases for search. While the ASR problem has been studied extensively for over fifty years, it is far from solved. There has been much progress and ASR technology is now in widespread use; however, there is still a considerable gap between human and machine performance, particularly in adverse conditions.

The performance of any speech recognition system can be improved by choosing proper symbols for representation. Characters of the language are chosen as symbols for the signal-to-symbol transformation module of our speech-to-text system being developed for the Indian language Hindi. The aim here is to emulate human processes as much as possible at the signal-to symbol transformation stage itself. In this case, the expert systems approach permits a clear distinction between the domain knowledge and the control structure needed to manipulate the knowledge. A number of speech recognition systems for continuous speech have been with varied success. The main drawback in these systems is that they use a simple approach for signal-to-symbol transformation with some abstract units as symbols, thereby increasing the complexity at higher levels of processing. Recent efforts [2, 5] try to improve the performance of signal-to-symbol transformation using speech specific knowledge.

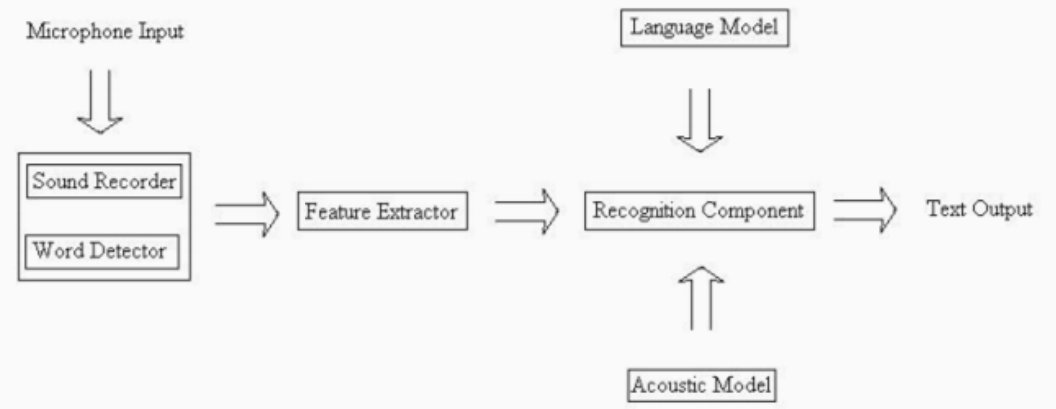

Figure 1: Block Diagram of Speech Recognition

Approach

Step One

Sound Recording and Word detection component is responsible for taking input from microphone and identifying the presence of words. Word detection is done using energy and zero crossing rate of the signal.

Step Two

Feature Extraction component generated feature vectors for the sound signals given to it. It generates Mel Frequency Cepstrum Coefficients and Normalized energy as the features that should be used to uniquely identify the given sound signal.

Step Three

Recognition component is the most important component of the system and is responsible for finding the best match in the knowledge base, for the incoming feature vectors.

Step Four

Knowledge Model: The component consists of Word based Acoustic. Acoustic Model has a representation of how a word sounds.

Linear Predictive Coding (LPC)

Linear predictive coding (LPC) is a tool used mostly in audio signal processing and speech processing for representing the spectral envelope of a digital signal of speech in compressed form, using the information of a linear predictive model. It is one of the most powerful speech analysis techniques, and one of the most useful methods for encoding good quality speech at a low bit rate and provides extremely accurate estimates of speech parameters.

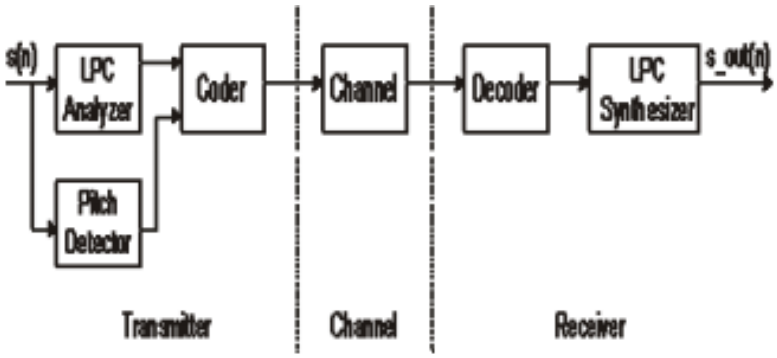

Fig. 2 Block diagram of an LPC vocoder

Figure 2: Block Diagram of an LPC Vocoder

It has two key components: analysis or encoding and synthesis or decoding [2]. The analysis part of LPC involves examining the speech signal and breaking it down into segments or blocks. Each segment is than examined further to find the answers to several key questions:

- Is the segment voiced or unvoiced?
- What is the pitch of the segment?
- What parameters are needed to build a filter that models the vocal tract for the current segment?

All vocoders, including LPC vocoders, have four main attributes: bit rate, delay, complexity, quality. Any voice coder, regardless of the algorithm it uses, will have to make tradeoffs between these attributes.

The first attribute of vocoders, the bit rate, is used to determine the degree of compression that a vocoder achieves. Uncompressed speech is usually transmitted at 64 kb/s using 8 bits/sample and a rate of 8kHz for sampling. Any bit rate below 64 kb/s is considered compression. The linear predictive coder transmits speech at a bit rate of 2.4 kb/s, an excellent rate of compression.

Delay is another important attribute for vocoders that are involved with the transmission of an encoded speech signal. Vocoders which are involved with the storage of the compressed speech, as opposed to transmission, are not as concern with delay. The general delay standard for transmitted speech conversations is that any delay that is greater than 300 ms is considered unacceptable. The third attribute of voice coders is the complexity of the algorithm used. The complexity affects both the cost and the power of the vocoder. Linear predictive coding because of its high compression rate is very complex and involves executing millions of instructions per second. The final attribute of vocoders is quality. Quality is a subjective attribute and it depends on how the speech sounds to a given listener.

Feature Extraction of Spoken Words using LPC

Feature Extraction refers to the process of conversion of sound signal to a form suitable for the following stages to use. Feature extraction may include extracting parameters such as amplitude of the signal, energy of frequencies, etc

Linear prediction is a good tool for analysis of speech signals. Linear prediction models the human vocal tract as an *infinite impulse response* (*IIR*) system that produces the speech signal. For vowel sounds and other voiced regions of speech, which have a resonant structure and high degree of similarity overtime shifts that are multiples of their pitch period, this modeling produces an efficient representation of the sound. Figure 2 shows how the resonant structure of a vowel could be captured by an IIR system.

LPC analyzes the speech signal by estimating the formants, removing their effects from the speech signal, and estimating the intensity and frequency of the remaining buzz. The process of removing the formants is called inverse filtering, and the remaining signal is called the residue. The numbers which describe the formants and the residue can be stored or transmitted somewhere else. LPC synthesizes the speech

signal by reversing the process: use the residue to create a source signal, use the formants to create a filter (which represents the tube), and run the source through the filter, resulting in speech. Because speech signals vary with time, this process is done on short chunks of the speech signal, which are called frames. Usually 30 to 50 frames per second give intelligible speech with good compression.

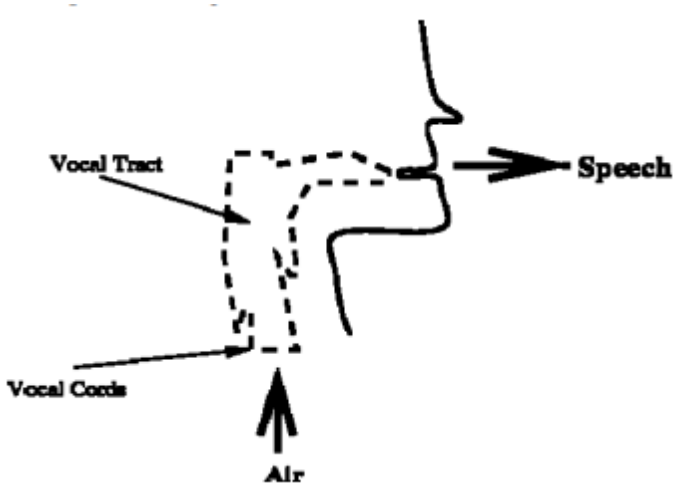

Figure 3: Physical Model of Human.

The basic problem of the LPC system is to determine the formants from the speech signal. The basic solution is a difference equation, which expresses each sample of the signal as a linear combination of previous samples. Such an equation is called a linear predictor, which is why this is called Linear Predictive Coding. The coefficients of the difference equation (the prediction coefficients) characterize the formants, so the LPC system needs to estimate these coefficients. The estimate is done by minimizing the mean-square error between the predicted signal and the actual signal.

Experimental Results and Observations

MATLAB software has various inbuilt functions to implement audio functions and coefficient evaluation. Spoken words of speaker were stored in a bank and their LPC coefficients were determined along with energy and zero crossover detection as well [3]. Later the words were spoken by another speaker whose phonemes and accent were sent to fuzzy analyzer for correct interpretation in intelligent way using fuzzy approach discussed above. Expert system was tested on 120 utterances in English spoken by two male speakers. It was observed that with just two parameters (total energy and first linear prediction coefficient) along with their fuzzy thresholds, spoken words were identified with more than 90% accuracy.*ess*

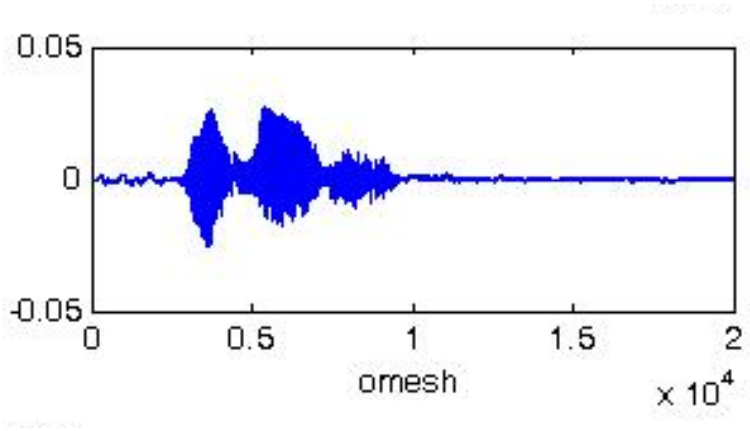

Figure 4: Wave Plot for word ‘omesh’

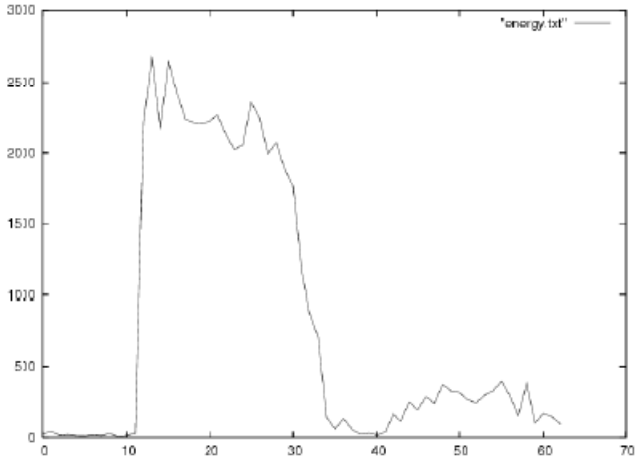

Figure 5: Energy plot for spoken word “omesh” in vernacular language

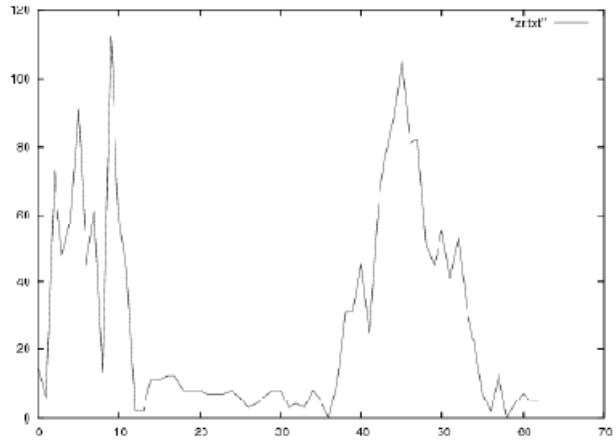

Figure 6: Zero Crossing plot for spoken word “omesh” in vernacular language

Conclusions

Linear Predictive Coding is an analysis/synthesis technique to lossy speech compression that attempts to model the human production of sound. Linear Predictive Coding achieves good bit rate which makes it ideal for secure telephone systems. Secure telephone systems are more concerned that the content and meaning of speech, rather than the quality of speech, be preserved. The trade off for LPC's low bit rate is that it does have some difficulty with certain sounds and it produces speech that sound synthetic. LPC encoders break up a sound signal into different segments and then send information on each segment to the decoder. The encoder send information on whether the segment is voiced or unvoiced and the pitch period for voiced segment which is used to create an excitement signal in the decoder. Vernacular language work is not concluded yet. Hence this paper is light on such approach for enhancing recognition power of intelligent techniques along with feature extraction. Experimental results also confirm the same.

Acknowledgment

The Authors place on record their grateful thanks to the authorities of Rungta College of Engineering for providing all the facilities for accomplishing this paper.

References

[1] M. Forsberg. 2003. Why Speech Recognition is Difficult. Chalmers University of Technology.

[2] L. R. Rabiner and R. W. Schafer, Digital Speech Processing. Prentice-Hall, 1978.

[3] J. R. Deller, J. H. L. Hansen, J. G. Proakis. Discrete-time Processing of Speech Signals. IEEE Press. 1993.

[4] Thiang, Suryo Wijoyo. Speech Recognition using LPC and Artificial Neural Network for Controlling the movements of Robot, 2011

[5] N.Uma Maheswari, A.P.Kabilan, R.Venkatesh "Speech Recognition system based on phonemes using neural networks". JCSNS International Journal of Computer Science and Network Security, Vol.9 No.7, July 2009

[6] Asim Shahzad, Romana Shahzadi, Farhan Aadil "Design and software implementation of Efficient speech recognizer". International Journal of Electrical & Computer Sciences IJECS-IJENS Vol. 10.

[7] R. Rodman, Computer Speech Technology. Artech House, Inc. 1999, Norwood, MA 02062.

[8] C. H. Lee; F. K. Soong; K. Paliwal "An Overview of Speaker Recognition Technology". Automatic Speech and Speaker Recognition: Advanced Topics. Kluwer Academic Publishers, 1996, Norwell, MA.

[9] S. R. Jang, "Neuro Fuzzy Modeling Architectures, Analyses and Applications". University of California, Berkeley, Canada, 1992.

[10] P. P. Bonissone, "Adaptive Neural Fuzzy Inference Systems (ANFIS): Analysis and Applications", technical report, GE CRD, Schenectady, NY USA, 2002.
[11] P. Eswar, S.K. Gupta, C. Chandra Sekhar, B. Yegnanarayana and K. Nagamma Reddy, An acoustic phonetic expert for analysis and processing of continuous speech in Hindi, in *Proc. European Conf. on Speech Technology*, Edinburgh, vol. 1 (1987) 369-372.
[12] J.P. Haton, Knowledge based approach in acoustic phonetic decoding of speech, in: H. Niemann, M. Lang and G. Serger, Eds., *Recent Advances in Speech Understanding and Dialog Systems*, NATO-ASI Series, vol. 46, (1988) 51-69.
[13] D.H. Klatt, Review of the ARPA speech understanding project, *J. Acoustic. Soc. Amer.* 62(6) (1978) 1345-1366.
[14] W.A. Lea, Ed., *Trends in Speech Recognition* (Prentice Hall, Englewood Cliffs, N J, 1980).
[15] R.De Mori, A. Giordana, P. Laface and L. Saitta, Parallel algorithms for syllable recognition in continuous speech, *IEEE Trans. Pattern Analysis" Machine Intelligence.*7(1) (1985). 55-68
[16] D. O'Shaughnessy, *Speech Communication - Human and Machine* (Addison-Wesley, Reading, MA, 1987).
[17] B. Yegnanarayana, C.C. Sekhar, G.V.R. Rao, P. Eswar and M. Prakash, A continuous speech recognition system for Indian languages, in: *Proc. Regional Workshop on Computer Processing of Asian Languages*, Bangkok (1989) 347-356.

Author Biography

Omesh Wadhwani is a B.E graduate from Kavikulguru Institute of Technology and Science, Ramtek with the specialization in Information Technology. He owns more than two year of Teaching experience.

Currently, He is doing his Master of Technology in Digital Electronics from Rungta College of Engineering and Technology, Bhilai. His area of interest are Speech processing and Pattern Recognition, Image Processing, Database Management systems and Artificial Intelligence.