International Journal of Advanced Research in IT and Engineering

ISSN: 2278-6244

PROPOSED A METHOD BASED ON FUZZY LOGIC TO DETECT THE DRIVER'S VOICE COMMANDS AND APPLIED TO AUDIO SYSTEM IN A NOISY ENVIRONMENT

Navid Samimi Behbahan*

Zohreh Mousavinasab*

Abstract: T*his research investigates the recognition of limited letters in the noisy environment. This issue would be considered, when the driver listening to the music and driving the car simultaneously, which cause to reduce the concentration of driver on driving the car. The letters that driver would use are Next (next track), Previous (previous track), Louder (increase the volume), Lower (decrease the volume). In order to describe any frame, the LPCC constants and the sound energy with the first and second differentiations (39 characterizations in total) have been used.*

---

*Sama Technical and Vocational Training College, Islamic Azad University, Omidiyeh Branch, Omidiyeh, Iran.

1- INTRODUCTION

The value of creating the technology to combine and recognize voices is enormous. Speaking is the most popular way to communicate fast and effective between humans. Writing and typing the letters with keyboards and other electrical controllers that have several keys and buttons can be replaced by the voice recognition technology. But, this new technology needs to be improved for commercial purposes. Voice recognition can make the computer easier for physically disabled people with good hearing and talking abilities. Voice combination not only uses the voice recognition techniques, but also can be used as a friendly outgoing device to diagnose different sounds and key words in order to replace the vision signs (such as traffic lights and etc.) and hearing signs (such as alarms and etc.) with these key words in different situations.

It should be considered that the progressing in voice recognition technology, not only covers the DSP area, but also, needs to have a good knowledge in artificial intelligence and artificial neural networks areas. Using these various sciences not only doesn't make to find the best and ideal [1].

Using several scientific areas to investigate the voice recognition technique not only doesn't make it hard to investigate, but also, increases the chance of achieving the ideal and efficient system. The voice recognition technology is the new technique to distinguish messages and voice commands in this area.

In the field of voice pattern recognition using the fuzzy logic, there isn't enough publications in Iran, and the limited publications in this area only focused on general investigations and the introduction for these types of technique. The results conducted by this study are applicable and outcomes by the software with Matlab coding. The results have been represented in graphs and tables at the end of this research.

2- DATA RANGE

The range of words is one of the important factors in determining good quality discrete speech recognition. The purpose of this research is to achieve recognition in the range of four words. The word range of this research includes the words "previous" (means previous track), "next" (means next track), "low" (means low voice) and "high" (means high voice). Every word was repeated twenty times, ten of which were in noiseless environment and ten others in noisy environment. Afterward the data were categorized into the trained and experimental data. They are used as the main data in this part of the research project of Persian speech processing.

3- FEATURE EXTRACTION

Linear prediction Coding (LPC) is an alternative method for spectral envelope estimation. This method is also known by the names, all-pole model, or the autoregressive (AR) model. It has good intuitive interpretation both in time domain (adjacent samples are correlated) and in frequency domain (all-pole spectrum corresponding to the resonance structure) [2].

The signal s[n] is predicted by a linear combination of its past values. The predictor equation is defined as

$$\tilde{s}[n] = \sum_{k=1}^{p} a_k s[n-k]$$

Here $s[n]$ is the signal, a_k are the predictor coefficients and $\tilde{s}[n]$ is the predicted signal. The prediction error signal, or residual, is defined as

$$e[n] = s[n] - \tilde{s}[n]$$

The coefficients a_k are determined by minimizing the residual energy $E[(e[n])^2$ using the Levinson-Durbin algorithm. As shown in Fig.1, $s[n]$ is the speech signal, $e[n]$ is the voice source (glottal pulses), and $H(z)$ is the response of the vocal tract filter.

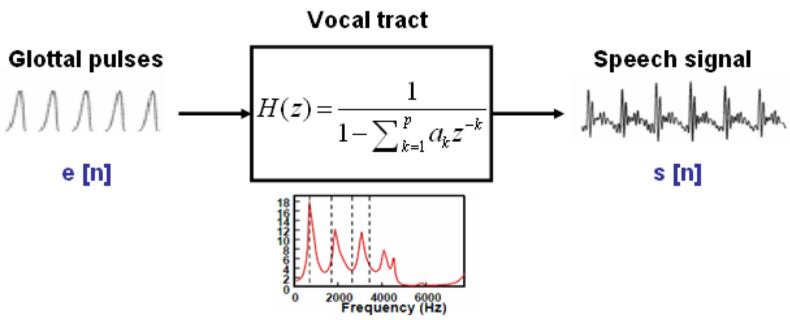

Fig.1: speech signal

$$e[n] = s[n] - \tilde{s}[n]$$
$$= s[n] - \sum_{k=1}^{p} a_k s[n-k]$$

$$E(z) = S(z)[1 - \sum_{k=1}^{p} a_k z^{-k}]$$

$$H(z) = \frac{S(z)}{E(z)}$$

Thus the spectral model, representing the vocal tract is

$$H(z) = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}}$$

The predictor coefficients a_k are rarely used as features but they are transformed into the more robust Linear Predictive Cepstral Coefficients (LPCC) features.

A recursive algorithm proposed by Rabiner and Juang can be used for computing the cepstral coefficients from the LPC coefficients [3].

However, unlike MFCC, the LPCC are not based on perceptual frequency scale, such as Mel-frequency scale. This led to the development of the Perceptual Linear Predictive (PLP) analysis.

4- FUZZY LOGIC

Sound recognition process contains fuzzy logic set creation process (Fig.2) and identification process (Fig.3). At the beginning of fuzzy logic set creation process the signals are sampled and normalized [4]. Afterwards data are converted through the Hamming window. Next data are converted into a frequency band through the fast Fourier transform. The fast Fourier transform creates feature vectors. Fuzzy logic set creation process and identification process is based on the same signal processing algorithms. The difference between them is a sequence of execution. All feature vectors are averaged in fuzzy logic set creation process. Two averaged feature vectors are created. Afterwards it is converted into fuzzy logic set. Fuzzy logic set creation process contains following steps: sampling, quantization, normalization, filtration, windowing, feature extraction (two averaged feature vectors) and fuzzy logic set formation [5][7].

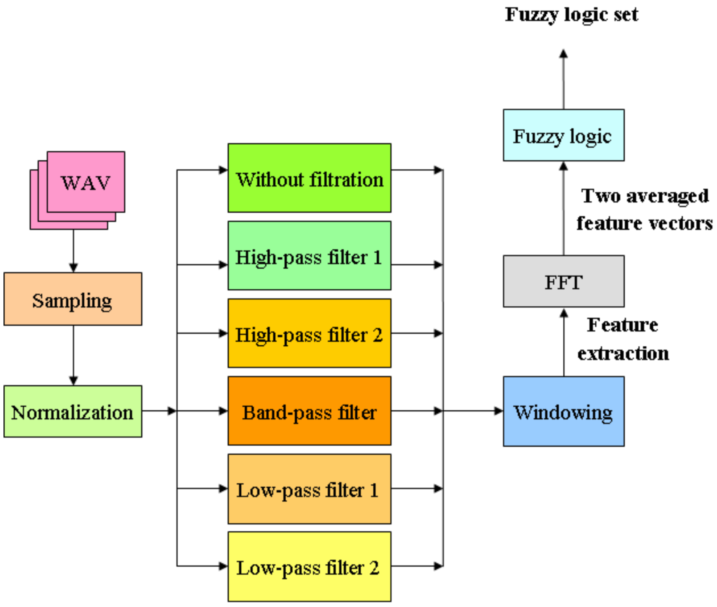

Fig.2: Fuzzy logic set creation process

Classification is used in the identification process. It is based on fuzzy logic. There was applied fuzzy logic as a classifier. To obtain results of recognition, it compares feature vector of new sample with averaged feature vector with the help of fuzzy logic functions. Identification process contains following steps: recording of acoustic signal, sound track division, sampling, quantization, normalization, filtration, windowing, feature extraction, classification. [7]

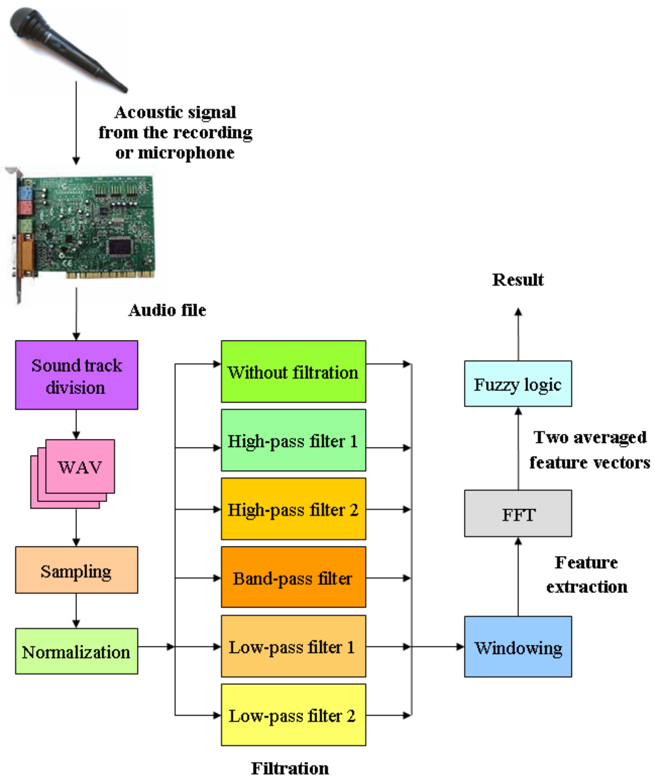

Fig.3: Identification process

4.1. ACOUSTIC SIGNAL RECORDING

The sound card with analogue-digital converter is able to record, process and replay sound. The recording of the acoustic signal is the first part of identification process. Acoustic signal is converted into digital data (wave format) by the microphone and the sound card. This wave file contains following parameters: sampling frequency is 16000 Hz, number of bits is 16, and number of channels is 1 (mono).[7]

4.2. SOUND TRACK DIVISION

Application divides sound track into sound fragments. It divides data. Next it creates new wave header. Afterwards new wave header is copied. Then new wave header is added to each chunk of data. New wave files are obtained. These files are used in the identification process. There are following advantages of such solution: precise determination of sound appearing, precise sound identification, and application does not have to allocate as much memory in identification process.[7]

4.3. SAMPLING

Sampling is a technique to convert an analog signal into a digital signal. It periodically samples an input signal and transforms into a sequence of intensity values. Sampling frequency is basic parameter. Sampling frequency is 16000 Hz in sound recognition application (Fig.4). [7]

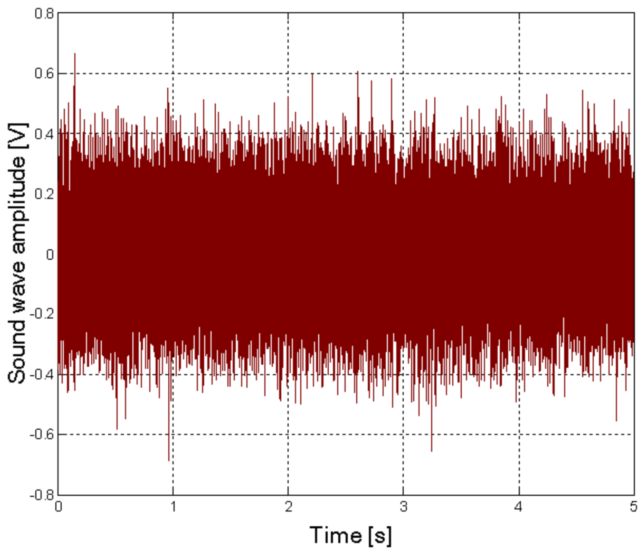

Fig.4: Sound of dc machine with shorted coils for five seconds before normalization

4.4. QUANTIZATION

Quantization is a technique to round intensity values to a quantum so that they can be represented by a finite precision. Precision of sample values is specific to number of bits. Common applied number of bits is 8 or 16. Sound recognition application uses 16 bits because it gives better precision. There is a choice of number of bits depending on quantity of input data and calculations speed in sound recognition process. The compromise is important to obtain good results in short time. [7]

4.5. NORMALIZATION

In sound recognition application, the normalization is the process of changing of the amplitude of an audio signal. There is a possibility that some sounds aren't recorded at the same level. It is essential to normalize the amplitude of each sample in order to ensure, that feature vectors will be comparable. All samples are normalized in the range [-1.0, 1.0]. In method the amplitude maximum of the samples is found and then each sample is divided by this maximum. [7]

4.6. FILTRATION

Filtration is a very efficient way of removing the unwanted noise from the spectrum. The filtration is used to modify the frequency domain of the input sample. The filtration is not necessary to sound recognition. However the usage of this can improve the efficiency of the sound recognition. [7]

4.7. WINDOWING

Windowing is a technique used to shape the time portion of measurement data, to minimize edge effects that result in spectral leakage in the FFT spectrum. By using window functions, the spectral resolution of frequency domain will be increased. There are different types of window functions available, each with their own advantage. The Hamming window is used to avoid distortion of the overlapped window functions. [7]

4.8. FAST FOURIER TRANSFORM

The sequence of frequency of a signal obtained by FFT becomes the basis for extracting of the frequency-domain features. It is applied instead of discrete Fourier transform because of shorter time of calculations. Obtained coefficients create feature vectors which are used in calculations. [7]

4.9. CLASSIFICATION

Difference between sounds depends on differences in ordered sequence [6]. Classification uses feature vectors and fuzzy logic functions in the identification process. It compares different values of feature vectors. It compares feature vectors with the help of fuzzy logic functions (feature vector of investigated sample, feature vector of specific category). Fuzzy logic functions use amplitude of the sample to determine probability (Fig. 5–6). If probability of determined fuzzy function is greater than 0.5 and then function is chosen (240 functions). There are some points where probability is 0.5 and then one of fuzzy function is chosen. Category including the most correct values is the result of the identification. [7]

$$p(x) = \begin{cases} 0 & x < a_1 \\ \dfrac{x - a_1}{a_2 - a_1} & a_1 \le x < a_2 \\ \dfrac{a_3 - x}{a_3 - a_2} & a_2 \le x < a_3 \\ 0 & x \ge a_3 \end{cases} \qquad (1)$$

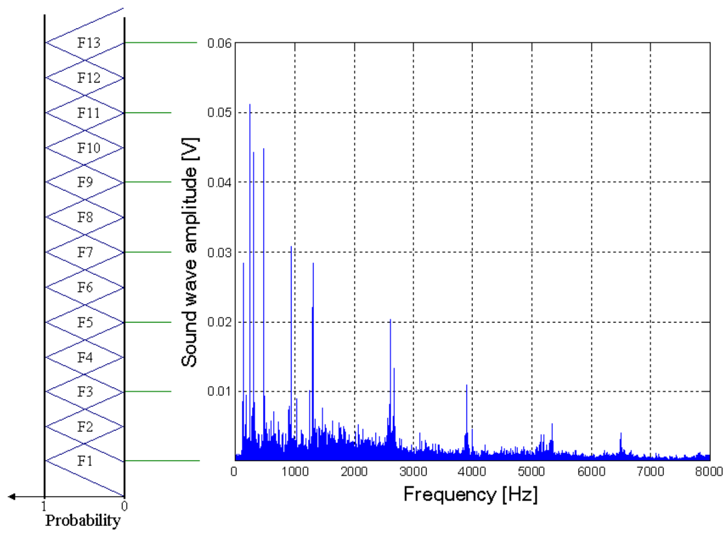

Fig.5: Frequency spectrum of sound of faultless dc machine for ten seconds after normalization with fuzzy logic functions.

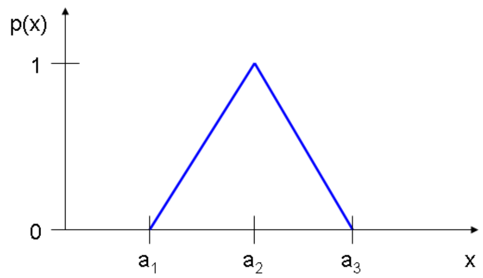

Fig.6. Fuzzy logic function.

5- CONCLUSION

In this research, a hundred men and women with different ages has been chosen randomly, and asked them to repeat four words (Next, Previous, Louder and Lower), which causes to collect and prepared the database with four hundred voice recorded files. The LPCC coefficients and their related differentials have been extracted (39 characterizations). This database divided to four different classes and 100 Fuzzy rules have been derived by processing these data (400 samples). The input data as discussed before have been held in these two different classes by derivation of their properties and comprised with Fuzzy rules. The detection rate by 10 causeways is 96.33.

REFERENCES

[1] R. K. Aggarwal and M. Dave, Using Gaussian Mixtures for Hindi Speech Recognition System, International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 4, No. 4, December, 2011, pp.157 .

[2] Shue Y. and Iseli M., "The role of voice source measures on automatic gender classification", in proc. of IEEE International Conference on acoustics, Speech and Signal Processing, Las Vegas, pp. 4493-4496, 2008.

[3] Rakesh K., Dutta S. and Shama K., "Gender Recognition using speech processing techniques in LABVIEW", International Journal of Advances in Engineering & Technology, vol. 1, no. 2, pp. 51-63, May 2011.

[4] Bondarenko, I.Yu. and Fedyaev, O.I., “The analysis of efficiency of the fuzzy pattern matching method for isolated word recognition”, Proc. of the 6th International Conference “Intellectual Analysis of the Information” IAI 2006, pp. 20-27. Kiev, 2006.

[5] HoseinNezhad, R., Moshiri., B and Eslambolchi, P., “Fusion of Spectrograph and LPC Analysis for Word Recognition: A New Fuzzy Approach”, Proc. of the 7th International Conference on Information Fusion, pp. 449-454. Stockholm, Sweden, 2004. Online: http://www.fusion2004.foi.se/papers/IF04-0449.pdf, accessed on 15 April 2008.

[6] L. R. Rabiner and R. W. Schafer. Digital Processing of Speech Signals. Prentice Hall, Englewood Cliffs, New Jersey, 1978.

[7] Glowacz A., Glowacz W. Sound recognition of DC machine with application of FFT and fuzzy logic, Prace Naukowe Instytutu Maszyn, Napędów i Pomiarów Elektrycznych Politechniki Wrocławskiej. Studia i Materiały, Vol. 62, Iss. 28, 2008, pp. 498-505.

[8] Glowacz A., Glowacz W. DC machine diagnostics based on sound recognition with application of FFT and fuzzy logic, Przeglad Elektrotechniczny, Vol. 84, Iss. 12, 2008, pp. 43-46.