

# SELECTIVE NOISE FILTERING OF SPEECH SIGNALS USING AN ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM AS A FREQUENCY PRE-CLASSIFIER

SACHIN LAKRA<sup>1</sup>, T. V. PRASAD<sup>2</sup>, G. RAMAKRISHNA<sup>3</sup>

<sup>1</sup>Research Scholar, Computer Sc. & Engg., K L University, Vijayawada, AP, India  
sachinlakra@yahoo.co.in

<sup>2</sup>Dean, Research & Planning, Chirala Engineering College, Chirala, AP, India  
tvprasad2002@yahoo.com

<sup>3</sup>Professor, Computer Sc. & Engg., K L University, Vijayawada, AP, India  
ramakrishna\_10@yahoo.com

## Abstract

The paper relates to the filtering of a noise signal present in a speech signal. Specifically, the use of an Adaptive Neuro-Fuzzy Inference System (ANFIS) to classify the frequencies present in a speech signal into three fuzzy sets, that is, those for low frequencies, voice frequencies and high frequencies is discussed in this work. Following the pre-classification step, the low frequencies are filtered which comprise the noise component in the speech signal. The pre-classifier was applied prior to the use of various FIR/IIR filters for reducing the noise present in a speech signal. The paper presents the use of an ANFIS for pre-classification of frequencies in a speech signal followed by application of a noise filter to individual or multiple classes of frequencies. It provides evidence for substantial improvement in the quality of the speech signal.

**Keywords:** Adaptive Neuro-Fuzzy Inference Systems; Frequency pre-classifier.

## 1. INTRODUCTION

Noise filtering is used to remove the noise component present in a signal through the use of a noise filter. Every signal received by a machine from a signal source also receives a noise component, if present in the environment. The noise signal causes the deterioration of the source signal thereby reducing the comprehension of the source signal. Specifically, if a source signal is a speech signal, which is received by a computer from a human speaker, through the use of a microphone, then noise from the environment can cause the deterioration of the speech signal, thereby reducing the accuracy of the recognition of the spoken words by the computer.

The objective of this research work is to develop a pre-classifier for other noise filters so as to enhance the speech quality. The objective was to develop the pre-classifier so as to verify the suitability of soft computing methods for noise removal from speech signals. The criterion for the research work to be successful was to achieve a significant improvement in the Signal-to-Noise ratio of the input speech signal for the low class of frequencies.

The noise signal to be removed was white noise from a ceiling fan recorded along with speech samples.

Filters of various categories are used to reduce the noise present in a speech signal. One such filter is the finite impulse response (FIR) filter. This filter is able to remove noise from the entire speech signal with certain effectiveness. However, the filter also causes deterioration of the speech content of the signal while removing noise from the speech content.

The entire content of a speech signal can be classified into three categories, namely, low frequency content, voice or speech frequency content and high frequency content. Voice frequencies are in the range of 20 Hz to 20,000 Hz, which is also the human audible range of sounds in terms of frequencies. Most of the noise is present in the low and high frequency contents of a speech signal, and lie below and above the limits of the voice frequency range, respectively.

The speech signal, comprising the speech and the noise components, can be filtered in such a manner

that the speech content of the signal remains intact, while a FIR filter is used to selectively reduce the noise in the non-speech portions of the signal only. This requires the use of a method to segregate the voice portions, and the noise-only portions of the speech signal. The algorithm is inspired from the method used by the human brain to recognize speech when there is noise in the environment. The human brain does this by separating low frequency noise such as the hum of a fan in a room, from the speech signal of a speaking human being.

The ANFIS, a combination of a neural network and a fuzzy system, acts as the mechanism by which the separation into low, voice and high frequencies is done.

The contribution of this paper is to develop a generic pre-classifier for 8 IIR/FIR filters. The

problem that the tool developed will solve is to mimic the way the brain removes noise by segregating the input speech signal into frequency sets and extracting only the voice frequencies.

## 2. METHODOLOGY

The authors have found that the use of an ANFIS to classify frequencies greatly improves the quality of the speech signal when followed by selectively filtering the low frequencies. Further, this step of pre-classifying frequencies can be applied before using any noise filter.

The noisy speech signal is input to a filter which yields the noise-filtered speech signal as output.

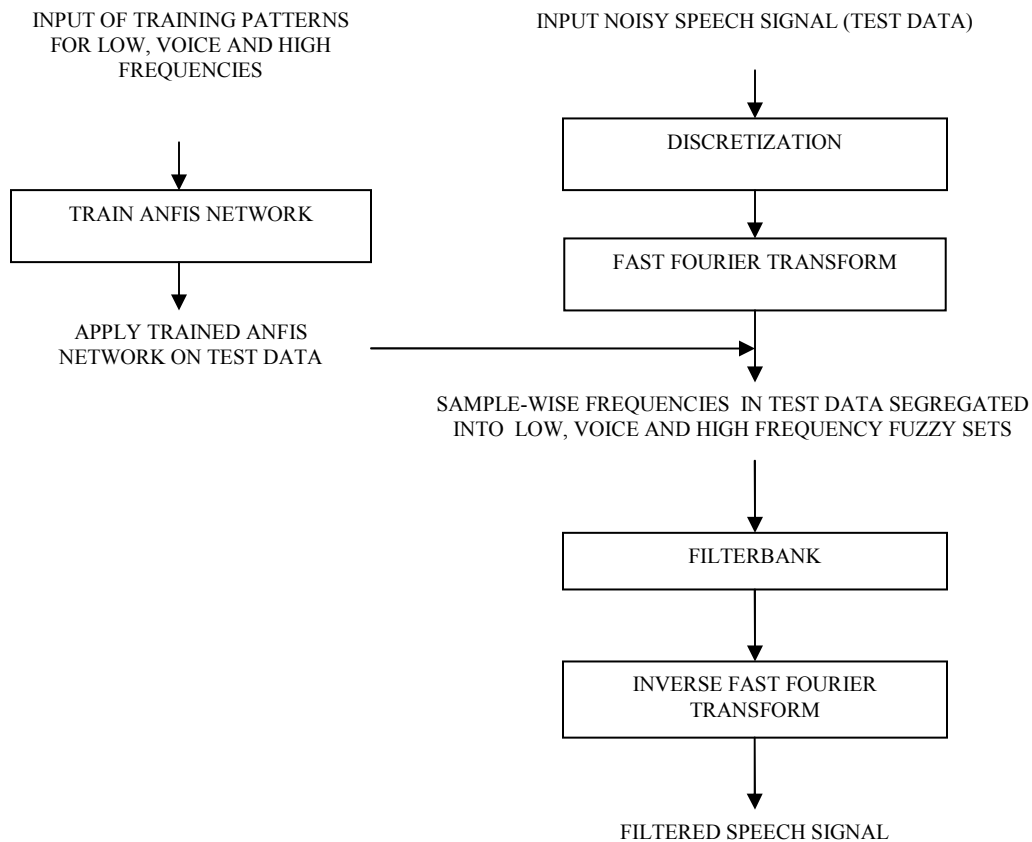


Figure 1: Methodology Followed While Developing The Anfis Based Pre-Classifer.

A set of frequencies are identified as belonging to one of the three crisp sets, namely, low-frequency, voice-frequency and high-frequency. Each set is assigned a set number, namely, 1 for low, 6 for

voice and 10 for high frequencies. A training pattern is then created with the following structure:

< Input:frequency | Output:set number >

The frequency in each training pattern is the input and the set number is the output. There is one training pattern for each frequency. These patterns are divided into training patterns and checking patterns in a pre-defined ratio. The entire training set, including both types of patterns, is given as input to an ANFIS for training, yielding results. The methodology used for the pre-classifier is as shown in Fig.1.

An ANFIS is a neuro-fuzzy system which is capable of creating and learning fuzzy rules from crisp training patterns. The rules, which hold relevance to the training patterns, can then be applied on any test data. The output is the classification of the test data into the fuzzy sets corresponding to the crisp sets input as the crisp training patterns.

A noisy speech signal, is given as test data to the algorithm, which transforms it into discrete samples representing the input signal.

A fast Fourier Transform is applied to transform the discrete samples from a time-domain representation to a frequency-domain representation.

At this stage the trained ANFIS is applied to the frequency-domain representation of the input speech signal and each of the discrete samples is classified into one of the fuzzy sets, namely, low, voice or high.

A filterbank is then applied to each frame of samples. The filterbank consists of an FIR filter for the low fuzzy set and no filter for the voice fuzzy set or the high fuzzy set. Another way of applying the algorithm, is with another filterbank consisting of a FIR filter for the low fuzzy set, an IIR filter for the voice and the high fuzzy sets. Other ways of applying the pre-classifier are also possible being based on a combination of any FIR filter and any IIR filter out of a set of 8 filters, as listed in Table 3.

The filtered discrete samples are then transformed back to the time-domain representation using an inverse fast Fourier Transform.

The output is the selectively filtered speech signal.

Fig. 2 depicts a tool developed to perform the method presented in Fig. 1. The experimental setup consisted of input of speech signals, infused with noise signals, to the tool. The noise signal was a ceiling fan running in the background, giving the

component of white noise to the signal. The frequency range of the white noise signal was between 0 Hz and 40 Hz.

The speech signals were recorded in a room using a laptop microphone, along with a ceiling fan running in the background. The processing procedure of the speech samples collected is as described in Fig. 1. The input variables and their ranges are as described in Table 1 and Table 2.

The tool, as depicted in Fig. 2, shows the inputs and results of the processing of a noisy speech signal. Each speech sample of the input signal is converted to the frequency domain and then, each frequency is segregated into three fuzzy sets as shown in Fig. 2. The segregation is done by a pre-trained ANFIS network as described in Fig. 1. Fig. 2 shows the results of training according to the following graphs:

- a. Input Signal.
- b. Membership functions for Low, Voice and High Frequencies for training data and checking data.
- c. Error curves.
- d. Results of training the ANFIS.
- e. Results of Testing the ANFIS.
- f. Filtered Output Signal.

A set of list boxes provide the options for selecting the IIR/FIR filters for each of the 3 fuzzy sets.

### 3. RESULTS AND DISCUSSION

The variables given to the ANFIS and the associated filters for carrying out the experiments, performed as part of this paper, are as in Table 1 and Table 2, respectively.

The speech signal given as input to the ANFIS during experiments consisted of a female speaker's recorded voice with a noise signal of low frequency received during recording from a fan running at high speed in the recording room. Therefore, the results shown in Table 3 correspond to only the improvements made when the method was applied to low frequency noise.

Other related work is as presented in Ref. [2-7, 9-11, 15]. Certain related applications are as described in Ref. [1, 5, 13-14]. Methods where soft computing methods have been applied for noise filtering are as at [1, 8, 12]. Other applications or research work, however, do not apply soft computing methods as applied in this paper.

Table 1: Ranges To Which Input Frequencies Have Been Mapped For Being Classified By The ANFIS (no. of epochs = 700).

Input for ANFIS	Low Fuzzy Set	Voice Fuzzy Set	High Fuzzy Set
Lower Threshold	0.0	1.1	6.0
Upper Threshold	1.1	6.0	10.0

Table 2: Variables And Their Ranges Given As Input To Various Filters Associated With The ANFIS (frame size = 25).

Inputs for Filters	Low Fuzzy Set	Voice Fuzzy Set	High Fuzzy Set
Lower Limit	0 Hz	70 Hz	4000 Hz
Upper Limit	40 Hz	3900 Hz	10000 Hz
Attenuation	5 dB	1 dB	1 Db

The results of experiments, performed using the tool, are as presented in Table 3.

Table 3: Comparative Results Of Experiments

Noise Filter	Signal-to-Noise Ratio (in dB) for Low Fuzzy Set	
	Without ANFIS	With ANFIS Pre-classifier
FIR	3.4667 (Attenuation NA)	24.3785 (Attenuation NA)
FIR Bandpass Equiripple	-3.1685	17.5621
FIR Bandpass Kaiser	-2.7102	20.1604
IIR Bandpass Butterworth	-1.5869	20.1624
IIR Bandpass Chebyshev Type I	-0.1319	20.1325
IIR Bandpass Chebyshev Type II	9.3555	23.9061
IIR Bandpass Elliptic	8.6385	21.2072
Kalman	-2.0491 (Attenuation NA)	13.4985 (Attenuation NA)

The choice of set numbers 1, 6 and 10 for specifying the low, voice and high crisp sets of frequencies is based on the need to provide a sufficient demarcation between frequencies.

The number of epochs required for training the ANFIS is purely based on experiments in which the desired decrease in Root Mean Square Error is obtained, leading to the correct formation and learning of rules by the ANFIS.

The range of frequencies in the low frequency set was 0Hz – 40 Hz, whereas, the training patterns were 1600 in number. This is done by introducing decimal values for frequencies between any two whole numbered frequencies, that is, between 34Hz and 35Hz, there are frequencies such as 34.025, 34.05, 34.075, 34.1, 34.125, ... and so on. This needs to be done to improve the training of the ANFIS.

The results of testing depict that there are no frequencies in the high set when human beings speak, whereas most of the speech samples include noise in the low set and where there is spoken content the frequencies are in the voice set.

The output signal needs to be examined carefully in comparison with the input signal to observe the subtle changes that have occurred as a result of the filtering. The notable changes are only in the parts of the signal where there is no speech.

As observable from Table 3 the research goal of achieving significant improvements in the low category of frequencies for various filters have been met. There is a marked improvement in the SNR.

#### 4. CONCLUSIONS

The paper provides the use of an ANFIS as a pre-classifier of a speech signal, as a preliminary step before filtering noise.



The paper establishes a process for filtering of a speech signal by using an ANFIS for selecting low frequency and high frequency samples, followed by the use of an FIR/IIR filter.

A similar process is possibly followed by the brain, although no experiments have been done on the validity with respect to the brain. The human brain is able to classify frequencies using fuzzy sets and is able to lessen the noise and recognize the frequency as a low, voice or high frequency if it lies in the range from 20-20,000 Hz.

## 5. FUTURE DIRECTIONS

The filter can be included as a preliminary step for any type of noise filter besides FIR/IIR filters for enhancement of speech signals.

## REFERENCES

- [1] Aritsuka, T., Amano, A., Hataoka, N., & Ichikawa, A. (1993). *U.S. Patent No. 5,185,848*, Noise reduction system using neural network, Washington, DC: U.S. Patent and Trademark Office.
- [2] Catté, F., Lions, P. L., Morel, J. M., & Coll, T. (1992). Image selective smoothing and edge detection by nonlinear diffusion, *SIAM Journal on Numerical analysis*, 29(1), 182-193.
- [3] Doğançay, K., & Tanrikulu, O. (2001). Adaptive filtering algorithms with selective partial updates. *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, 48(8), 762-769.
- [4] Glavieux, A., Laot, C., & Labat, J. (1997, September). Turbo equalization over a frequency selective channel. In *Proc. Int. Symp. Turbo Codes* (Vol. 962102).
- [5] Goldberg, R. G., Rosen, K. H., Sachs, R. M., & Winthrop III, J. A. (1999). *U.S. Patent No. 5,970,446*, "Selective noise/channel/coding models and recognizers for automatic speech recognition", Washington, DC: U.S. Patent and Trademark Office.
- [6] Ibrahim, B. B. (1994). Direct sequence spread spectrum matched filter acquisition in frequency-selective Rayleigh fading channels. *Selected Areas in Communications, IEEE Journal on*, 12(5), 885-890.
- [7] Meurers, T., Veres, S. M., & Elliott, S. J. (2002). Frequency selective feedback for active noise control. *IEEE control systems magazine*, 22(4), 32-41.
- [8] Morgan, D. P., & Scofield, C. L. (1991). *Neural networks and speech processing* (pp. 329-348). Springer US.
- [9] Nishimura, Dwight G. "Multiple measurement noise reducing system using artifact edge identification and selective signal processing." U.S. Patent No. 4,499,493. 12 Feb. 1985.
- [10] Pok, G., Liu, J. C., & Nair, A. S. (2003). Selective removal of impulse noise based on homogeneity level information. *Image Processing, IEEE Transactions on*, 12(1), 85-92.
- [11] Porter, Jack E. "Apparatus and methods for the selective addition of noise to templates employed in automatic speech recognition systems." U.S. Patent No. 4,933,973. 12 Jun. 1990.
- [12] Sachin Lakra, T.V. Prasad, G. Ramakrishna, "Speech Signal Filters based on Soft Computing Techniques: A Comparison", Proceedings of International Congress on Computer Applications and Computational Science 2010, Singapore, 4-6 December, 2010, pp. 1031-1035.
- [13] Trajkovic, Miroslav, Srinivas Gutta, and Eric Cohen-Solal. "Active noise canceling headset and devices with selective noise suppression." U.S. Patent Application 09/825,045.
- [14] Wittkop, T., & Hohmann, V. (2003). Strategy-selective noise reduction for binaural digital hearing aids. *Speech Communication*, 39(1), 111-138.
- [15] Xu, Y., Weaver, J. B., Healy Jr, D. M., & Lu, J. (1994). Wavelet transform domain filters: a spatially selective noise filtration technique. *Image Processing, IEEE Transactions on*, 3(6), 747-758.

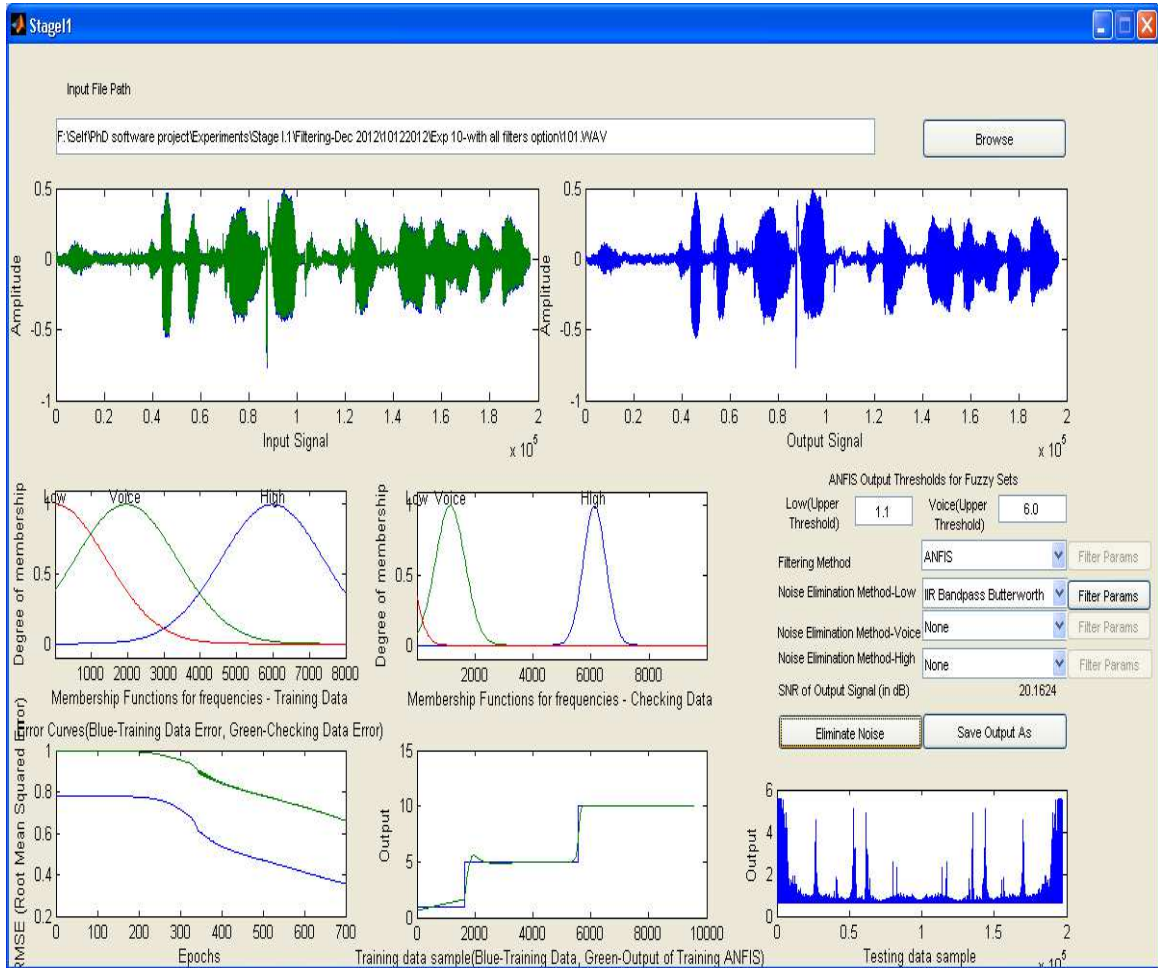


Figure 2: The Tool Developed For Implementing The Algorithm In Matlab.