

Fuzzy ARTMAP Technique for Speech Noise Reduction

BOONRUANG MARUNGSRI and ANANT OONSIVILAI
 School of Electrical Engineering, Institute of Engineering
 Suranaree University of Technology
 111 University Ave., Muang District, Nakhon Ratchasima, 30000
 THAILAND
 bmsv@sut.ac.th , anant@sut.ac.th

Abstract: - This paper presents an approach to reduction noise of speech voice commanded automatic wheel chair using a technique based on the fuzzy ARTMAP neural network (FAMNN). The measurable output noisy speech with 5dB, 1dB, -1dB, -5dB and -10dB SNR level is obtained as the contaminated signal of the interference to compare with the output data of the filter. The white noise source is acquired as the input. Finally, after training, the fuzzy ARTMAP output (i.e. estimated interference) was demonstrated. Then the estimated information signal is calculated as the difference between the measured signal and the estimated interference. The fuzzy ARTMAP could do a practically superior situation in adaptive de-noising of a speech voice commanded automatic wheel chair system with nonlinear characteristics.

Key-Words: - speech, noise, noise reduction, fuzzy ARTMAP neural network, FAMNN

1 Introduction

Recently, technologies have been continuously developed in order to made comfortable daily life. Technologies not only developed for normal peoples but also developed for handicapped peoples too. Handicapped peoples have right to take care. By this reason, we have developed a speech voice commanded automatic wheel chair for the handicapped people as shown in Fig. 1. Command voices are “go ahead”, “go back”, “turn left”, “turn right” and “stop” in Thai language.

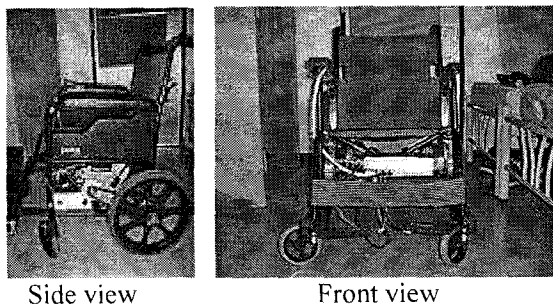


Fig. 1 Speech voice commanded automatic wheel chair.

This automatic wheel chair received command voice from the commander via microphone. The receiving command voice was used to compare with the reference command voices. If the receiving voice is same or near the same with the reference voice then the automatic wheel chair done by using the receiving command voice.

Since we live in a natural environment where noise is inevitable and ubiquitous, speech signals can seldom be recorded in pure form and are generally contaminated by acoustic background noise. As a result, the microphone signals have to be “cleaned up” with digital signal processing tools before they are stored, transmitted, or played out. The cleaning process, which is often referred to as either noise reduction or speech enhancement, can be achieved in many different ways.

In recent years, much research has been undertaken into noise reduction to improve the performance of speech communication and recognition systems[1–10]. Noise reduction suppress environmental noises, improving the speech quality and increasing the recognition accuracy. Thus, noise reduction has been of increased interests for many researchers.

2 Fuzzy ARTMAP Neural Network

The Fuzzy ARTMAP neural network (FAMNN) has been introduced and detailed description extensively by Carpenter et al. in 1992 [11]. The architecture of FAMNN based on Adaptive Resonance Theory (ART). The family of ART NN [12] includes the fuzzy ART mapping (ARTMAP) NN (FAMNN), the Fuzzy min-max NN, the laterally primed adaptive resonance theory (LAPART), the evidence integration for dynamic predictive mapping (ART-EMAP), and the Gaussian ARTMAP. The FAMNN have been successfully applied in many task such as data mining [13], remote sensing [14], recognition of hand written

characters [15], and signature verification [16] etc. For completeness in the following sections, we present only the necessary details in this section.

One advantage is that FAMNN is faster to train than other neural networks due to the small number of training epochs required by the network to “learn” the input data. FAMNN is considered fast even among members of the ARTMAP family due to the computationally “cheap” mapping between inputs and outputs. Also, the classification results of FAMNN are easily interpretable. Furthermore, compared to standard nearest neighbor techniques which are also commonly used, FAMNN requires less memory since it uses a compressed representation of the data, and for the same reason FAMNN requires less classification time. The existence of memory reduction algorithms does not diminish this advantage of the FAMNN since these algorithms can be used for both nearest neighbor techniques and the FAMNN.

The FAMNN is a supervised network which is composed of two Fuzzy ART modules, ART_a and ART_b , as shown in Fig. 2. Those modules are linked together via an inter-ART module, F^{ab} , called a *map field*. The map field is used to realize the *match tracking rule*, whereby the vigilance parameter of ART_a in crease in response to a predictive mismatch at ART_b . Match tracking reorganizes category structure so that predictive error is not repeated on subsequent presentations of the input.

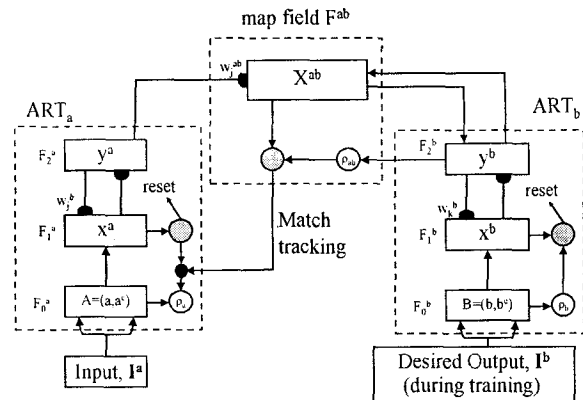


Fig. 2 FAMNN processor consists of two fuzzy ART modules linked by a map field.

As shown in Fig. 2, variables in ART_a and ART_b module are designed by subscripts or superscripts a and b . Inputs to each module are in the complement code form, $I^a = A = (a, a^c)$ for ART_a module and $I^b = B = (b, b^c)$ for ART_b module, respectively.

For ART_a module, $x^a = (x_1^a, \dots, x_{2Ma}^a)$ represent the F_1^a output vector, $y^a = (y_1^a, \dots, y_{2Ma}^a)$ represent the F_2^a

output vector and $w_j^a = (w_{j1}^a, w_{j2}^a, \dots, w_{j2Ma}^a)$ represent the j^{th} weight vector.

For ART_b module, $x^b = (x_1^b, \dots, x_{2Mb}^b)$ represent the F_1^b output vector, $y^b = (y_1^b, \dots, y_{2Mb}^b)$ represent the F_2^b output vector and $w_k^b = (w_{k1}^b, w_{k2}^b, \dots, w_{k2Mb}^b)$ represent the k^{th} weight vector.

For the map field, let $x^{ab} = (x_1^{ab}, \dots, x_{2Nb}^{ab})$ represent the F^{ab} output vector and $w_j^{ab} = (w_{j1}^{ab}, w_{j2}^{ab}, \dots, w_{j2Nb}^{ab})$ represent the weight vector from the j^{th} F_2^a node to F^{ab} . Vector x^a, y^a, x^b, y^b and x^{ab} are set into 0 between input presentation.

The map field F^{ab} is activated whenever one of the ART_a or ART_b categories is active. If node J of F^{ab} is chosen, then its weights w_j^{ab} activate F^{ab} . If node K of F_2^b is active, then the node K in F^{ab} is activated by 1-1 pathways between F_2^b and F^{ab} .

$$x^{ab} = \begin{cases} y^b \wedge w_j^{ab} & \text{If the } J^{th} F_2^a \text{ node is active and } F_2^b \text{ is active} \\ w_j^{ab} & \text{If the } J^{th} F_2^a \text{ node is inactive and } F_2^b \text{ is active} \\ y^b & \text{If the } F_2^a \text{ is inactive and } F_2^b \text{ is active} \\ 0 & \text{If the } F_2^a \text{ is inactive and } F_2^b \text{ is inactive} \end{cases} \quad (1.1)$$

Due to composed of two Fuzzy ART module, Fuzzy ARTMAP module vectors [17] based on two separately distance criteria, *match* and *choice*. The match function is defined by

$$S_j(I) = \frac{|I \wedge w_j|}{|I|} \quad (1.2)$$

where w_j is an analog-valued weight vector associated with cluster j , \wedge denotes the fuzzy AND operator, $(p \wedge q) = \min(p, q)$, and the norm $|\dots|$ is defined by $|p| = \sum_i |p_i|$. The choice function is defined by

$$T_j(I) = \frac{|I \wedge w_j|}{\alpha + |w_j|} \quad (1.3)$$

where α is a small constant. Increasing α biases the search more towards modules with large w_j . Input vector (I) is assigned to the category with maximizes $T_j(I)$ while satisfying $S_j(I) \geq \rho$, where the vigilance, ρ , is a constant, $0 \leq \rho \leq 1$. The fuzzy ARTMAP learning rule is given by

$$w_{ji}^{new} = \begin{cases} w_{ji}^{old} & ; w_{ji} \leq I_i \\ w_{ji}^{old} - \beta(w_{ji}^{old} - I_i) & ; w_{ji} > I_i \end{cases} \quad (1.4)$$

where $0 < \beta \leq 1$. Only the weights of the Cluster to which I has been assigned are updated. All w_{ij} are initially set to 1.

The map field is fundamentally a look-up table, retrieving an analog-valued weight w_{jl}^{ab} when module a node J and module b node L are active. Note that only one node of each module is active at a given time. If $w_{jl}^{ab} < \rho^{ab}$ the vigilance of module a , r^a , is raised until node J becomes inactive (and some other node becomes active). This process is repeated until $w_{jl}^{ab} \geq \rho^{ab}$. When the next input is presented, r_a is returned to its baseline value. All w_{jl}^{ab} are initially set to 1. During learning, when nodes J and L become active and $w_{jl}^{ab} \geq \rho^{ab}$, all w_{jl}^{ab} , $l \neq L$, are reduced in value (typically set to 0).

3 Principles of Speech Noise Reduction

In general, the noisy speech signal is given by $x(k) = s(k) + n(k)$, where $s(k)$ and $n(k)$ represent the clean speech and noise signal, respectively.

Adaptive Noise Cancellation is used to remove noise from useful signals. This is a useful technique where a signal is submerged in a noisy environment. The process of using the FAMNN for solving the adaptive noise cancellation problem is given in Fig 3. The information signal x was recorded with corrupting noise n that was generated from another noise source through an unknown nonlinear process. The noise n is Gaussian with zero mean and unit variance.

The purpose of adaptive noise cancellation is to produce an anti-wave whose magnitude is exactly the same as that of the unwanted noise and whose phase is exactly opposite. The scheme of adaptive noise cancellation system is shown as Fig. 3. The primary input source receives the desired signal $s(k)$ with corrupting noise $y(k)$. The corrupting noise $y(k)$ is generated by the noise source $n(k)$. The received signal of primary input is thus

$$x(k) = s(k) + y(k) \tag{1.5}$$

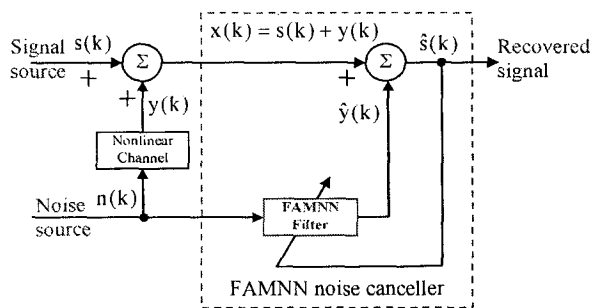


Fig. 3 The adaptive noise cancellation system

A secondary (reference) input source receives a noise $n(k)$ uncorrelated with the signal source $s(k)$ but correlated with the corrupting noise $y(k)$. This secondary input source provides the reference input to the adaptive noise canceller. The $n(k)$ is used by an adaptive process to generate an output $\hat{y}(k)$ (the output of the proposed FAMNNF) that a replica of $y(k)$. The output is then subtracted from the primary input $x(k)$ to recover the desired signal $s(k)$. The basic assumptions for the adaptive noise cancellation system include:

1. The $s(k)$, $n(k)$ and $y(k)$ are stationary zero-mean processes.
2. The $s(k)$ is uncorrelated with $n(k)$ and $y(k)$.
3. The $n(k)$ is correlated with $y(k)$.
4. The $\hat{y}(k)$ are uncorrelated with $s(k)$.

From Fig. 3, it follows

$$\hat{s}(k) = s(k) + n(k) - \hat{y}(k) \tag{1.6}$$

By squaring equation (2), we have

$$\hat{s}^2(k) = s^2(k) + (y(k) - \hat{y}(k))^2 + 2s(k)(y(k) - \hat{y}(k)) \tag{1.7}$$

If we take expectation on both sides, the following equation is obtained.

$$E[\hat{s}^2(k)] = E[s^2(k)] + E[(y(k) - \hat{y}(k))^2] \tag{1.8}$$

The objective of adaptive noise cancellation is to minimize $E[(y(k) - \hat{y}(k))^2]$, which is equivalent to minimize $E[\hat{s}^2(k)]$ by observing equation (1.8). If $E[(y(k) - \hat{y}(k))^2]$ approaches to zero, then the remaining error $\hat{s}(k)$ is exactly the same as the desired signal $s(k)$.

Thus, the signal $E[(y(k) - \hat{y}(k))^2]$ will be used as the error signal to turn the parameters of the proposed FAMNN filter.

3 Simulation Results

Base on the purposed techniques, command voices for the automatic wheel chair in Thai language, which are “go ahead”, “go back”, “turn left”, “turn right” and “stop”, were used to confirm the effectiveness. The measurable output noisy speech with 5dB, 1dB, -1dB, -5dB and -10dB SNR level is taken as the contaminated version of the interference, since the noise is not directly available. The white noise source is taken as the input. The simulation is shown in Fig. 4 – Fig. 8. Each recovered signal are agree with each original signal.

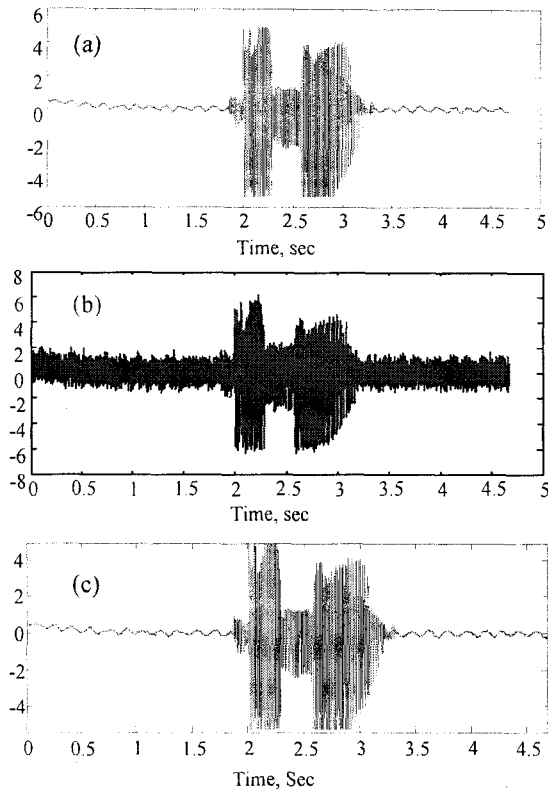


Fig. 4 Results of noise reduction for “go ahead” in Thai language, (a) Clean speech, (b) Noisy Speech (5 dB) and (c) Speech after noise reduction.

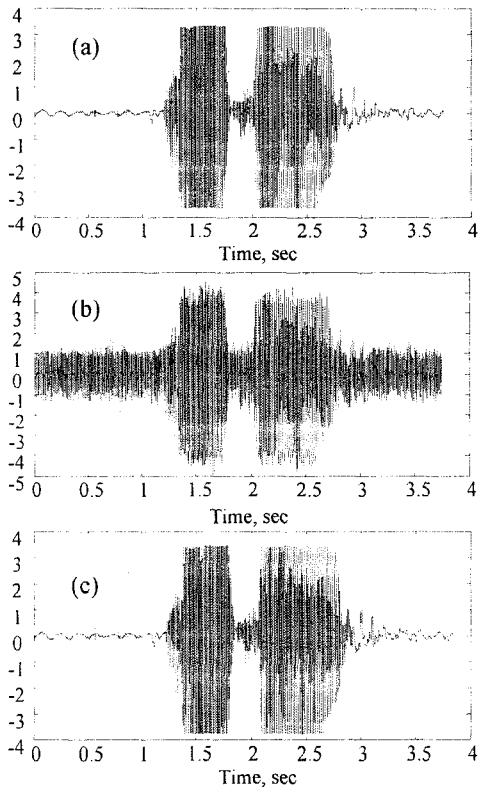


Fig. 6 Results of noise reduction for “turn right” in Thai language, (a) Clean speech, (b) Noisy Speech (-1 dB) and (c) Speech after noise reduction.

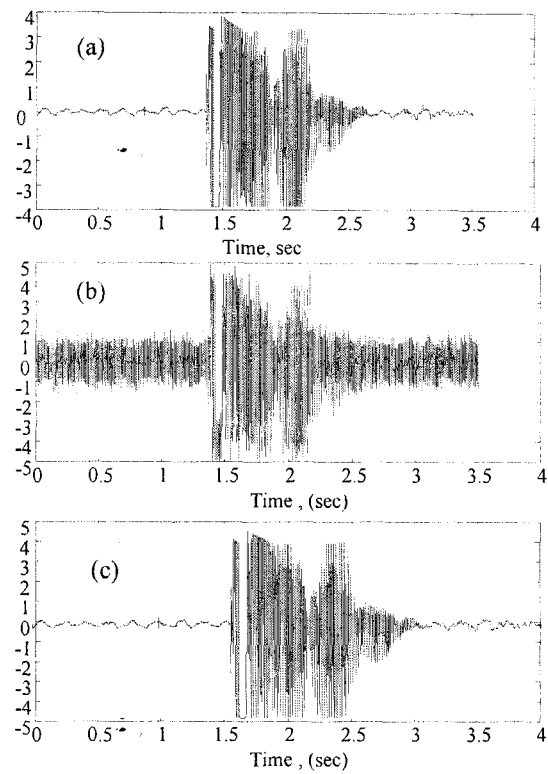


Fig. 5 Results of noise reduction for “go back” in Thai language, (a) Clean speech, (b) Noisy Speech (1 dB) and (c) Speech after noise reduction.

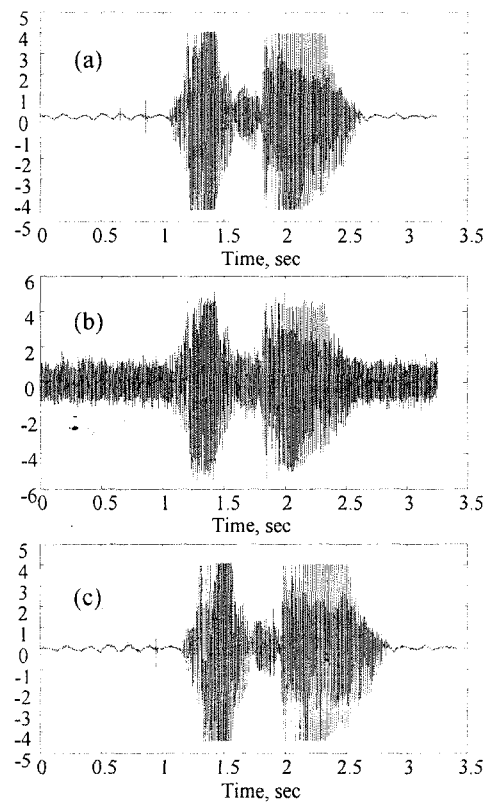


Fig. 7 Results of noise reduction for “turn left” in Thai language, (a) Clean speech, (b) Noisy Speech (-5 dB) and (c) Speech after noise reduction.

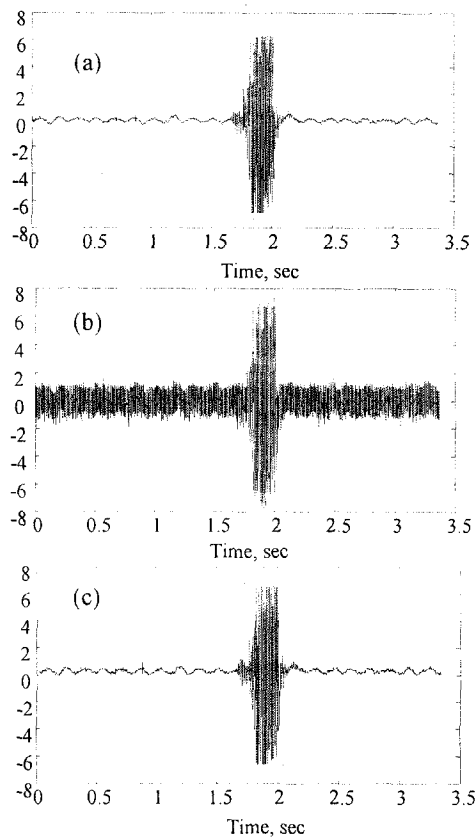


Fig. 8 Results of noise reduction for “stop” in Thai language, (a) Clean speech, (b) Noisy Speech (-10 dB) and (c) Speech after noise reduction.

4 Conclusion

From the simulation results, the effort to enhance speech with FAMNN filtering scheme for noise cancellation is acceptable. The automatic wheel chair worked well when using both the original and recovered signals. The effectiveness of purposed noise reduction techniques was confirmed. Furthermore, the purposed noise reduction technique can be applied in real application.

References:

[1] J. Thevaril, H. K. Kwan, “Speech Enhancement using Adaptive Neuro-Fuzzy Filtering”, *Proc. of 2005 Int. Sym. on Intelligent Signal Processing and Communication Systems*, 13-16 December 2005, pp. 753 – 756.

[2] C. T. Lin, “Single-channel Speech Enhancement in Variable Noise-level Environment”, *IEEE Trans. on System, Man and Cybernetics Part A*, Vol. 33, No. 1, January 2003, pp. 137 – 143.

[3] T. W. Lee and K. S. Yao, “Speech Enhancement by Perceptual Filter with Sequential Noise Parameter Estimation” *Proc. of IEEE Int. Conf.*

on Acoustics, Speech and Signal Processing, Vol. 1, 17 – 21 May 2004, pp. 693 – 696.

[4] L. Singh and S. Sridharan, “Speech enhancement using pre-processing”, *Proc. of IEEE Region 10 Annual Conf. on Speech and Image Technologies for Computing and Telecommunications*, Vol. 2, 2 – 4 December 1997, pp. 755 – 758.

[5] J. D. Chen, J. Benesty, Y. T. Huang and S. Doclo, “New insights into the noise reduction Wiener filter”, *IEEE Trans. on Audio, Speech and Language Processing*, Vol. 14, No.4, July 2006, pp. 1218 – 1234.

[6] C. Plapous, C. Marro, and P. Scalart, “Improved Signal-to-Noise Ratio Estimation for Speech Enhancement”, *IEEE Trans. on Audio, Speech and Language Processing*, Vol. 14, No.6, November 2006, pp. 2098 – 2108.

[7] H. B. D. Sorensen, “Speech recognition in noise using a self-structuring noise reduction model and hidden control models”, *Int. Joint Conf. on Neural Networks*, Vol. 2, 7-11 June 1992, pp. 279 – 284.

[8] N. Sasaoka, K. Sumi,; Y. Itoh and K. Fujii, “A new noise reduction system based on ALE and noise reconstruction filter”, *IEEE Int. Sym. on Circuits and Systems*, Vol. 1, 23-26 May 2005, pp. 272 – 275.

[9] G. S. Kang and D. A. Heide, “Acoustic noise reduction for speech communication: (second-order gradient microphone)” *IEEE Int. Sym. on Circuits and Systems*, Vol. 4, 30 May-2 June 1999, pp. 556 – 559.

[10] K. Kroschel, “Methods for noise reduction applied to speech input systems”, *Proc. of VLSI and Computer Peripherals. VLSI and Microelectronic Applications in Intelligent Peripherals and their Interconnection Networks*, Vol. 2, 28-12 May 1989, pp. 82 – 87.

[11] G.A. Carpenter, S. Grossberg, N. Markuzon, J.H. Reynolds, and D.B. Rosen, “Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps”, *IEEE Trans. on Neural Networks*, Vol. 3, No. 5, 1992, pp. 698-713.

[12] D. Chralampidis, T. Kasparis and M. Georgiopoulos, “Classification of noisy signals using fuzzy ARTMAP neural networks”, *IEEE Trans. on Neural Networks*, Vol. 12, No. 5, September 2001, pp. 1023 – 1036.

[13] S. Sumathi, S. Sivanandam and N. Jagadeeswari, “Design of Soft Computing Models for Data Mining Applications”, *Indian Journal of Engineering and Materials Sciences*, Vol. 7, No. 3, 2000, pp. 107-121.

- [14] G.A. Carpenter, M.N. Gjaja, S. Gopal and C.E. Woodcock, "ART Neural Networks for Remote Sensing: Vegetation Classification from Landsat TM and Terrain Data", *IEEE Trans. on Geosciences and Remote Sensing*, Vol. 35, No. 2, 1997.
- [15] Murshed, N.A., Bortolozzi, F., and Sabourin R., "A Cognitive Approach to Signature Verification. International Journal of Pattern Recognition and Artificial Intelligence", *Special issue on Bank Cheques Processing*, Vol. 11, No. 7, 1997, pp.801-825,
- [16] Tay, Y. H., Khalid, M., Tan, K. K., and Yusof, R., "Hand-Written Postcode Recognition by Fuzzy ARTMAP Neural Network", *COSTAM National Science Congress*, Genting Highlands, November, 1997.
- [17] B. Matthias, and C. E. Sadik, "Optoelectronic Fuzzy ARTMAP Processor", *Journal of Optical Computing*, Vol. 10, March 1995, pp. 213-215.
- [18] C. Plapous, C. Marro and P. Scalart, "Improved Signal-to-Noise Ratio Estimation for Speech Enhancement", *IEEE Trans. on Audio, Speech and Language Processing*, Vol. 14, No. 6, November 2006, pp. 2098 – 2108.
- [19] S. N. Wu and J. S. Wang. "An adaptive recurrent neuro-fuzzy filter for noisy speech enhancement" *Proc. of IEEE Int.Joint Conference on Neural Networks*, Vol.4, 25-29 July 2004, pp. 3083 – 3088.