음성신호를 이용한 감정인식 모델설계

김이곤, 김서영, 하종필

여수대학교 전기공학과

Design of Emotion Recognition Using Speech Signals

Yigon Kim, S. Y. Kim, J. P. Ha

Dept. of Electrical Engineering. Yosu National University

Yosu 554-749, Korea

ABSTRACT

Voice is one of the most efficient communication media and it includes several kinds of factors about speaker, context emotion and so on. Human emotion is expressed in the speech, the gesture, the physiological phenomena(the breath, the beating of the pulse, etc). In this paper, the method to have cognizance of emotion from anyone's voice signals is presented and simulated by using neuro-fuzzy model.

키워드

감정인식, 음성, 웨이브렛, 퍼지

Ⅰ. 서론

최근 생활환경의 눈부신 발전에 따라 보다 인간중심의 문제해결 방법의 추구로 인하여 사회 모든 분야에서 인공지능에 대한 관심이 집중되고 있다. 더 나아가 인간 지능에 대한 연구가 진전됨에 따라, 인간의 비 지적인 감정, 무의식에 대한 연구가 점차 요구되고 있다. 그 중에서도 인간의 감정에 관한 연구는 여러 분야에 이용될 수 있는데, 대표적인 것은 인간과 기계의 능동적인 인터페이스다. 능동적인 인터페이스는 사용자에게 메뉴를 제공하고 사용자의 명령에 반응하는 형식이 아닌, 사용자의 행동, 감정, 심리 상태를 파악하고 그에 따라서 능동적으로 대응하는 형식을 갖는다. 따라서 이러한 능동적인 양방향성 인터페이스 기술은 기본적으로 사용자의 감정상태 등을 파악, 예측하는 것이 필수적이며, 이를 체계적으로 접근하기 위해서는 사용자의 현재의 감정상태와 감정상태 변화에 대한 인식 모델이 필요하다.

감정이란, 외부의 물리적 자극에 의한 감각, 지각으로부터 인간의 내부에 야기되는 심리적 체험으로, 쾌적함, 분노, 불쾌감, 불편함 등의 복합적인 감정을 말한다. 이와 같은 인간의 감정을 정성, 정량적으로 측정 평가하여 이를 제품이나 환경 설계에 응용함으로써 인간의 삶을 보다 편리하고 안전하게 하고자 하는 연구가 새로운 분야로 각광받고 있다. 국내에서는 1992년 G7 후보과제의 하나로 선정된 후 삼성, LG, 대우 등의 가전사를 중심으로 제품의 감성설계를 위한 감성공학 연구조직을 설치하여 연구중이며, 대학에서도 관련된 연구가 활발히 진행되고 있다. 일본은 1990년부터 98년까지 200억엔 규모의 통산성 대형국책프로젝트의 하나로 '인간감각 계측응용기술개발 프로젝트"를 수행하였으며, 영국의 럿보루우 대학의 CES, 네덜란드 아이트호벤의 인간감각연구소(IPO)등에서는 인간의 감각연구결과를 제품에 응용하기 위한 연구가 활발하며, 미국에서는 MIT대학을 중심으로 많은 연구가 수행되고 있다.

인간의 심리적 감정상태를 인식하는 방법에는 여러 가지 접근방법이 있으나 대부분 표현된 언어로부터 감정을 인식하는 방법과, 표정 및 제스처로부터 인식하는 밥법, 육체의 생화학적 변화를 통해 감지하는 방법, 그리고 상대의 음성을 통해 인식하는 방법등이 있다[1-6]. 각각의 방법들은 특징을 갖고 있지만 표현된 언어의 경우 사용된 언어에 내포된 의미를 분석하는 것으로 의도된 감정을 전달하는 것으로 수동적 인식방법이다. 따라서 무의식의 상태에 대한 감정인식에는 한계가 있다. 표정이나 제스처로부터 감정인식은 시각적 방법으로 장치의 복잡함과 경제적 제한이 많지만 능동적 인식방법으로 양방향성 인터페이스가 가능한 기술이다. 생리학적 변화를 이용한 감지방법은 정확성의 측면에서는 장점을 갖고 있는 방법이지만 거의가 착용해야 하는 부자연스러움으로 인하여 자연스러운 감정인식에는 어려움이 있다. 따라서 자연스럽고 간편하며 적용의 제한성이 없는 음성 신호를 이용한 인식방법이 최근 연구되고 있다. 음성 신호를 이용한 감정인식에 관련된 종래의 연구에서는 음성에 포함된 감정의 특징으로, 음성의 고저, 강약, 리듬과 템포 등 음율적인 비 언어 특징을

이용하였다[4]. 본 연구에서도 여기에 기초하여 음성신호로부터 감정인식을 위한 새로운 특징추출 방법을 제안하고 그 타당성을 실험을 통해 입증하였다.

II. 음성신호로부터 감정특징추출

감정 인식은 여러 방법으로 접근할 수 있으나 본 연구에서는 특정인의 음성인식을 통해 현재의 감정상태를 인식하는 것으로, 일상의 대화중의 음성에서 화자가 느끼는 감정을 감지하는 감정인식 모델을 제시하고자 한다.

감정 인식에서 가장 어려운 문제중의 하나는 매체로부터 실제로 감정을 나타내는 정보를 추출하는 것이다. Yanaru는 사람의 감정 표현 모델에 대한 수학적 모델을 제시한 바 있다[1,2]. Yanaru의 논문은 P.Plutchik과 P.T.Young의 심리학 연구 결과를 토대로 하고 있는데 Yanaru가 가정한 R. Plutchick과 P.T. Young의 이론은 인간은 적은 수의 기본 감정을 갖고 있으며, 다른 감정들은 이 감정의 복합적인 조합으로 이루어지고, 이 기본 감정들은 복잡하고 다양한 형태로 표현될 수 있는데, 어떤 특정의 기본감정에 대하여 강도의 크기는 같으나 방향이 다른 기본감정이 존재한다고 하였다. 그에 따르면 이러한 기본감정은 감정요소의 형태, 강도의 크기, 감정요소가 제시된 시간에 따라서 영향을 받아 복잡한 감정으로 표출된다.

인간의 감정 표현에 대한 모델을 제시한 Yanaru는 인간의 감정상태와 감정요소를 Plutchik의 가설에 따라서 8개의 기본감정으로 나누어 표현하고, 각 기본 감정은 모두 자신의 감정추론기관이 있어서, 8개의 기본 감정으로 나눠진 현재의 감정상태와 감정요소 중 자신의 강도를 입력으로 하여 감정변화를 추론하게 하였다. 그러나 이 방법은 현재의 상태와 과거의 상태에 대한 연속적인 감정의 상태를 표현하는데는 장점이 있으나 복잡하며 적용에 제한이 있다.

본 연구에서는 단순하게 화자의 음성을 통해 현재의 감정상태를 인식함으로써 적용의 제한성을 없애고 동시에 인식모델의 단순화를 얻고자 하였다. 이를 위해 인간의 감정을 Yanaru가 제안한 8개의 기본 감정을 기준으로 표현하였으며, 인간의 감정은 연구의 목적을 간략하게 하기 위해 독립된 형태로 표현하였다. 감정을 나타내는 특징으로는 한국의 전통 음악인 국악의 창에서 인간의 희로애락을 표현하는 방법으로 음의 고저와 장단을 기본으로 하여, 음성의 에너지 분포의 크기 및 변화 폭, 분포 대역의 형태, 음의 길이를 감정특징 파라메터로 선정하여 분석하였다. 이를 위하여 "아! 그렇습니까?"에서 음의 장단으로 "아"의 지속정도를 음의 장단표현으로 분석하고, 전체 음의 Wavelet분석으로 각 주파수 대역 별 분석과 이들 신호를 다시 스펙트럼을 통해 각 주파수 대역의 음의 에너지 분포 형태와 음의 고저를 판단하도록 분석하였다. 분석된 값을 종합하여 8개의 감정에 대한 특징 데이터를 구성하였다.

III. 실험장치구성 및 인식모델설계

데이터 취득은 그림 1.과 같이 마이크로폰은 독일 Brüel Kjæl Type 2671이고, LPF는 NF Electronic Instrument FV-664(fc는 15KHz)이다. A/D는 AD2838로 삼홍데이터 시스템제품을 사용하였으며, 이때 Sampling Time은 40KHz로 하여 데이터를 획득하였다. 그림과 같이 마이크로부터 전기적 신호로 변환된 신호를 전단증폭기를 통해 증폭하여 A/D변환기 입력에 인가한다. A/D변환된 신호는 PC를 통해 취득되어 분석알고리즘을 통해 분석하도록 구성하였다.

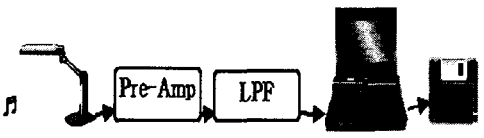

그림 1. 음성 취득 장치 구성도

본 연구에서 얻어진 음성신호를 특정인의 감정에 대한 음성 특징을 얻기 위하여 특정구문 "아! 그렇습니까?"를 8개의 감정을 제시하여 말하게 하였으며, 각각 3회 반복하게 하여 데이터를 취득하였다. 기본감정은 Yanaru가 사용하였던 joy, sadness, expectancy, surprise, anger, fear, hate, acceptance로 분류하여 특정인에 대한 실험을 행하였다. 5인에 대하여 각각의 실험을 수행하였으며, 감정표현의 다양성을 위해 각자의 판단에 의해 감정을 표현하게 하였다. 이를 위해 다른 사람의 감정 표현을 위한 발성을 전혀 알지 못하도록 독립된 상태에서 실험을 진행하였다.

IV. 분석 알고리즘 및 실험

실험구성도 및 하드웨어는 그림 1.과 같이 구성하였으며, 5명의 화자를 이용하여 8개의 감정에 대한 단어리스트를 전달하고 각자에게 적합한 감정표현을 음성으로 하도록 부탁하였다. 문장을 요구에 대한 응답으로 "아! 그렇습니까?"의 반문형으로 하였다. 각자가 자기의 주관대로 감정을 표현할 수 있도록 자율적으로 반복하여 훈련한 후에 같은 문장을 3회 반복하였으며, 3회 반복된 문장을 대상으로 분석하였다. 특징데이터 획득방법은 음의 길이(장단), 에너지분포(음의 강도, 음의 고저)로 표현되는 국악의 원리를 분석 기초로 하였다. 그리고 음성의 음색에 의한 분석을 추가하기 위해 음성의 주파수 분석을 통해 정량적 분석을 수행하였다. 주파수 대역에 대한 음성의 에너지 분포를 실시간 분석하기 위하여 멀티필터링 기법인 wavelet기법으로 시간-주파수관계를

분석하여 정량화 하였다. 본 논문의 관점은 DWT(Digital Wavelet Transform)의 이론이 아니므로 논하지 않았다(필요한 독자는 참고 문헌을 참조바람)[14,15]. 음성을 각 레벨 대역으로 분석하여 각 대역에 분포 된 에너지 비와 음의 강도를 중심으로 정량화 하였다.

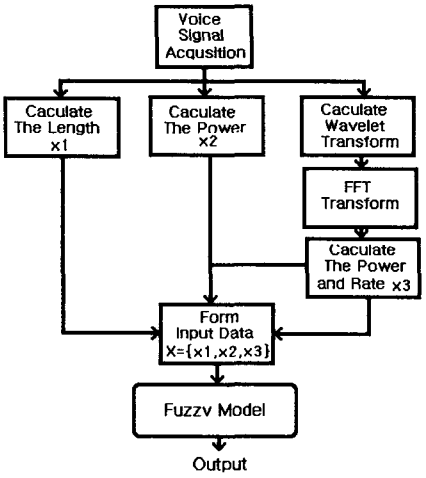

그림 2. Flow Diagram of Emotion recognition Algorithms

그림 2.는 감정인식모델의 학습시스템의 구성도로서 음성신호의 분석 알고리즘의 흐름과 모델의 학습구조를 보여주고 있다. 음성신호 데이터를 입력으로 하여 이 신호를 일정 음절 동안 적분 값을 계산하고, 다시 원래의 신호를 Wavelet 알고리즘을 이용하여 적합한 채널의 신호로 분류한다. 그리고 분류된 각 채널의 각 음절 신호의 구간 적분 값을 계산한다. 이렇게 계산된 Wavelet 변환전의 신호 적분 값과, 이 값에 대한 각 채널 신호의 적분값의 비를 음성신호에 대한 정량화된 감정인식 특징데이터로 한다. 마지막으로 정량화된 패턴 데이터를 이용하여 감정인식 퍼지 모델을 학습하도록 하였다.

본 연구에서는 그림 3.과 같이 성우로부터 감정을 표현한 학습신호를 취득하였으며 각 감정을 표현한 신호를 시간영역 소스신호로 보여주고 있다.

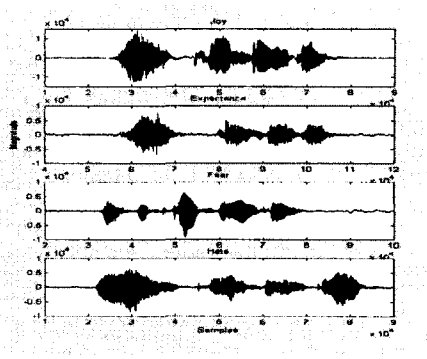

그림 3. Samples of Voice Signals

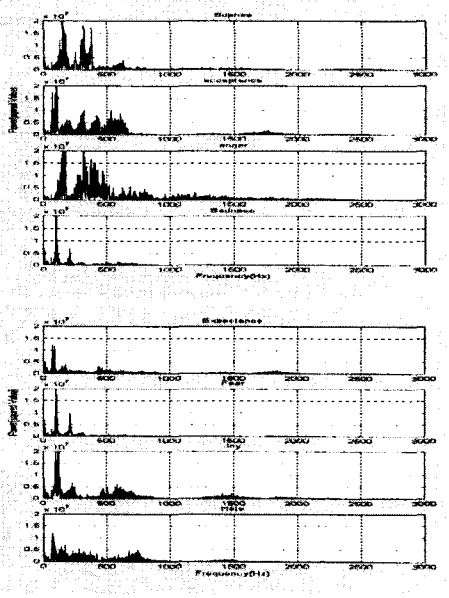

그림 4. FFT Spectrums of Sample's Voice signals.

FFT를 이용하여 분석한 결과를 그림 4.와 같이 얻었다. 그림으로부터 주파수 대역에 분포된 형태가 각각 특징을 갖고 있음을 확인할 수 있다. 같은 음절의 경우에서도 서로 다른 분포형태를 갖고 있음을 확인할 수 있다. 그러나 신호의 분포 형태를 특징 데이터로 하기에는 복잡하고 정형화하기가 대단히 복잡하다. 각 개인과 감정의 표현 방법에도 서로의 차이가 존재하므로 보다 정형화 할 수 있는 방법을 얻기 위하여 Wavelet를 이용하였다. 신호를 Wavelet변환(db.5, Level 8)을 이용하여 분석한 결과 그림 5.와 같은 결과를 얻었다.

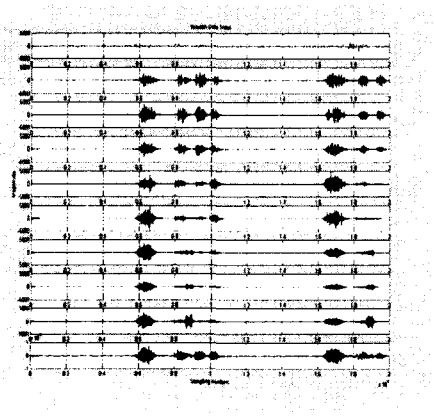

그림 5. Result signals of Wavelet Transform(Fear)

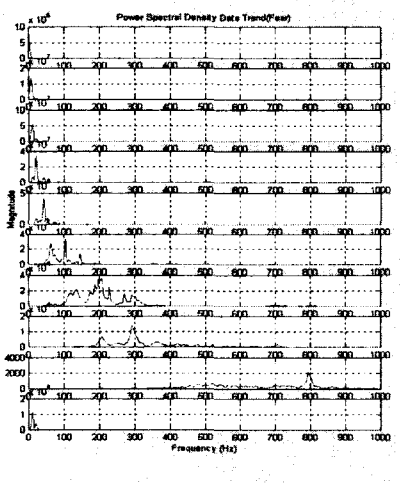

그림 6. FFT Spectrums of Result signal of Wavelet Transform

그림 4.에서와 같이 감정의 종류에 따라 각 주파수대역에 분포되는 음성신호의 크기가 다르게 나타남을 알 수 있다. 그림 6.은 그림 5.의 웨이브렛 변환된 각 채널 신호를 FFT를 이용하여 분석한 스펙트럼 분포를 나타내고 있다.

주파수 스펙트럼 분석결과로부터 저주파성분과 고주파성분의 에너지 분포 비를 분석할 수 있다. 통해 8개의 기본감정을 표현한 음성신호를 분석한 결과는 아래 표 1.과 같다. mV단위로 취득된 신호의 강도를 일정구간 에너지 밀도크기로 나타낸 음성의 Intensity(강도)와 발음 중에서 음의 길이는 "아! 그렇습니까?"에서 "아!" 부분의 길이를 샘플링 수로 결정하였다.

표 1. Characteristic data of voice signals

Characteristics / Emotions	Intensity (x1)	Period of tones (x2) (Samples,40KHz)
Surprise	284	7000
Anger	442	6000
Sadness	203	11000
Expectancy	166	10000
Acceptance	308	17000
Joy	312	13000
Hate	200	12000
Fear	162	8000

표 2. Characteristic data(x3) of voice signals (in Wavelet)

Channel / Emotions	S	a8	d1	d2	d3	d4	d5	d6	d7	d8
Surprise	2.7e9	37.2	0.05	0.25	0.75	5.25	27.6	40.6	24.5	28.8
Anger	6.1e9	0.07	0.06	0.3	1.6	8.4	32.2	41.0	16.34	0.17
Sadness	6.7e8	5.8	0.05	0.16	0.4	1.6	3.88	22.6	48.8	13.5
Expectancy	3.58e8	10.7	0.21	1.27	3.94	6.0	8.32	9.4	32	28.6
Acceptance	1.9e9	0.97	0.04	0.38	2.07	10.43	25.76	15.6	31.5	13.6
Joy	2.1e9	4.68	0.11	0.51	1.76	4.86	8.51	19.47	48.9	8.59
Hate	6.2e8	1.93	0.06	0.56	3.28	13.39	19.44	19.77	26.33	15.75
Fear	5.14e8	3.95	0.02	0.09	0.3	0.99	5.55	30.65	47.95	11.08

표 2.에서는 학습데이터에 대하여, Wavelet을 이용하여 각주파수대역으로 변환된 신호들을 다시 전력스펙트럼(Power Spectrum)으로 분석된 특징 벡터를 나타내고 있다. 9개의 대역으로 분석하도록 하였다.

V. 감정인식 모델의 설계 및 고찰

다음은 이 학습데이터를 이용하여 음성신호로부터 감정을 인식할 수 있는 인식모델을 ANFIS를 이용하여 설계하였다[8-13]. 그림 7.은 ANFIS를 이용하여 설계된 감정인식 모델이다. 설계된 모델을 성우에 의해서 연출된 8개의 감정데이터를 이용하여 실험하였으며 그림 8.과 같은 결과를 확인하였다.

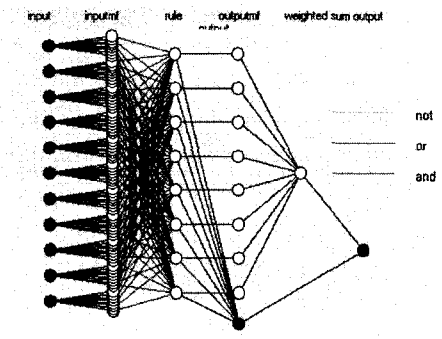

그림 7. Structure of Nero-Fuzzy Model

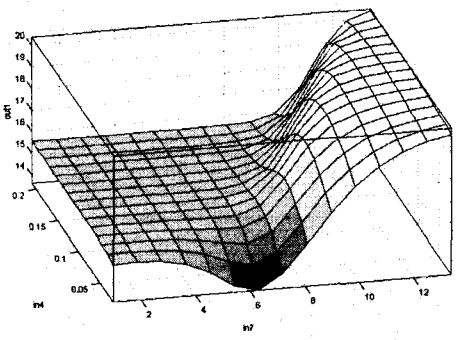

그림 8. Surface of learned Nero-Fuzzy Model

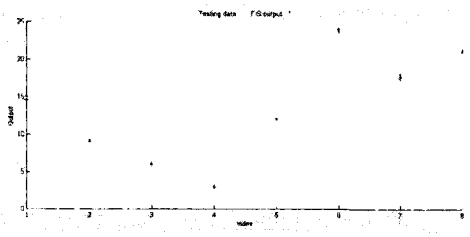

그림 9. Result of Simulation

본 실험에서는 퍼지-신경망을 이용하여 각 개인의 음성으로부터 감정인식이 가능한 감정인식 모델을 설계하였다. 특징 데이터를 이용하여 설계된 모델을 학습하고, 인식이 가능한지의 여부를 알기 위하여 학습데이터와 실험데이터를 분리 실험하여 결과를 고찰하였다. 시뮬레이션 결과를 그림 9.에 나타내었다. 각 감정에 대한 정량화된 값을 수치데이터로 정하였으며 무작위로 3으로부터 각각의 값에 차를 3으로하여 24까지 정하였다. 결과와 같이 선택된 학습데이터를 취득하였던 성우 중 1명을 선택하여 실험한 결과이다. 약 10% 이내의 오차를 보였으며 학습데이터 분석과 같이 유사한 특징을 갖는 감정의 경우는 오차를 상대적으로 많음을 확인하였다.

Ⅵ. 결 론

본 연구에서는 개인의 감정을 인식하는 감정인식 모델에 대하여 제안하였다. 제안한 모델의 감정요소를 갖는 입력 파라미터는 개인의 음성신호로부터 음의 강약과 음절의 지속시간(장단), 그리고 음성의 주파수 대역별 에너지 분포형태로 하였다. 얻어진 입력 파라미터를 이용하여 진단모델을 설계하고 학습데이터를 얻은 특정인의 음성을 실측 데이터로 하여 실험한 결과 90%이상의 인식 결과를 얻었다. 한정된 집단에 국한되었으나, 데이터 수집에 따라 다양한 계층, 다양한 집단의 특징을 반영할 수 있는 감정인식모델의 개발이 가능함을 확인하였다. 향후 다양한 집단에 대한 보다 적은 수의 인식 파라미터에 대한 연구를 수행하고자한다.

감사의 글

이 논문은 과학기술부, 과학재단 지정 지역협력센터인 여수대학교 설비자동화 및 정보시스템 연구개발센터와 포항제철소의 연구비 지원에 의해 연구되었음.

참고문헌

[1] Torao Yanaru, Naruki Shirahama, Kaori Yoshida, Masahiro Nagamatsu, "An Emotion Processing System Based on Fuzzy Interference and Subjective Observations." Information Sciences 101(3-4): 217-247 (1997)

[2] Torao Yanaru, Naruki Shirahama, Kaori Yoshida, Masahiro Nagamatsu,"An Emotion Processing System Based on Subjective Observation," Proceedings of the Fifth International Conference on Neural Information Processing (1998) pp.475-477

[3] Jennifer Healey, Rosalind Picard,"Digital Processing of affective signals", picard@media.mite.edu

[4] Tsuyoshi Moriyama 'and Shinji Ozawa,"A Mweasurement of Human Vocal Emotion Using Fuzzy Control."

[5] L.W.Cambell, D.A.Becker, A.Azarbayjani, A.F. Bobick, and A.Pentland, "Invariant feature for 3-D gesture recognition", in proceedings, International Conference on Automatic face and Gesture Recognition, Pp 157-162, Killington,VT, 1996, IEEE

[6] Elias Vyzas and Rosalind W. Picard," Off line and Online Recognition of Emotion Expression from Physiological Data", MIT Media Laboratory, e-mail:Picard@Media.mit.edu

[7] Medsker, L., Hybrid Neural Networks and

Expert Systems. New York: Kluwer, 1994
[8] Mozetic, I., Model-Based Diagnosis: A Overview. Page 419-430 of: Advanced Topics in Artificial Intelligence. Springer-Verlag, 1992
[9] M. Braee and D. A. Rutherford, "Fuzzy Relations in a Control Setting", Kybernetes Vol. 7, No. 3, pp. 185-188, 1978.
[10] E. M. Scharf and N. J. Mandic, "The Application of a Fuzzy Controller to the Control of a Multi-Degree-Freedom Robot Arm", in Industrial Applications of Fuzzy Control, M. Sugeno, Ed. Amsterdam: North-Holland, pp. 41-62, 1985.
[11] W. J. M.. Kicker and E. H. Mamdani, "Analysis of a Fuzzy Logic controller", Fuzzy Sets an Systems Vol. 1, No. 1, pp. 29-44, 1978.
[12] W. J. M.. Kickert, "Futher Analysis and Application of Fuzzy Logic Control", Interal Rep. F/WK2/75, Queen Mary College, London, 1975.
[13] R. Parekh, J. Yang, V. Honavar," Constructive Neural-Network Learning Algorithms for Pattern Classification", IEEE Trans. on Neural-Network ,Vol.11, No.2, pp. 436-451,March,2000
[14] T. K. Sarkar, C. Su, R. Adve, M.Sala zar-Palma, L. Garcia-Castillo, Rafael R. Boix,"A Tutorial on Wavelets from an Electrical Engineering Perspective, Part 1: Discrete Wavelet Techniques."IEEE Antennas and Propagation Mag.,Vol.40, No.5, pp49-69, Oct.1998
[15] F. X. Canning, J. F. School," Diagonal Preconditioners for The EFIE Using a Wavelet Basis," IEEE Trans. on Ant. and Prop.,Vol. 44,No.9,pp.1239-1246, 1996