J. Biomed. Eng. Res.
Vol. 23, No. 3, 197-205, 2002

다단계 신경회로망을 이용한 후두질환 감별진단 시스템의 개발

전계록 · 김기련* · 권순복* · 예수영* · 이승진* · 왕수건**

부산대학교 의과대학 의공학교실, 부산대학교 대학원 의공학협동과정*, 부산대학교 의과대학 이비인후학교실**

(2002년 2월 20일 접수, 2002년 5월 30일 채택)

A Basic Study on the Differential Diagnostic System of Laryngeal Diseases using Hierarchical Neural Networks

Gye-Rok Jeon, Gi-Ryon Kim*, Soon-Bok Kwon*, Soo-Young Ye*, Seung-Jin Lee*, Soo-Geun Wang**

Dept. of Biomedical Engineering, College of Medecine, Pusan National University

*Dept. of Interdisciplinary program in Biomedical Engineering, Pusan National University

**Dept. of Otolaryngology, College of Medecine, Pusan National University

(Received February 20, 2002. Accepted May 30, 2002)

요 약 : 본 연구에서는 잡음이 존재하는 공간에서 획득한 음성신호로부터 후두질환을 감별진단 할 수 있는 분류기를 구현하였다. 이를 위해 후두질환 환자로부터 수집한 /아/ 모음에 잡음을 혼입하여 음성 신호를 획득하였고, 여러 가지 후두질환을 감별진단 할 수 있는 파라미터를 추출하였으며, 이를 입력으로 하는 계층적 신경회로망을 구성하여 후두질환을 감별진단 하도록 하였다. 감별진단용 분류기는 다섯 단계의 계층적 신경회로망으로 구성하였다. 첫 번째 신경회로망은 정상·양성 후두질환과 악성 후두질환을, 두 번째 신경회로망은 정상과 양성 후두질환을 감별진단 하도록 하였다. 그리고 세 번째 신경회로망은 양성 후두질환 중 후두용, 성대결절, 후두마비를 감별진단 하도록 하였으며, 네 번째와 다섯 번째 신경회로망은 성문암 1-4기를 감별진단 하도록 구성하였다. 분류기에 적용된 신경회로망은 다층퍼셉트론 구조로써 역전파 알고리듬으로 학습시켰으며, 선형변환, 표준점수변환 등 전처리과정을 적용하여 분류기의 성능을 개선하였다.
후두질환의 감별진단 결과 후두용 88.23%, 정상, 성대결절, 후두마비 100%, 성문암 1기 90%, 성문암 2-4기 100%의 감별진단율을 관찰할 수 있었다.

Abstract : The objectives of this paper is to implement a diagnostic classifier of differential laryngeal diseases from acoustic signals acquired in a noisy room. For this purpose, the voice signals of the vowel /a/ were collected from patients in a soundproof chamber and got mixed with noise. Then, the acoustic parameters were analyzed, and hierarchical neural networks were applied to the data classification.
The classifier had a structure of five-step hierarchical neural networks. The first neural network classified the group into normal and benign or malign laryngeal disease cases. The second network classified the group into normal or benign laryngeal disease cases. The following network distinguished polyp, nodule, palsy from the benign laryngeal cases. Glottic cancer cases were discriminated into T1, T2, T3, T4 by the fourth and fifth networks. All the neural networks were based on multilayer perceptron model which classified non-linear patterns effectively and learned by an error back-propagation algorithm. We chose some acoustic parameters for classification by investigating the distribution of laryngeal diseases and pilot classification results of those parameters derived from MDVP. The classifier was tested by using the chosen parameters to find the optimum ones. Then the networks were improved by including such pre-processing steps as linear and z-score transformation.
Results showed that 90% of T1, 100% of T2-4 were correctly distinguished. On the other hand, 88.23% of vocal polyps, 100% of normal cases, vocal nodules, and vocal cord paralysis were classified from the data collected in a noisy room.
Key words : Laryngeal disease, Diagnosis of differential disease, Noise, Classifier, Acoustic signal

서 론

후두질환을 조기에 치료받지 아니하여 악성 후두질환으로 전이되면 생존을 위하여 광범위한 절제가 불가피하게 되며, 장

통신저자 : 전계록, 602-639부산시 서구 아미동 1가 10번지
부산대학교병원 의공학교실
Tel. (051)240-7535, Fax. (051)240-7534
E-mail. grjeon@hyowon.pusan.ac.kr

시간의 음성 재활 훈련을 받아야 한다. 그러나 정상적인 음성의 회복은 불가능하여 삶의 질에 엄청난 저하를 초래하게 된다. 여러 가지 후두질환에 있어서 음성의 변화가 주된 증상이기 때문에 경험이 풍부한 이비인후과 임상의들은 환자의 음성만으로도 후두의 병변 부위를 예측하기도 하며, 이러한 음성의 변화를 후두질환의 진단 및 조기진단에 이용하기도 한다. 왜냐하면 음향학적 검사는 비침습적이며, 검사가 간편할 뿐 아니라 피검자의 시간적 · 공간적 제약을 최소화할 수 있는 장점을 지니고 있기 때문이다.

여러 가지 후두질환을 감별진단하기 위해서는 후두의 상태를 잘 반영할 수 있는 음향학적 변수를 구하는 것이 중요하며, 이러한 연구가 많은 연구자에 의해 수행되었다. 1980년 Horri[1]는 피치동요와 관련된 변수를 이용하여 후두질환 진단에 관한 연구를 수행하였으며, 1982년 Yumoto[2], 1986년 Kasuya[3]는 병적 음성의 잡음 성분을 이용하여 후두질환 감별진단에 관한 연구를 수행하였다. 그리고 1995년 Alwan[4]은 음성 장애의 정도를 측정하기 위해 시 · 주파수 합성법에 의한 변수에 관한 연구를 수행하였으며, 1997년 Ciocea[5]는 장애음성을 감별진단하기 위해 포만트-면적사상을 이용하였다. 국내에서는 1999년 왕[6]이 방음실에서 녹음된 음성을 MDVP로 분석된 33개의 변수를 이용하여 누적도수곡선(cumulative frequency curve)에 의한 후두암을 진단하고자 하였고, 1999년 조[7]는 켑스트럼을 이용한 새로운 변수를 개발하고 집단선별검사를 시행하여 90%이상의 선별율을 보였으며, 2001년 조[8]와 김[9]은 신경회로망을 이용하여 ARS 장애음성의 식별에 관한 연구를 수행하였다. 상기와 같이 선행 연구자들에 의해 후두질환 감별진단용 파라미터들이 개발되었으나, 여러 가지 후두질환별 특징과 변수들의 상관관계에 대한 분석이 미흡하였으며 대부분 방음실 혹은 조용한 공간에서 녹음된 음성을 이용하여 분석하였다. 그리고 사용된 감별진단 파라미터도 2~3 개로 극히 제한적이었으며, 파라미터들이 나타낼 수 있는 범위의 한계성 때문에 새로운 후두질환 감별진단 파라미터의 개발에 관한 지속적인 연구가 요구된다.

최근에 건강에 대한 관심이 고조되면서 자신의 건강 상태를 쉽게 진단할 수 있는 제품의 보급이 증가하는 추세에 있다. 일반 가정이나 직장에서 피검자의 음성신호를 획득하고 후두질환을 감별진단 할 수 있는 시스템을 개발함으로써 병원 내원이 어려운 직장인들과 노인들이 언제든지 컴퓨터를 이용해 후두질환 검진을 받을 수 있다면 그 활용성은 매우 크다고 할 수 있다. 하지만 피검자의 원음에 녹음 환경에서 존재하는 배경잡음이 혼입되거나 마이크, 스피커, 사운드카드, 그리고 여러 가지 녹음 장비들의 주파수 특성으로 인해 녹음된 음성신호에 왜곡현상이 발생할 수 있으므로 잡음 환경에 보다 강인한 감별진단 파라미터의 추출과 분류기의 구현이 이루어져야 한다.

따라서 본 연구에서는 신경회로망을 적용하여 여러 가지 후두질환을 감별진단 할 수 있는 분류기를 구현하였고, 구현한 분류기에 음성신호를 입력하여 감별진단 된 후두질환과 임상의가 진단한 진단결과를 비교 · 검토하여 분류기의 임상 적용 가능성을 검증하였다. 그리고 이를 근거로 잡음이 존재하는 공간에서 유용한 후두질환 감별 진단 시스템을 구현하고자 하였다. 이를 위하여 음성신호의 획득, 후두질환 감별진단용 특징 파라미터 추출, 계층적 신경회로망을 이용한 분류기 구현, 그리고 임상적용 가능성을 검토하였다.

이론해석

1. 후두질환과 발성

후두는 음성의 발생과 호흡을 위하여 가장 중요한 역할을 하는 장기이다. 후두 내의 성대가 적당히 긴장하고 협착되어 호기 시에 공기가 아래에서 위로 올라와 성대를 진동시킴에 의해 음파가 형성되며, 형성된 음파는 인두와 구강으로 구성된 성도를 거쳐 공명되어 음성이 생성된다. 정상인의 경우 성대의 전후 · 측에 걸쳐 거의 균질한 구조를 지닌다. 이러한 균질성이 손상되면 성대 내부에서 다른 파장의 진동이 발생하여 음의 왜곡이 발생할 뿐만 아니라 성대의 진동을 어렵게 한다[10]. 후두질환 환자는 후두에 병변이 생겨 발성 장애를 호소하는 경우가 대부분이며, 후두의 병변으로는 후두염, 성대결절, 후두용, 후두마비, 성문암, 후두 외상 등이 있다.

2. 음성신호의 분석

음성은 자음과 모음으로 구성되는데, 자음은 잡음 특성을 지니며 앞뒤에 수반되는 모음에 따라 많은 변화를 보이기 때문에 음성분석시 주로 유성음인 모음만을 추출하여 그 특성을 관찰하게 된다. 음성 신호중 유성음 구간을 검출하기 위해서는 주로 10msec ~ 30msec 구간을 한 단위로 하는 단구간 에너지와 영교차율 등이 적용된다.

단구간 에너지 함수(En)는 음성신호의 크기가 변화하는 성질을 반영하며 다음 식과 같이 정의된다.

$$E_n = \sum_{m=-\infty}^{\infty} [s(m)W(n-m)]^2 = \sum_{m=-\infty}^{\infty} s^2(m)h(n-m) \qquad (1)$$

여기서, s(m)은 샘플링된 음성신호, W(n-m) 또는 h(n-m)= $W^2(n)$은 분석하고자하는 구간의 창을 나타낸다.

유성음은 성도에서 공명 현상에 의해 발생되어 에너지로 나타나므로 이 에너지를 유성음 추출 파라미터로 사용하고 있다. 단구간 에너지와 더불어 흔히 쓰이는 방법으로 단구간 평균 영교차율을 이용하는 방법이 있다. 유성음은 영교차율이 적고, 무성음은 고주파 잡음 성분에 의해 매우 크게 나타난다. 음성신호의 주파수 정보를 간단히 알아볼 수 있는 단구간 평균 영교차율(Zn)은 다음과 같이 정의된다.

$$Z_n = \sum_{m=-\infty}^{\infty} |sgn[s(m)] - sgn[s(m-1)]| W(n-m) \qquad (2)$$

여기서, sgn[s(m)]은 음성신호의 부호, W(n-m)은 분석하고

자하는 구간의 창을 의미한다.

그리고 음성의 주기적 성질을 나타내는 피치는 음성으로부터 여러 가지 의미를 추출하기 위한 중요한 요소로 작용한다. 피치란 말하는 사람의 의도, 정서상태, 성대의 이상유무 등에 의해 변화하며, 음성신호의 주기 및 주파수가 지니는 의미 이상으로 중요한 의미를 지닌다. 음성신호의 피치를 검출하기 위하여 사용되는 시간영역 계산방법은 음성신호의 주기성을 강조한 뒤 결정논리에 의해 피치를 찾는 방법으로서 병렬처리법, AMDF(average magnitude difference function)법, 그리고 자기상관법 등이 사용된다. 주파수영역 계산법은 음성 스펙트럼상의 고조파 간격을 측정하여 유성음의 기본 주파수를 검출하는 방법으로서 고조파 분석법, Lifter법, 그리고 Combo-filtering 법이 사용되고 있다. 또한 시간주파수 혼성법으로는 켑스트럼법, 스펙트럼 비교법 등이 사용된다. 장애음성은 정상음성보다 훨씬 많은 잡음 성분을 내포하고 있기 때문에 피치 주기가 선명하지 못하다[11].

3. 음성신호 분석용 파라미터

장애 음성을 분석하기 위해 임상에서 적용되는 특징파라미터에는 장·단기 주파수 동요 측정치인 Jitt, RAP, PPQ, vFo, 장·단기 진폭동요 측정치인 ShdB, Shim, APQ, vAm, 소음과 연관된 측정치인 NHR, VTI, SPI, 그리고 떨림과 연관된 측정치인 FTRI, ATRI, Fftr, Fatr 등이 있다. 상기 특징파라미터를 세부적으로 설명하면 다음과 같다. Jitt는 단구간별 피치값의 상대적인 변화를 나타내고, RAP와 PPQ는 각각 3 개와 5개의 피치값을 평균한 음성신호에서의 변이값을 나타내며, vFo는 기본주파수의 상대적인 변이를 나타낸다. 또한 ShdB는 구간별 진폭값의 변이를 dB로 나타내고, Shim는 구간별 진폭값의 상대적인 동요비율을 나타내며, APQ는 11 구간의 값을 평균한 값의 진폭동요비율을 나타내고, vAm은 정점진폭의 상대적인 표준편차를 의미한다. 그리고 NHR는 70~4500Hz 사이의 배음 성분의 에너지와 1500~4500Hz 범위의 배음 외의 성분음 에너지의 평균비율을 나타내고, VTI는 70~4500Hz 사이의 주파수대역상 배음 에너지에서 2800~5800Hz 범위의 배음 외의 성분이 차지하는 비율을 나타내며, SPI는 70~1600Hz의 저주파 성분음의 에너지와 1600~4500Hz의 고주파 성분음 에너지의 평균비율을 나타낸다. FTRI는 주파수떨림 강도지수로서 분석구간 내 가장 강한 저주파 변조성분이 전체 주파수에서 차지하는 상대비율을 나타내고, ATRI는 진폭떨림 강도지수로서 가장 강한 저주파 변조성분이 전체 진폭에서 차지하는 상대비율을 나타내며, Fftr은 주파수 변조된 저주파 요소 중 가장 에너지가 큰 주파수를 나타내고, Fatr은 진폭 변조된 저주파 요소 중 가장 에너지가 큰 주파수를 나타낸다[12].

실험 방법

1. 실험 대상

본 연구를 수행하기 위하여 후두질환군과 정상대조군을 표 1과 같이 선별하였다. 먼저, 실험군은 1995년 12월부터 2000년 3월까지 음성장애로 부산대학교병원 이비인후과에 내원한 환자 중에서 후두질환 검사결과 질환이 확인되었으나 음성검사가 가능한 환자들을 선별하였다. 즉 양성 후두질환군의 후두용(polyp) 37예, 성대결절(nodule) 15예, 후두마비(palsy) 10예 등이었다. 그리고 악성 후두질환군은 병기별로 성문암 1기 20예, 성문암 2기 20예, 성문암 3기 20예, 성문암 4기 20예 이었다. 다음으로, 정상 대조군은 조음이 정상이며 후두병변이 없고 정상 발성이 가능한 건강한 성인 남녀 15명을 무작위 추출하여 대상으로 하였다.

2. 실험 환경 및 음성 신호의 획득

잡음이 존재하는 공간에서 후두질환 환자의 음성을 수집하는 것은 많은 시간이 소요될 뿐만 아니라 환자군의 추출도 어렵기 때문에, 본 연구에서는 음성검사를 목적으로 방음실에서 획득한 음성신호에 잡음을 혼입하여 분석용 음성신호를 획득하였다. 이를 위하여 방음실에서 환자를 마이크 앞에서 15㎝가량 거리를 둔 의자에 편안한 자세로 앉게 한 후 가능하면 평상시와 같은 높이와 크기로 약 3초간 /아/모음을 지속적으로 발성하게 하였다. 이러한 과정을 3회 이상 실시하여 Digital Audio Recorder(DAT, DTC-59ESJ, Sony, Japan)에 녹음한 뒤 후두질환 유형별로 분류하여 데이터베이스화하였다. 기록된 음성신호는 음성 검사용으로써 분석 장비의 특성상 50㎑로 샘플링되고 16bit로 양자화되었다. 왜냐하면 음성 신호를 분석하도록 프로그램화된 Kay사(Kay Elemetrics Corp. USA)의 Computerized Speech Lab 4300b(CSL) 기종의 Multi-Dimensional Voice Program(MDVP)가 50㎑의 샘플링 레이트를 가진 음성신호만을 지원하기 때문이다. 따라서 일반 PC에서 음성 신호를 원활하게 재생하고 녹음할 수 있도록 하기 위해 decimation과정과 interpolation 과정을 수행하여 50㎑의 샘플링 레이트로 기록된 음성신호를 48㎑의 샘플링 레이트로 변환하여 PC에 저장하였다. 또한 기본 모음 중 /아/만으로도 후두질환을 감별할

표 1. 실험 대상군
Table 1. Experiment subject group

Control groups / Data	Normal	Benign laryngeal disease			Malign laryngeal disease			
		Polyp	Nodule	Palsy	T1	T2	T3	T4
Training data	10	20	10	5	10	10	10	10
Testing data	5	17	5	5	10	10	10	10
Total	15	37	15	10	20	20	20	20

수 있으므로 다른 모음 즉, /이/,/우/, /에/, /오/ 모음을 제외시켰다[13]. 그리고 사용자(end user)의 시스템 환경의 다양성을 고려하여 방음실의 잡음이 배제된 음성신호를 PC에서 정규화하여 재생한 후, 다른 PC를 사용하여 녹음함으로써 실제로 발생할 수 있는 여러 가지 잡음을 혼입하였다. 피검자의 발성 강도와 주변 환경에 따라 혼입된 잡음의 양은 다소 차이가 발생하지만 획득된 음성신호는 전체적으로 23.8493±3.09 dB SNR을 나타내었다. 이 때 재생과 녹음에 사용된 PC의 사운드카드는 SoundBluster와 Soundtrack이었고, 마이크는 방음실과 동일한 다이나믹 마이크를 사용하였으며, 음성신호의 재생 및 녹음을 위하여 음향신호 분석 소프트웨어인 Cool Edit 96(Syntrillium Software Co., USA)을 사용하였다. 음성신호를 획득하는 과정은 그림 1과 같다. 이러한 과정으로 획득된 음성은 일반적인 방음실 데이터와는 큰 차이를 보이는데 기본적인 샘플링 레이트나 대역폭이 다를 뿐 아니라 백색잡음(white noise)과 같은 여러 가지 배경 잡음과 채널 특성에 의한 왜곡현상이 발생할 수 있다. 하지만 음성신호의 주파수 대역폭을 20Hz~37KHz 라고 볼 때 사운드카드, 마이크, 스피커 등으로 구성된 본 연구의 음성 신호 획득 시스템에서는 방음실 마이크를 제외하고는 음성 신호의 왜곡은 발생하지 않는다고 볼 수 있다. 엘리어싱(aliasing)으로 인하여 문제시되는 방음실 마이크는 20KHz의 대역폭을 지니고 있지만, 본 연구가 원음을 재현하기보다는 획득된 음성 신호를 이용하여 특징 파라미터를 추출하는 것이기 때문에 무시해도 무방할 것이다.

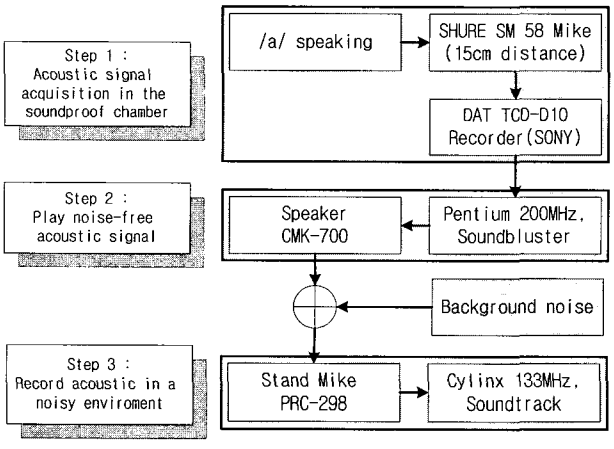


그림 1. 음성 신호 획득 구성도
Fig. 1. Block diagram of acoustic signal acquisition

3. 파라미터 추출

사람의 발성은 성대에서 원음을 만들어 성도의 변화를 거쳐 원하는 음성을 만든다. 따라서 원음의 발생기관인 후두에 병변이 있을 때 후두의 떨림 수나 진동폭에 불규칙성이 개입되기 때문에 이러한 동요를 적절히 측정함으로써 정상인과 후두질환 환자들의 후두 상태를 쉽게 구분할 수 있게 된다. 따라서 후두질환 감별진단에 있어 가장 중요한 요소는 후두의 상태를 잘 반영할 수 있는 음향학적 변수를 구하는 것이다.

본 연구에서 음성 신호의 특징 파라미터를 추출하는 과정은 임상에서 후두의 상태를 관찰하기 위하여 33개의 음향 변수를 추출하도록 프로그램화된 Kay사의 CSL 기종의 음성분석용 소프트웨어인 MDVP를 이용하였다. 그리고 MDVP에서 제공하는 여러 가지 변수에 대해 분석을 수행한 후 여러 가지 후두질환과 연관성을 검토하여 본 연구에 적합한 특징 파라미터들을 추출하였다. 즉, 장·단기 주파수 동요측정치인 Jitt, RAP, PPQ, vFo, 장·단기 진폭 동요 측정치인 ShdB, Shim, APQ, vAm, 소음과 연관된 측정치인 NHR, VTI, SPI, 떨림과 연관된 측정치인 FTRI, ATRI, Fftr, Fatr 등을 추출하여 사용하였다.

그러나 상기에서 언급한 MDVP 분석 파라미터만으로는 후두질환의 감별진단이 용이하지 못하였다. 따라서 본 연구에서는 MDVP에서 사용된 파라미터를 산술 연산하고 조합하여 새로운 파라미터를 추출하였다. 즉, Jitt/Shim와 같은 새로운 파라미터를 생성하여 후두질환의 감별진단에 이용하였다. 왜냐하면 하나의 파라미터만으로는 질환의 군집성이 떨어지지만, 두 가지 이상의 파라미터를 산술 연산하여 조합시키면 탁월한 군집성을 나타낼 수도 있기 때문이다. 또한 감별 파라미터 각각의 가중치에 의해 감별율의 차이가 발생하므로, 감별진단이 용이하도록 가중치를 부여하고 산술 조합을 수행하여 새로운 파라미터를 추출하였다.

추출된 파라미터들은 여러 가지 검증 과정을 거쳐 감별파라미터로서의 타당성을 검증한 후 여러 가지 후두질환을 감별진단하기 위한 파라미터로 선정하였다. 우선 추출된 파라미터들의 여러 가지 후두질환별 분포도를 관찰한 후 우수한 군집성을 나타내는 파라미터들을 선택하였다. 그리고 선택된 파라미터를 입력층으로 하는 신경회로망을 구성하고 분류 결과를 관찰한 후 질환별 감별율이 높은 파라미터만을 선택하였다[14]. 마지막으로 선택된 파라미터의 통계적 유의성(significance level : p-value)을 검증하기 위해 T검증(T-test)과 일원변량분석(ANOVA)을 사용하였으며, 유의수준(criterion for significance)을 0.05로 하고 질환군간의 유의적인 차이를 나타내는 파라미터만을 선택하여 후두질환 감별 파라미터로 선정하였다. T-test는 두 질환군의 표본의 평균과 분포를 기초로 하여 두 질환군의 평균간의 차이와 그 유의도를 검증하기 위해 사용하는 통계기법이며, ANOVA는 두 개 이상 질환군의 표본의 평균치를 통해 이들 표본이 각 질환군의 모집단을 대표하고 있는지를 밝히고 질환군의 모집단 평균간의 차이를 검증하는데 사용되는 통계방법이다[15].

4. 분류기 구현

본 연구에서는 보다 효율적으로 후두질환을 감별진단 할 수 있는 분류기를 구현하기 위해 신경회로망을 다단계 다층구조로 구성한 계층적 신경회로망을 제안하였다. 왜냐하면 음성신호는 비선형적인 구조를 지니고 있어 일반적인 다층 신경회로망을 사용하면 감별율의 향상이 어려우며, 또한 많은 양의 데이터가

요구되어 학습에 상당한 시간이 요구되기 때문이다. 이러한 문제점을 해결하기 위하여 각 질환별로 우수한 파라미터를 선정하여 계층적 신경회로망을 구성한 후 이들로부터 최종결과를 판단하도록 하는 후두질환 감별진단 분류기를 구현하였다. 이를 위하여 구성한 신경회로망은 다층 퍼셉트론 구조로서 역전파 알고리듬으로 학습시켰다.

후두질환을 감별진단하기 위하여 제안한 계층적 신경회로망의 구성도는 그림 2와 같다. 그림 2에서는 NN1에서부터 NN5까지 5개의 신경회로망을 계층적으로 연결하여 질환 감별이 용이하도록 구성하였다. NN1에서는 정상·양성 후두질환과 악성 후두질환을 감별하도록 하였고, NN2에서는 정상과 양성 후두질환을 감별하도록 하였다. 그런 후 NN3에서는 양성 후두질환군 중 성대폴립, 성대결절, 성대마비를 감별하도록 하였고, NN4에서는 악성 후두질환 중 성문암 1기에서부터 3기까지와 성문암 4기를 감별하도록 하였다. 마지막으로 NN5에서는 성문암 1기, 2기, 3기를 각각 감별진단 하도록 하였다[16]. 분류기의 가장 상위층인 NN1에서 정상·양성 후두질환과 악성 후두질환을 감별진단 한 이유는 후두암의 조기진단이 후두 기능의 보존 및 치료를 위해서 무엇보다도 중요하며, 두 질환군들 즉, 정상·양성 후두질환과 악성 후두질환의 파라미터 분포특성이 다른 질환군에 비해 명확하게 구분되기 때문이다.

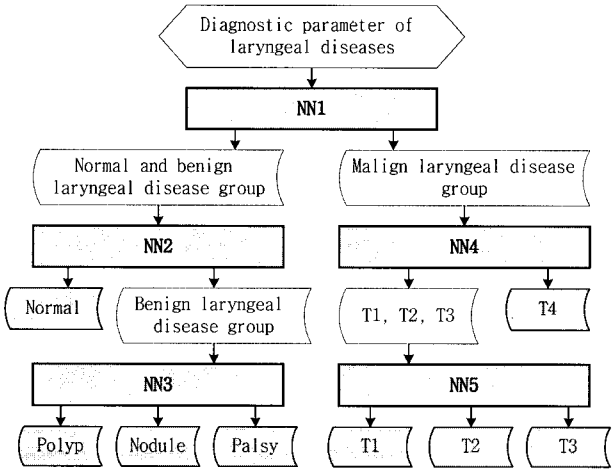


그림 2. 제안된 계층적 신경회로망
Fig. 2. Proposed hierarchical neural network

본 연구에서 후두질환의 감별진단에 사용한 계층적 신경회로망의 각 감별 파라미터가 아무리 통계적 특성이 양호하고 질환별 분류율이 우수하다고 하더라도 감별진단 오류를 피할 수는 없다. 따라서 이러한 문제점을 해결하기 위하여 발견법적 방법에 의하여 입력 파라미터와 은닉층 뉴런의 개수를 결정하고, 선형변환, 표준점수변환 등 분류기의 전처리과정을 적용하여 감별진단율을 향상시켰다.

5. 시스템 구현

실험 내용을 토대로 후두질환 감별진단용 시스템을 구현하였으며, 시스템의 순서도는 그림 3에 나타내었다. 먼저, 피검자는 음성녹음용 프로그램의 안내에 따라 /아/음을 3초간 발성하게 되며, 발성된 음성은 48kHz로 샘플링되고 16bit로 양자화되어 저장된다. /아/ 음의 1 초에 해당하는 샘플 구간을 추출해내고, 추출된 구간의 음성 데이터는 여러 가지 방법에 의하여 감별진단용 특징 파라미터로 계산된다. 그리고 시간 영역과 주파수 영역에서 계산된 특징 파라미터는 계층적 신경회로망에 입력되고, 계층적 신경회로망의 각 단계별로 후두질환을 감별진단 하도록 하였다. 이러한 시스템 구축을 위하여 획득된 음성 데이터 중 학습 데이터를 임의 추출하여 계층적 분류기를 단계별로 학습과정을 수행하였으며, 계층적 신경회로망의 총 학습시간은 44분 46초이었다. 또한 구현된 시스템을 사용하여 여러 가지 후두질환을 감별진단하기 위해서는 대략 68초가 소요되었다. 이러한 시스템을 구현하기 위한 일련의 과정들은 Matlab 5.3 (MathWorks Inc., USA)을 이용하여 수행하였다.

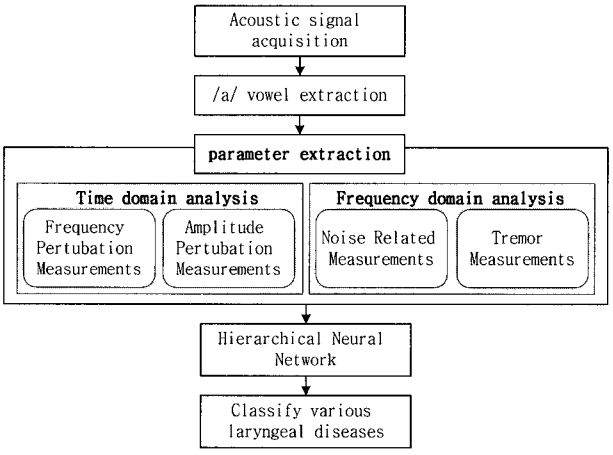
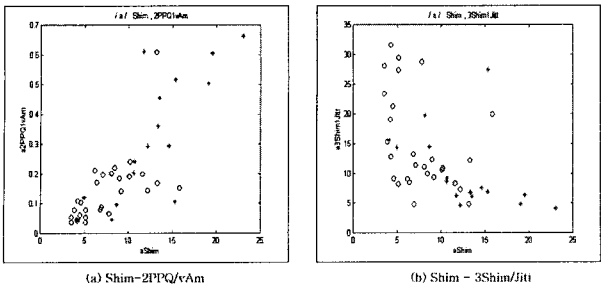


그림 3. 후두질환 감별진단용 시스템의 구성도
Fig. 3. Flowchart of differential diagnostic system of differential laryngeal diseases

결 과

여러 가지 후두질환을 감별진단 할 수 있는 파라미터를 선정하기 위해 세 가지 과정을 수행하였다. 첫째, 추출된 MDVP 분석 파라미터들의 후두질환별 분포도를 관찰한 후 우수한 군집성을 나타내는 파라미터를 선택하였다. 그림 4는 정상·양성 후두질환군과 악성 후두질환군을 감별진단 할 수 있는 파라미터를 추출하기 위해 Shim, 2PPQ/vAm, 3Shim/Jitt에 따른 해당 질환군 데이터의 분포도를 나타낸 것이며, 동일 파라미터에 대한 질환군의 군집성뿐만 아니라 질환군과 파라미터와의 상관관계를 잘 표현해주고 있다. 이와 같은 분포도 조사에 의해 하나의 신경회로망으로는 여러 가지 후두질환을 감별할 수 없음을 알 수 있었기 때문에, 계층적 신경회로망의 질환 분류 단계별로 파라미터를 선택하였다. 둘째, 군집성이 우수한 파라미터를 입력층으로 하는 신경회로망을 구성하고 분류 결과를 관찰



(a) Shim-2PPQ/vAm (b) Shim - 3Shim/Jitt

그림 4. 정상 · 양성후두질환군과 악성후두질환군의 감별 파라미터에 따른 데이터 분포 (a) Shim-2PPQ/vAm, (b) Shim-3Shim/Jitt
Fig. 4. Distribution of normal and benign laryngeal diseases group and malign laryngeal diseases group in conjunction with parametric relation. (a) Shim-2PPQ/vAm, (b) Shim-3Shim/Jitt.(legend : o : normal, benign, * : malign)

하였다. 표 2는 각 분류단계별 파라미터를 3 개씩 조합하여 신경회로망으로 분류한 결과 중 대체적으로 우수한 결과만을 정리하여 나타낸 것이며, 이를 토대로 여러 가지 후두질환군의 감별 파라미터를 선별하였다. 입력 파라미터를 3 개씩 사용한 신경회로망도 높은 감별율을 나타내었지만, NN2의 경우는 정상군을 제대로 감별해내지 못함을 관찰할 수 있었다. 셋째, 선별된 감별 파라미터들의 통계적 유의성을 평가하기 위해 SPSS사(SPSS Inc, USA)의 SPSS for window 10.0이라는 통계 프로그램을 사용하여 유의수준을 0.05라 하고 T-test와 ANOVA를 수행하였다. 표 3은 계층적 신경회로망의 입력층으로 사용되는 여러 가지 감별 파라미터들에 T-test와 ANOVA를 수행한 후 통계적 유의성을 관찰하여 정리하였다. NN1, NN2, NN4에서는 T-test와 ANOVA 모두 감별 파라미터들의 p-value가 대부분 유의수준을 만족하지만, NN3와 NN5에서는 그렇지 못해 감별 파라미터로서의 타당성이 다소 결여되었다는 것을 알 수 있었다.

여러 가지 다양한 후두질환을 효율적으로 감별진단하기 위해 계층적 신경회로망을 사용하였으며 질환군을 분류하는 특징 파라미터가 분류 단계별로 상이하였기 때문에 많은 특징 파라미터가 필요하였다. 후두질환의 분류 단계별 특징 파라미터와 감별진단 결과는 표 4에 나타내었다. 각 분류 단계별 특징 파

표 2. 신경회로망에 의한 파라미터별 감별 결과의 비교
Table 2. Comparing classification result by neural networks using various parameter

Neural Network	applied parameters	Hitting rate(percent,%)
NN1	Shim, 2PPQ/vAm, 3Shim/Jitt	Normal and benign laryngeal disease group : 100, Malign laryngeal disease group : 100
NN2	Jitt, ATRI, Shim	Normal group : 80, Benign laryngeal disease group : 100
	RAP, Fftr, PPQ	Normal group :80, Benign laryngeal disease group : 91.3
NN3	vFo, vAm, SPI	Polyp : 100, Nodule : 100, Palsy : 100
	Jitt, vAm, RAP	Polyp : 94.1, Nodule : 60, Palsy : 100
NN4	vAm, APQ/Jitt, Jitt/Shim	T1 · T2 · T3 : 100, T4 : 100
	vAm, APQ/Jitt, vAm/Jitt	T1 · T2 · T3 : 100, T4 : 100
NN5	ShdB, FTRI, Fftr	T1 : 90, T2 : 100, T3 : 100

표 3. 감별 파라미터들의 통계적 유의성 검증(p=0.05)
Table 3. Verification of statistical significance of various parameters at the 0.05 level by T-test and ANOVA

Neural Networks	Parameter	p value		Neural Networks	Parameter	p value	
		T-test	ANOVA			T-test	ANOVA
NN1	Shim	0.010	0.001	NN3	vAm	〈 0.093	0.014
	2PPQ/vAm	0.005	0.001		Jitt	〈 0.207	0.443
	3Shim/Jitt	0.040	0.051		RAP	〈 0.150	0.411
NN2	Jitt	0.005	0.124	NN4	vAm	0.001	0.011
	ATRI	0.041	0.039		APQ/Jitt	0.042	0.037
	Fftr	0.005	0.016		Jitt/Shim	0.001	0.008
	PPQ	0.002	0.085		vAm/Jitt	0.001	0.156
	RAP	0.007	0.125	NN5	ShdB	〈 0.102	0.276
	Shim	0.002	0.031		FTRI	〈 0.080	0.258
NN3	vFo	〈 0.713	0.465		Fftr	〈 0.027	0.217
	SPI	〈 0.513	0.417				

라미터는 해당 후두질환군에 대해 군집성과 분류율이 우수한 것만을 선별하여 적용하였다. 정상·양성 후두질환군과 악성 후두질환군을 감별진단 하는 NN1에서는 2PPQ/vAm, 3Shim/Jitt, Shim, 정상군과 양성 후두질환군을 감별진단하는 NN2에서는 Jitt, ATRI, Fftr, PPQ, RAP, Shim를 사용하였다. 양성 후두질환군 중 후두용, 성대결절, 후두마비를 감별진단 하는 NN3에서는 vFo, SPI, vAm, Jitt, RAP를 사용하였으며, 악성 후두질환군 중 성문암 1기, 성문암 2기, 성문암 3기, 성문암 4기를 감별진단하는 NN4와 NN5에서는 vAm, APQ/Jitt, Jitt/Shim, vAm/Jitt와 ShdB, FTRI, Fftr을 각각 사용하였다. 각 분류단계별 후두질환의 감별 진단 결과는 모든 단계에서 90 %이상의 감별진단율을 보였다. 각 단계별 감별진단율은 해당질환군의 데이터를 입력하였을 때 해당질환으로 감별되는 비율을 의미한다. 예를 들어 정상인의 음성신호를 입력하였을 때 정상이라고 감별하는 비율을 나타내는데, 표 4에서 NN2의 경우와 같이 정상과 양성 후두질환의 데이터를 입력하였을 때 정상을 정상으로 감별하는 비율과 양성 후두질환을 양성 후두질환으로 감별하는 비율을 각각 나타내었다.

여러 가지 후두질환을 감별진단한 결과는 표 5와 같으며, 구현된 시스템에 음성신호를 입력하여 감별진단된 후두질환과 임상의가 진단한 결과를 비교·검토하여 분류기의 임상 적용 가능성을 검증하고자 하였다. 후두용 환자 2 명이 정상으로, 성문암 1기 환자 1 명이 성문암 2기로 잘못 감별진단 된 것으로 나타났으나, 전체적으로는 95.8%의 감별진단율을 보여 매우 우수한 것으로 판단되었다.

표 6은 본 연구에서 실험한 결과와 선행 연구로서 조[7]가 방음실에서 수집한 장애음성과 조[8]와 김[9]이 ARS 시스템에서 획득한 장애음성을 이용하여 감별진단한 결과를 비교하여 나타내었다. 사용된 감별 파라미터와 녹음 환경이 선행 연구와 상이하기 때문에 비교하기가 힘들지만 감별진단한 결과만을 보

표 4. 분류단계에 따른 특징 파라미터와 감별진단 결과
Table 4. Diagnosis result and acoustic parameter of five-step hierarchical neural networks

Neural network	Control groups	Diagnostic parameters	Diagnosis result			
			Success	Fail	Total	Hitting rate(%)
NN1	Normal and benign laryngeal disease	2PPQ/vAm, 3Shim/Jitt, Shim	32	0	32	100 %
	Malign laryngeal disease		40	0	40	100 %
NN2	Normal	Jitt, ATRI, Fftr, PPQ, RAP, Shim	5	0	5	100 %
	Benign laryngeal disease		25	2	27	92.6 %
NN3	Polyp	vFo, SPI, vAm, Jitt, RAP	17	0	17	100 %
	Nodule		5	0	5	100 %
	Palsy		5	0	5	100 %
NN4	T1, T2, T3	vAm, APQ/Jitt, Jitt/Shim, vAm/Jitt	30	0	30	100 %
	T4		10	0	10	100 %
NN5	T1	ShdB, FTRI, Fftr	9	1	10	90 %
	T2		10	0	10	100 %
	T3		10	0	10	100 %

표 5. 여러 가지 후두질환의 감별진단 결과
Table 5. Diagnosis result of differential laryngeal diseases

Simulation / Original	Normal	Benign laryngeal disease			Malign laryngeal diseases				Diagnosis result			
		Polyp	Nodule	Palsy	T1	T2	T3	T4	Success	Fail	Total	Hitting rate
Normal	5								5	0	5	100 %
Polyp	2	15							15	2	17	88.23%
Nodule			5						5	0	5	100 %
Palsy				5					5	0	5	100 %
T1					9	1			9	1	10	90 %
T2						10			10	0	10	100 %
T3							10		10	0	10	100 %
T4								10	10	0	10	100 %

표 6. 선행 연구들과의 감별진단율 비교
Table 6. Comparing diagnosis result between this study and leading study

Case	This study	Jo's study[7]	Jo's study[8]	Kim's study[9]
Recording environment	noisy room	Soundproof room	ARS system	ARS system
Normal vs. Abnormal	97.22 %	95.45 %	92.57 %	94.11 %
Benign vs. Malign	97.01 %	91.67 %	·	·
Total	95.83 %	94.11 %	92.57 %	94.11 %

면 본 연구의 결과에 비해 방음실과 ARS 시스템에서 획득한 장애음성의 감별진단율이 떨어짐을 알 수 있다. 왜냐하면 선행 연구에서 제한된 파라미터만을 사용했을 뿐만 아니라 ARS 데이터의 경우 잡음 성분이 많고, 전화선 자체의 시간왜곡현상 등으로 인해 정확한 데이터의 수집이 어려우며, 음성 녹음시 적용된 샘플링 레이트가 다르기 때문이다. 샘플링 레이트는 PC에서 48㎑를 가지는데 비해 방음실의 경우 50㎑, ARS 시스템의 경우 11㎑의 샘플링 레이트를 가지므로, 대역폭의 차이로 인해 같은 음성 데이터라 할 지라도 서로 다른 파라미터 값을 가지게 된다. 본 연구에 사용된 음성 데이터는 많은 잡음을 포함하고 있지만 질환분류 단계별로 감별 진단에 도움이 되는 파라미터를 추가함으로써 감별진단율을 향상시킬 수 있었다.

고 찰

제안된 인터넷 환경 기반하에서 계층적 신경회로망의 분류기에 의하여 최종적으로 감별진단된 결과는 정상 100%, 후두용 88.2%, 성대결절 100%, 후두마비 100%의 감별진단율을 나타내었다. 그리고 악성후두질환의 감별진단율은 성문암 1기 90%, 성문암 2기 100%, 성문암 3기 100%, 성문암 4기 100%로 관찰되었다. 그러나 계층적 분류기의 최상위 층에 속하는 NN1에서 오류가 발생할 경우 NN1의 하위 층에 구성되어 있는 NN2와 NN4에서는 NN1의 오류를 상속받기 때문에 오류가 발생하게 된다. 이와 마찬가지로 NN2와 NN4의 하위 층에 구성되어 있는 NN3와 NN5에서도 이러한 영향을 상속받게 된다. 본 연구에서도 NN3의 하위 층에 위치한 후두용에서 2개, NN5의 하위 층에 위치한 성문암 1기에서 1개의 감별오류가 발생하였다. 오류발생 원인을 분석한 결과 성문암 1기의 경우에는 자체적으로 발생한 오류로 기인되나, 후두용에서는 계층적 분류기 중 상위 층인 NN2에서 발생한 오류가 상속되어 나타났음을 관찰할 수 있었다. 이와 같이 다단계 다층 구조로 구성된 계층적 분류기는 각 단계별로 구성되는 분류기들의 형태·구조학적인 특징과 다양한 동적 특성을 고려하여 배치하여야 하며, 하위 구조에 속하는 각 분류기들은 상위 구조로부터 속성과 변수 및 연산을 상속받게 된다. 따라서 계층적 분류기를 사용하여 여러 가지 후두질환을 감별진단 할 수 있는 분류기를 구성할 때도 이러한 특징을 반영하여 분류기를 구현하여야 한다.

계층적 분류기의 각 단계별 입력 파라미터가 상이하지만 Shim, PPQ는 정상음성과 장애음성을 감별진단하기에 유용한 감별 파라미터이고, Jitt, RAP, vFo는 양성 후두질환을, Fftr은 후두질환을 감별진단하는데 전반적으로 유용한 감별 파라미터라고 판단되었다. 그리고 정상·양성 후두질환군과 악성 후두질환군의 감별진단은 진폭 동요 요인과 주파수 동요 요인 등이, 정상과 양성 후두질환의 감별진단은 진폭 동요 요인과 떨림이, 양성 후두질환의 감별진단은 주파수 동요 요인이, 악성 후두질환은 떨림의 정보가 관련성이 많은 것으로 판단된다.

결 론

본 연구에서는 신경회로망을 적용하여 여러 가지 후두질환을 감별진단 할 수 있는 분류기를 구현하고, 구현된 분류기에 음성 신호를 입력하여 감별진단된 후두질환과 임상의가 진단한 결과를 비교·검토하여 분류기의 임상적용 가능성을 검증하고자 하였으며, 도출된 결과를 근거로 배경잡음 하에서 후두질환을 감별진단 할 수 있는 시스템을 구현하고자 하였다. 이를 위하여 먼저 환자군과 정상군의 /아/ 모음을 녹음하고 잡음을 혼입하여 분석용 음성신호를 획득하였다. 수집된 음성신호를 사용하여 MDVP 프로그램에서 제공하는 여러 가지 특징 파라미터를 추출하였으며, 추출한 특징 파라미터는 여러 가지 후두질환에 대한 군집성 조사, 신경회로망을 사용한 분류율 관찰, T-test와 ANOVA를 사용한 통계적 유의성 검증 과정을 수행하여 양호한 특징 파라미터만을 해당질환을 감별진단하는 파라미터로 사용하였다. 그리고 여러 가지 후두질환을 효율적으로 분류하기 위해 선정된 특징 파라미터들을 입력으로 하는 신경회로망을 사용하여 다 단계 다층 구조를 지니는 계층적 분류기를 구현하였다. 이러한 일련의 과정을 통하여 구현된 후두질환 감별 진단용 분류기를 사용하여 감별진단 된 여러 가지 후두질환과 임상의가 진단한 결과를 비교하면, 정상 100%, 후두용 88.2%, 성대결절 100%, 후두마비 100%, 성문암 1기 90%, 성문암 2기 100%, 성문암 3기 100%, 성문암 4기 100%의 감별진단율이 관찰되었다. 잡음 환경에 적합한 감별 파라미터를 추출하여 분류기에 적용한 결과 높은 감별진단율을 나타냄으로써 구현된 후두질환 감별진단용 분류기가 잡음이 존재하는 공간에서 사용될 수 있음을 확인할 수 있었다. 향후 연구에서는 성문암 4기의 피치 검출이 어려우므로 피치검출 과정이 필요없는 파라미터의 개발이 필요하고, 잡음 환경에 대한 다양한 실험과 풍부한 데이터를 이용하여 성별, 나이별, 보다 많은 후두질환에 대해서도 감별진단에 대한 연구가 수행되어야 할 것으로 사료된다.

참 고 문 헌

1. Y. Horri, "*Vocal shimmer in sustained phonation*", Journal of Speech and Hearing Research, Vol. 23, pp. 202-209, 1980
2. E. Yumoto, W.J. Gould, T. Baer, "*Harmonic-to-noise ratio as an index of the degree of hoarseness*", J Acoust Soc Am, Vol. 71, pp. 1544-50, 1982
3. H. Kasuya, S. Ogawa, K. Mashima, S. Ebihara, "*Normalized noise energy as an acoustic measure to evaluate pathologic voice*", J.Acoust. Soc. Am, Vol. 80(5), 1986
4. Abeer Alwan, Philbert Bangayan, Jody Kreiman, and Christopher Long, "*Time and Frequency Synthesis Parameters of Severely Pathological Voice Qualities*", Proc. ICPhS 95, vol 2, pp. 250-253, 1995
5. S. Ciocea, J. Schoentgen, L. Crevier-Buchman, "*Anlaysis of Dysarthric Speech by meas of Formant-To-Area Mapping*", Proc. Eurospeech 97, vol 4, pp. 1799-1802, 1997
6. 왕수건, 양병곤, 조철우, 박현민, 권순복, 김태섭, 홍현석, "병적 음성의 감별 진단을 위한 음향학적 변수", 대한이비인후과학회지, vol. 42, pp. 1561-1567, 1999
7. 조철우, 양병곤, 왕수건, "음향신호의 분석에 의한 후두질환의 진단에 관한 연구", 한국음성과학회지, Vol. 5, No. 1, pp. 151-165, 1999
8. 조철우, 김광인, 김대현, 권순복, 김기련, 김용주, 전계록, 왕수건, "신경회로망을 이용한 ARS 장애 음성의 식별에 관한 연구", 한국음성과학회지, Vol. 8, No. 2, pp. 61-71, 2001
9. 김기련, 김용주, 권순복, 김광인, 김대현, 양병곤, 왕수건, 전계록, "자동응답시스템을 이용한 후두질환 특징 파라미터 추출", 대한의료정보학회지, Vol. 6, No. 4, pp. 175-184, 2001
10. Daniel R. Boone, 음성과 음성치료, 서울, 원미사, pp 13-98, 1996
11. L.R. Rabiner, R.W. Schafer, Digital Processing of Speech Signal, New Jersey, Prentice Hall, pp. 116-455, 1998
12. 고도흥, 정옥란 외 16인, 음성 및 언어 분석기기 활용법, 서울, 한국문화사, pp. 165-188, 2001
13. 조철우, 김대현, "Cepstrum방법과 신경회로망을 이용한 정상, 양성종양, 악성종양 상태의 식별에 관한 연구", 한국음향학회 학술대회 논문집, Vol. 17, No. 2, pp. 151-155, 1998
14. T.H.Martin, B.D.Howard, B. Mark, Neural network design, Boston, PWS Publishing Company, pp. 3.1-12.50, 1996
15. Taik Sup Auh, Quantitative Analysis of Social Science Data SAS/SPSS PC+, Seoul, NANAM Publishing House, pp. 179-216, 1994
16. 김기련, 김용주, 권순복, 전계록, 왕수건, "음성신호를 이용한 후두질환 감별진단 알고리듬 개발", 대한의용생체공학회 추계학술대회, 제23회, pp. 10, 2000