

신경회로망을 이용한 ARS 장애음성의 식별에 관한 연구*

Classification of Pathological Voice from ARS using Neural Network

조철우** · 김광인** · 김대현** · 권순복*** · 김기련***
 김용주*** · 전계록**** · 왕수건*****
 C-W Jo · K-I Kim · D-H Kim · S-B Kwon · K-R Kim
 Y-J Kim · K-R Jun · S-G Wang

ABSTRACT

Speech material, which is collected from ARS(Automatic Response System), was analyzed and classified into disease and non-disease state. The material include 11 different kinds of diseases. Along with ARS speech, DAT(Digital Audio Tape) speech is collected in parallel to give the bench mark.

To analyze speech material, analysis tools, which is developed local laboratory, are used to provide an improved and robust performance to the obtained parameters. To classify speech into disease and non-disease class, multi-layered neural network was used. Three different combinations of 3, 6, 12 parameters are tested to obtain the proper network size and to find the best performance.

From the experiment, the classification rate of 92.5% was obtained.

Keywords : ARS, Pathological Voice, Neural Network, Discrimination, Laryngitis.

1. 서론

후두질환은 치료하기에 상당히 까다로운 질병 중 하나이다. 가벼우면 약간의 통증만을 유발하며 약물로도 치료가 가능하지만, 심각할 경우에는 생명을 위협할 뿐만 아니라 완치된다 하더라도 목소리를 잃게 되는 수도 있다. 하지만 어떤 질병이든 조기 치료가 중요하듯이, 후두질환 역시 조기에 치료한다면 상당한 완치율을 기대할 수 있다. 이러한 후두질환의 조기치료를 위해 음성을 이용한 후두질환 감별은 음성신호를 음향분석 기기를 통해 분석하므로 검사가 신속, 간편하고, 환자에게 전혀 고통이 없다.[1]

음성신호를 녹음하는 방법에 있어서 몇 가지 방법이 있지만 그 중 방음실에서 정확하고

* 본 연구는 한국과학재단 목적기초연구(1999-2-20600-002-3) 지원으로 수행되었음.

** 창원대학교 제어계측공학과

*** 부산대학교 일반대학원 의공학 협동과정

**** 부산대학교 의과대학 의공학교실

***** 부산대학교병원 이비인후과

깨끗한 음성을 녹음하는 것은 음성신호분석에 있어서 가장 정확한 분석을 할 수 있도록 해준다. 그러나 방음실이 있는 곳은 한정되어 있으므로, 현대인의 바쁜 생활 속에서 이런 곳을 찾는다는 것은 상당히 어렵다. 그래서 이에 보다 수월한 방법으로 전화기를 이용한다. 전화기를 이용하여 음성을 녹음한다면 방음실에서 녹음한 음성보다 음성의 질이 떨어지지만, 장소에 구애받지 않고 진단 받을 수 있다는 커다란 장점을 가질 수 있다.

본 논문은 녹음된 음성으로 분석된 파라미터를 신경회로망을 이용하여 식별하였다. 장애음성에 신경망을 이용하여 식별한 예[1][2]는 이미 있고, 그 식별율은 90% 이상의 값을 가진다. 그리고 후두질환에 대해 신경망으로 식별한 예 역시 존재한다.[3] 그러나 ARS 음성 데이터는 일반적인 음성 데이터와는 큰 차이를 보인다. 기본적인 샘플링 레이트나, 대역폭, 잡음성분의 포함정도, 또는 전화선 특유의 시간왜곡 등이 일반적으로 녹음한 음성과 많은 차이가 있다.

본 논문은 우선 장애음성의 수집과 수집된 음성의 특성파라미터의 분석에 대해서, 그리고 신경회로망을 이용한 식별 결과에 대해서 이야기한다.

2. 장애음성의 수집

장애음성수집은 일반적으로 방음실이나 조용한 실험실 환경에서 수집해야 하나, ARS 환경을 생각했기 때문에 비교적 조용한 방에서 전화기를 사용하여 녹음하였다.



그림 1. 음성 녹음 방법

환자는 일반적인 가정집에서 전화를 가정한다는 하에 6 평 정도되는 방안에서 일반적으로 사용하는 전화기를 사용하여 발성하였다. 그림 1.은 음성이 녹음되는 과정이다. 음성은 전화선을 통해 내부 전화망을 타고 ARS 보드에 와서 컴퓨터의 WAV 파일 형식으로 녹음되었다. 이 경우는 WAV 파일은 ARS 보드에서 지원하는 11,025 Hz의 샘플링 레이트에 8 bit의 샘플링 비트의 모노사양을 가지게 된다. 일반적으로 전화선은 약 300 Hz에서 3,400 Hz의 대역폭을 가지기 때문에 충분한 샘플링 레이트이다. 이렇게 저장된 음성은 크기는 남성과 여성, 그리고 각 병명별로 정상(normal), 양성후두질환에서 낭종(Intra cordal cyst), 라인케 부종(Reinke's edema), 후두염(Laryngitis), 성대결절(Vocal nodule), 성대마비(Vocal cord palsy), 용종(Vocal polyp), 기타, 그리고 악성후두질환에서 성대암 1기(T1), 2기(T2), 3기(T3)의 모두 3 종류의 11 개의 병명으로 분류된다. 실험에 사용된 음성은 정상 38, 장애음성 119, 모두 157 개의 음성이고 남녀의 구별은 없다.

3. 특성 파라미터

장애음성의 특성을 나타내는 파라미터는 다양하지만 본 논문에서는 MDVP에서 사용되는 33 가지의 파라미터 중 11 가지와 그 외 캡스트럼에 관련된 파라미터 1 가지를 합쳐서 12 가지의 파라미터를 선택하였다.[4] 33 개의 파라미터 중 선정기준은 퍼센트(%) 값, 즉 상대적인 값으로 나타나는 값들을 선정하였다. 기본 주파수 F_0 [Average Fundamental Frequency]나 기본 진폭값 A_0 [Average Fundamental Amplitude]같은 것 역시 음성의 특징을 나타내 주는 파라미터지만, 성별, 나이 등에 따라 사람마다 상당한 편차를 보여준다. 그렇기 때문에 병명에 따른 사람들의 특징 파라미터를 구하기에는 부적합하다. 그러므로 가능한 각 사람의 고유 특징이 나타나지 않는 파라미터를 선택하여야 한다. 파라미터를 특성에 따라 분류해 보면, 기본 주파수의 정보에 관한 값 (STD [Standard Deviation of the Fundamental Frequency]), 주파수 동요에 관한 값 ($Jitt$ [Jitter Percent], RAP [Relative Average Perturbation], PPQ [Pitch Period Perturbation Quotient], $sPPQ$ [Smoothed Pitch Period Perturbation Quotient], vF_0 [Coefficient of Fundamental Frequency Variation]), 진폭 동요에 관한 값($Shim$ [Shimmer Percent], APQ [Amplitude Perturbation Quotient], $sAPQ$ [Smoothed Amplitude Perturbation Quotient], vAm [Coefficient of Amplitude Variation]), 소음에 관한 값(NHR [Noise-to-Harmonic Ratio], VTI [Voice Turbulence Index], SPI [Soft Phonation Index]),[5] 그리고 캡스트럼에 관련된 값(HNR [Harmonic-to-Noise Ratio : Residual])이 있다.[6]

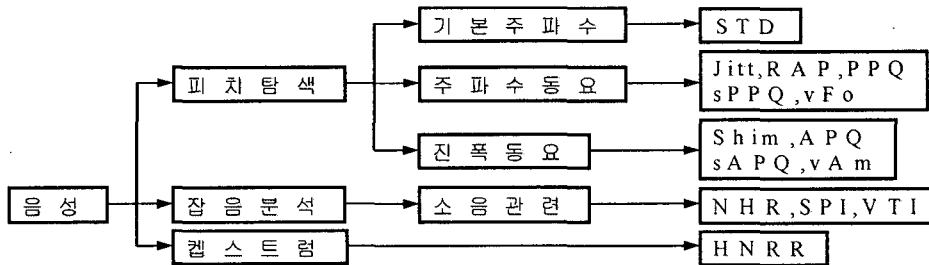


그림 2. 분석 파라미터

3.1 피치 및 연관된 파라미터들

피치성분을 정확히 구하는 것은 정확한 음성분석을 위해 필수적이다. 잘못된 피치값으로 파라미터들을 구하면 그 파라미터들은 의미가 없어지므로 이 과정은 아주 중요하다. 본 논문에서는 에너지 보간법을 이용한 피치 추출방법을 이용하여 개별피치를 구하였다. 이 방법은 시간영역에서 직접 신호에 연속적으로 일정한 윈도우를 씌워 신호의 처음부터 끝까지의 에너지를 구한다. 이렇게 구한 에너지 값은 음성신호에서 마치 저대역 통과 필터를 씌운 것과 같이 신호의 대략적인 파형만을 그리게 된다. 여기서 각각의 피크의 간격이 피치의 값이 된다. 이 방법은 개별 피치값을 구할 수 있기 때문에 피치의 누락으로 인한 파라미터값의 예러가 줄어드는 특성을 가진다.

이렇게 구해진 피치의 값들은 그 자체로도 많은 정보를 가지고 있지만, 직접적으로 사용

하기에는 그다지 적합하지 않다. 그래서 이 피치를 이용한 또 다른 파라미터, 예를 들면 *STD*, *Jitt*, *Shim* 등의 파라미터를 구하여 식별에 적용한다.

*Jitt*는 피치주기의 변화율을 나타내는데 사용되는 파라미터로, 연속적인 피치주기 사이의 평균변화율을 나타낸다. 이와 유사한 파라미터들로는 *RAP*, *PPQ*, *sPPQ* 등이 있는데 차이점은 *Jitt*는 단 2 개의 피치값을 사용하는데 반해 *RAP*는 3 개, *PPQ*는 5 개, *sPPQ*는 55 개의 피치값을 사용한다. 그러므로 *Jitt*의 경우에는 잡음에 영향을 많이 받지만, 피치의 섬세한 변화를 알 수 있고, *sPPQ* 쪽으로 갈수록 잡음에는 영향을 많이 받지 않지만, 피치의 섬세한 변화는 알 수 없다는 장단점이 있다. 그리고 이 피치값들의 표준편차를 나타내는 파라미터인 vFo 가 있다.

*Shim*는 진폭의 변화율을 나타내는데 사용하는 파라미터로 연속적인 진폭변화율의 평균값을 나타낸다. 이와 유사한 파라미터들로는 *APQ*, *sAPQ* 등이 있는데, 차이점은 *Shim*의 경우에는 단 2 개의 진폭값을 사용하는데 반해 *APQ*는 11 개 *sAPQ*는 55 개의 진폭값을 사용한다. 이것은 위의 피치의 경우와 유사하다. 이 파라미터들 역시 *Shim*의 경우에는 잡음에 많은 영향을 받지만 진폭의 섬세한 변화를 알 수 있고, *sAPQ* 쪽으로 갈수록 잡음에 영향을 많이 받지 않지만 진폭의 섬세한 변화를 알 수 없다는 장단점이 있다. 그리고 이 진폭값들의 표준편차를 나타내는 파라미터인 vAm 이 있다.

3.2 잡음성분 관련 파라미터

잡음성분의 분석은 주파수영역에서 이루어진다. 잡음성분에 관련된 파라미터는 3개가 있는데 셋 다 비슷한 파라미터이지만 계산하는 대역폭과 하모닉 성분에 따라서 달라진다. 우선 음성신호를 약 81.92 ms의 길이로 자른다. 이때 실제 데이터의 개수는 2^k 이 되도록 맞추어야 한다. 예를 들면 50 KHz의 음성인 경우 $0.08192 \times 50,000$ 의 계산을 하면 4,096 개의 데이터가 나오고, 25 KHz일 경우는 2,048 개의 데이터가 나온다. 25 KHz와 50 KHz의 사이 값인 30 KHz 같은 경우에는 큰 쪽으로 계산하여, 4,096 개의 데이터가 나온다. 그러므로 ARS 음성 같은 경우에는 11,025 Hz이므로 1,024 개의 데이터를 사용한다. 그리고 실제의 샘플링은 12.5 KHz가 된다. 그 후 6,000 Hz의 22차의 저대역필터를 거친 후 12.5 KHz의 신호로 다운 샘플링을 해야 하지만 ARS 음성의 경우는 그보다 낮은 샘플링 레이트를 가지기 때문에 이 과정은 생략하게 된다. 그리고 이 데이터들을 FFT한 후 파워스펙트럼을 구한다. 마지막으로 구한 파워스펙트럼의 데이터에서 하모닉 에너지와 인하모닉(노이즈) 에너지로 분리한다. 세 가지 파라미터는 모두 위의 과정을 거친 후에 계산이 된다. 우선, *NHR*은 노이즈와 하모닉 에너지의 비로써 1,500-4,500 Hz의 하모닉 에너지의 값과 70-4,500 Hz의 인하모닉(노이즈) 에너지의 비이다. *VTI*는 음성의 난조를 나타내는 파라미터로, *NHR*과 비슷하지만 그 범위가 달리 2,800-5,800 Hz의 하모닉 에너지와 70-4,500 Hz의 인하모닉 에너지의 비이다. *SPI*는 연발성(鞭發聲)을 나타내는 파라미터로 저주파(70-1,600 Hz)에서의 하모닉 에너지와 고주파(1,600-4,500 Hz)에서의 하모닉 에너지의 비이다.

3.3 캡스트럼 분석 파라미터

음원의 분석을 위해서는 선형예측분석으로부터 구해진 예측오차신호로부터 캡스트럼을

구한 뒤 성도필터성분과 주기성분, 잡음성분을 분리해 낸다. 이 결과로부터 *HNR*(Harmonic to Noise Ratio)를 구한다. 이렇게 하여 구한 파라미터를 *HNRR*(Harmonic-to-Noise Ratio: Residual)이라고 한다.

*HNRR*은 음성의 음원에서의 하모닉 성분과 잡음성분의 비를 추정한 것으로 볼 수 있으므로, 정확하지는 않지만 음원의 변화를 대략적으로 나타내주는 파라미터로 볼 수 있다. 이론적으로 위의 *NHR*과 비슷하지만 *NHR*의 경우는 FFT의 결과로 구한 파라미터이기 때문에 성도의 성분이 남아있으므로 *HNRR*과는 차이가 있다.

4. 식별 실험 및 결과

4.1 장애음성의 식별방법

패턴인식방법으로는 대량의 복잡한 데이터를 병렬처리 하는데 유용한 방법 중의 하나인 신경회로망을 사용한다. 왜냐하면 신경회로망은 데이터의 특성을 잘 알지 못하더라도 회로망 자체에서 데이터의 특성을 살려서 식별할 수 있는 특징을 가지고 있으므로, 대량의 복잡하고 그 특징을 분명히 파악할 수 없는 데이터라 할지라도 데이터의 특성이 고려되어서 식별이 가능하다.

본 논문에서 일반적인 Back-propagation을 사용하였다. 식별하고자 하는 것은 정상과 비정상의 두 개의 상태에 대해서이므로 다른 특별한 모델을 필요로 하지 않는다. 구성은 3 개, 6 개, 12 개의 입력을 가지고 3 개의 layer를 갖고 2 개의 출력을 갖도록 되어 있다. 출력벡터는 두 개의 값으로 나오는데 두 값 중 첫 번째 값이 크면 정상, 두 번째 값이 크면 비정상으로 식별한다. 각 layer의 전달함수는 1 번째 layer에서는 Hyperbolic Tangent Sigmoid를 사용하였고, 2 번째 layer에서는 Log Sigmoid, 3 번째 layer에서는 Linear 함수를 사용하여 학습하였다.

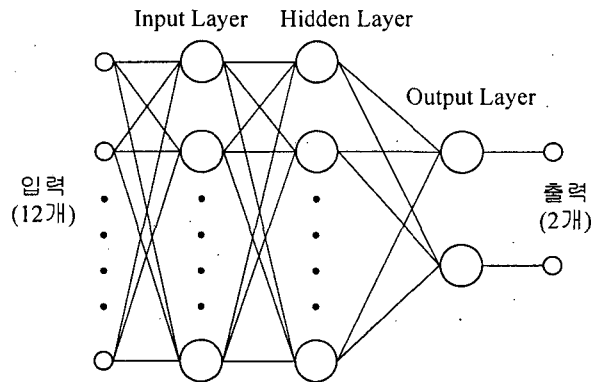


그림 3. 신경회로망의 구성

4.2 신경망의 결과

신경회로망의 식별에 쓰인 데이터는 위에서 수집한 11,025 Hz의 샘플링 레이트에 8 bit의 샘플링 비트의 모노사양을 가지는 WAV 파일이다. 개수는 정상음성 16 개와 장애음성 76 개, 총 92 개의 데이터이다. 식별은 정상음성과 장애음성 두 가지로 하였고 /아/를 약 1 초간 발음한 음성들 중 정상인 데이터 중 무작위로 2/3, 비정상인 데이터 중 무작위로 2/3를 선출한 데이터로 학습에 사용했고, 나머지 1/3을 식별에 사용했다. 이때 정상음성의 개수가 장애음성의 개수에 비해 너무나 많이 차이가 나기 때문에 정상음성을 두 배로 만들어 32 개의 정상음성과 76 개의 장애음성으로 식별실험을 하였다.

실험은 입력 파라미터의 개수를 달리하여 다섯 번에 걸쳐 실행하였다. 다섯 번에 걸쳐서 실험한 이유는, 신경망은 훈련시키는 데이터의 순서에 따라 결과가 달라지기 때문이다. 실험 데이터가 많을 경우에는 그 정도가 무시할 정도로 작아지지만, 실험데이터가 작을 경우에는 결과의 폭이 상당히 커진다.

우선 *Jitt*와 *Shim* 그리고 *HNRR*의 3 개의 파라미터만으로 식별 실험을 하였다. 과거 이 세 개의 파라미터로 Kay사의 Voice Disorders Database의 데이터를 이용하여 식별한 예 [1][2]가 있기 때문에 ARS를 통하여 잡음이 많이 섞인 음성이 방음실에서 녹음한 음성과 비교했을 때 식별률이 어떻게 차이가 나는지 비교해 볼 수 있다. 대략적으로 방음실에서 녹음한 음성 쪽이 ARS로 녹음한 음성보다 노이즈가 적고 대역폭이 크기 때문에 보다 낮은 결과가 예측된다.

두 번째로 전 파라미터를 사용하여 식별 실험을 하였다. 일반적으로 신경망은 입력파라미터가 많을수록 식별 시간은 오래 걸리지만 식별률은 높아진다. 때문에, 12 개의 파라미터를 사용해 식별하면 가장 좋은 결과가 나오리라고 기대할 수 있다.

세 번째는 6 개의 파라미터를 선별하여서 식별하였다. 이 경우는 위에서 선택한 12 개의 파라미터 중 파라미터의 군집도를 통해서 식별시 도움이 된다고 판단한 파라미터(*Jitt*, *Shim*, *hnr*, *nhr*, *vti*, *spi*)만을 선별한 다음 식별하였다.

표 1. 2. 3은 각각 3 개의 파라미터, 6 개의 파라미터, 12 개의 파라미터를 가지고 식별 실험을 한 결과이다. 그리고 식별실험의 식별회수는 모두 똑같으며 모두 10^{-15} 이하의 MSE (Mean Squar Error)를 가지므로 충분히 훈련되었다고 생각할 수 있다.

우선 3 개의 파라미터에 대한 식별율을 살펴보면, Training 데이터의 식별율이 100%가 되지 않는다. 이것은 충분히 훈련되기는 했지만, 완벽하게 훈련이 되지 않았다는 것을 의미한다. 그에 반해 6 개, 12 개의 파라미터의 경우를 살펴보면 모두 100%로 주어진 데이터에 한해서는 완벽하게 훈련되었음을 볼 수 있다. 이것이 의미하는 바는 3 개의 파라미터로는 정상과 비정상의 식별하는데 부족하다는 것이다. 다시 말해 *Jitt*와 *Shim* 그리고 *HNRR*만으로는 정상과 비정상음성의 특징을 확실하게 나타내 주지는 못한다는 것이다. 그리고 훈련된 신경망으로 식별한 Test 데이터의 결과는 보면 1차 82.86%, 2차 88.57%, 3차 88.57%, 4차 88.57%, 5차 77.14%로써 평균 85.14%이다. 이 식별률은 다른 두 식별율(6 개의 파라미터 : 92.57, 12 개의 파라미터 : 88.57)과 비교했을 때 가장 떨어진다.

그리고 6 개의 파라미터에 대한 식별률과 12 개의 파라미터에 대한 식별률을 비교해보면, 6 개의 파라미터를 사용한 신경망의 식별률이 더 좋을 수 있다. 기본적으로 많은 파라

미터를 가지고 식별한 결과가 더 좋지만, 그렇지 않은 이유는 식별에 사용된 12 개의 파라미터 전부가 정상음성과 비정상 음성을 구분하는데 도움을 주지는 않는다는 것이다. 분명 12 개의 파라미터들은 음성의 특성을 나타내주지만, 그 특성들이 정상음성과 비정상음성을 구분하지 못하고 공통되는 특징을 나타낼 수도 있다. 그렇게 되면 인식률은 자연히 떨어지게 될 뿐이다.

표 4는 본 논문에서 실험한 결과와 선행연구에서 방음실의 데이터를 이용하여 수집한 음성의 식별결과이다. 선행연구 1은 과거 본 연구실에서 Kay사의 Disordered Voice Database의 음성데이터에서 3 개의 파라미터(Jitter, Shimmer, HNR)를 사용하여 식별한 결과이고, [1] 선행연구 2는 같은 데이터로 타 연구실에서 MDVP에서 구할 수 있는 26 개의 파라미터로 신경망을 이용하여 식별한 결과이다.[2] 두 실험 모두 ARS 음성을 사용하지 않았기 때문에 본 논문의 식별결과와 직접적으로 비교할 수는 없지만, 상대적인 평가는 가능하다. 우선 3 개의 파라미터에 대한 식별률은 동일한 파라미터를 사용한 선행연구 1과 비교했을 때 상대적으로 떨어진다. 그것은 ARS 데이터의 경우 잡음성분이 많고, 전화선 자체의 시간왜곡현상 등으로 인한 정확한 데이터의 수집이 어려움에 기인한다. 게다가 방음실의 경우는 25 KHz의 샘플링 레이트를 가지지만, ARS의 경우에는 11 KHz의 샘플링 레이트를 가지므로, 대역폭의 차이로 인해 같은 음성데이터라 할 지라도 서로 다른 파라미터 값을 가지게 된다. 이러한 단점을 보완하기 위해 본 논문에서는 식별에 도움이 되는 파라미터를 추가함으로써 인식률을 높였다. 12 개의 파라미터 전부를 사용했을 경우의 실험결과는 선행연구 1의 결과와 비슷한 결과를 보이고, 6 개의 파라미터를 사용했을 경우에는 더 뛰어난 식별 결과를 보이게 된다. 그러나 선행연구 2와 비교했을 때에는 미치지 못한다. 이는 선행연구 2에서 사용된 26 개의 파라미터에 대해 상당히 부족한 6 개의 파라미터만을 사용했을 뿐더러, 음성데이터 역시 높은 샘플링 주파수를 가지는 방음실의 데이터에 비해 상대적으로 낮은 샘플링 주파수를 가지는 데이터이므로 식별률이 떨어진다. 그러나 사용된 파라미터의 개수나 음성데이터의 우위에도 불구하고 선행연구 2의 식별결과가 월등히 뛰어나다고 할 수는 없으므로 식별시간이나 실제 시스템 구성의 편의성에 생각한다면 본 논문의 식별방법이 상당히 유용함을 알 수 있다.

표 1. 3 개의 파라미터에 대한 식별률

		Test Data		Training Data	
		정 상	비정상	정 상	비정상
1차	Success	8	21	20	49
	Fail	2	4	2	2
	Total	10	25	22	51
	식별률	82.86%		94.52%	
2차	Success	10	21	22	44
	Fail	0	4	0	7
	Total	10	25	22	51
	식별률	88.57%		90.41%	

		Test Data		Training Data	
		정 상	비정상	정 상	비정상
3차	Success	10	16	20	49
	Fail	0	9	2	3
	Total	10	25	22	51
	식별률	88.57%		94.52%	
4차	Success	8	23	20	51
	Fail	2	2	2	0
	Total	10	25	22	51
	식별률	88.57%		97.26%	
5차	Success	10	17	22	36
	Fail	0	8	0	15
	Total	10	25	22	51
	식별률	77.14%		79.45%	

표 2. 6 개의 파라미터에 대한 식별률

		Test Data		Training Data	
		정 상	비정상	정 상	비정상
1차	Success	10	23	22	51
	Fail	0	2	0	0
	Total	10	25	22	51
	식별률	94.29%		100%	
2차	Success	8	23	22	51
	Fail	2	2	0	0
	Total	10	25	22	51
	식별률	88.57%		100%	
3차	Success	8	24	22	51
	Fail	2	1	0	0
	Total	10	25	22	51
	식별률	91.43%		100%	
4차	Success	10	25	22	51
	Fail	0	0	0	0
	Total	10	25	22	51
	식별률	100%		100%	
5차	Success	8	23	22	51
	Fail	2	2	0	0
	Total	10	25	22	51
	식별률	88.57%		100%	

표 3. 12 개의 파라미터에 대한 식별률

		Test Data		Training Data	
		정 상	비정상	정 상	비정상
1차	Success	10	20	22	51
	Fail	0	5	0	0
	Total	10	25	22	51
	식별률	85.71%		100%	
2차	Success	8	23	22	51
	Fail	2	2	0	0
	Total	10	25	22	51
	식별률	88.57%		100%	
3차	Success	8	23	22	51
	Fail	2	2	0	0
	Total	10	25	22	51
	식별률	88.57%		100%	
4차	Success	8	24	22	51
	Fail	2	1	0	0
	Total	10	25	22	51
	식별률	91.43%		100%	
5차	Success	8	23	22	51
	Fail	2	2	0	0
	Total	10	25	22	51
	식별률	88.57%		100%	

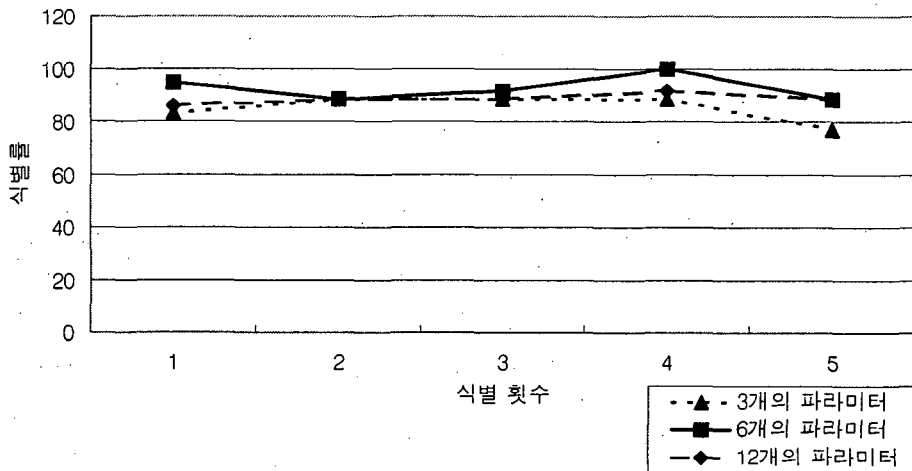


그림 4. 3 개, 6 개, 12 개의 파라미터에 대한 식별률의 비교

표 4. 전체 식별률

	종류	3 개	6 개	12 개	선행연구 1	선행연구 2
식별률	Test Data	85.14%	92.57%	88.57%	88.64%	94.87%
	Training Data	91.23%	100%	100%	90.9%	100%

5. 결 론

본 논문에서는 ARS를 이용하여 국내 후두질환환자의 음성을 수집하고, 수집한 데이터로 12 개의 파라미터로 분석하였다. 그렇게 분석된 파라미터들로 각각 3,6,12 개로 나누어 신경망을 이용하여 정상과 비정상음성을 식별하였다. 식별에 사용된 신경망은 back-propagation을 사용한 3-layer, 2 출력을 가지는 신경회로망이다. 실험결과 6 개의 파라미터를 이용한 결과가 가장 높게 나왔고, 그때의 평균 식별률은 92.57%이다. 이 결과는 충분히 정상음성과 질환음성을 이 방법에 의해 구분이 가능함을 알 수 있었다.

차후의 연구에서는 기존의 데이터베이스를 확장하여 더 많은 장애음성을 수집하여 일반성을 높이고 ARS로부터 오는 신호의 음향적 특성을 고려한 파라미터를 도출할 필요가 있다. 또한 단순히 정상음성과 질환음성만을 구분하는 것이 아니라 보다 세부적인 분류방법을 개발하여 유용성을 높이고자 한다.

참 고 문 헌

- [1] 김대현, 조철우, "장애음성의 분류방법에 관한 연구" 제15회 음성통신 및 신호처리 워크샵, pp388-391, 1998.
- [2] Godino-Llorente, Santiago Aguilera-Navarro*, Carlos Hernandez-Espinosa**, Mercedes Fernandez-Redondo**, Pedro Gomez-vilda+, (*LTR, **Dpto. de Informatica, +Facultad de Informatica) "On The Selection of Meaningful Speech Parameters Used by a Pathologic/Nonpathologic Voice Register Classifier", Eurospeech '99, Volume 1, Page 563-566.
- [3] 김광인, 조철우, 김대현, *왕수건, *전계록, *안시훈, *김기련, *김용주. (창원대학교, *부산대학교), "ARS와 신경회로망을 이용한 장애음성의 수집, 분석 및 식별에 관한 연구", 제 13회 신호처리 합동학술 대회, pp955-958, 2000.
- [4] Operations Manual, 'Disordered Voice Database', Model 4337, version 1.03, Kay Elemetrics Corp, 1994.
- [5] 조철우, 김대현, 김광인, *왕수건, *전계록, *안시훈, *김기련, *김용주. (창원대학교, *부산대학교), "다양한 수집방법에 의한 장애음성 분석도구의 구현에 관하여", 제 17회 음성통신 및 신호처리 합동학술대회 논문집, pp211-214, 2000.
- [6] 조철우, 김대현, "Cepstrum 방법과 신경회로망을 이용한 정상, 양성종양, 악성종양 상태의 식별에 관한 연구", 한국음향학회 추계학술발표대회 논문집 제18권, pp.399-402, 1998.

접수일자: 2001. 4. 17.

게재결정: 2001. 5. 28.

▲ 조철우

경남 창원시 사림동 9번지
창원대학교 제어계측공학과
Tel : +82-279-7552 , H/P : 011-580-7552
E-mail : cwjo@sarim.changwon.ac.kr

▲ 김광인

경남 창원시 사림동 9번지
창원대학교 제어계측공학과
H/P : 011-842-1833
E-mail : karisman@hitel.net

▲ 김대현

경남 창원시 사림동 9번지
창원대학교 제어계측공학과
H/P : 011-562-6047
E-mail : midas03@taegu.net

▲ 권순복

부산광역시 서구 아미동 1-10번지
부산대학교 일반대학원 의공학협동 박사과정
Tel : +82-051-240-7543
E-mail : sbkwon@intizen.com

▲ 김기련

부산광역시 서구 아미동 1-10번지
부산대학교 일반대학원 의공학협동 박사과정
Tel : +82-051-819-2866
E-mail : geniusgr@hanmail.net

▲ 김용주

부산광역시 서구 아미동 1-10번지
부산대학교 일반대학원 의공학협동 석사과정
Tel : +82-051-819-2866
E-mail : fid2@orgio.net

▲ 전계록

부산광역시 서구 아미동 1-10번지
부산대학교 의과대학 의공학교실 교수
Tel : +82-051-240-7535
E-mail : grjeon@hyowon.pusan.ac.kr

▲ 왕수건

부산광역시 서구 아미동 1-10번지
부산대학교병원 이비인후과 교수
Tel : +82-051-240-7331
E-mail : Wangsg@pusan.ac.kr