

장애음성의 분류방법에 관한 연구

김대현, 조철우

창원대학교 제어계측공학과

On the Classification of the Pathological Speech

Dae-Hyun Kim, Cheol-Woo Jo

Department of Control & Instrumentation Engineering

cwjo@sarim.changwon.ac.kr

요약

본 논문에서는 jitter, shimmer 및 캡스트럼 방식의 음원분석에 의한 파라미터를 이용하여 장애음성을 진단, 식별하는 방법을 제안한다. 먼저 통계적 처리결과를 바탕으로 식별에 유효한 파라미터들을 선택하고 이들 파라미터들을 이용하여 최종 진단한다. 식별방법으로는 신경회로망을 이용한다. 입력파라미터로는 jitter, shimmer, HNRR을 사용한다. 신경회로망은 1 은닉층을 갖는 3-layer 신경회로망을 사용한다. 실험결과 효과적으로 정상음성과 장애음성의 구분이 가능하였다[1]

1. 서론

최근 들어 인간의 건강에 대한 관심이 점점 증가하고 있다. 이비인후과 영역에서는 성대의 질환을 음성의 음향적 특성을 분석하여 질병의 징후를 발견하려는 시도가 여러 곳에서 이루어지고 있다. 이러한 시도는 질병의 정확한 진단이 목적이기보다는 질병의 가능성을 사전에 발견하여 질병의 상태가 깊어지지 않도록 예방한다는 차원에서 필요하다. 일반적으로 성대의 질환을 진단하기

위해서는 성대를 직접 들여다 볼 수 있는 내시경과 같은 기구를 이용하여 직접 보는 것이 가장 효과적이라고 한다. 실제로 의사들도 이러한 방법을 많이 사용하고 있다. 그러나 이 방법은 정확한 진단이 가능한 반면 숙달된 전문의사와 전문적인 기구가 필요하여 일반의사나 환자 자신이 질병의 유무를 판단하기에는 적합한 방법이 아니다. 이러한 목적으로는 환자의 음성만에 의하여 질병의 유무, 종류를 진단하려는 시도가 곳곳에서 이루어지고 있다. 성대의 질병은 많은 경우 성질(Voice Quality)의 변화를 수반하지만 그렇지 않은 경우도 있다. 목소리만으로 질병을 진단할 수 있는 경우는 질병으로 인해 음성의 음향적 특징이 변화된 경우에 한한다.

본 논문에서는 이와 같은 음향적 분석법에 의한 성대질환의 진단을 목적으로 하는 파라미터로 사용되는 jitter, shimmer 및 캡스트럼 방식의 음원분석에 의한 파라미터인 HNRR[1]의 조합으로 신경회로망을 이용하여 정상음성과 장애음성을 구분하는 실험을 행하고 그 결과를 고찰하였다.

2. 장애음성의 식별파라미터

장애음성의 음향적 특성을 나타내는 파라미터로는 여러 가지 종류가 있다. 그 중에서도

통계적으로 분석에 많이 사용되어지고 변별력이 있는 파라미터로 jitter, shimmer가 대표적인 파라미터이다.[2] 그리고 기존의 파라미터의 신뢰성 문제로 인해 새로 제안된 캡스트럼 방식의 음원분석에 의한 파라미터인 HNRR(Harmonic to-Noise Ratio : Residual)를 장애음성 식별에 사용하려고 한다.

Jitter는 피치주기의 변화율을 나타내는데 사용되어지는 파라미터로 연속적인 피치주기 사이의 평균 퍼센트 변화로 나타낼 수 있다.

$$Jitter = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |P(i) - P(i+1)|}{\frac{1}{N} \sum_{i=1}^N P(i)} \quad (2-1)$$

식2-1에서 $P(i)$ 는 i 번째 피치의 주기이고, N 은 측정된 피치의 개수이다.

Shimmer는 jitter와는 달리 peak사이의 크기에 대한 변화율을 나타낸다.

$$Shimmer = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A(i) - A(i+1)|}{\frac{1}{N} \sum_{i=1}^N A(i)} \quad (2-2)$$

식2-2에서 $A(i)$ 는 i 번째 peak사이의 크기를 나타낸다.

마지막으로 HNRR은 음성에서 선형예측분석으로부터 구해진 예측오차신호로부터 캡스트럼을 구한 뒤 하모닉 성분과 잡음성분을 분리하여 그 비율을 나타낸 파라미터이다.[1][3] 선형실험에서 HNRR은 jitter나 shimmer에 비해 높은 변별력을 갖는 것이 확인되었다.

이러한 세 가지 파라미터를 이용하여 서로간의 통계적 처리결과를 바탕으로 장애음성과 정상음성사이의 관계 및 효과적인 식별방법을 찾고자 한다. 그 식별방법으로 신경회로망을 이용한다.

3. 패턴인식방법

근래 들어 인간의 두뇌가 대량의 복잡한 데이터를 병렬 처리할 수 있을 뿐만 아니라

학습능력이 있다는 사실에 근거하여 새로운 패턴인식의 방법으로 신경회로망이 자주 사용되고 있다.

신경회로망은 뉴런 즉 처리할 요소의 전이함수, 회로망의 구조 즉 layer의 수 및 처리할 요소들 간의 연결상태 및 연결강도를 주어진 문제해결에 적절하게 조정하는 학습 규칙들에 따라 여러 가지의 형태가 있다. 대표적인 것을 들면 Hopfield network, perceptron, Boltzmann machine, error propagation model, Grossberg model 및 neocognition 등이 있다. 이들에서 패턴인식 및 식별에 직접적으로 관련되는 모델들은 Hopfield network, Grossberg model, perceptron, Neocognitron 등이 있으며, 그중 perceptron은 feed-forward 연결구조를 가지며 패턴 식별의 기능을 갖는 간단한 형태의 신경회로망이다. 초기의 단층 구조의 perceptron은 그 구조의 간결성과 특히 수학적으로 증명된 학습 규칙으로 많은 관심의 대상이었으며, 그후 Rumelhart 등에 의하여 back-propagation의 학습 알고리즘이 제안됨으로써 다층구조의 perceptron은 그 계산기능이 크게 확장되어 많은 문제에 응용될 수 있게 되었다.[4]

Back-propagation은 일종의 gradient 탐색 방법으로서 기대되는 출력층의 node들에 대하여 비용함수를 계산하여 이를 감소하는 방향으로 가중치를 조정하여 다음에 한층 낮은 층의 node에 대하여 에러를 계산하여 이를 감소하도록 가중치를 조절하는 식으로 반복한다.

본 논문에서는 3개의 계층구조를 갖는 신경회로망을 구성하고 back-propagation 방법에 의해서 훈련시킨 다음 정상음성과 장애음성의 식별에 사용하였다.

4. 실험 및 검토

본 논문에서는 Kay의 Disordered Speech Database[5]를 대상으로 정상음성과 장애음성을 식별하는 실험을 수행하였다. 이 데이

장애음성의 식별에 관한 연구

터베이스에서 정상음성 53개와 장애음성 중 3가지 병명의 데이터 127개를 선택하여, 전체 180개의 데이터에 대해 실험하였다. 실험에 사용한 각 음성은 모음 /아/를 일정한 시간동안 발음한 것으로 샘플링 주파수는 25kHz이며, 양자화 비트수는 16bit로 되어있다.

그림1은 전체적인 실험방법에 대해 간략하게 나타낸 것이며, 먼저 정상음성과 장애음성에 대한 파라미터로 jitter, shimmer, HNRR을 구해 각 파라미터들을 일정한 조합으로 신경회로망의 입력이 된다.

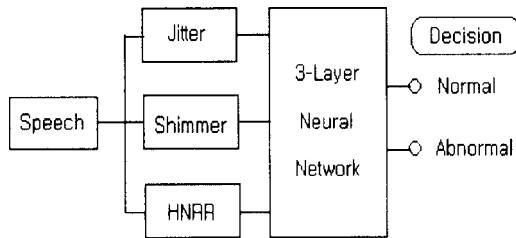


그림 1 전체적인 실험방법

신경회로망은 그림2에 주어진 것과 같이 입력층과 출력층 사이에 1개의 은닉층을 가지는 다층구조 신경망을 사용하였다. 입력층은 세가지 파라미터의 조합이고, 은닉층에는 3개의 뉴런을 사용하였으며, 원하는 출력값은 정상과 비정상(장애음성)으로 구분하기 위해 1과 0으로 하였다. 신경망의 학습을 위해서는 역전파 학습알고리즘을 이용하였다.

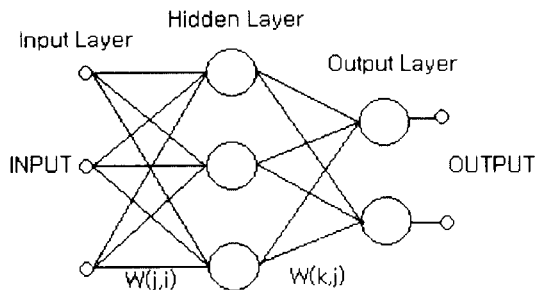


그림 2 다층 구조 신경회로망

신경회로망에 사용한 전체적인 입력데이터 중 훈련을 위해 2/3를 훈련데이터로 사용하

고 나머지 1/3은 테스트 데이터로 사용하였다. 훈련데이터와 테스트데이터에 대한 에리의 수를 표1에 나타내었으며, 표2에서는 정상음성과 장애음성에 대한 식별률을 나타낸 것이다.

표1. 데이터의 Error의 수

	Train Data				Test Data			
	JS	SH	HJ	JSH	JS	SH	HJ	JSH
1	.	3	3	3	.	4	7	10
-1	6	4	3	2	3	1	.	2

(1 : 정상->비정상, -1 : 비정상->정상
J : Jitter, S : Shimmer, H : HNRR)

표2. 데이터의 식별률(단위 : %)

	JS	SH	HJ	JSH
Train Data	96.43	95.83	96.42	97.02
Test Date	96.51	94.19	91.86	86.05

정상음성과 장애음성이 뚜렷한 경계를 가지고 식별되지 않기 때문에 back-propagation을 이용하여 원하는 에리에서 가장 근접할 때까지 훈련을 계속하여 에리의 범위가 더 이상의 변화가 없을 경우 훈련을 그만두는 방법을 선택하였다. 훈련데이터에서 장애음성이 정상음성으로 식별되는 경우가 다소 많지만, 식별률이 거의 95%이상으로 구분이 가능한 결과를 얻을 수 있었다. 테스트데이터를 가지고 다시 테스트를 해 본 결과 두가지 파라미터를 가지고 판별을 하는 경우가 식별률이 좋으며, 세가지 모두를 가지고 판별을 할 경우는 식별률이 낮음을 알 수 있다. 그리고 테스트데이터의 경우는 훈련데이터와는 달리 정상음성이 장애음성으로

판별되는 경우가 많은 것을 알 수 있다.

그러나 표1,2에서 알 수 있듯이 정상음성과 장애음성의 정확한 구분이 힘들지만 신경회로망을 이용하여 90%이상의 식별률로 두 음성간의 구분이 자동으로 이루어질 수 있음을 알 수 있었다.

5. 결론

본 논문에서는 정상음성과 장애음성을 구분하기 위해 3-layer의 신경회로망을 이용하였으며 신경회로망의 입력으로는 장애음성식별파라미터인 jitter, shimmer, HNRR을 이용하였다. 신경회로망의 학습을 위해서는 back-propagation을 사용하였다. 실험결과 세 가지 파라미터를 이용하여 효과적으로 정상음성과 장애음성간의 구분이 가능하다는 것을 입증할 수 있었다.

차후의 연구에서는 보다 많은 종류의 질병에 대하여 정상음성과의 식별 및 각 질병간의 구분에 대해서도 실험을 계속 진행할 예정이다. 또한 악성종양 환자의 음성수집 및 국내의 정상인과 장애음성환자의 음성을 수집하여 동일한 실험을 계속할 예정이다.

감사의 글

본 연구는 한국학술진흥재단의 1997 학제간 연구기 지원에 의한 연구결과와 일부입니다. 지원에 감사드립니다.

참 고 문 헌

1. 조철우, 김대현, "음원분석을 통한 장애음성의 음향적 특성분석에 관한 연구", 한국음향학회 하계학술발표대회 논문집, pp163-166, 1998, 7
2. Operations Manual, "Multi-Dimensional Voice Program(MDVP)", Model 4305, Kay Elemetrics Corp, 1993
3. B.Yegnanarayana, C.d'Alessandro,

V.Darsinos, "An Iterative Algorithm for Decomposition of Speech Signals into Periodic and Aperiodic Components", IEEE trans. on Speech and Audio Processing, Vol.6, No 1, Jan. 1998

4. R.P.Lippman, "An Introduction to computing with Neural Nets", IEEE ASSP Magazine, Vol.4. No 2, pp4-20, April, 1987

5. Operations Manual, "Disordered Voice Database", Model 4337, version 1.03, Kay Elemetrics Corp, 1994