Wavelet 변환과 신경회로망을 이용한 후두의 양성종양의 식별에 관한 연구

김대현, 조철우

창원대학교 제어계측공학과

Classification of Pathological Speech Signals Using Wavelet Transform and Neural Network

Cheol-Woo Jo, Dae-Hyun Kim
Dept. of Control and Instrumentation Eng.
Changwon National University, Changwon, Kyeongnam, 641-773 Korea
cwjo@sarim.changwon.ac.kr

요약

본 논문에서는 웨이브렛 변환에서 구해진 파라미터와 신경회로망을 이용하여 후두의 양성종양과 정상상태를 구분하는 실험을 행하였다. 식별 파라미터로는 웨이브렛변환으로부터 도출된 ECS 파라미터와 jitter, shimmer 등 이용하였으며 신경회로망은 한 개의 은닉층을 갖는 다층구조 신경망을 이용하였다. 신경망의 입력으로는 세가지 파라미터의 조합을 두개 또는 세개를 입력하여 각각의 경우의 식별율을 조사하였다 실험결과 75%에서 93%에 이르는 식별율을 얻었다.

1. 서론

최근들어 인간의 건강에 대한 관심이 점점 증가하고 있다. 이비인후과 영역에서는 성대의 질환을 음성의 음향적 특징을 분석하여 질병의 징후를 발견하려는 시도가 여러 곳에서 이루어 지고 있다. 이러한 시도는 질병에 대한 정확한 진단이 목적이라기 보다는 질병의 가능성을 사전에 발견하여 질병의 상태가 걷잡을 수 없도록 예방하자는 차원에서 필요하다. 일반적으로 성대의 질환을 진단하기 위해서는 성대를 직접 들여다볼 수 있는 내시경과 같은 기구를 이용하여 직접 보는 것이 가장 효과적이라고 한다. 실제로 의사들은 이러한 방법을 많이 사용하고 있다. 그러나 이 방법은 정확한 진단이 가능한 반면 숙련된 전문 의사와 고가의 기구가 필요하여 전문의가 아닌 일반의사와 환자자신이 질병의 유무를 판단하기에는 정확한 방법이 아니다. 이와 같은 질병진단의 목적으로 환자의 음성만을 사용하여 질병의 유무나 종류등을 판별하려는 시도가 곳곳에서 이루어 지고 있다. 성대의 질병의 경우 많은 경우 성질(Voice Quality)의 변화를 수반하지만 그렇지 않은 경우도 있다. 목소리만으로 질병을 진단할 수 있는 경우는 질병으로 인해 음성의 음향적 특성이 변화된 경우에 한한다.

본 논문에서는 이와 같은 음향적 분석법에 의해 성대의 질환가능성을 판정하려는 시도의 하나로 신경회로망을 이용하여 정상음성과 양성종양과를 구분하는 실험을 행하고 그 결과를 고찰하였다.

2. 장애음성의 음향적 특성

성대장애환자의 음성은 성대의 장애로 인하여 발성기관의 유원부분이 변형되면서 음성의 특성이 변하게 된다. 특성의 변화는 성대의 장애부위에 따라서 나타나게 되는데 대개의 변화는 성대부위의 종양, 경직화 등이 원인이 되어 불규칙적인 변화를 가져온다.

성대장애환자의 음성의 음향적 특성은 정상인에 비하여 피치의 불규칙성, 잡음성분의 증가, 고주파성분의 증가 등으로 나타난다. 이러한 변화를 묘사하기 위하여 제안된 기존의 파라미터들로는 jitter, shimmer, NHR 등이 있다 이러한 파라미터들 중 어느 것도 장애음성의 특징을 명확히 구분해 주지는 못하고 있기 때문에 여러가지 파라미터를 복합적으로 사용하여 진단에 사용하고 있

다.[1][2]

특히 기존의 파라미터들의 경우 피치주기를 먼저 측정하여 주기의 불규칙성등을 계산하고 있기 때문에 피치의 측정이 잘못된 경우 파라미터의 값은 신뢰성이 떨어지게 된다. 장애음성의 경우는 피치가 두 배가 되든지 반이 되는 경우가 많이 일어나며 육안으로도 구분이 되지 않을 정도로 불규칙한 경우가 많다. 장애음성을 진단하는 방법중의 하나로 음질을 이용하여 환자음성을 평가하는 방법도 있다.

본 연구에서는 환자의 성대의 질병상태를 자동으로 진단할 수 있게 하기 위한 연구의 일환으로 신경회로망에 의한 장애음성중 양성종양을 식별하는 실험을 신경회로망을 이용하여 시도하였다.

본 실험에서는 장애음성의 종류를 양성종양으로 제한하였다.

3.식별파라미터

식별파라미터로는 기존의 파라미터인 jitter, shimmer 와 웨이브렛변환결과로부터 구해진 ECS를 사용하였다.

Jitter는 피치주기의 변화를 나타내는 데 사용하는 파라미터로 연속적인 피치주기사이의 평균 퍼센트 변화로 나타낼 수 있으며, 정상음성보다 장애음성에서 더 커짐을 알 수 있다.

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

식(1)에서 P(i)는 i번째 피치주기이고 N은 측정된 피치의 개수이다.

Shimmer는 jitter와는 달리 peak 사이의 크기에 대한 변화율을 나타낸다

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

식(2)에서 A(i)는 i번째 peak 사이의 크기를 나타낸다. 그러나 jitter와 shimmer는 정의식에서 보는 바와 같이 피치주기 특징의 에러에 의해 영향을 받는다.

마지막으로 사용한 파라미터는 웨이브렛 변환을 이용한 대역에너지의 비이다. 이 파라미터를 ECS(Energy Consistency Slope)라고 한다. ECS는 아래와 같은 과정에 의해서 구해진다.

웨이브렛은 일종의 시간-주파수 영역 변환이다. 이 변환은 다양한 시간-주파수 영역의 해상도를 제공해 주기 때문에 필터링이나 신호압축등에 자주 사용되고 있다.

DWT(Dyadic Wavelet Transform)의 경우에는 저주파 대역에서는 낮은 주파수 해상도를 보여주며 고주파대역에서는 높은 주파수 해상도를 보여준다. 식(3)은 웨이브렛 변환의 정의식이다.

$$WT(a,b) = \int f(t)\Psi_{a,b}(t)dt \quad (3)$$

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \cdot h\left(\frac{t-b}{a}\right)$$

그림 1은 본 실험에서 사용된 웨이브렛 변환의 구조를 보여준다.

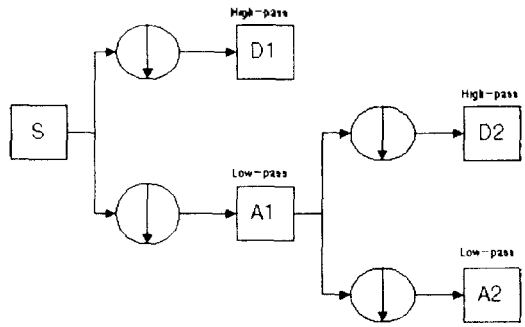

그림 1. 웨이브렛 변환

이 변환결과로부터 ECS 파라미터는 다음과 같은 과정을 거쳐 구해진다.

먼저 2 레벨 웨이브렛 변환을 거쳐 저대역과 고대역부분을 분리한다 분석한 결과에서 보면 저대역 스케일 신호는 고대역 스케일신호에 비해 주기성분이 강하면서 상대적으로 잡음성분이 아주 작고, 또한 에너지가 매우 크다. 고대역 스케일성분은 원래 음성이 갖고 있는 성분중 잡음성분에 가깝다고 볼 수 있다. 이와 같은 음성신호의 잡음성분을 측정하는 파라미터로 NHR이 제안되어 있기는 하지만 이는 음성신호에서 70-4500Hz 사이의 하모닉성분과 1500-4500Hz 사이의 잡음성분의 비를 FFT한 결과로부터 구한 것으로 ECS와 의미적으로는 유사하지만 차이가 있다.

ECS를 제안한 이유는 장애음성의 식별에 있어서 jitter, shimmer 등 파라미터가 피치추출의 정확도에영향을 받기 때문에 피치검출율에 영향을 받지 않는 파라미터를 찾기 위해서이다.

$$E_N(i) = \frac{E_i}{E_T}(i = 1,2,3) \quad (4)$$

식(4)는 N 번째프레임의 i 번째 스케일의 에너지 비이다. E_T는 N 번째 프레임의 총 에너지이다. ECS 는 $E_N(1)$, $E_N(2)$, $E_N(3)$의 값을 잇는 직선의 기울기를 구한 것이다.

ECS 값의 분포는 잡음이 많은 음성의 경우 상대적으로 고주파대역의 에너지가 커져서 기울기가 완만해 지고 잡음성분이 적은 음성의 경우 고주파대역의 에너지가 작아지므로 기울기가 급해진다.

이 파라미터의 유효성을 확인하기 위해 정상음성과 양성종양음성으로부터 각 파라미터를 구하여 평균치와 표준편차값을 구한 것이 표 1 에 나타나 있다

표 1. 각 파라미터의 평균과 표준편차분포

파라미터	구분	평균	표준편차
Jitter	정상	0.6157	0.4377
	종양	1.9645	2.4118
Shimmer	정상	2.2055	0.9242
	종양	6.5540	4.1706
NHR	정상	0.1158	0.0124
	종양	0.1945	0.1241
ECS	정상	-0.4985	0.0010
	종양	-0.4850	0.0356

표 1 에서 보면 Jitter 와 Shimmer 의 경우 정상과 양성종양 상태간에 뚜렷한 차이를 보이고 있으며 NHR 과 ECS 는 ECS 가 약간 나은 분리도를 보이고 있음을 그래프를 통해 확인할 수 있었다. NHR 의 경우는 파라미터를 구하는 과정이 ECS 에 비하여 복잡하고 하모닉 성분을 추출하는 과정에서의 오차발생가능성이 있으므로 식별실험에 ECS 를 사용하기로 하였다.

그림 2 는 정상 및 양성종양상태에서의 각 스케일별 에너지의 변화를 그린 것이다. o 표는 양성종양의 경우, +표는 정상음성의 경우를 나타낸 것이다.

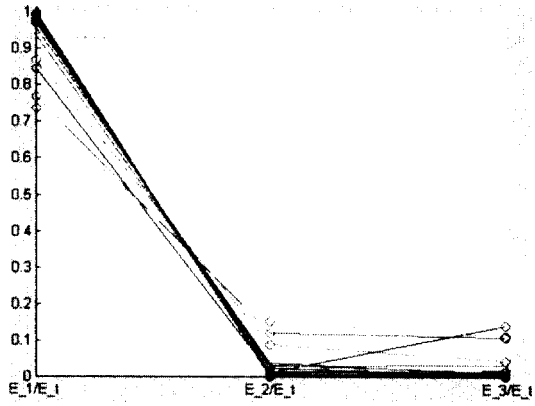

그림 2 스케일별 에너지의 분포

(* : normal voice, o : pathological voice)

4. 식별실험

앞서 제시한 세가지 파라미터(jitter, shimmer, ECS)를 이용하여 식별실험을 행하였다.

실험에 사용한 음성데이타는 KAY 의 Disordered Speech Database[3]를 대상으로 정상음성과 장애음성을 식별하는 실험을 수행하였다. 이 데이터베이스에서 정상음성 53 개와 장애음성 68 개 중 121 개를 택하였는데 장애음성은 4 가지 병명을 선택하였다. 선택기준은 성대의 이상이 뚜렷하다고 판단되는 질병만을 포함하였다.

각 음성데이타는 모음 /아/를 일정한 시간동안 발음한 것으로 표본화 주파수는 25KHz 이며 16 비트로 양자화 되어 있다.

식별을 위하여 3 층의 신경망 회로를 구성하였고 입력 파라미터로는 Jitter, Shimmer, ECS 의 세가지를 조합하여 사용하였다. 파라미터를 조합한 이유는 어느 한 파라미터만을 사용할 경우는 파라미터를 조합한 경우보다 식별율이 떨어질 수 있음을 파라미터의 분포도에서 확인하였기 때문이다. [4]

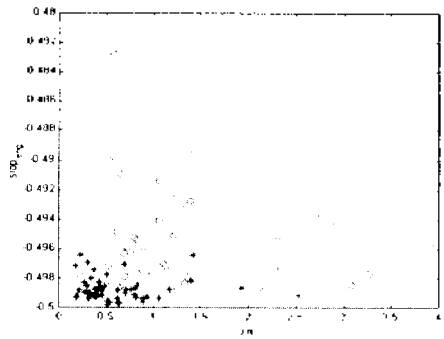

그림 3. 파라미터의 조합에 의한 식별율개선사례

그림 3 은 파라미터 조합에 의한 식별율 개선의 사례를 보여주는 파라미터 분포도이다. 그림 3 에서는 Jitter.와 ECS 의 파라미터 분포를 보여준다. 그림에서 o 는 종양상태의 음성을 나타내며 +는 정상음성을 나타낸다. Jitter.나 ECS 파라미터 하나만으로 구분하는 것보다 두개가 조합될 경우 식별이 훨씬 용이함을 직관적으로 알 수 있다

표 2 는 본 논문에서 수행한 식별실험의 결과이다. 신경망의 훈련은 실험자료의 2/3 을 이용하여 실행하고 시험은 나머지 1/3 의 음성시료를 이용하여 실행하였다.

실험결과로부터 우리는 Jitter 와 Shimmer 의 조합이 가장 높은 식별율을 보이는 것을 알 수 있으나 시험용 자료에대하여는 Shimmer 와 ECS 의 조합이 가장 높은 식별율을 보임을 알 수 있다. 그러나 제 3 의 시료를 이용하여 식별실험을 행하였을 경우는 Jitter-Shimmer 의 조합의 식별율이 매우 낮아짐을 알 수 있다. 또 세개의 파라미터를 모두 사용하였을 경우는 식별율이 중간성도에 머물고 있다.

이와 같은 결과는 훈련용 시료와 시험용 시료의 전체 갯수가 적고 시료특성의 균일성이 확보되지 못한 것이 이유라고 판단된다.

5. 결론

본 실험에서는 웨이브렛 변환결과로부터 구한 파라미터와 기존의 파라미터인 Jitter, Shimmer 를 이용하여 정상음성과 양성종양을 구분하는 실험을 행하였다. 실험결과 식별율은 71%에서 93%까지 파라미터의 조합에 의해 다양한 결과가 나왔으며 식별결과로부터는 제안된 파라미터의 유효성을 확인할 수는 없었다. 훈련데이타에서는 식별율이 높게 나왔다. 차후 보다 보편성 있는 시료가 훈련용 시료로 확보되어야 함을 나타내 준다. 또 기존의삽음성분 식별 파라미터인 NHR 에 비해 유사한 기능을 하면서도 구하는 과정이 단순한 파라미터로서 ECS 를 이용할 수 있음은 확인되었다.

차후 보다 많은 실제 장애음성에 대한 시료확보와 피치검출에 민감하지 않은 파라미터에 대한 지속적인 연구가 계속되어야 할 것이다.

감사의글

본 논문은 1997 년도 한국학술진흥재단 학제간연구 '음향신호의 분석에 의한 후두질환의 진단에 관한 연구'의 연구결과의 일부입니다. 지원에 감사드립니다.

참고문헌

[1] F.Plante, H.Kessler,B.Cheetham, J.Earis, "Speech Monitoring of Infective Laryngitis", Proceedings of ICSLP'96, pp.749-752, Philadelpia, 1996.

[2] M.N.Vieira, F.R.McInnes, M.A.Jack, "Robust F0 and Jitter Estimation in Pathological Voices",I Procedings of ICSLP'96, pp.745-748, 1996

[3] "Disordered Voice Database", Version 1.03, Kay Elemetrics Corp., 1994.

[4] C.W.Jo, D.H.Kim, "Analysis of Disordered Speech Signal Using Wavelet Transform", accepted to be presented, ICSLP'98, Sidney, 1998

[5] 김대현,조철우, "장애음성의 분류방법에 관한 연구",제 15 회 음성통신 및 신호처리 워크샵논문집, pp.388-391, 1998

표 2. 신경망에 의한 식별 결과(J:Jitter, S:Shimmer, E:ECS)

		Train Data				Test Data			
		JS	JE	SE	JSE	JS	JE	SE	JSE
에러수	정상→비정상	4	17	5	6	8	12	1	5
	비정상→정상	2	4	8	4	2	1	2	3
식별률(%)		93.33	76.66	85.55	88.88	78.26	71.74	93.48	82.61