

장애 음성 신호처리

조철우

창원대학교 제어계측공학과

Signal Processing of Disordered Speech

Cheol-Woo Jo

Dept. of Control & Instrumentation Engineering, Changwon National University

cwjo@sarim.changwon.ac.kr

요 약

본 논문에서는 음성신호처리 기법을 이용하여 장애음을 진단, 개선하는 데 필요한 다양한 신호처리방법에 대하여 다루고자 한다. 음성장애중 성대장애를 중심으로 신호에 나타나는 현상과 이를 이용한 신호처리 방법들을 소개하며 응용사례로 음성을 이용한 성대질환의 진단에 관한 내용을 소개한다.

1. 서 론

음성신호는 인간의 의사소통에 있어서 가장 중요한 수단이다. 이러한 음성에 장애가 올 경우 많은 불편을 초래하게 된다.

후두악성종양은 최근 항암약물치료, 방사선치료, 수술적치료 등의 발전으로 인하여 5년 생존률이 70% 정도에 이르지만, 진행된 암의 경우 후두의 광범위 절제 및 전결제가 불가피하여 음성의 보존이 어렵워 이로 인한 환자의 삶의 질(Quality of Life)에 많은 장애를 초래하게 된다. 따라서 후두암의 조기진단은 생존률의 향상은 물론이고 후두기능을 보존할 수 있어 후두암의 치료에 있어 가장 중요한 요소이다.

이런 관점에서 집단검진(Mass Screening Test)은 비교적 발생빈도가 타 종양보다 낮은 후두암에서 질환을 조기에 발견할 수 있는 좋은 방법이라 할 수 있다. 또한 후두질환의 가장 흔한 증상이 애성 및 음성의 변화가 주이기 때문에 경험이 많은 이비인후과 의사의 경우는 환자의 음성만으로도 후두의

병변을 예측하기도하며, 이러한 지각적 병적 음성의 분류를 통한 진단법은 많은 보고자들이 연구하여 진단에 이용되기도 한다. 이러한 사실은 음성의 음향학적 분석에 기초를 둔 screening test의 개발 가능성을 제시하였으며, 이러한 음향학적 screening test는 비침습적이고, 녹음을 통해서 음향분석 기기를 이용하여 분석하므로 검사가 간편하여 피검자의 시간적, 공간적 제약을 최소화 할 수 있는 장점이 있다.

본 논문에서는 이러한 장애음을 이용한 후두질환의 진단법에 관련한 여러가지 신호처리 기법과 사례를 살펴본다.

2. 장애음성

음성에 장애가 오는 경우는 성대장애의 경우, 근육장애의 경우등 다양한 경우가 있지만 여기서는 성대장애의 경우에 한정해서 이야기한다. 다양한 종류의 장애에 따라 서로 다른 특성의 변화가 일어나게 되므로 다른 처리방법이 필요하게 된다. 음성장애의 종류에는 원목소리, 공명장애, 소리의 크기 및 높낮이 장애, 경련성 발성장애, 암과 음성장애 등이 있다.[1]

3. 장애음성의 특성

장애음성은 성대부위의 종양등 여러가지 원인에 의해 정상적인 목소리보다 주파수, 시간축에서의 변형이 초래되어 부자연스러운 음성이 발생되게 된다. 대표적인 특성으로 허스키한 목소리는 후두의 염증, 성대의 혹,

성대마비, 후두암 등이 원인이 되어 발생한다고 한다.[1] 많은 경우 성대부위의 염증은 성대의 피부의 운동을 둔화시켜 피치주기를 불규칙하게 하거나 진동의 세기를 불규칙하게 하고, 때로는 고주파성분의 잡음을 증가시키는 효과를 가져온다.

4. 장애음성의 특징추출 파라미터

장애음성의 특징을 정량적으로 측정하기 위한 시도로 여러가지 파라미터가 제안되고 있다. 앞서 언급한 장애음성의 불규칙한 특성 때문에 이러한 조건을 극복하면서 음성의 특징을 추출할 수 있는 여러가지 파라미터가 제안되고 있다.

사람의 발성은 성대에서 원음을 만들어 성도의 변화를 거쳐 원하는 음성으로 만든다. 따라서, 원음의 발생기관인 성대에 결절이나 혹, 또는 마비가 있을 때 성대의 떨림 수나 진동 폭에 불규칙성이 개입되기 때문에 이러한 동요를 적절히 측정함으로써 정상인과 비정상인의 성대 모양을 쉽게 구분할 수 있게 된다. 이러한 현상을 나타내 주는 여러 가지 파라미터들이 제안되어 있다. 이들 파라미터는 측정하고자 하는 성질에 따라서 다음과 같은 그룹으로 나뉘어진다. 장, 단기 주파수 동요추정값으로는 Jita, Jitt, RAP, PPQ, sPPQ, VF0 등이 있고 장, 단기 진폭동요추정값에 관련한 파라미터는 ShdB, Shm, APQ, sAPQ, vAm 등이, 유성단절관련 측정값은 DUB, NVB, DSH, NSH 등이, 유성불규칙성과 연관된 값은 DUV, NU 등이, 소음과 연관된 값은 NHR, VTI, SPI 등이, 떨림과 연관된 값은 FTRI, ATRI, Fftr. 등이 있다. 각 파라미터에 관한 상세한 설명은 참고문헌[2]에 나와있다.

대표적인 파라미터로는 Jitter와 Shimmer, NHR이 있다. 이들 파라미터는 음성의 기본 주기, 즉, 피치 주기를 바탕으로 단위 구간내에서의 평균피치의 변화 또는 평균진폭의 변화를 측정한다. 그러나 많은 수의 장애음성의 경우 기존의 피치추출 알고리즘으로는 정확한 피치의 추출이 어려운 경우가 많다. 따

라서 정확한 피치추정이 가능하도록 변형된 방법을 사용한다. 가장 간단한 방법의 하나는 피치추출 프레임의 구간을 늘이는 것이다. 이렇게 함으로써 자기상관함수등에서 나타나는 주기성이 더욱 강조되어 피치를 구하기 쉽게 만들어 준다.

Jitter는 다음과 같이 구해진다.

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

이 식에서 보면 Jitter는 평균 피치 값에 대한 피치의 평균 변동률이다.

Shimmer는 다음과 같은 식으로 주어진다.

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

이 식에서 보면 Shimmer는 평균 진폭값에 대한 평균 진폭변동률이다.

NHR(Noise to Harmonic Ratio)은 음성 파형에서 70~4500Hz범위의 하모닉 성분의 에너지에 대한 1500~4500Hz범위의 하모닉 성분이 아닌 성분(잡음성분)의 에너지 사이의 비율로 나타내며, 일반적으로 분석신호에서의 잡음성분의 존재를 평가하는데 사용된다. NHR이 증가하면 잡음성분이 증가함을 나타낸다.

5. 장애음성의 진단 사례

장애음성을 이용한 계측과 진단에의 활용 사례로 후두질환을 신호처리 방법에 의해 진단하는 경우를 소개한다.

먼저 병원에서 환자와 정상인을 대상으로 후두질환환자의 음성을 수집하였다.

본 연구에서 사용된 화자 데이터베이스의 수집은 부산대학교 병원의 방음실에서 편인한 자세로 앉게 한 후 마이크 앞에서 15cm 가량 거리를 두도록 한 후 가능하면 평상시와 같은 음높이와 크기로 약 3초간/우/, /오/

표 1. 양성, 악성, 정상음성에서 각 파라미터의 평균과 표준편차

	Malignant		Benign		Normal	
	Average	Standard deviation	Average	Standard deviation	Average	Standard deviation
Age	56.31	5.14	42.83	8.89	27.12	4.55
F0	169.53	42.14	168.21	43.84	157.46	50.90
PFR	12.77	9.00	5.22	2.70	2.80	1.12
Jita	308086.03	1109282.16	174.62	164.15	59.91	34.20
Jitt(%)	6.47	5.27	2.93	2.85	0.97	0.68
RAP(%)	3.76	3.11	1.63	1.62	0.56	0.43
PPQ(%)	4.08	3.37	1.72	1.74	0.52	0.36
vFo(%)	16.04	15.34	4.0	2.92	1.90	1.10
ShdB(dB)	1.25	0.72	0.96	0.66	0.31	0.09
Shi(%)	14.19	7.95	10.55	6.94	3.61	0.98
APQ(%)	11.04	6.68	7.89	6.08	108.74	531.31
vA(%)	20.45	5.76	16.42	6.35	11.54	5.32
NHR	0.47	0.36	0.16	0.05	0.27	0.70
VTI	0.17	0.11	0.06	0.03	0.79	3.66
SPI	10.66	7.31	15.04	11.48	8.36	4.61

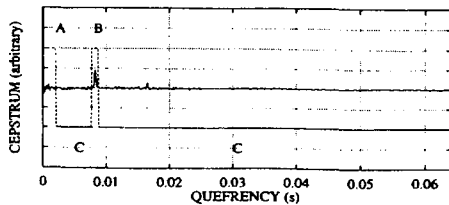


그림 1 Cepstrum에서의 피치, 잡음성분의 분리 /아/, /에/, /이/를 지속적으로 발생하도록 하였으며 이를 3회 이상 실시하여 Sony사의 Digital Audiotape Record(DAT, DTC-59ESJ, Sony, Japan)에 녹음한 뒤 이를 질병 유형별로 분류하고 데이터 베이스화 한 후 녹음된 음성을 Kay사(Kay Elemetrics Corp. USA)의 Computerized Speech Lab 4300B(CSL)기종의 음성분석 소프트웨어인 Multi-Dimensional Voice Program(MDVP)를 이용하여 총 33개의 변수에 대한 분석을 하여 각 음성의 변수의 특성을 추출하였다. 이 결과를 이용하여 여러 개의 파라미터 중 변별력이 있는 파라미터를 찾기 위하여 통계학적인 분석을 행하고 그 중 가장 변별력이 있다고 판단되는 파라미터를 이용하여 식별에 사용한다. 표1은 이러한 통계분석의 결과이다.

분석결과를 살펴볼 때 질병분류에 사용할만

한 파라미터로서는 진폭변화를 나타내는 Shimmer와 기본주파수 변화를 나타내는 Jitter가 이용될 수 있으며, 소음성을 진단하는 NHR도 기여도가 높다.

이들 파라미터외에 별도로 켈프스트럼을 이용한 파라미터를 제안하여 사용하였다. 이 파라미터는 선형예측분석에서의 오차신호로부터 켈프스트럼을 구한 신호를 이용하여 잡음성분과 피치성분의 비를 구한 것으로 그림1과 그림2에서 보는 바와 같이 정상음성과 장애 음성간을 잘 식별해 준다.

이와 같은 파라미터들을 이용하여 정상, 양성종양, 악성종양 상태를 구분한다. 식별기로는 다층 신경회로망이 사용되었다. 먼저 입력 파라미터로 Jitter, Shimmer 및 HNRR를 구하여 각 파라미터들을 2,3개씩 조합하여 가장 식별률이 좋은 파라미터의 조합을 찾았다. 신경회로망은 한 개의 입력층과 한 개의 출력층 그리고 한 개의 은닉층을 갖는다.

그리고 이러한 신경회로망이 연속적으로 2단계로 적용되었다. 첫 번째 단계에서는 정상과 종양 간을 구분하였다. 두 번째 단계에서는 양성종양과 악성종양간을 구분하였다. 이러한 실험결과 첫 번째 단계에서는 Jitter-Shimmer-HNRR의 세 파라미터의 조합이 가장 좋은 식별율을 보였고 약 90%의 정

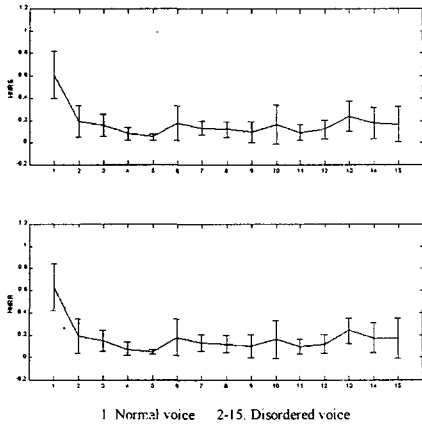


그림 2 HNRR, HNRS의 통계분석결과

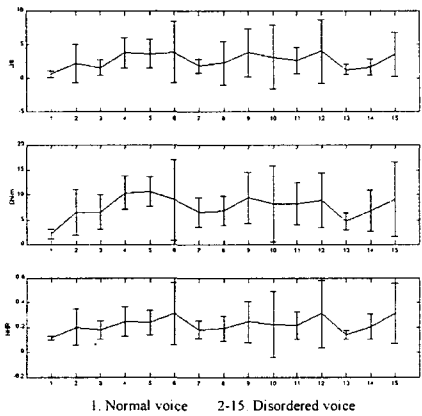


그림 3 Jitter, Shimmer, NHR의 통계분석결과

확도를 보였다. 정상음성과 장애음성간의 식별에서 훈련데이터의 경우는 세 가지 파라미터를 모두 사용한 경우가 88.64%로 가장 식별율이 높은 것으로 판정되었고 이렇게 훈련된 신경회로망을 이용한 나머지 데이터를 이용한 신경회로망이 가장 높은 식별율을 나타내었다. 그러나 세 가지 파라미터를 모두 사용한 경우도 90.9%로 높았기 때문에 양쪽의 경우 모두 높은 식별율을 보인 세가지 조합이 유효할 것으로 판단된다. 표 12의 경우 양성파와 악성종양의 구분실험에서는 전반적으로 HNRR과 Jitter의 조합이 높은 식별율을 보였는데, 현재 악성과 양성종양이 환자데이터의 수가 부족하기 때문에 신뢰성은 많이 떨어지지만 충분한 양의 데이터가 확보된다면 더욱 신뢰성있는 식별이 가능함을 보여 주었다.

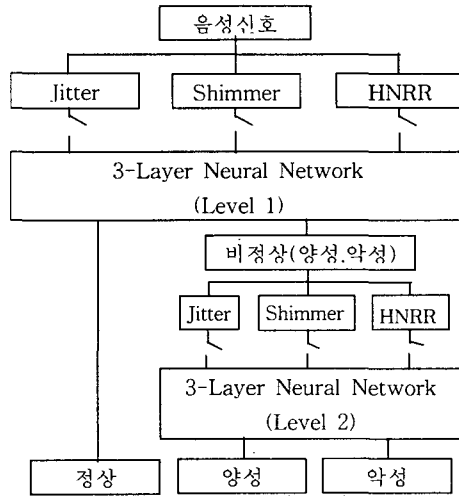


그림4. 신경회로망을 이용한 정상, 양성, 악성종양의 식별도

두 단계의 신경회로망을 이용한 실험으로부터 우리는 정상, 양성, 악성종양의 구분이 신경회로망과 파라미터의 적절한 조합에 의해 상당한 정확도를 가지고 수행될 수 있음을 확인하였다. 현 상태에서는 종양데이터의 수가 절대적으로 부족한 상태이지만 앞으로 자료가 충분히 확보된다면 본 실험에서 사용한 방법이 환자진단의 선행단계로 사용이 가능함을 보여주었다.

6. 결 론

본 논문에서는 장애음성신호처리의 파라미터들과 이를 이용하여 후두질환을 예측, 진단하는 방법에 관하여 소개하였다. 연구결과에 의하면 상당한 정확도로 질환의 예측이 가능함을 보였다. 이러한 방법은 향후 후두질환의 조기진단에 유용하게 사용될 수 있을 것으로 생각된다. 그러나 자료수집에서의 어려움 때문에 많은 자료의 수집과 추가적인 분석이 요구되고 있다.

참고문헌

[1] 최홍식, "음성언어 의학과 음성외과 수술", 음성과학회 논문집, pp.29-34, 1997
 [2] Operations Manual, "Multidimensional Voice Program(Model4305)", Kay Elemetrics Corp., 1993
 [3] Cheol-Woo Jo, Dae-Hyun Kim, "Classification of Pathological Speech into Normal/Benign/Malignant State", Eurospeech'99, Budapest, Hungary, 1999