

음향신호의 분석에 의한 후두질환의 진단에 관한 연구*

A Study on the Diagnosis of Laryngeal Diseases
by Acoustic Signal Analysis

조철우** · 양병곤*** · 왕수건****

(Cheol-Woo Jo · Byong-gon Yang · Soo-Geon Wang)

ABSTRACT

This paper describes a series of researches to diagnose vocal diseases using the statistical method and the acoustic signal analysis method. Speech materials are collected at the hospital. Using the pathological database, the basic parameters for the diagnosis are obtained. Based on the statistical characteristics of the parameters, valid parameters are chosen and those are used to diagnose the pathological speech signal. Cepstrum is used to extract parameters which represents characteristics of pathological speech. 3 layered neural network is used to train and classify pathological speech into normal, benign and malignant case.

Keywords: diagnosis, disease, acoustic signal

1. 서론

후두악성종양은 최근 항암약물치료, 방사선치료, 수술적치료 등의 발전으로 인하여 5년 생존률이 70 % 정도에 이르지만, 진행된 암의 경우 후두의 광범위 절제 및 전절제가 불가피하여 음성의 보존이 어려워 이로 인한 환자의 삶의 질(quality of life)에 많은 장애를 초래하게 된다. 따라서 후두암의 조기진단은 생존률의 향상은 물론이고, 후두 기능을 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(Mashima 등). 또한 후두질환의 가장 흔한 증상이 애성 및 음성의 변화이기 때문에 경험이 많은 이비인후과

* 이 논문은 1997년 한국학술진흥재단의 공모과제 연구비(학제간연구)에 의하여 연구되었음.

** 창원대학교 공과대학 제어계측공학과

*** 동의대학교 인문대학 영어영문학과

**** 부산대학교 의과대학 이비인후과

의사의 경우는 환자의 음성만으로도 후두의 병변을 예측하기도하며, 이러한 지각적 병적 음성의 분류를 통한 진단법은 많은 연구자들이 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 최소화할 수 있는 장점이 있다.

이런 검사법의 개발에 가장 중요한 요소로서 후두의 상태를 잘 반영할 수 있는 음향학적인 변수를 구하는 것으로 1961년 Lieberman 등이 후두질환의 진단에 있어 pitch perturbation factor가 잘 반영한다고 보고한 이래 많은 보고자들이 후두의 병변과 연관된 변수를 연구하고, 변수와 연관지어 후두 병변시 음성변화를 초래하는 기전에 관한 연구를 보고하였다.

후두병변과 관련된 이런 변수는 pitch period의 평균과 강도, pitch period, peak amplitude perturbation, vocal noise, cycle to cycle waveform variation, average frequency spectrum characteristic, transitional wave form characteristics 등의 음향학적인 특성을 주로 이용하였다. 이런 음향학적 변수에 대한 연구가 진행되면서 이를 기초로 하여 후두질환의 감별진단에 이용하고자 연구가 진행되었는데 1987년 Koike 등이 APQ(Amplitude Perturbation Quotient), PPQ(Pitch Period Perturbation)를 이용한 후두질환의 감별을 연구하였고 1986년 Laver, 1987년 Mashima 등이 PPQ, APQ, Vocal noise를 이용한 screening test를 개발하기 위한 예비 연구를 발표하였다. 그러나 과거의 연구는 음향분석기기의 미발달로 인한 음향분석이 어려웠으며, 사용된 변수가 2-3개로 극히 제한적이었으며, 정성적 비교에 의존하였고 임상에 실용화하기 위한 소프트웨어나 진단기기의 개발에 관한 연구가 진행되지 않아 실제 임상에 실용화되지 못하였다. 따라서 이런 난점을 해결하기 위하여서는 병적 음성에 대한 음성의학적, 음성학적 연구와 더불어 음성공학적인 연구가 동반되어야만 가능할 것으로 생각되어 음성공학, 음성학, 음성의학분야의 학제간 연구를 수행하게 되었다.

본 연구의 목적은 음성신호의 음향적 특성을 여러 가지 파라미터에 의해 통계적으로 분석하고, 이들 중 정상, 양성, 악성종양을 식별할 수 있는 파라미터를 추출하여 자동식별의 가능성을 알아보하고자 하는 데 있다.

2. 기존파라미터에 대한 고찰

음향신호에 의한 식별에 앞서 기존의 파라미터들에 대해 요약하면 다음과 같다. 본 연구에서는 현재 임상용으로 가장 널리 사용되고 있는 Kay사(Kay Elemetrics Corp. USA)의 Computerized Speech Lab 4300B(CSL) 기종의 음성분석 소프트웨어인 Multi-Dimensional Voice Program(MDVP)에서 분석 가능한 파라미터들을 중심으로 살펴본다. 사람의 발성은 성대에서 원음을 만들어 성도의 변화를 거쳐 원하는 음성으로 만든다. 따라서, 원음의 발생기관인 성대에 결절이나 혹, 또는 마비가 있을 때 성대의 떨림수나 진동폭에 불규칙성이 개입되기 때문에 이러한 동요를 적절히 측정함으로써 정상인과 비정상인

의 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 주파수동요 측정값으로는 Jita(Absolute Jitter), Jitt(Jitter Percent), RAP(Relative Average Perturbation), PPQ, sPPQ(Smoothed Pitch Period Perturbation), VF0(Coefficient of Fundamental Frequency Variation) 등이 있고 장, 단기 진폭동요값 측정에 관련한 파라미터는 ShdB(Shimmer in dB), Shim(Shimmer Percent), APQ, sAPQ(Smoothed Amplitude Perturbation Quotient), vAm(Coefficient of Amplitude Variation) 등이, 유성단절관련 측정값은 DVB(Degree of Voice Breaks), NVB(Number of Voice Breaks), DSH(Degree of Sub-Harmonic components), NSH(Number of Sub-Harmonic segments) 등이, 유성불규칙성과 연관된 값은 DUJ(Degree of Voiceless), NUV(Number of Unvoiced Segments) 등이, 소음과 연관된 값은 NHR(Noise to Harmonic Ratio), VTI(Voice Turbulence Index), SPI(Soft Phonation Index) 등이, 떨림과 연관된 값은 FTRI(Frequency Tremor Intensity Index), ATRI(Amplitude Tremor Intensity Index), FFTR(F0-Tremor Frequency) 등이 있다. 각 파라미터에 관한 상세한 설명은 참고문헌(KAY, 1993, 1994)에 나와 있다.

여러 파라미터들 중에서도 성대의 질병진단 등에 많이 사용되어지고 있는 파라미터로는 Jitter, Shimmer, NHR 등이 있다. 이 세 가지 파라미터를 기본으로 여러 파라미터가 구해진다. 이 세 가지 파라미터에 대한 간단한 내용을 구체적으로 알아보면, Jitter는 성대 진동의 각 주기마다의 시간의 불규칙성 즉, 피치주기의 변화율을 나타내는데 사용되어지는 파라미터로 연속적인 피치주기 사이의 평균 퍼센트 변화로 나타낼 수 있다. 정상적인 음성일 경우 1%이하로 나타난다고 한다. 식 (2-1)과 같은 식으로 나타낼 수 있다.

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

식(2-1)에서 $P(i)$ 는 i 번째 피치의 주기이고, N 은 측정된 피치의 개수이다.

Shimmer는 Jitter와는 달리 성대진동의 주기마다의 강도(Amplitude)의 불규칙성 즉, peak사이의 크기에 대한 변화율을 나타낸다. Shimmer는 %나 dB로 나타낼 수 있다. 식(2-2)과 같이 정의한다.

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

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

NHR(Noise to Harmonic Ratio)은 음성 파형에서 70~4500 Hz범위의 하모닉 성분의 에너지에 대한 1,500~4,500 Hz범위의 하모닉 성분이 아닌 성분(잡음성분)의 에너지 사이의 비율로 나타내며, 일반적으로 분석신호에서의 잡음성분의 존재를 평가하는데 사용된다.

NHR이 증가하면 잡음성분이 증가함을 나타낸다.

3. 자료수집과정 및 사전실험

본 연구에서 사용된 화자 데이터베이스의 수집은 부산대학교 병원의 방음실에서 편안한 자세로 앉게 한 후 마이크 앞에서 15 cm 가량 거리를 두도록 한 후 가능하면 평상시와 같은 음높이와 크기로 약 3초간 /우/, /오/, /아/, /에/, /이/를 지속적으로 발성하도록 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.

한국인 화자에 대해 다섯 개의 모음을 발성하게 한 이유는 서로 다른 모음의 경우 추출되는 파라미터에 대한 연관성을 조사하여 어떤 음성을 차후 측정에 사용하여야 하는지를 결정하고자 함이다. 이 중에서 조음위치에 따라 /아, 우, 이/를 선택하여 파라미터의 변화를 관찰하였다. 장애음성과는 별도로 정상인으로부터 수집된 세 개의 모음을 사용하여 사전실험을 행하였다. 사전실험의 목적은 장애음성의 진단에 필요한 모음의 종류를 결정하기 위한 것이다. 만약 여러 개의 모음을 통한 파라미터의 특성이 차이가 거의 없다면 하나의 모음을 대표음으로 하여 실험하여도 되므로 실험과정을 간소화할 수 있다.

피험자는 청력에 이상이 없고 발성에 장애가 없는 음성학자가 방음 장치된 방에서 마이크와 적절한 거리를 유지하고 발음했다. 동일한 화자가 모음 /아, 우, 이/를 발음했을 때 측정된 분석결과는 상호 어떤 관계를 보이는가를 조사했다.

표 1은 동일화자에 대한 모음별 파라미터 값들의 상관계수를 나타낸 것이다. 각 파라미터에 대한 동일화자가 서로 다른 모음을 발성했을 때의 파라미터의 값들은 표 2에 자세하게 나타내었다.

표 1. 동일화자가 서로 다른 모음을 발성했을 때 파라미터 값의 상관계수

	아	우	이
아	1		
우	0.984	1	
이	0.988	0.998	1

표 2. 동일화자가 서로 다른 모음을 발성했을 때 파라미터 값

파라미터	아	우	이	파라미터	아	우	이
Fo	150.969	153.236	156.125	vFo	0.00848	0.01746	0.02156
To	6.624	6.528	6.408	ShdB	0.177	0.129	0.27
Fhi	152.858	161.238	166.556	Shim	0.0207	0.01494	0.02572
Flo	147.885	131.475	132.837	APQ	0.01884	0.00919	0.01718
STD	1.281	2.676	3.367	sAPQ	0.04206	0.02027	0.0373
PFR	1	4	5	vAm	0.12642	0.06303	0.14819
Fftr	3.81	3.008	5	NHR	0.1249	0.1159	0.148
Tsam	0.88	1.1	1.019	VTI	0.0285	0.0144	0.0751
Jita	12.892	44.361	43.615	SPI	6.5044	17.9021	4.6291
Jitt	0.00195	0.0068	0.00681	FTRI	0.0039	0.00335	0.00428
RAP	0.00095	0.00388	0.00379	SEG	27	34	32
PPQ	0.00117	0.00363	0.00375	PER	131	167	157
sPPQ	0.00459	0.00743	0.0062				

두 가지 모두 아주 높은 상관계수를 보임으로써 모음이 달라지더라도 측정값의 차이는 거의 없다고 보아야 할 것이다. 실제 파라미터 가운데는 SPI와 같이 모음의 포먼트 구조에 민감한 변수가 있어서 모음 /우/에서 17.9, 모음 /아/에서 6.5, 모음 /이/에서 4.6과 같은 많은 차이를 보이고 있지만 나머지 파라미터들은 대체로 성대의 떨림의 규칙성 여부를 점검하는 값이기 때문에 별로 큰 차이를 기대할 수 없을 것이다.

이번에는 동일한 화자가 모음의 음계를 높이거나 내렸을 때 파라미터 값의 변화는 어떠한지를 측정했다. 모음 /아/를 아주 낮은음 (Fo=116 Hz)에서 중간음 (Fo=131 Hz), 고음 (Fo=151 Hz)으로 높였을 때의 각각의 파라미터 측정값의 상관계수와 상대적인 비율을 주로 반영하는 파라미터간의 상관계수를 표 3에서 나타내준다.

각 파라미터에 대한 동일화자가 모음 /아/를 다른 높이로 발성했을 때의 파라미터의 값들은 표 4에 자세하게 나타내었다.

표 3. 동일화자가 모음 /아/를 다른 높이로 발성했을 때 파라미터 값의 상관계수

	고음	중간음	저음
고음	1		
중간음	0.997	1	
저음	0.998	0.997	1

여기에도 매우 높은 상관계수를 보여서 발성노력의 차이도 별로 파라미터값에 영향을 미치지 않음을 알 수 있다. 성대노력의 반영은 성대의 떨림 속도의 변화를 나타내어 기본 주파수값의 차이는 있을지라도 파라미터들이 상대적인 비율값을 나타내기 때문에 이것도 큰 차이가 없는 것이 당연하다. 한가지 더 지적할 것은 절대값인 Jita에서는 저음일수록 높았으나 비율을 나타내는 Jitt에서는 비슷한 값을 보이고 있으며 나머지 파라미터에서도 거의 동일한 값을 나타내고 있다.

표 4. 동일화자가 모음 /아/를 다른 높이로 발성했을 때 파라미터 값

파라미터	고음	중간음	저음	파라미터	고음	중간음	저음
Fo	150.92	130.59	116.55	vFo	0.00525	0.00581	0.00631
To	6.63	7.66	8.58	ShdB	0.161	0.158	0.181
Fhi	152.44	132	118.85	Shim	0.01867	0.01829	0.0205
Flo	148.19	128.97	114.81	APQ	0.01712	0.01486	0.01727
STD	0.79	0.76	0.74	sAPQ	0.03063	0.03543	0.0438
PFR	1	2	2	vAm	0.09573	0.05518	0.08983
Tsam	0.86	0.72	0.86	NHR	0.1234	0.1334	0.1423
Jita	12.46	18.61	22.1	VTI	0.0436	0.0675	0.0876
Jitt	0.00188	0.00243	0.00258	SPI	4.5736	4.6842	3.9617
RAP	0.00091	0.0013	0.00127	SEG	27	22	27
PPQ	0.00122	0.00141	0.00159	PER	128	93	98
sPPQ	0.00282	0.00474	0.00531	PER	131	167	157

다음으로는 한 모음을 네 번 발성했을 때 파라미터간의 관계는 어떤지를 살펴보았다. 피치값을 적절히 조정하여 거의 동일한 값이 되도록 하였다. 이러한 방법을 사용하는 것은 음성신호처리를 할 때 기계상의 일관성이 있는지를 규명해 보기 위해서 필요하다. 모음 /아/를 네 번 발성하여 MDVP로 분석한 결과값들을 상호 비교한 결과는 표 5와 같다.

각 파라미터에 대한 동일화자가 모음 /아/를 네 번 발성했을 때의 파라미터의 값들은 표 6에 자세하게 나타내었다.

표 5. 동일화자가 모음 /아/를 네번 발성했을 때 파라미터 값의 상관계수

	첫번째	두번째	세번째	네번째
첫번째	1			
두번째	0.991	1		
세번째	0.989	0.997	1	
네번째	0.999	0.984	0.982	1

여기서도 매우 높은 상관도를 보였다. 따라서, 지금까지의 실험에서는 피험자가 동일한 발음을 했을 때 정확히 동일한 음성분석 결과를 보여주고 있음을 알 수 있다. 특히, 상대적 변수인 비음값을 주로하는 MDVP의 파라미터의 분석자료는 차이가 거의 없음을 알 수 있다. 결국, 모든 모음을 다 분석하더라도 동일한 결과가 나온다면, 실험에서 모음 /아/ 하나만을 제한하여 분석하더라도 괜찮다는 결론을 내릴 수 있다.

따라서, 본 연구에서는 KAY에서 제작한 CD-ROM 자료에서 모음 /아/를 중심으로 이들 파라미터와 질병간의 관계를 추적해 보고 또한 부산대학교 병원에서 수집한 자료에서 MDVP의 파라미터와 각종 성대주위의 질병 사이의 관계를 조사해보고자 한다. 특히, 환자의 특징을 나타내는 파라미터가 무엇이고 어떤 임계값이 있는지를 조사함으로써 앞으로 구성할 음성질환 감별 소프트웨어 기초자료를 수집하고자 한다.

표 6. 동일화자가 모음 /아/를 네번 발성했을 때 파라미터 값

파라미터	첫번째	두번째	세번째	네번째	파라미터	첫번째	두번째	세번째	네번째
Fo	130.053	130.59	129.926	129.642	vFo	0.0062	0.0058	0.0077	0.0049
To	7.689	7.658	7.697	7.714	ShdB	0.222	0.158	0.202	0.145
Fhi	132.17	131.996	132.767	131.062	Shim	0.0254	0.0183	0.0234	0.0165
Flo	128.502	128.966	127.389	127.91	APQ	0.0221	0.0149	0.0179	0.0133
STD	0.799	0.759	0.999	0.633	sAPQ	0.0361	0.0354	0.028	0.0333
PFR	2	2	2	2	vAm	0.0545	0.0552	0.0698	0.0656
Fftr	2.985				NHR	0.1256	0.1334	0.1438	0.1205
Fatr	2.581	5.556		8.511	VTI	0.049	0.0675	0.0477	0.0566
Tsam	0.952	0.722	0.739	1.034	SPI	6.8699	4.6842	4.8348	4.9422
Jita	17.868	18.609	33.17	15.16	FTRI	0.0025			
Jitt	0.0023	0.0024	0.0043	0.002	ATRI	0.0211	0.0224		0.0029
RAP	0.0012	0.0013	0.0026	0.001	SEG	30	22	23	32
PPQ	0.0013	0.0014	0.0023	0.0012	PER	122	93	95	132
sPPQ	0.004	0.0047	0.0054	0.0031					

4. 통계분석 및 신호처리법에 의한 판별실험

4.1 통계분석결과의 검토

통계분석에 사용된 장애음성시료는 한국화자 데이터베이스에서 악성후두종양군은 1996년 10월부터 98년 5월까지 후두악성종양이 의심되어 병원을 방문한 환자중 후두조직검사상 편평상피세포암으로 확진된 환자 중 음성검사가 가능하였던 25명을 대상으로 하였으며 이들의 원발부위에 따른 분포는 성문암이 21례, 성문상부암이 3례, 경성문암이 1례였으며, 병기별로는 T1b 10례, T2 4례, T3 5례, T4 3례였다(표 7). 양성후두질환군은 97년 10월부터 98년 5월까지 애성을 주소로 병원을 내원한 환자 중 검사상 양성후두질환으로 확인된 33례를 대상으로 하였으며, 질환별 분포는 성대결절 3례, 용종 27례, 성대구 1례, 낭종1례, 후두염 1례였으며 성대마비환자는 제외하였다(표 8). 정상대조군으로 조음이 정상이며 후두에 병변이 없고 정상 발성이 가능한 성인 남녀 각각 남자 29례, 여자 6례로 35명을 대상으로 하였다(표 9).

표 7. 악성종양군의 병기별, 연령별 분포

age \ type of ca.	glottic				supraglottic	transglottic	total
	T1b	T2	T3	T4			
40-49	1	0	1	0	0	0	2
50-59	7	2	2	2	1	0	14
60-69	1	1	1	1	2	1	7
70-79	1		1		0	0	2
total	10	4	5	3	3	1	25

표 8. 양성후두질환군의 성별, 연령별 분포

	10-19		20-29		30-39		40-49		50-59대		total
	M	F	M	F	M	F	M	F	M	F	
Nodule	1		1					1			3
Polyp				3	6	5	10	2	1		27
Sulcus	1										1
Laryngitis					1						1
Cyst							1				1
Total	2		1	3	7	5	11	3	1		33

표 9. 정상대조군의 성별, 연령별 분포

	10-19	20-29	30-39	40-49	50-59	total
male	14	5	1	5	4	29
female	4	2				6
total	18	7	1	5	4	35

부산대학교 병원에서 분석한 자료의 평균과 표준편차를 모음 /아/ 발성에 한정하여 처리해 보았다. benign는 양성종양, malignant는 악성종양, normal은 정상음성에 해당하는 값을 나타낸다. 분석자료에서 일부 파라미터가 3표준편차 이상의 명백히 그룹을 벗어나는 값들은 제거하고 평균을 구했음을 밝혀둔다. 표 10에 나타내었다.

지금까지의 분석결과를 살펴볼 때 질병분류에 사용할만한 파라미터로서는 진폭변화를 나타내는 Shimmer와 기본주파수 변화를 나타내는 Jitter가 이용될 수 있으며, 소음성을 진단하는 NHR도 기여도가 높다.

그러나 양성과 악성에 따른 분류는 매우 비슷한 음향적 결과를 초래하는 질병들이 섞여있는 경우를 많이 발견하였는데, 이러한 문제를 극복하기 위해서는 기존의 질병 분류에 의해 묶어서 처리하기보다는 훈련된 의사의 청각적 판단을 기준으로 질병의 심한 정도를 재분류하여야 할 것이다. 질병의 정도와 음향적 파라미터의 관계를 재정립함으로써 어떤 질병에는 어떤 음향적 변수가 가장 영향력을 주고 있는지를 밝히고, 정상인과 비정상인

경계값을 구해 이를 음성질환 판단에 활용하는 것이 가장 바람직할 것으로 여겨진다.

표 10. 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

4.2 신호처리법에 의한 분석에 의한 판별실험

신호처리법에 의한 분석에서는 기존의 파라미터인 Jitter, Shimmer와 캡스트럼을 이용한 파라미터인 HNRR를 사용하였다. Jitter, Shimmer에 대하여는 2장에서 설명한 바 있으며 여기서는 HNRR에 관하여 설명한다. HNRR은 기존의 파라미터와의 통계적 특성비교에서 기존에 제안된 파라미터에 비해 변별력이 우수하였으므로 식별파라미터로 사용하였다.

장애음성은 대체적으로 성대질환에 의해 음의 발생에 영향을 미치기 때문에 음원(Voice source)이 변하게 된다. 그래서 본 논문에서 이러한 음원의 변화를 관찰하기 위해 캡스트럼을 이용하여 음원을 하모닉(Harmonic) 성분과 잡음(Noise) 성분의 분리가 가능한 파라미터를 유도하였다.

먼저 음성생성의 모델로서 음성신호를 표현하면

$$s(t) = e(t) * v(t) = (p(t) + r(t)) * v(t) \quad (4-1)$$

식(4-1)에서 $s(t)$ 는 음성신호를 나타내고, $v(t)$ 는 성도(vocal tract)의 임펄스 응답을, $e(t)$ 는 여기신호(excitation signal)를, $p(t)$ 는 여기신호에서 주기적인(periodic, 엄밀히 말하자면 완전한 주기적인 신호가 아니기 때문에 준주기적(quasiperiodic이라고도 함) 부분을, $r(t)$ 는 여기신호에서의 비주기적인(aperiodic) 부분을 나타낸다.

주파수 도메인(domain)으로 나타내면

$$\begin{aligned}
 S(w) &= |S(w)|e^{j\theta(w)} \\
 &= [P(w) + R(w)]V(w) \\
 &= (|P(w)|e^{j\theta_p(w)} + |R(w)|e^{j\theta_r(w)})|V(w)|e^{j\theta_v(w)}
 \end{aligned}
 \tag{4-2}$$

식(4-2)에서 $S(w), P(w), R(w), V(w)$ 는 $s(t), p(t), r(t), v(t)$ 의 푸리에 변환(Fourier transform)한 것이다.

위 식들에서 음성신호에서 성도필터 성분과 여기신호를 분리하기 위해 캡스트럼방법을 사용하였다. 캡스트럼은 원신호를 푸리에 변환을 한 뒤, 변환된 신호의 절대치를 대수계산을 하여 다시 역푸리에변환을 한다. 식으로 나타내면 다음과 같다.

$$\log|S(w)| = \log|P(w) + R(w)| + \log|V(w)| \tag{4-3}$$

$$c(t) = IDFT[\log|S(w)|]. \tag{4-4}$$

계산과정 나타내면 그림 1과 같다.

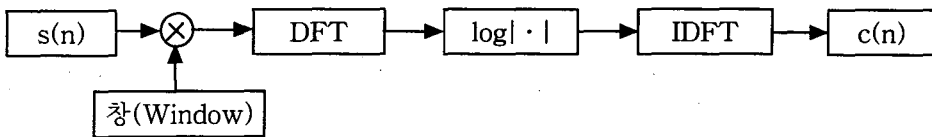


그림 1 캡스트럼 분석

입의 음성신호로부터 캡스트럼 분석을 통해 얻어진 신호 $c(t)$ 는 그림 2과 같다.

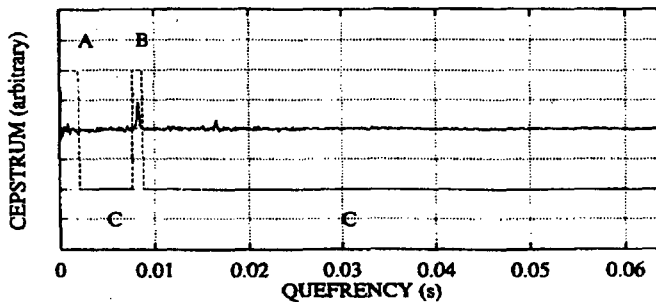


그림 2. 캡스트럼 신호($c(t)$) 및 하모닉성분의 분리

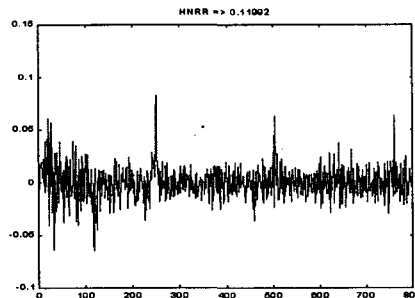
그림 2에서 A부분은 성도의 특성을 나타내는 부분이고, B와 C의 두 부분이 여기신호를 나타낸다. B는 여기신호의 주기적인 성분 즉, 하모닉(Harmonic)성분을 나타내고, C는 여기신호에 있어 비주기적인 성분인 잡음(Noise)성분을 나타낸다. 그래서 A를 먼저 Liftering(캡스트럼 신호에서 필터링)을 통하여 분리한 후 그 다음 B와 C성분을 같은 방법으로 분리하였다.

위와 같은 방법으로 얻어진 파라미터는 음성신호를 선형예측분석으로부터 구해진 예측오차 신호로부터 Cepstrum을 구한 뒤 성도필터 성분과 여기신호를 분리하고, 분리된 여기신호를 위에서 설명된 과정을 통하여 주기성분, 잡음성분을 분리해 낸다. 이 결과로부터 하모닉성분과 잡음성분의 비를 구한다. 예측오차신호가 성도특성을 제거하였다고는 하지만 아직도 성도성분이 완전히 제거되지 않고 남아있다고 볼 수 있기 때문에 확실히 성도성분의 영향을 좀 더 확실히 제거한 후 음원성분을 분리하기 위하여 예측오차를 cepstrum 변환하고 음원성분을 분리하였다. 이렇게 하여 파라미터를 구하는 제안식을 HNRR(Harmonic-to-Noise Ratio : Residual)이라고 한다. 식(4-5)과 같이 정의된다.

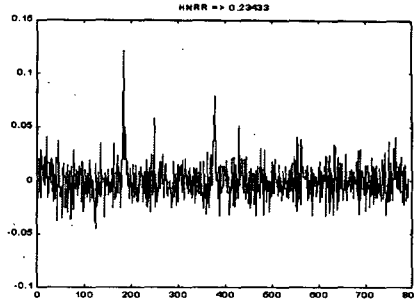
$$\text{HNRR} = \frac{\text{Harmonic Component Energy of Residual Signal}}{\text{Noise Component Energy of Residual Signal}} \quad (4-5)$$

이 파라미터로부터 대략 다음과 같은 성질을 유추할 수 있다. HNRR은 음성의 음원에서의 하모닉 성분과 잡음 성분의 비를 추정한 것으로 볼 수 있고 HNRS는 원신호에서의 하모닉 성분과 잡음 성분의 비를 추정한 것으로 볼 수 있다. 이러한 점에서 이들 파라미터는 음원의 변화를 대략적으로 나타내주는 파라미터로 볼 수 있다. 기존에 제안된 유사한 파라미터로는 NHR이 있는데 이는 원음성에서 70-4,500 Hz사이의 하모닉 성분과 1,500-4,500 Hz사이의 잡음성분의 비를 FFT결과로부터 구한 것으로 성도의 성분을 제거하지 않고 구하였기 때문에 여기서 제안한 파라미터들과는 약간 차이가 있다.

제안한 파라미터들의 특성은 다음과 같다. 그림 3은 각 경우의 캡스트럼분석 파형을 보여준다. 그림에서 정상음성의 경우는 주기성분이 장애음성에 비해 강조됨을 볼 수 있다. 정상음성의 경우 HNRS나 HNRR 모두 장애음성의 경우에 비해 작은 값을 가진다. 이는 환자음성의 경우 부가되는 잡음성분에 기인하는 것으로 판단된다.



(a) 정상음성의 경우 Cepstrum신호와 파라미터값(HNRR : 0.11992)



(b) 장애음성의 경우Cepstrum 신호와 파라미터값(HNRR : 0.23433)

그림 3. 정상음성과 장애음성에서의 Cepstrum신호와 HNRR

음향적 분석법에 의한 성대질환의 진단을 목적으로 하는 파라미터로 사용되는 Jitter, Shimmer와 켈스트럼방식의 음원분석에 의한 파라미터인 HNRR의 조합으로 신경회로망을 이용하여 정상, 양성종양, 악성종양을 구분하는 실험을 행하고 그 결과를 고찰하였다.

그림 4는 정상과 양성, 악성을 분류하기 위한 실험방법에 대해 간략하게 나타낸 것이다.

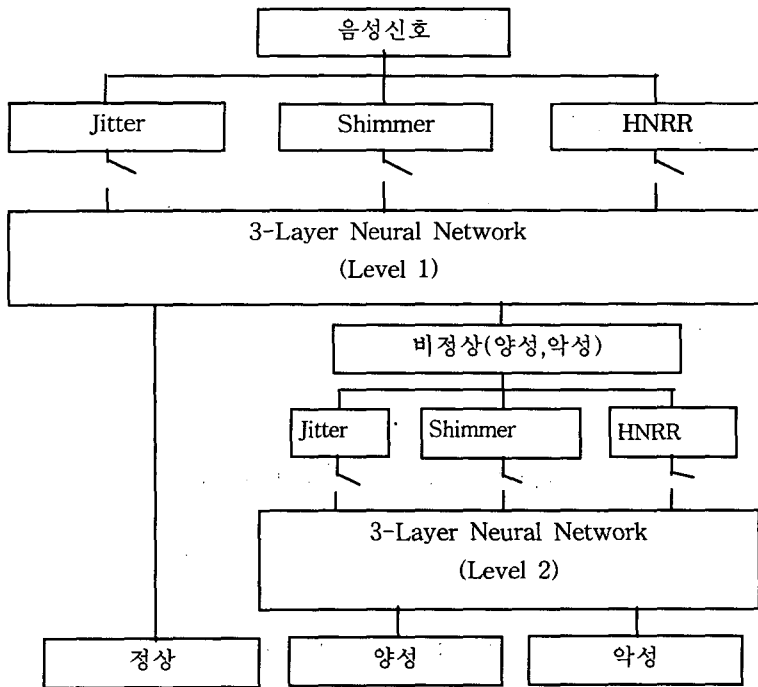


그림 4. 분류실험방법(정상, 양성종양, 악성종양)

먼저 입력 파라미터로 Jitter, Shimmer, HNRR을 구해 각 파라미터들을 2개씩 또는 3개의 일정한 조합으로 단계 1의 신경회로망에 입력하여 정상음성과 장애음성(양성, 악성)

을 구분한다. 단계 1의 결과에서 식별된 장애음성을 다시 단계 2의 신경회로망을 통과시켜 양성종양과 악성종양으로 분류한다. 사용한 신경회로망은 3개의 은닉층을 갖는 다층 퍼셉트론 신경망을 사용하였다.

표 11. 단계 1에서의 에러의 수와 식별율

		Jitt-Shim		Shim-HNRR		HNRR-Jitt		Jitt-Shim-HNRR	
		정상	비정상	정상	비정상	정상	비정상	정상	비정상
Train Data	정상	21	1	19	3	18	4	21	1
	비정상	6	16	2	20	3	19	4	18
식별율(에러/전체)		84.09 % (7/44)		88.64 % (5/44)		84.09 % (7/44)		88.64 % (5/44)	
Test Data	정상	11	.	9	2	9	2	10	1
	비정상	1	10	1	10	2	9	1	10
식별율(에러/전체)		95.45 % (1/22)		86.36 % (3/22)		81.81 % (4/22)		90.9 % (2/22)	

표 12. 단계 2에서의 에러의 수와 식별율

		Jitt-Shim		Shim-HNRR		HNRR-Jitt		Jitt-Shim-HNRR	
		양성	악성	양성	악성	양성	악성	양성	악성
Train Data	양성	11	.	9	2	9	2	11	.
	악성	.	11	0	11	.	11	.	11
식별율(에러/전체)		100 % (0/22)		90.9 % (2/22)		90.9 % (2/22)		100 % (0/22)	
Test Data	양성	6	.	6	.	5	1	6	.
	악성	3	3	3	3	.	6	6	.
식별율(에러/전체)		75 % (3/12)		75 % (3/12)		91.67 % (1/12)		50 % (6/12)	

표 11의 정상음성과 장애음성간의 식별에서 훈련 데이터의 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 세 가지의 조합이 유효할 것으로 판단된다. 표 12의 경우 양성 과 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.

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 보여 주었다.

7. 결 론

본 연구에서는 음성신호에 의해 정상음성과 양성종양, 악성종양간의 분류를 행하는 실험을 행하여 좋은 결과를 얻었다. 통계적 방법에 의한 식별실험에서는 잘 알려진 대로 Jitter, Shimmer가 가장 효과적인 식별 파라미터로 판명되었다. 식별을 위한 음성시료의 종류로는 모음의 종류에 따른 파라미터의 편차가 거의 없는 것으로 관찰되었기에 한 종류의 모음 데이터만으로도 유효한 결과를 얻을 수 있음을 알 수 있었다.

신호처리에 의한 식별실험에서는 jitter, shimmer, HNR등의 파라미터와 두 단계의 다단계신경회로망을 이용하여 정상과 이상 및 양성과 악성으로의 분류를 시도하여 주어진 데이터의 범위 내에서 약 80 %이상의 식별율을 얻었다.

이와 같은 연구를 통하여 얻어진 결과는 차후 성대의 질환을 조기진단하여 질병을 예방하는 데 이용될 수 있다.

참 고 문 헌

- 김대현, 조철우. 1998. "장애음성의 분류방법에 관한 연구." 제15회 음성통신 및 신호처리 워크샵(KSCSP 98) 15(1), 338-391.
- 김대현, 조철우. 1998. "Wavelet변환과 신경회로망을 이용한 후두의 양성종양의 식별에 관한 연구." 한국음향학회 학술발표대회 논문집 17(2), 395-398.
- 조철우, 김대현. 1998. "음원분석을 통한 장애음성의 음향적 특성분석에 관한 연구." 한국음향학회 학술발표대회 논문집 17(1), 163-166.
- 조철우, 김대현. 1998. "Cepstrum방법과 신경회로망을 이용한 양성종양, 악성종양 상태의 식별에 관한 연구." 한국음향학회 학술발표대회 논문집 17(2), 399-402.
- 조철우, 김대현. 1998. "Diagnosis of Pathological Speech Signals Using Wavelet Transform." 한국음성학회 음성과학지 4(2), 17-24.
- B. Yegnanarayana, C. d'Alessandro, V. Darsinos. 1998. "An Iterative Algorithm for Decomposition of Speech Signals into Periodic and Aperiodic Components." *IEEE trans. on Speech and Audio Processing* 6(1).
- Cheol-Woo Jo, Dae-Hyun Kim. 1998. "Diagnosis of Pathological Speech Signals Using Wavelet Transform." *Proc. ITC-CSCC*, 657-660.
- Cheol-Woo Jo, Dae-Hyun Kim. 1998. "Analysis of Disordered Speech Signals Using Wavelet Transform." *Proc. ICSLP* 7, 3039-3042.
- Cox, N., Ito, M. & Morrison, M. 1989. "Data labeling and sampling effects in harmonic-to-noise ratios." *Journal of the Acoustical Society of America* 85(5), 2165-2178.
- Iwata, S. 1972. "Periodicities of pitch perturbations in normal and pathological larynges." *Laryngoscope* 82, 87-96.
- Kay Elemetrics Ltd. 1993. "Multidimensional Voice Program(MDVP)-Model 4305." *Operations Manual*.
- Kay Elemetrics Ltd. 1994. "Disordered Voice Database-Model 4337." version 1.03, *Operations Manual*.

- R. P. Lippman. 1987. "An Introduction to Computing with Neural Nets." *IEEE ASSP Magazine* 4(2), 4-20.
- Yasuo Endo and Hideki Kasuya. 1996. "A Stochastic Model of Fundamental Period Perturbation and Its Application to Perception of Pathological Voice Quality." *Proc. ICSLP 2*, 772-775.

접수일자 : '99. 2. 14.

게재결정 : '99. 3. 15.

▲ 조 철 우

경남 창원시 사림동 9번지
창원대학교 공과대학 제어계측공학과(우: 641-773)
Tel: (0551) 279-7552
e-mail: cwjo@sarim.changwon.ac.kr

▲ 양 병 곤

부산시 부산진구 가야동 산 24
동의대학교 인문대학 영어영문학과(우: 614-714)
Tel: (051) 890-1227 (O), 896-7636 (H)
Fax: (051) 890-1209
e-mail: bgyang@hyomin.donggeui.ac.kr

▲ 왕 수 진

부산시 서구 아미동 1가 10
부산대학교 의과대학 이비인후과(우: 602-739)
Tel: (051) 240-7328,7331 (O), 553-4089 (H)
Fax: (051) 248-1248