

병적 음성과 정상 음성의 음향학적 파라미터 분포에 대한 통계적 분석

김용주, 권순복, 김기런, 신민철, 조철우*, 전계록**, 왕수건***
부산대학교 일반대학원 의공학 협동과정, 부산대학교 의과대학 의공학 교실*,
창원대학교 공과대학 제어계측학과, 부산대학교병원 이비인후과**
전화 : 051-240-7535 / 핸드폰 : 016-581-9788

An analysis of a statistical difference of acoustic parameters' distribution between normal voice and pathological voice

Yongju Kim, Soonbok Kwon, Kiryun Kim, Minchul Shin, Chulwoo Jo*, Gyerok Jeon**,
Soogeun Wang***
Dept. of Interdisciplinary program in Biomedical Engineering, Pusan National University
*Dept. of Biomedical Engineering,
**Dept. of Biomedical Engineering, College of Medicine, Pusan National University
***Dept. of Otolaryngology, College of Medicine, Pusan National University
E-mail : fid2@orgio.net

Abstract

The most basic means of communication among humans is a voice. Without speaking of voice technologies, we found it is important and convenient to use a voice in everyday life. But, in consideration to speech recognition systems, we can't always desire a normal voice input as input signal to the system. Generally speaking, a pathological voice as against a normal, which is a voice with a problem in the larynx, could be also special case of input voice. Of course, but the distortion of a speech signal by environmental effects i.e., noise or transmission channel was a raised problem, we will take up a pathological voices with laryngeal disease which is essential distortion factor in voice. Also, we are to find out the difference of acoustic parameters distribution between normal and pathological voice by a statistical method in our research.

I. 서론

인간의 가장 기본적인 의사소통 수단인 음성은, 굳이 음성을 매개로 하는 기술을 말하지 않더라도, 그 편리함과 중요성은 이미 일상 생활에서 느끼고 있는 사실이다. 하지만, 음성 인식 시스템에 있어서 사람의 목소리는 입력 신호가 되는 것으로서, 정상적인 음성 입력만을 기대할 수는 없다. 여기서, 정상적인 음성이란, 순수하게 병적 음성의 반대의 의미를 가지는 것으로 해석하면 될 것이다. 물론, 잡음이나 전송 채널 등의 환경적인 요인에 의한 음성 신호 왜곡이 일반적으로 대두된 문제점이지만, 본 연구에서는 음성의 본질적인 왜곡 요인인 후두 질환을 가진 병적인 음성에 대해 언급하고, 음향학적 파라미터 분포의 특성을 정상 음성과 통계적으로 비교·분석하여 차이점을 알아보려고 한다.

본 연구의 대 전제는, 정상 음성과 병적 음성에 대하여, 두 집단 각각의 음향학적인 파라미터 분포가 서로 다르다는 가정이다. 따라서, 각 음성 집단의 파라미터 분포를 통계적으로 처리하여 결과를 분석하여, 가정이 참이라는 결론을 이끌어 내는 것이다. 아울러, 병적 음성 인식을 위한 기초적이고 객관적인 연구 자료를 제공하고, 정상 음성의 인식을 위한 전처리와는 다른, 병적 음성의 전처리에 대해 간략히나마 언급하고자 한다.

II. 음향학적 파라미터 추출

2.1 음성 data 수집

(1) 음성 data 수집 과정

본 연구에 사용된 병적 음성 data는 2000년 1월부터 2001년 2월까지 부산대학교 병원 이비인후과에서 병명을 확인 받은 85개 음성 data를 대상으로 하였고, 정상 음성은 후두 질환이 없는 사람을 대상으로 40개의 음성 data를 수집하였다. 음성 data 수집 방법은, 방음실에서 환자에게 /a/ 발성을 3회 하게 하고, 이를 5kHz의 sampling rate와 16 bit로 양자화 하여 DAT (Digital Audiotape Recorder, DTC-59 ESJ, Sony, Japan)에 녹음하였다.

(2) 구축된 음성 data base 현황

수집된 음성 data는 정상 40개와 다음의 4가지 질환 즉, 후두암(glottic cancer) 21개, 성대 용종(polyp) 40개, 성대 마비(palsy), 라인케 부종(Reinke's edema) 11개 로서 다음 표 1로 정리하였다.

표 1. 병적 음성 data 현황

병명	환자수
cancer	21
polyp	40
palsy	13
Reinke's Edema	11

2.2 음향학적 파라미터 추출

(1) 음향학적 파라미터 추출 과정

본 연구에서는 현재 이비인후과에서 가장 널리 사용되고 있는 Kay사(Kay Elemetrics Corp. USA)의 CSL(Computerized Speech Lab 4300B) 기종의 음성 분석 소프트웨어인 MDVP (Multi-Dimensional Voice Program)를 이용하여 33개의 파라미터의 값을 구하였다.

음성 data의 수집에서부터 음향학적 파라미터의 추출에 이르는 과정은 그림 1과 같다.

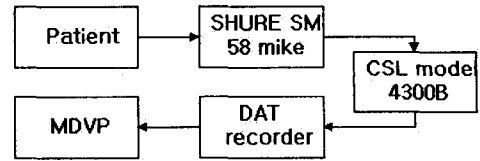


그림 1. 음성 data 수집 및 파라미터 추출 과정

(2) 18개 음향학적 파라미터

MDVP를 이용하여 음향학적 파라미터들을 추출하면 총 33개의 파라미터들이 나오게 되는데, 우리는 그 중 상대적 의미 즉, 질환별 분류에 대한 가능성이 있는 17개의 파라미터만을 사용하여 그 분포를 알아보고자 한다.1)

그리고, 33개의 파라미터 중에서 본 연구에 사용한 18개 파라미터와 그에 대한 간략한 설명은 표2로 나타내었다.

표 2. 본 연구에 사용된 17개 파라미터

Parameters		Unit
APQ	Amplitude Perturbation Quotient	%
ATRI	Amplitude Tremor Intensity Index	%
Fatr	Amplitude Tremor Frequency	Hz
Fftr	Fo Tremor Frequency	Hz
FTRI	Fo Tremor Intensity Index	%
Jita	Absolute Jitter	us
Jift	Jitter percent	%
NHR	Noise-to-Harmonic Ratio	
PPQ	Pitch Period Perturbation Quotient	%
RAP	Relative Average Perturbation	%
ShdB	Shimmer in dB	dB
Shim	Shimmer percent	%
SPI	Soft Phonation Index	
STD	Standard Deviation of the	Hz
vAm	Peak Amplitude Variation	%
vFo	Fundamental Frequency Variation	%
VTI	Voice Turbulence Index	

(3) 각 파라미터의 특성

위의 파라미터가 가지는 의미를 살펴보면, 기본 주파수의 통계적 특성, 장·단기 주파수 동요 요인, 장·단기 크기 동요 요인, 잡음 요인, 성대 떨림의 동요 요인의 5가지의 카테고리 분류를 할 수 있다. 파라미터들을 위의 분류로서 재정리하면 다음의 표 3으로 나타낼 수 있겠다.

1) 예를 들어, Fundamental frequency(Fo) 같은 파라미터는 후두 질환에 대한 상대적 의미가 없다. 다시 말하면, 음성의 피치는 사람의 고유 특성으로서 후두 질환과는 관계가 거의 없다.

병적 음성과 정상 음성의 음향학적 파라미터 분포에 대한 통계적 분석

표 3. 특성에 따른 파라미터 분류

Categories	Parameters
Fundamental Frequency Information Measurement	STD
Short and Long Term Frequency Perturbation Measurement	Jitt, Jita, RAP, PPQ, vFo
Short and Long Term Amplitude Perturbation Measurement	ShdB, Shim, APQ, vAm
Noise Related Measurement	NHR, VTI, SPI
Tremor Related Measurement	FTRI, ATRI, Fftr, Fatr

III. 통계 처리

(1) 통계적 분석 방법

통계적 분석을 위하여 Microcal사 (Microcal Software Inc., USA)의 Microcal Origin 이라는 데이터 분석용 프로그램을 이용하였다. 분석 메뉴 중, One-way ANOVA라는 일원 변량 분석 방법을 사용하여, 유의 수준을 0.05로 하고 두 집단간의 차이가 유의한 지를 검증하였다.

(2) 통계 결과

여기서는 표 3에 분류되어 있는 카테고리 중에서 대표되는 파라미터 4개를 선택하여 다음의 그림들로 결과를 나타내어 보았다.

One-Way ANOVA on col(A) -> col(B):

Data	Mean	Variance	N
A	1.53783	0.57441	40
B	6.03447	20.72594	85

F = 38.36234
p = 8.06994E-9

At the 0.05 level,
the means are significantly different.

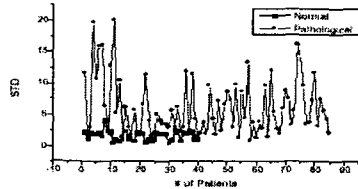


그림 2. STD에 대한 통계 결과

One-Way ANOVA on col(A) -> col(B):

Data	Mean	Variance	N
A	34.90303	401.36298	40
B	177.30526	17632.17168	84

F = 45.27094
p = 5.91547E-10

At the 0.05 level,
the means are significantly different.

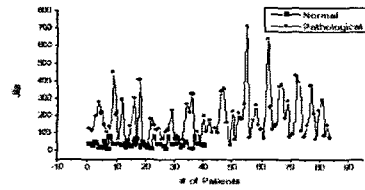


그림 3. Jita에 대한 통계 결과

One-Way ANOVA on col(A) -> col(B):

Data	Mean	Variance	N
A	2.35215	0.79952	40
B	7.35061	11.09912	85

F = 86.75443
p = 6.66134E-16

At the 0.05 level,
the means are significantly different.

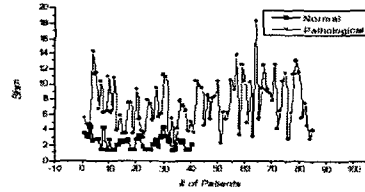


그림 4. Shim에 대한 통계 결과

One-Way ANOVA on col(A) -> col(B):

Data	Mean	Variance	N
A	0.12437	1.93139E-4	40
B	0.18431	0.00373	85

F = 37.43322
p = 1.16126E-8

At the 0.05 level,
the means are significantly different.

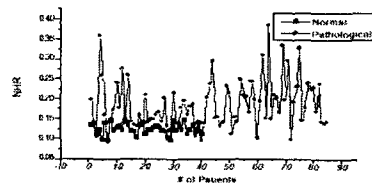


그림 5. NHR에 대한 통계 결과

IV. 결론 및 고찰

이상의 결과로부터, 다음과 같은 사항을 유추할 수 있다.

첫째, 통계적 분석 결과에 의해 정상 음성과 병적 음성이 가지는 음향학적 파라미터의 분포가 유의한 차이를 보이므로, 위의 파라미터들로부터 음성 질환을 예측할 수 있을 것이고, 둘째로, 병적 음성은 고 주파수 대역의 성분을 정상보다 상대적으로 많이 가진다. 따라서, 음성 인식에 있어서 중요한 MFCCs (Mel-Frequency Cepstral Coefficients)와 같은 파라미터를 계산하기 전에, 전처리 과정이나 filter bank의 수정이 필요함을 알 수 있다. 마지막으로, 병적 음성의 파라미터 분포를 보면, 그 표준 편차가 정상에 비해 상당히 크다는 것을 알 수 있다. 즉, spectral envelope를 이용해 인식을 할 경우, 정상 음성보다 더 강인한 인식 시스템을 구현해야 할 것이다.

본 논문은 객관적인 결과를 가지고, 음성 인식이 가지는 일반적인 문제에 접근하려 했다. 앞으로, 여러 가지 연구 결과를 토대로, 병적 음성이 가지는 특성과 일반적인 음성 신호의 인식 문제에 대한 연구가 계속 되어야 할 것이다.

V. 참고 문헌

- [1] 왕수건 외 7인, "병적음성의 감별진단을 위한 음향학적 변수", 대한 이비인후과 학회지, 제 42권, 1999.
- [2] 조철우 외 2인, "음향신호의 분석에 의한 후두질환의 진단에 관한 연구", 대한 음성과학회지, 제 5권 제 1호, 1999. 4
- [3] Operation Manual, "Multi-Dimensional Voice Program (MDVP) Model 4305", Kay Elemetrics Corp., 1993
- [4] 오택섭, "사회과학 데이터 분석법", 나남출판, 1994