

양성후두 질환 음성에 대한 여러 기존 피치검출 알고리즘의 성능 평가

·장승진, ··최성희, ·김효민, ···최홍식, ·윤영로
·연세대학교, 보건과학대학 의공학과
··위스콘신대학 이비인후/두경부 전문외과
···연세대학교 의학대학 이비인후과
e-mail : highnoon@yonsei.ac.kr

Performance Assessment of Several Established Pitch Detection Algorithms in Voices of Benign Vocal Fold Lesions

·Seung-Jin Jang, ··Seong-Hee Choi, ·Hyo-Min Kim, ···Hong-Shik Choi, ·Young-Ro Yoon
·Dept. of Biomedical Engineering, College of Health & Science, Yonsei University
··Div. of Otolaryngology-Head and Neck Surgery, Dept. of Surgery, University of Wisconsin-Madison, WI, USA
···Dept. of Otolaryngology, College of Medicine, Yonsei University

Abstract

Robust pitch estimation is an important study in many areas of speech processing. In voice pathology, diverse statistics extracted from pitch were commonly used to test voice quality. In this study, we compared several established pitch detection algorithms (PDAs) for verification of adequacy of the PDAs. In the database of total pathological voices of 99 and normal voices of 30, an analysis of errors related with pitch detection was evaluated between pathological and normal voices, or among the types of pathological voices such as benign vocal fold lesions; polyp, nodule, and cysts. Consequently, it is required to survey the severity of tested voice in order to obtain accurate pitch estimates.

I. 소개

음성언어 병리학에서 피치를 기반으로 유도되는 주파수 변동(jitter) 및 진폭 변동(shimmer)과 같은 통계 수치들은 음성 질환의 정도를 평가하거나 음성의 쉰(hoarseness)소리와 같은 음성의 질을 평가하는 용도로 주로 이용된다[1,2]. 하지만, 정상인의 발성에 비하여 음성질환자가 가지는 음성의 패턴과 주기가 상이하기 때문에, Titze[3]는 음성신호를 Type 1, 2, 3과 같이 3가지로 분류하여 오직 Type 1의 음성신호일 경우에 한해서 피치 검출의 유효성이 신뢰될 수 있다고 주장하였다. 하지만, 대부분의 피치 검출 알고리즘들은 음성 신호 구간에 대하여 거의 주기적(quasi-periodic)이고 안정적인(stationary) 신호가 존재한다는 가정으로

인하여 실제 음성질환자의 음성에 대한 피치 검출 결과의 신뢰성이 문제되고 있다[4,5]. 그로 인하여, 많은 음성언어 병리학자들이 피치 검출 알고리즘으로부터 도출된 음성 평가의 통계적 수치에 대한 신뢰성에 문제를 제기하고 있는 추세이다[6]. 본 논문에서는 음성언어 장애 분석을 하기에 알맞은 피치 검출 알고리즘을 찾아내기 위하여, 몇 가지 잘 알려진 PDAs에 대한 피치 검출 성능을 평가하였다. 정상인과 음성질환자의 음성을 대상으로 한 피치 검출 성능 분석 연구를 기반으로 여러 기준의 피치 검출 오류에 대한 평가 분석을 기반으로 병적인 음성에 대한 기존의 피치 검출 알고리즘에 대한 신뢰성을 평가하는 것이 목적이다.

II. 관련 연구

2.1 Time domain analysis

시간 영역 기반의 방법들로 Autocorrelation(AC)[7,8], Average Magnitude Difference Function(AMDF)[9], 그리고 YIN[10] 알고리즘을 선택했다. AC 알고리즘의 주요한 점은 피치를 검출하기 위한 에너지 문턱치가 피치 검출에 많은 연관을 갖는다는 것이다. 일반적으로 사용되는 고정된 문턱치 대신 에너지 크기의 변화에 따른 가변 에너지 문턱치를 개발하였다.

2.2 Frequency domain analysis

주파수 영역 기반의 방법들로 Simple Inverse Filter Tracking(SIFT)[11]과 Cepstrum(CEP)[12] 알고리즘을 조사하였다.

2.3 Alternative domain analysis

시간, 주파수 도메인이 아닌 다른 분석 도메인을 이용한 연구로 웨이블릿을 이용한 PDA (WAV)[13]와 시간 delay를 이용한 m차의 상태 공간 거리차 분석(CHA) [14] PDA를 선택했다.

III. 연구방법

3.1 Speech Signal Database

연세대학교, 영동세브란스 병원 음성언어의학연구소에서 2003년부터 2007년 사이에 실제 99명의 음성질환자의 음성과 비교군을 위한 정상인 30명의 정상 음성을 대상으로 피치 검출 능력을 평가하였다. 병적음성은 후두에 외형적인 변형이 발생한 양성후두질환(용종, 낭종, 결절) 음성만을 대상으로 검사하였다. 음성 data는 잡음이 없는 환경에서 컴퓨터화된 음성녹음장비를 (Kay Elemetrics CSL) 이용하여 음성신호와 성문파형신호를 동시에 수집하였다. 모든 음성신호와 EGG data는 22kHz로 샘플링 되어졌고 16bit 분해능을 갖고 있다. 실험 프로토콜에 의해 2-3초간 유지되는 모음([al, lel, lil, lol, lul]) 샘플을 분석하였다.

3.2 Error Measures

각 알고리즘의 성능 평가를 위한 피치 검출 오류 5개 타입을 아래와 같이 정의했다.

- X2(doubling) error :

$$\frac{\text{computed pitch estimate} - \text{reference pitch estimate}}{\text{reference pitch estimate}} > 0.2$$

- /2(halving) error :

$$\frac{\text{computed pitch estimate} - \text{reference pitch estimate}}{\text{reference pitch estimate}} < -0.2$$

- G(gross) error : 반감 오류와 배가 오류의 합

- F(fine) error : 유성음으로 판명된 기간 동안의 계산된 피치추정과 참조 피치추정의 차의 절대 값의 평균

- S(standard deviation) error : F error에 의해 정의된 절대 값의 차이의 표준 편차

V. 결론 및 향후 연구 방향

표 1. 정상인을 대상으로 한 피치 검출 오류

구분	G	X2	/2	F	S
AC	0.3837	0.2731	0.1106	1.3883	3.0915
AMDF	0.4372	0.0667	0.3704	1.3829	2.1622
YIN	0.7315	0.1924	0.5390	14.0413	2.7886
CEP	1.5993	1.4496	0.1497	2.3922	4.9028
SIFT	0.8160	0.6428	0.1732	4.3534	4.2623
WAV	0.2472	0.1473	0.0999	0.9703	2.2854
CHA	1.4448	0.4831	0.9617	2.4554	8.4912

표 2. 양성후두질환자를 대상으로 한 피치 검출 오류

구분	G	X2	/2	F	S
AC	1.6041	0.8824	0.7218	3.0093	6.3882
AMDF	0.9433	0.6747	0.2686	1.9022	3.5404
YIN	3.7361	2.9225	0.8136	15.5253	5.3519
CEP	2.4332	2.4043	0.0290	3.2720	4.7061
SIFT	3.3400	2.4814	0.8586	2.5511	4.6667
WAV	1.1055	1.0384	0.0672	2.5511	4.6667
CHA	4.1187	2.0839	2.0347	6.0618	13.3203

최적의 윈도우 구간을 기반으로 한 각 PDA들의 피치 검출 오류는 표 1, 2와 같다. 각각 정상인과 양성후두질환자를 대상으로 분석한 결과 정상인의 경우, G 오류에 있어 CEP와 CHA 나쁜 결과를 보인 반면 AC와 WAV가 좋은 결과를 보였다. G 오류를 제외한 구간에서 검출된 F 오류의 경우 거의 대부분이 비슷한 수준을 보인 반면 YIN 기반의 PDA에 대한 오류는 실제 피치와 상당히 큰 차이를 발생함을 알 수 있다. 종합적으로 분석한 결과 정상인을 대상으로 한 PDA 검출 결과는 실제 사람의 눈으로 피치를 분석하는 방법과 유

사한 WAV 기반의 PDA가 가장 우수한 성능을 보이는 것으로 판명되었다. 반면 음성질환자의 경우에 있어 정상인의 경우와 비교해 볼 때 모든 PDA에 있어 G 오류가 대부분 크게 증가함을 알 수 있었다. 하지만 F 오류의 경우에는 정상인의 경우와 비교해 볼 때 감소한 경우가 발생하거나 증가하더라도 크게 증가하지 않았음을 알 수 있다 (CHA는 제외). 표 2의 결과로 미루어 볼 때, 양성후두질환자의 음성이 정상인보다 불규칙하고 피치 반감, 배가와 같은 오류가 더 많이 발생하기 때문에 G 오류가 더 높아졌음을 유추할 수 있다. 종합적인 분석에 의해 양성후두질환자의 경우 G 오류는 YIN, SIFT, CHA가 가장 나쁜 결과를 보이고, AMDF가 좋은 결과를 보였다. F 오류의 경우에도 마찬가지로 AMDF가 좋은 결과를 보였으며, 이는 실제 피치를 검출하기 위한 반주기 검출 알고리즘으로 피치 배가와 같은 오류가 감소된 것이라 예상된다.

Acknowledgment

This study was supported by a grant of the Korea Health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea (A020602).

참고문헌

- [1] F. Klingholz, F. Martin, "Quantitative spectral evaluation of shimmer and jitter", J. Speech Hear. Res., vol. 28, June 1985, pp. 169-174.
- [2] S. Feijoo, C. Hernandez, "Short-term stability measures for the evaluation of vocal quality", J. Speech Hear. Res., vol. 33, June 1990, pp. 324-334.
- [3] IR. Titze, Workshop on Acoustic Voice Analysis: Summary. Statement. Iowa City, IA: National Center for Voice and Speech; May 1995.
- [4] M.P. Karnell, R.S. Scherer, L. Fischer, "Comparison of acoustic voice perturbation measures among three independent voice laboratories", J. Speech hear. Res., vol. 34, 1991, pp. 781-790.
- [5] M.P. Karnell, K.D. Hall, K. Landahl, "Comparison of fundamental frequency and perturbation measurements among three analysis systems", J. Voice, vol. 4, 1995, pp. 383-393.
- [6] S. Bielamowicz, J. Kreiman, B.R. Gerratt, M.S. Dauer, and G.S. Berke, "Comparison of voice analysis systems for perturbation measurements", J. Speech Hear. Res., vol. 39, 1996, pp. 126-134.
- [7] M.M. Sondhi, "New methods of pitch determination", IEEE Trans. Audio Electroacoust., Vol. 16, June, 1968, pp. 262-266.
- [8] L.R. Rabiner, "On the use of autocorrelation analysis for pitch detection", IEEE Trans. Acoust. Speech Signal Process., Vol. 25, no. 1, Feb. 1977, pp. 24-33.
- [9] J.R. Deller, Jr., J.G. Proakis, J.H.L. Hansen, Example short-term features and applications, Discrete-Time Processing of Speech Signals, Macmillan, New York, 1993.
- [10] A. D. Cheveigne and H. Kawahara. Yin: A fundamental frequency estimator for speech and music. Journal of the Acoustical Society of America, vol. 111, no. 4, 2002, pp. 1917-1930.
- [11] J. D. Market, "The SIFT algorithm for fundamental frequency estimation", IEEE Trans. Audio Electroacoust., vol. 20, Dec 1972, pp.367--377.
- [12] R. W. Schafer and L. R. Rabiner, "System for automatic formant analysis of voiced speech", Acoust. Soc. Amer., vol. 47, Feb 1970, pp. 634--648.
- [13] S. Kadambe, GF. Bourdeaux-Bartels, "Application of the Wavelet transform for pitch detection of speech signals", IEEE Trans Inf. Theory , vol. 38, no. 2, 1992, 917 --24.
- [14] D.E. Terez, "Robust pitch determination using nonlinear state-space embedding", Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02), Vol. 1, pp. 345-348.