

피치 반감 배가를 유발하는 병적인 음성 분석을 위한 강인한 피치 검출 알고리즘

¹장승진, ²최성희, ¹김효민, ³최홍식, ¹윤영로
¹연세대학교, 보건과학대학 의공학과
²위스콘신대학 이비인후/두경부 전문외과
³연세대학교 의학대학 이비인후과

Robust Pitch Detection Algorithm for Pathological Voice inducing Pitch Halving and Doubling

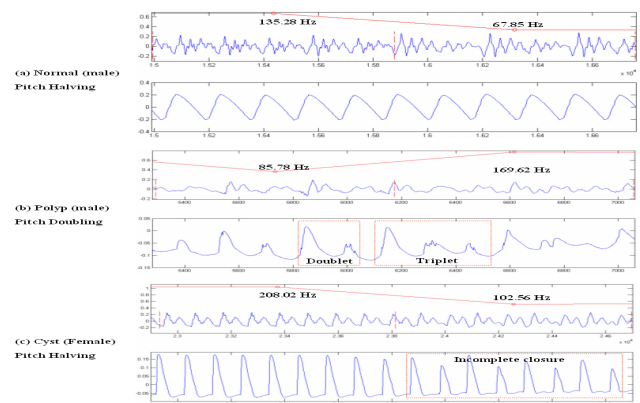
¹Seung-Jin Jang, ²Seong-Hee Choi, ¹Hyo-Min Kim, ³Hong-Shik Choi, ¹Young-Ro Yoon
¹Dept. of biomedical engineering, College of Health & Science, Yonsei University

²Div. of Otolaryngology-Head and Neck Surgery, Dept. of Surgery, University of Wisconsin-Madison, WI, USA

³Dept. of Otolaryngology, College of Medicine, Yonsei University

Abstract - In field of voice pathology, diverse statistics extracted from pitch estimation were commonly used to assess voice quality. In this study, we proposed robust pitch detection algorithm which can estimate pitch of pathological voices in benign vocal fold lesions. we also compared our proposed algorithm with three established pitch detection algorithms: autocorrelation, simplified inverse filtering technique, and nonlinear state-space embedding methods. 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.

According to the results of pitch errors, gross pitch error showed some increases in cases of pathological voices; especially excessive increase in PDA based on nonlinear time-series. In an analysis of types of pathological voices classified by aperiodicity and the degree of chaos, the more voice has aperiodic and chaotic, the more growth of pitch errors increased. Consequently, it is required to survey the severity of tested voice in order to obtain accurate pitch estimates.



〈그림 1〉 피치 반감, 배가가 발생하는 음성신호 사례 (a)정상(남성)신호에 발생하는 피치 반감, (b)용종(남성)신호에 발생하는 피치 배가, (c)낭종(여성)신호에 발생하는 피치 반감 - 각 그래프의 상단과 하단은 각각 speech signal과 Electroglographic signal이며, autocorrelation 피치 검출 알고리즘에 의한 분석 결과임

1. 서 론

음성질환 연구 분야에 있어 피치 검출로부터 유도된 피치의 주파수 변동(jitter)과 피치의 진폭 변동(shimmer)과 같은 통계적 파라미터들은 음성질환의 정도를 평가하거나, 음성의 권(hoarseness)소리와 같은 음성의 질을 평가하는 용도로 주로 사용된다[1,2]. 일반적인 피치 검출 알고리즘에 있어 제한된 신호구간에서 음성 파형은 주기적(periodic)이고 안정적(stationary)이라는 가정 하에 각각의 검출 도메인에서 최소한의 주기적인 패턴을 분석하여 피치를 검출하게 된다. 그러나 실제 음성신호는 제한된 구간에서 결코 안정적이지도 않고 근사주기적인(quasi-periodic) 신호가 아닌 경우가 발생한다.

Titze[3]는 음성신호를 Type 1, 2, 3과 같이 3가지로 분류하여 오직 Type 1의 음성신호일 경우에 한해서 피치 검출의 유효성이 입증될 수 있다고 주장하였다. 각각의 Type 분류에 대한 정의를 살펴보면, Type 1 신호는 분석 구간에서 양적인 변화 없는 거의 주기적인 신호를 의미한다. Type 2의 음성 신호는 기본주파수(fundamental frequency)의 에너지에 근접하는 변조된 주파수 또는 서브주파수를 갖는 신호에 의해 분석 구간에서 변화가 발생하는 신호로 정의된다. Type 3 신호는 전혀 주기적인 신호를 알 수 없는 신호로 정의된다. 다양한 상용화된 음성 분석 소프트웨어에 내장된 피치 검출 알고리즘들이 동일한 음성 신호 분석에 대하여 제 각각 상이한 피치 검출 결과 값을 산출한다는 연구 결과들이[4,5] 발표되었다. 그로 인하여, 많은 음성언어 병리학자들이 피치 검출 알고리즘으로부터 도출된 음성 평가의 통계적 수치에 대한 의의가 가장 되고 있는 추세이다[6].

본 연구에서는 그림 1과 같이 Type 2 음성 신호에서 자주 발생하는 피치 반감 또는 배가에 대해서도 정확하게 피치를 검출해내는 강인한 피치 검출 알고리즘을 제안하였다. 또한, 피치 검출 성능의 우수성을 입증하기 위하여 기존의 3가지 알고리즘들(autocorrelation, simplified inverse filtering technique, and nonlinear state-space embedding)과 비교하여 그 우수성을 입증함을 연구의 목적으로 한다.

2. 본 론

2.1 강인한 피치 검출 알고리즘

2.1.1 Fast Orthogonal Search 알고리즘

Fast Orthogonal Search (FOS) 알고리즘은 1987년 Korenberg[7]에 의해 처음 개발된 알고리즘으로서 비선형 차분 방정식의 모델에 대한 계수와 구조를 결정하기 위한 용도로 사용되어 왔다. FOS 알고리즘은 white와 colored noise에 강인한 뿐만 아니라 누설되거나 잘못 위치된 신호에 대해서도 매우 강한 알고리즘으로 증명되었다. 그 외에도 16차의 다항식으로 모델링된 정적 비선형 모델에 대해서도 매우 안정적으로 동작함을 입증할 뿐만 아니라, FFT보다 더 좋은 분해능을 보이는 것으로 판명되었다. 그 원리를 간단히 살펴보면, 만약 아래와 같은 비선형 차분 방정식이 존재한다고 가정할 때,

$$y(n) = F[y(n-1), \dots, y(n-k), \dots, x(n), \dots, x(n-l)] + e(n)$$

이 방정식은 다음과 같이 간단하게 요약될 수 있다.

$$y(n) = \sum_{m=0}^M a_m p_m(n) + e(n) \quad (a_m : \text{coefficients}, p_m : \text{candidate functions})$$

일반적으로 비선형 시스템을 위한 효과적인 방정식은 A 행렬이 대칭 양정행렬이라는 가정 하에 Cholesky 재구성 방법을 사용하여 풀 수 있으며, 이러한 가정으로 아래와 같이 식을 변형할 수 있다.

$$y(n) = \sum_{m=0}^M g_m w_m(n) + e(n)$$

g_m 과 $w_m(n)$ 은 각각 직교(orthogonal) 전개를 위한 계수와 목시적인 직교 함수들을 의미한다. 위의 식을 풀기 위한 일반적인 식은 Gram-Schmidt orthogonalization에 의한 방법이며 아래와 같이 전개할 수 있다.

$$w_m(n) = p_m(n) - \sum_{r=0}^{m-1} \alpha_{mr} w_r(n) \quad \text{where } g_m = \frac{\sum_{n=1}^N y(n)w_m(n)}{\sum_{n=1}^N (w_m(n))^2}$$

하지만 위의 $w_m(n)$ 을 구하는 것은 매우 계산적으로 많은 시간이 소요되므로, $w_m(n)$ 을 계산하지 않고 수정된 Cholesky 재구성 방법을 이용하여 수식을 전개하면 아래와 같이 구할 수 있다.

$$g_m = \frac{C(m)}{D(m,m)}, \quad m=0, \dots, M$$

이 때, 각각의 변수들은 아래와 같이 정의된다.

$$D(0,0) = 1$$

$$D(m,0) = \overline{p_m(n)}, \quad m=1, \dots, M$$

$$D(m,r) = \overline{p_m(n)p_r(n)} - \sum_{i=0}^{r-1} \alpha_{ri} D(m,i), \quad m=1, \dots, M, \quad r=1, \dots, m$$

$$\alpha_{mr} = \frac{D(m,r)}{D(r,r)}, \quad m=1, \dots, M, \quad r=0, \dots, m-1$$

또한, 부가적으로 다음과 같은 변수들의 정의를 필요로 한다.

$$C(0) = \overline{y(n)}$$

$$C(m) = \overline{y(n)p_m(n)} - \sum_{r=0}^{m-1} \alpha_{mr} C(r)$$

이 때, 비선형 차분방정식의 후보 함수 p_m 를 아래와 같이 대치하면,

$$p_{2m}(n) = \sin\left(\frac{2\pi f_m n}{N}\right)$$

$$p_{2m+1}(n) = \cos\left(\frac{2\pi f_m n}{N}\right)$$

아래와 같은 계산에 의하여 분석 신호에 대한 주파수 밀도에 함수를 구할 수 있게 된다.

$$F(f_m) = \sqrt{a_{2m}^2 + a_{2m+1}^2}$$

$$\phi(f_m) = \tan^{-1}\left(\frac{a_{2m+1}}{a_{2m}}\right)$$

2.2 주파수밀도 분해능 분석을 위한 차수 결정

주파수 밀도에 대한 분해능은 후보 함수 p_m 의 m 차수를 증가시킬수록 늘어나게 된다. 하지만 메모리, 시간적 소모를 줄이기 위하여 mse가 적정 문턱치까지 변화하지 않으면 수행을 중단하게 한다. 아래식과 같이

$$\overline{\epsilon^2(n)} = \overline{y^2(n)} - \sum_{m=0}^M g_m^2 \overline{w_m^2(n)}$$

mse는 다음과 같이 구해 질 수 있고 이는 cos과 sin 두 후보함수의 pair로 주파수 밀도를 구하기 때문에 같이 아래와 같이 연결지어 E_m 을 구하며 문턱치의 값은 0.1e-4로 설정하여 수행하였다.

$$E_m = g_m^2 \overline{w_m^2(n)} = g_m^2 D(m,m)$$

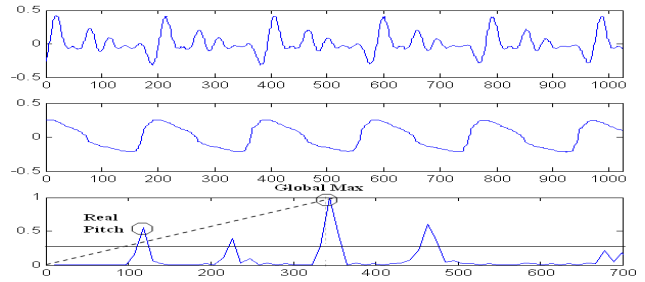
$$E_m = E_{2k} + E_{2k+1}$$

2.2 피치 검출 방법

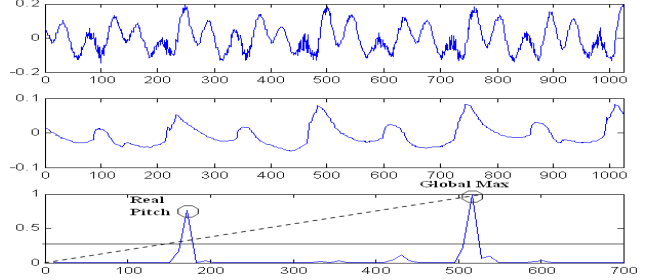
피치 검출방법은 0~700Hz 사이의 피치검출 대역에 대하여 검사한다. 검사 구간에서 가장 큰 피크를 우선 검출 후 이를 기준으로 0점에 대한 기울기 직선보다 큰 피크가 위치하는 위치를 검출하여 가장 큰 피크가 검출되면 그 값을 실제 피치로 결정한다. 이러한 이유는 harmonic 성분이 많이 포함된 음성신호 분석에서 FOS는 일반적으로 해당 harmonic 성분의 검출로 인해 실제 피치 피크보다 더 큰 값의 피크 값을 보이기 때문이다. 또한 피크 검출의 정확성을 높이기 위해 문턱치 0.2 이상의 피크구간에 대하여 메디안 값을 구하여 피치를 검출하였다(그림 2-3).

2.2.1 피치검출 오류 분석

정상인(30)과 양성후두질환자(99)를 대상으로 2초 이상의 |a| 모음에 대하여 제안된 알고리즘과 비교되는 알고리즘을 대상으로 피치 검출 에러를 분석하였다. 실제 전문가에 의해 반자동으로 실제 피치에 해당하는 기준 피치를 검출 후 비교한 결과 표 1과 같은 결과를 도출하였다. 분석 X2는 배가 발생 오류, /2는 반감 발생 오류이고 (빈도수 계산), Fine은 기준 피치에 대한 오차 평균, S는 표준 편차 오류이다. 피치 반감, 배가 오류가 발생할 경우 Fine 오류 계산에는 적용하지 않았다.



〈그림 2〉 정상인 피치 검출 분석



〈그림 3〉 플립한자 피치 검출 분석

〈표 1〉 정상인과 후두질환자를 대상으로 한 피치검출 성능분석 결과

	Normal Voices (n=30)				Pathological Voices (n=99)				
	Error	X2(f,%)	/2(f,%)	Fine(%)	S(%)	X2 (f,%)	/2 (f,%)	Fine (%)	S (%)
AC	1.05	4.04	3.12	2.09	2.58	14.23	14.87	21.14	
SIFT	1.09	3.47	4.98	3.71	7.73	5.91	13.76	19.04	
NSSE	0.29	2.49	2.94	2.29	2.81	7.37	9.15	8.18	
FOS	0.05	1.04	2.99	2.13	1.75	1.81	3.21	4.13	

3. 결 론

후두질환자를 대상으로 평가한 결과 FOS를 기반으로 한 피치 검출 알고리즘이 다른 기존의 알고리즘에 비하여 피치 반감, 배가 오류가 훨씬 적게 나타남을 알 수 있다. 또한 정상인을 대상으로도 좋은 피치 검출 결과가 발생함을 알 수 있었다.

Acknowledgment

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

[참 고 문 헌]

- [1] F. Klingholz, F. Martin, "Quantitative spectral evaluation of shimmer and jitter", J. Speech Hear. Res., vol. 28, pp. 169-174, 1985.
- [2] S. Feijoo, C. Hernandez, "Short-term stability measures for the evaluation of vocal quality", J. Speech Hear. Res., vol. 33, pp. 324-334, 1990.
- [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, pp. 781-790, 1991.
- [5] M.P. Karnell, K.D. Hall, K. Landahl, "Comparison of fundamental frequency and perturbation measurements among three analysis systems", J. Voice, vol. 4, pp. 383-393, 1995.
- [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, pp. 126-134, 1996.
- [7] M.J. Korenberg, "Fast orthogonal identification of nonlinear difference equation and functional expansion models", in Proc. Midwest Symp, Circuit Sys. vol. 1, pp. 270-276, 1987.