음원 파라미터 모델과 인공신경망을 이용한 음성장애 검출*

Screening of Voice Disorder using Source Parameter Model and Artificial Neural Network

파벨 시틸** · 조 철 우*** · 미샤 파벨**
Pavel Chytil · Cheolwoo Jo · Misha Pavel

ABSTRACT

There is a number of clinical conditions that affect directly or indirectly the physical properties of the vocal folds and thereby the pressure waveforms of elicited sounds. If the relationships between the clinical conditions and the voice quality are sufficiently reliable, it should be possible to detect these diseases or disorders. The focus of this paper is to determine the set of features and their values that would characterize the speaker's state of vocal folds. To the extent that these features can capture the anatomical, physiological, and neurological aspects of the speaker they can be potentially used to mediate an unobtrusive approach to diagnosis. We will show a new approach to this problem supported with results obtained from two disordered voice corpora.

Keywords: pathological voice, neural network, disorder, pattern recognition

1. 서 론

음성의 생성은 화자의 인지학적, 신경학적, 물리적 상태에 영향을 받는다. 그리고 성대는 음성의 음원으로 중요한 역할을 한다. 성대에 발생하는 질환의 경우 음원에 영향을 주게 되어 음성의 특징을 변화시키게 된다. 바꾸어 말하면 음성의 특성을 측정함으로써 음원에 해당하는 성대의 질환을 역으로 추정하는 것도 가능하게 된다[1]. 기존의 음성분석에 의한 진단방법들은 음성신호로부터 주파수 또는 시간 영역에서의 불규칙성을 간접적으로 측정하는 방법을 통하여 파라미터를 추출하였다. 기존의 방법들에서 사용되어온 Jitter, Shimmer, NHR등과 같은 파라미터들은 성대의 운동을 직접적으로 측정하기 보다는 성대의 운동으로부터 발생하는 음성신호의 불규칙성을 간접적으로 측정할 수 있게 하는 파라미터들이었다.

또한 지금까지 음성의 음원을 수학적으로 모델링하고 이에 의해 음성신호로부터 음원을 추정하는 연구가 진행되어 왔다[2,3,4]. 만약 음원으로부터 직접 추출한 파라미터를 이용할 수 있다면 성대

* 본 논문은 2007년도 창원대학교 연구교수 연구비에 의하여 연구되었음.
** Biomedical Engineering Department, OHSU, USA
*** 창원대학교 메카트로닉스 공학부

부위의 장애를 더 정확하게 측정할 수 있을 것이라고 가정해 볼 수 있다. 음원 파라미터는 기존의 파라미터에 비해서 보다 근접한 성대운동을 묘사해 줌으로써 장애음성 진단에 유용할 것이라고 생각할 수 있다.

본 연구에서는 잘 알려진 음원모델을 적용하여 음원 파라미터를 구하고 이들 파라미터를 장애음성신호 분석에 적용하여 인공신경망에 의해 정상음성과 장애음성간의 식별을 시도하였다.

2. 분석방법

음원분석과정은 두 단계로 이루어져 있다. (1) 성도 특성의 측정과 피치주기의 측정 (2) 성대 모델에 최적의 파라미터 추정. 성도 모델은 일정한 구간동안 특성이 일정하다는 가정하에 일정 구간의 음성을 취하여 분석한다. 추정 방법은 선형예측 방법에 의해

성도의 특성을 추정한 뒤 선형예측 오차 신호로부터 성문파형을 복원해 낸다. 성문파형복원에는 IAIF방법을 적용한다[6]. 이 방법은 기존의 방법에서 성문파형을 추정할 때 발생하는 문제점을 여러 번의 다양한 차수의 선형예측을 통하여 감소시킴으로서 통상적인 피치 동기 분석에 버금가는 결과를 얻을 수 있다는 점에서 채용하였다. 일차 예측된 성문파형을 이용하여 성대 모델과 적합도를 비교한다. 성대모델의 파라미터들을 비교하며 가면서 가장 오차가 적은 파라미터를 주어진 음성의 성대모델로 구한다.

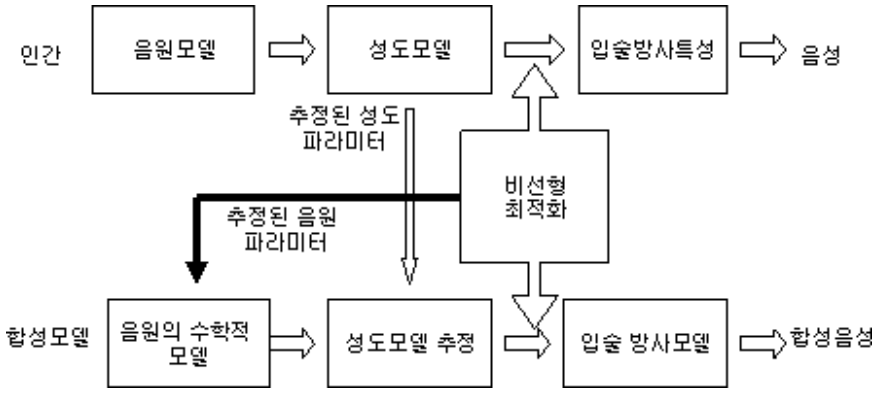

그림 1. 파라미터 추정과정의 블록도

최적 파라미터를 구하기 위해서 실제 음성 데이터와 모델로부터 성도 특성을 통과한 신호와의 상관도를 측정하여 상관도가 가장 높은 경우를 택하였다. 최적화 알고리듬으로는Nelder-Mead simplex검색방법을 적용하였다. <그림 1>은 처리과정에 대한 블럭도를 보인 것이다.

음원 모델로는 Fujisaki-Ljungqvist(FL)모델을 사용하였다. 성문파와 미분값은 부분적인 다항식으로 구성되어 있다. 이와 같은 다항식을 통해 각 파라미터들을 효과적으로 구현할 수 있다. FL모델은 3 개의 시간 관련 파라미터와 3개의 진폭관련 파라미터들로 구성되어 있다. <그림 2>는 FL모

델을 나타낸 것이다. 그림에서 파라미터 A는 다른 응용에서는 잘 사용되지 않지만 여기서는 성문의 열림점에서 두번째 여기가 일어나는 경우가 종종 있기 때문에 포함하였다. 성문의 닫힘에 있어서 둥근 형상은(D구간) 실제 음원 여기에서와 유사한 형태를 띠고 있다. 성문파의 미분값은 다음과 같이 표현할 수 있다.

$$E(t)=\begin{cases} A-\frac{2A+R\alpha}{R}t+\frac{A+R\alpha}{R^2}t^2 & 0<t\le R \\ \alpha(t-R)+\frac{3B-2F\alpha}{F^2}(t-R)^2-\frac{2B-F\alpha}{F^3}(t-R)^3 & R<t\le W \\ C-\frac{2(C-\beta)}{D}(t-W)+\frac{C-\beta}{D^2}(t-W)^2 & W<t\le W+D \\ \beta & W+D<t\le T \end{cases}, \tag{1}$$

여기서 W=R+F이며 a, b 는 다음 식으로 표시된다.

$$\alpha=\frac{4AR+6FB}{2R^2-F^2}, \tag{2}$$

$$\beta=\frac{CD}{D-3(T-W)}. \tag{3}$$

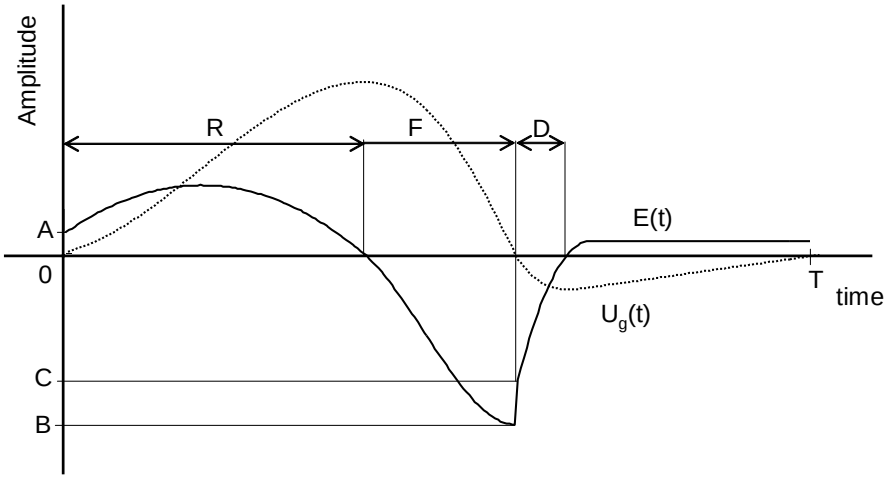

그림 2. FL모델에 의한 한 주기의 성문파 모델
Ug(t)와 미분값 E(t) 및 파라미터들

음원 추출을 위해서 다음과 같은 세 가지 다른 방법을 시도하였다.

1) LPC 분석을 통한 방법
2) Cepstrum을 이용한 방법
3) IAIF(Interactive Adaptive Inverse Filtering)방법

분석절차는 다음과 같다.

1) LPC분석에서는 음원 모델 추정을 선형예측 오차로부터 성문파를구한다[4].
2) Cepstrum 방법에서는 Cepstrum으로부터 구한 선형예측오차로부터 성문파를 구한다[5].
3) IAIF방법은 반복적 선형예측을 통해 성문파를 구한다[6].

적용한 세 가지 방법의 유효성을 판정하기 위해서 분석된 음원을 이용하여 재합성한 음성과 원래 음성간의 상관도 분석을 실시하였다. 상관도를 분석한 뒤 각 화자에 대한 상관도의 빈도수를 측정하였다. 전체 화자에 대하여 원래 음성과 지정된 방법에 의해 분석한 파라미터를 이용하여 재합성한 음성간의 상관도가 상대적으로 높은 화자의 수가 많은 분석 방법을 최적의 음원분석법으로 정할 수 있다.

식별을 위한 패턴인식 방법으로는 인공신경망을 이용하였다[7]. 사용한 인공신경망의 구조는 총 3 개 층으로 구성되었다. 1 개의 입력층, 1 개의 은닉층, 1 개의 출력층으로 구성되었으며 입력층의 갯수는 모델 파라미터의 개수와 동일하며 은닉층의 개수는 5 개에서 54 개까지 변화시켜 가며 적용하였다. 신경망은 전체 데이터 중에서 정상음성과 특정 장애간을 식별하도록 구성하였다. 본 실험에서는 신경망을 훈련할 때 정상음성과 훈련에 사용될 특정 장애음성 외에 다른 장애음성은 제외하고 각 질환과 정상음성간의 식별만을 시도하였다.

기존의 식별실험들에서는 정상음성 전체와 장애음성 전체 집단간을 구분하는 시도를 하고 있다. 그러나 본 실험에서는 정상음성과 특정 장애음성간의 식별을 하는 것을 목표로 하고 있다는 것이 기존의 연구와 다른 접근 방법이다. 이러한 시도는 향후 개별 질병의 식별기를 구현하기 위한 전단계로 볼 수 있다.

3. 장애음성 데이터베이스

실험에 사용된 장애음성 데이터베이스는 700 명 분의 화자로부터 녹음된 1400 개의 음성시료로 구성된 Kay Elemetrics의 Disordered Voice Database[8]와 100 명의 화자의 음성을 포함한 한국어 장애음성 데이터베이스[9]를 이용하였다.

Kay의 데이터베이스는 여러 가지 질환에 대한 장애음성이 53 명의 정상화자의 음성과 함께 녹음되어 있으며 모음의 지속발성과 문장음성이 포함되어 있으나 그 중에서 우리는 지속 발성된 /아/ 음성만을 사용하였다. 이는 기존의 연구에서 /아/ 음성만에의한변별력이다른음성의 경우와 유사하게 나왔기 때문이다. 한국어 장애음성 데이터베이스의 경우 41 명의 정상화자와 함께 28 명의 양성 후두질환, 31 명의 악성 후두질환의 경우를 포함하고 있으며 역시 각 화자당 하나씩의 /아/ 음성만을 사용하였다. 두 데이터베이스는 모두 16 KHz로 재 표본화하여 동일한 대역폭을 갖도록 수정후 통합하여 사용하였다. 해상도는 16 비트이다.

표 1. Kay-Elemetrics 코퍼스에서의 질병별 화자의 수

Disorders	No. of cases
Adductor spasmodic dysphonia	20
A-P squeezing	167
Paresis	22
Gastric reflux	54
Hyper-function	188
Paralysis	79
Polyp	56
Cyst	10
Nodules	23
Keratosis / Leukoplakia	30
Normal	53
Others	97

4. 결과 및 검토

<그림 3>은 역필터링에 의해 구한 FL모델와 원 음성과의 최대 상관값의 변화를 보인 것이다. 모델을 구할 때 모델과 원래 음성과의 상관도를 기준으로 구하였으므로 세 가지 방법중 가장 최대 오차가 적은 방법을 최적의 방법으로 간주할 수 있다. 이 그래프에서는 최대 상관값에 따른 화자의 수를 세 가지 분석방법에 대해 표시하였다. IAIF방법이 최대 상관도를 보인 화자의 수가 가장 많으므로 분석 결과가 다른 방법들 보다 비교적 정확하다고 판단할 수 있다.

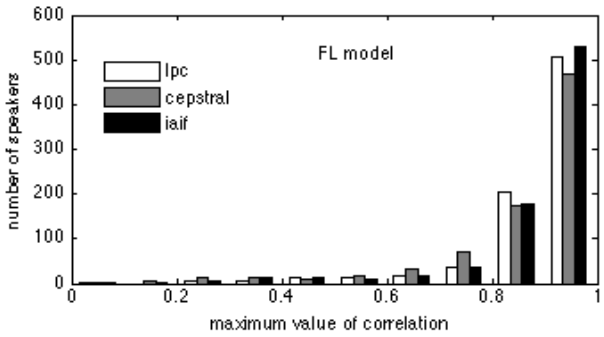

그림 3. 서로 다른 역 필터링 방법에 대한 FL모델과 원신호와의 상관도 분포

<그림 4>는 최대 상관도 상황에서 추출한 FL모델과 그 때의 여기신호를 비교한 것이다. <그림 5, 6>은 장애음성과 분석된 파라미터를 이용하여 합성한 음성을 서로 비교한 것을 각각 서로 다른 상관도의 경우에 제시한 것이다.

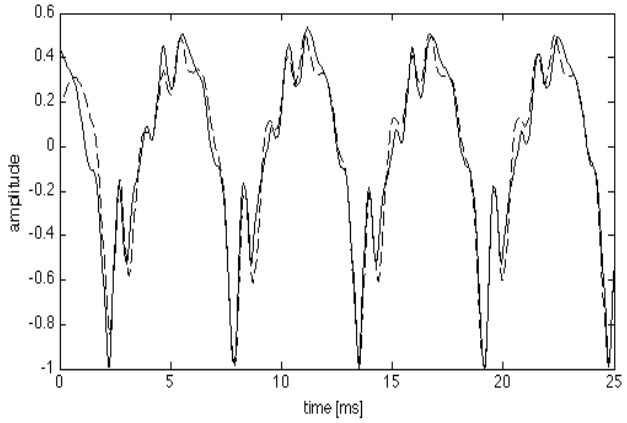

그림 4. 역필터링 방법에 의한 최대 상관도의 경우의 FL모델의 추정결과

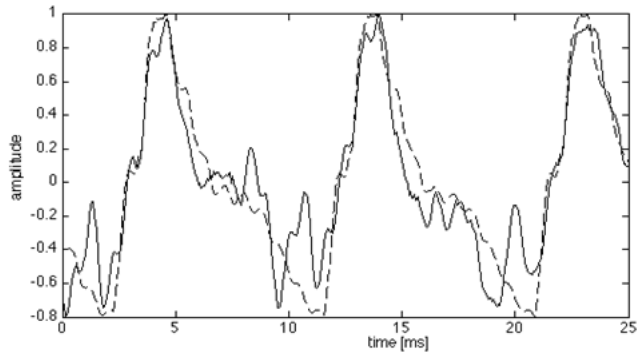

그림 5. 장애음성과 분석후 합성음의 비교, 실선이 원래음성이고 점선이 합성음성 (최대 상관도 0.804).

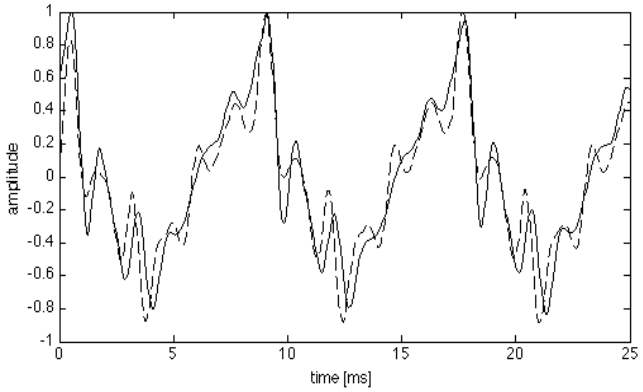

그림 6. 장애음성과 분석후 합성음의 비교, 실선이 원래음성이고 점선이 합성음성 (최대 상관도 0.932).

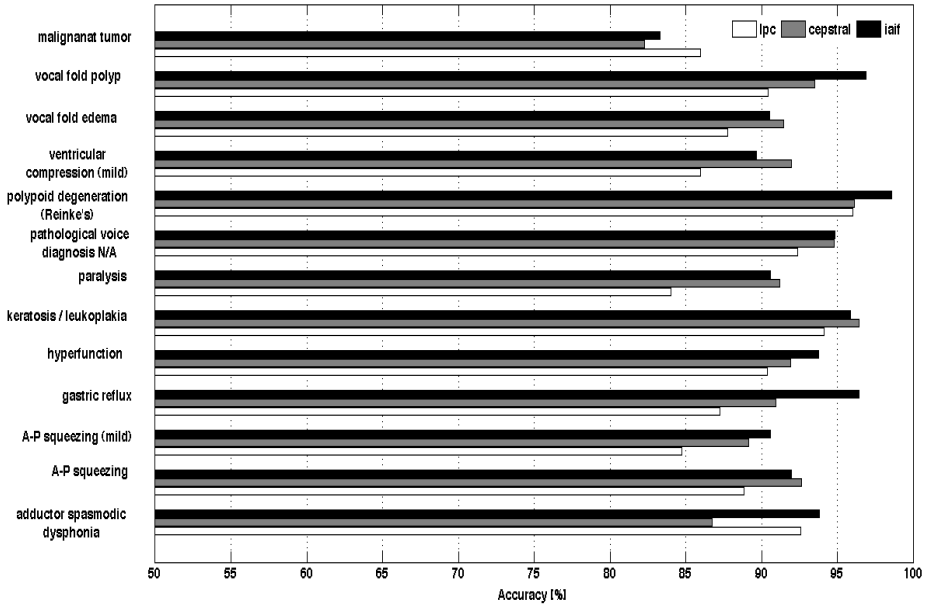

그림 7. 세 가지 역 필터링 모델에서의 FL모델을 이용한 장애 검출율

<그림 7>은 데이터의 각 분류별로 인공신경망에 의해 정상음성과 장애음성이 식별되는 정도를 세 가지 분석방법에 대해 표시한 것이다.

전체적으로 정상음성과 특정 질환을 가진 화자와의 식별율은 85% 정도의 정확도를 보이고 있다. 동일한 질환의 경우에는 분석방법에 따라 IAIF가 가장 높은 식별율을 보이고 있어서 적합한 분석방법이라는 것을 알 수 있다. 그러나 특정 질환에 따라서는 그렇지 못한 경우도 있었다.

A-P squeezing의 경우와 A-P squeezing-mild의 경우에서 보듯이 mild의 경우가 질병이 심해진 경우에 비해서 안좋은 정확도를 보여주고 있다. 이것은 mild한 경우가 질환의 증상이 뚜렸하지 않기 때문에 진단이 어렵다는 점과 일치한다.

개별 질환과 정상음성간의 식별율에 있어서 주목할만한 식별율을 얻었으므로 일반적인 식별시스템을 구축하고자 한다면 이들 신경망을 결합한 하이브리드 시스템의 구축을 통해 가능할 것으로 생각되며 여기에 본 분석결과들을 활용 할 수 있을 것으로 생각된다.

4. 결 론

본 연구의 결과를 통하여 실제 음원의 생리적 모델인 성대 펄스 모델 파라미터를 이용한 장애음성 식별 방법의 유효성의 가능성을 확인하였다. 성대 펄스 모델은 실제 발성기관의 구조를 모델링한 것이므로 질환으로 인한 성대의 변화와 관련성이 있다는 것을 확인할 수 있었다. 장애음성과 정상음성간의 평균 식별율은 LPC방법의 경우 88.7%, 켑스트럼 방법의 경우 90.8%이었고 IAIF방법

의 경우는 92.4%였다. 성대 모델로 FL을 사용하였을 경우 IAIF방법을 적용한 경우 가장 높은 식별율을 보였으므로 향후 성대모델에 의해 식별할 경우 IAIF방법이 가장 유효한 분석방법으로 판단된다.

음원파라미터의 측정에 의한 성대질환의 식별율은 기존의 파라미터들에 의한 성대질환의 식별율과 유사하였다. 그러나 기존의 방법들에서는 주로 장애음성 전체와 정상음성간의 식별을 위주로 하였으므로 직접적인 비교는 불가능하다. 그럼에도 불구하고 제안한 방법의 경우 개별 장애음성과 정상음성간의 식별을 시도해 보았다는 점에서 의의가 있으며 서로 다른 향후 제시한 방법에 의해 장애음성 상호간의 식별에 대한 시도를 해 볼 필요가 있다고 생각된다.

참 고 문 헌

[1] Gobl, C. 1989. “A preliminary study of acoustic voice quality correlates.” *Speech Transmission Laboratory Quarterly Progress and Status Report*, 4, 9-22.

[2] Rosenberg, A. E. 1971. “Effect of glottal pulse shape on the quality of natural vowels.” *JASA*, 49, 583-590.

[3] Fujisaki, H. & Ljunqvist, M. 1986. “Proposal and Evaluation of Models for the Glottal Source Waveform”, *IEEE International Conference on Audio, Speech, and Signal Proceeding* Tokyo, Japan.

[4] Qi, Y. & Bi, N. 1994. “A Simplified Approximation of The Four-Parameter LF Model of Voice Source”, *JASA*, 96(2), 1182-1185.

[5] Chytil, P. & Pavel, M. 2006. “Estimation of Vocal Fold Characteristics using a Parametric Source Model”, *Eleventh Australasian International Conference on Speech Science and Technology*, Auckland, NewZealand.

[6] Alku, P. 1992. “Glottal wave analysis with pitch synchronous Interactive Adaptive Inverse Filtering”, *Speech Communication* 11, 109-118.

[7] Duda, R. O., Hart, P. E. & Stork, D. H. 2000. *Pattern Classification, 2nd ed.* NewYork, Wiley-Interscience.

[8] Kay-Pentax, 1994. *Disordered Voice Database*, ver 1.03.

[9] Changwon National University. 1999. *Korean Disordered Voice Database. ver 1.0.*, Changwon National University, Korea.

접수일자: 2008. 5. 18
게재결정: 2008. 6. 9

▲ Pavel Chytil
Biomedical Engineering Department
OHSU, 20000 NW Walker Rd
Beaverton, OR 97006 USA
Tel: +1-503-418-9314 (O)
Fax: +1-503-418-9349
E-mail: pchytill@bme.ogi.edu

▲ 조철우
경남 창원시 사림동9 번지
창원대학교 공과대학 메카트로닉스 공학부 (우: 641-773)
Tel: +82-55-213-3662 (O) 82-55-262-6174 (H)
Fax: +82-55-262-5064 M/P: 011-580-7552
E-mail: cwjo@changwon.ac.kr

▲ Misha Pavel
Biomedical Engineering Department
OHSU, 20000 NW Walker Rd
Beaverton, OR 97006 USA
Tel: +1-503-418-9314 (O)
Fax: +1-503-418-9349
E-mail: pavel@bme.ogi.edu