

청각장애아 및 건청아 음성으로부터 성도 면적 추정

김세환*, 권오욱

충북대학교 제어계측공학과

Vocal Tract Area Estimation from Deaf and Normal Children's Speech

Se-Hwan Kim*, Oh-Wook Kwon

Department of Control and Instrumentation Engineering, Chungbuk National University

{shkim,owkwon}@chungbuk.ac.kr

Abstract

This paper analyzes the vocal tract area estimation algorithm used as a part of a speech analysis program to help deaf children correct their pronunciations by comparing their vocal tract shape with normal children's. Assuming that a vocal tract is a concatenation of cylinder tubes with a different cross section, we compute the relative vocal tract area of each tube using the reflection coefficients obtained from linear predictive coding. Then, obtain the absolute vocal tract area by computing the height of lip opening with a formula modified for children's speech. Using the speech data for five Korean vowels (/a/, /e/, /i/, /o/, and /u/), we investigate the effects of the sampling frequency, frame size, and model order. We compare vocal tract shapes obtained from deaf and normal children's speech.

I. 서론

청각장애인은 자신이 발성하는 소리를 들을 수 없기 때문에 자신이 올바른 발음을 하고 있는지 알 수가 없어서, 보통사람과는 다른 발음을 익히게 되어 의사소통에 큰 어려움이 있다. 따라서 특수학교에서 배우는 것 이외에도 혼자서도 학습할 수 있는 청각장애인 발음훈련기기의 필요성을 느끼게 되었다.

청각장애인은 음성을 발성할 때 정상인과 다른 성도 모양을 나타내기 때문에 이점을 이용해 정상인과 흡사한 성도 형태를 갖추기 위한 목적으로 청각장애인 발음 훈련기기가 연구되었고, 표준 발음의 성도 형태와 훈련자가 발음한 성도 형태를 동시에 표시하여 교정이 필요한 부분을 시각적으로 나타낸다. 하지만, 기본 주파수가 일정하지 않은 장애인의 경우 성대와 코에 진

동센서를 부착하여 데이터를 추출하기 때문에 특별한 하드웨어 장치가 필요하였다[1][2]. 이와 같은 발음 훈련기기는 시각적인 강화가 청각장애인의 음성 발성 개선에 좋은 효과를 나타내는 연구[4]와 상통한다.

성도 면적 추정은 초기에 Ladefoged와 Harshman 등이 X-선을 이용하여 성도 모양을 추정하였고, 이후에 Wakita[6] 와 Deng[7][8][9] 등은 음성 신호로부터 성도 면적 함수를 계산하여 인간의 성도 모양을 나타내었다.

본 논문에서는 청각장애인의 음성으로부터 성도 형태를 추정하여 청각장애인의 발음 교정을 보조하는 프로그램의 핵심 모듈인 성도 면적을 추정 하여, 원도우 크기, 모델 차수 등에 대한 영향을 분석한다. 언어의 교정에 따른 조기 교육의 중요성으로 청각장애 어린이를 대상으로서, 기존 성인에 대한 수식을 한국인 청각장애 어린이에 맞게 수정하였다.

II. 성도 면적 추정

1. Burg 알고리즘

Burg 알고리즘은 autoregressive (AR) parameters를 추출하기 위하여 Levinson recursion과 lattice 구조를 기초로 하여 반복적인 방법으로 반사계수(reflection coefficient)를 구하는 이론이다. 식을 유도하기 위해서, $s(n)$, $n=0,1,\dots,N-1$ 이고, $s(n)$ 을 음성신호라고 정의 하면, 차수 m 의 전향신호 및 후향신호는 (1)과 (2)로 주어진다.

$$\hat{s}(n) = -\sum_{k=1}^m a_m(k)s(n-k) \quad (1)$$

$$\hat{s}(n-m) = -\sum_{k=1}^m a_m^*(k)s(n+k-m) \quad (2)$$

그리고, 전향과 후향신호 에러 $f_m(n)$ 과 $g_m(n)$ 은 (3)과

(4)와 같이 정의 된다.

$$f_m(n) = s(n) - \hat{s}(n) \quad (3)$$

$$g_m(n) = s(n-m) - \hat{s}(n-m) \quad (4)$$

여기서, $a_m(k)$ 는 예측 계수이고, $0 \leq k \leq m-1$ 과 $m=1,2,\dots,M$ 으로 주어진다. Least square error는 (5)와 같다.

$$\varepsilon_m = \sum_{n=m}^{N-1} \left[|f_m(n)|^2 + |g_m(n)|^2 \right] \quad (5)$$

이 error는 예측계수를 선택함으로 최소화 하고, Levinson-Durbin recursion을 (6)과 같이 주어진 전제로 한다.

$$a_m(k) = a_{m-1}(k) + K_m a_{m-1}^*(m-k), \quad 1 \leq k \leq m-1 \quad (6)$$

$$1 \leq m \leq M$$

$K_m = a_m(m)$ 은 전조가 되어 실현된 lattice filter에 m 차 반사계수 이다. (6)은 $f_m(n)$ 과 $g_m(n)$ 을 치환한 것이고, 그 결과는 (7)에서 부터 주어진 전향과 후향 예측에러를 한 쌍의 order-recursive equation으로 나타난다.

$$f_m(n) = f_{m-1}(n) + K_m g_{m-1}(n-1) \\ g_m(n) = K_m^* f_{m-1}(n) + g_{m-1}(n-1), \quad (7)$$

$$m = 1, 2, \dots, M$$

만약 (6)을 (7)식으로 치환하고, 반사계수 K_m 에 관하여 ε_m 의 최소화를 행하여 실행한다면, 식(8)과 같이 반사계수를 얻을 수 있다.

$$K_m = \frac{-\sum_{n=m}^{N-1} f_{m-1}(n) g_{m-1}^*(n-1)}{\frac{1}{2} \sum_{n=m}^{N-1} \left[|f_{m-1}(n)|^2 + |g_{m-1}(n-1)|^2 \right]}, \quad (8)$$

$$m = 1, 2, \dots, M$$

2. 성도의 음향 튜브 모델

음성 신호로부터 성도의 단면적을 추정하기 위해서는 성도를 구간마다 면적이 다른 음향 튜브 모델을 연속적으로 나열한 비균일 모델을 이용하며, 그림 1과 같다. $U_m^+(t, d_m)$ 은 구간 m 상태의 성도에서 입술로 진행하는 음파이고, $U_m^-(t, d_m)$ 은 그 반대방향으로 진행하는 음파의 체적 속도(volume velocity)이다. d_m 은 성문에서 구간 m 과 $m+1$ 의 접점까지의 거리이다.

역 필터와 음향 튜브가 일치하려면 샘플링 주파수와 차수 M 과 $F_s = Mc/2l$ 의 조건은 만족해야한다. 여기서, $M=l/\Delta l$ 이고, c 는 음속, l 은 전체성도 길이, Δl 은 한 구간의 길이를 나타낸다.

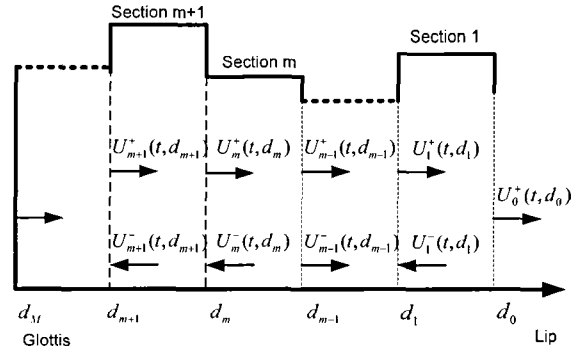


그림 1 성도 튜브 모델.

성도 모델의 마지막 단인 성도에서 면적을 $S_{M+1}=1$ 로 경계조건이 주어질 때, 성도에서 입술까지의 면적을 추출된 반사 계수로 (9)와 같이 성도 면적을 구할 수 있다[6].

$$S_m = S_{m+1} \frac{1 + K_m}{1 - K_m} \quad (9)$$

여기서, $m=M, M-1, \dots, 1$ 이고, K_m 은 반사계수, S_m 은 구간 m 에서의 성도 면적이다.

III. 실험 결과

1. 음성 데이터베이스

초등학교와 중학생으로 구성된 10명의 청각장애아동과 10명의 건청아동을 남녀 각 5명을 대상으로 조용한 사무실에서 녹음하였다. 순음 청력이 90dB 이상인 장애아동이 보청기 착용 상태에서 단모음 /아/, /에/, /이/, /오/, /우/ 5개와 장애아 발음평가에 이용되는 40단어[4]를 발성하였다. 비교 평가를 위하여 건청아동도 같은 내용을 발성하였다.

녹음은 Shure사의 SM12A 헤드셋 마이크를 사용하여, 샘플링 주파수 16kHz와 16비트 모노로 녹음하였다.



그림 2 성도 면적 추정 블록도.

2. 아동 성도면적 추정

마이크로폰으로 입력된 음성 신호는 계수를 1.0으로 설정하여 preemphasis와 hamming 윈도우를 거쳐 Burg 알고리즘으로 반사계수를 구하고, (9)를 이용하여 성도면적 비를 구한다. 입술 사이의 거리를 구하기 위한 포트 주파수들의 관계식을 C_n 으로 정의 하였고 [10], 영어 모음의 포트 주파수와 한국어 모음의 포

먼트 주파수의 차이를 고려해 실험에 의하여 결정하였다[3].

본 논문은 초중학생의 청각장애아동 및 건청아동을 대상으로 음성을 녹음하였으므로, (10)과 같이 성인과 어린이의 성도 길이에 따른 포먼트 위치의 차이를 비교 분석해 새로운 계수를 적용하였다.

$$X_1 = C_1/1.3 * F_2 + C_2/1.96 * F_2 * F_3 + C_3 * (1.3/1.4) * F_1/F_2 + (C_4 + 28)$$

$$C_1 = 0.300 * 10^{-3}$$

$$C_2 = -0.343 * 10^{-6} \quad (10)$$

$$C_3 = 4.143$$

$$C_4 = -0.174$$

각 구간에서 면적과 (10)에서 구한 입술 사이의 거리를 이용하여, 입술에서 성대까지의 면적과 거리의 비례식을 (11)와 같이 정의한다. 각 구간의 성도 면적을 성도에서 입술까지 구하였고, 역으로 입술에서부터 성도까지 각 구간의 면적과 입술 사이의 첫 번째 거리 X_1 를 이용하여, 대응하는 각 구간을 (11)과 같이 비례식을 적용함으로써 모든 구간의 X_m 을 구할 수 있다. 여기서, $m=M, M-1, \dots, 1$ 이다.

$$X_m = S_m * X_{m+1} / S_{m+1} \quad (11)$$

3. 정상아동 음성의 성도면적 추정 성능

다양한 샘플링 주파수, 윈도우 크기, 차수에 대하여 성도 면적 추정 실험을 수행한 결과, 샘플링 주파수 10kHz, 윈도우 크기 20ms, 차수는 8정도가 가장 좋은 성도 면적 결과를 나타내었다.

그림 3, 4는 2명의 정상아동에 대한 한국어 모음 /a/와 /i/에 대해서 (a)는 입력 파형, (b)는 LPC 스펙트럼, (c)는 성도 면적비, (d)는 정규화한 성도 면적비를 나타낸다.

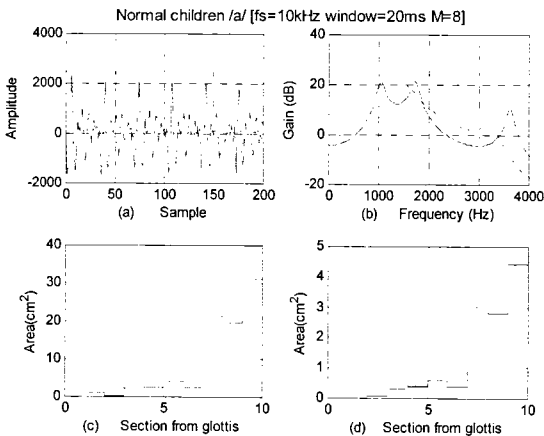


그림 3 정상아동 모음 /a/에서 (a)입력 파형, (b)LPC 스펙트럼, (c)성도 면적비, (d)정규화된 성도 면적.

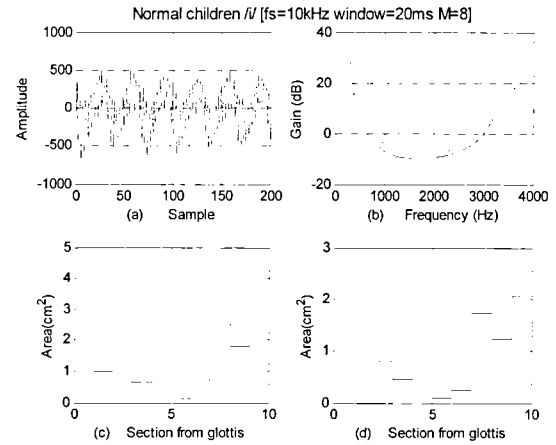
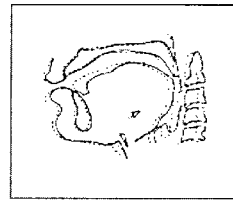
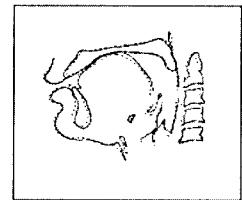


그림 4 정상아동 모음 /i/에서 (a)입력 파형, (b)LPC 스펙트럼, (c)성도 면적비, (d)정규화된 성도 면적.

그림 5, 6은 일본어 모음 /a/와 /i/를 MRI로 직접 촬영하여 그림으로 나타낸 연구 자료이고[11], 이와 같이 정상아동의 성도 면적 모양과 비슷한 형태를 확인할 수 있다.



/a/



/i/

그림 5 MRI사진 /a/.

그림 6 MRI사진 /i/.

4. 장애아동 음성의 성도면적 추정 성능

그림 7, 8은 청각 장애아동의 모음 /a/와 /i/로서 (a)는 입력 파형, (b)는 LPC 스펙트럼, (c)는 성도 면적비, (d)는 정규화한 성도 면적비를 나타낸다.

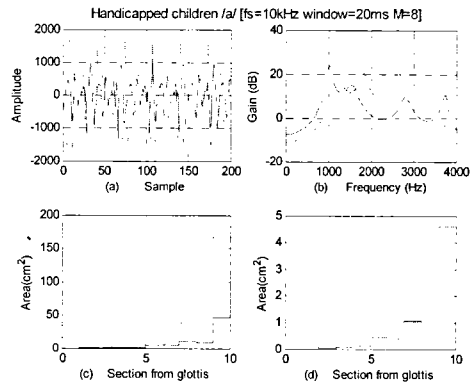


그림 7 장애아동 모음 /a/에서 (a)입력 파형, (b)LPC 스펙트럼, (c)성도 면적비, (d)정규화된 성도 면적.

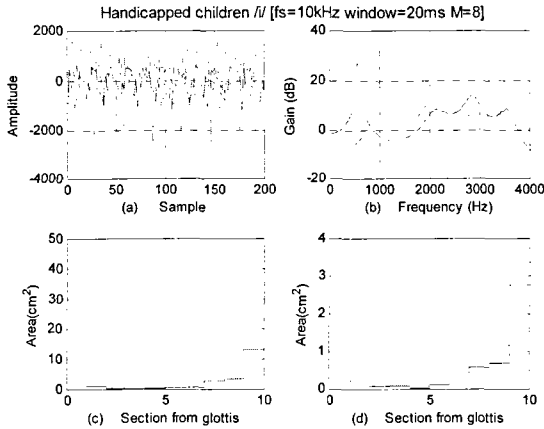


그림 8 장애아동 모음 /i/에서 (a)입력 파형, (b)LPC 스펙트럼, (c)성도 면적비, (d)정규화된 성도 면적.

정상아동의 성도 모델 특성은 비교적 잘 나타나어 졌지만, 장애아동의 경우는 분석결과 포먼트의 위치가 부정확해 상대적으로 면적의 수치가 현저하게 낮게 나타나서 성도 모델 추정에 어려움이 있었다. 특히, /i/의 발음은 (10)에 적용한 포먼트 계수를 수정할 필요가 있다.

IV. 결론

본 연구에서는 청각장애아동 발음 훈련기기를 개발하는 과정의 핵심 부분으로 정상아동과 청각장애아동의 성도를 모델링 하여 비교 분석하였고, 기존에 연구되었던 성인의 성도 모델 추출 과정에서 입술 높이에 따른 한국 어린이에 맞게 계수를 적용하고, 새로운 비례식을 사용하여 청각장애아동 및 정상아동의 성도를 모델링하였다. 음성 데이터베이스의 부족으로 추출된 성도 모델은 다소 특이점을 발견할 수 있지만, 기존에 연구된 X-ray나 MRI로 연구된[11] 조음기관 사진과 비교해 보면 비슷한 추세를 유지하고 있다. 특히, 청각장애아동의 경우는 부정확한 포먼트 추출과 새롭게 적용한 계수 값의 차이로 성도 모델링 하는데 어려움이 있었다.

이 연구를 더 보완 하여 음성만으로 정확한 성도 모델링으로 시각적으로 훈련할 수 있는 프로그램을 PDA 같은 휴대용 기기에 적용하여 공간적인 제약에서 벗어나서, 언제 어디서나 발음을 훈련할 수 있는 기기를 개발할 것이며, 그에 따라, 언어 습득의 조기교육에 맞추어 청각장애아동들의 발음 훈련도 이른 시기에 실시할 수 있어 언어 습득 효과를 기대한다.

감사의 글

“이 논문은 2005년도 교육인적자원부 지방연구중심대학 육성사업의 지원에 의하여 연구되었음.”

참고문헌

- [1] 김동준, 윤태성, 박상희, “청각 장애자용 발음 훈련 기기의 개발,” 한국자동제어학술회의논문집, Vol. 1 of 2, pp.746-751, 1991. 10.
- [2] 송제혁, 김동준, 박상희, “성도 면적 함수를 이용한 음성 인식에 관한 연구,” 대한의용생체공학학회논문집, Vol. 16, No. 3, pp.345-352, 1995. 9.
- [3] 박상희, 김동준, 이제혁, 윤태성, “조음도를 이용한 발음 훈련 기기의 개발,” 대한전기학회논문집, Vol. 41, No. 2, pp.209-216, 1992. 2.
- [4] 석동일, 조음 및 음운장애치료, 대구대학교출판부, 1999.
- [5] 여경연, 석동일, “시·청각적 강화가 청각장애 아동의 음성 및 발음 개선에 미치는 효과,” 언어치료연구, 제8권 제1호, 1999.11.
- [6] H. Wakita, “Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms,” IEEE Trans. Audio Electro-Acoustics, vol. 21, Oct. 1973.
- [7] H. Deng, M. Beddoes, R. Ward, “Estimating the vocal-tract area function and the derivative of the glottal wave from a speech signal,” Proc. EUROSPEECH, Geneva, Swiss, 2003.
- [8] H. Deng, R. K. Ward, M. P. Beddoes, M. Hodgson, “Estimating vocal-tract area functions from vowel sound signals over closed glottal phases,” Proc. ICASSP '04, pp. 1.589-1.592, Montreal, Canada, May. 2004.
- [9] H. Deng, R. K. Ward, M. P. Beddoes, M. Hodgson, “Effects of glottal and lip boundary conditions on vocal-tract area function estimates from speech signals,” proc. ICASSP '05, pp. 901 - 904, philadelphia, PA, March. 2005.
- [10] P. Ladefoged, R. Harshman, L. Goldstein, L. Rice, “Generating Vocal Tract Shapes from Formant Frequencies,” Acoustical Society of America, Oct. 1978.
- [11] T. Kitamura, H. Takemoto, K. Honda, “Difference in vocal tract shape between upright and supine postures: Observations by an open-type MRI scanner,” Acoust. Sci. & Tech., May, 2005.