

청각장애아동과 건청아동의 성도면적 추정 성능*

김세환(충북대), 김남(충북대), 권오욱(충북대)

<차 례>

- | | |
|-------------------|--------------------------|
| 1. 서론 | 3.1. 음성 데이터베이스 |
| 2. 성도면적 추정 | 3.2. 건청아동 음성의 성도면적 추정 성능 |
| 2.1. Burg 알고리즘 | 3.3. 장애아동 음성의 성도면적 추정 성능 |
| 2.2. 성도의 음향 튜브 모델 | |
| 2.3. 아동 성도 면적 추정 | |
| 3. 실험결과 | 4. 결론 |

<Abstract>

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

Se-Hwan Kim, Nam Kim, Oh-Wook Kwon

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, we 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 on the estimated vocal tract shape. We compare the vocal tract shapes obtained from deaf and normal children's speech.

* Keywords: Vocal tract area, Linear predictive coding, Reflection coefficient.

* 이 논문은 2005년도 교육인적자원부 지방연구중심대학 육성사업의 지원에 의하여 연구되었음

1. 서론

인간의 언어활동 중에 가장 큰 비중을 차지하는 듣기는 언어 학습의 가장 기본 단계로써 언어에 대한 많은 경험을 통해 말하기 능력이 점차 생성된다. 듣기 기능인 청각은 소리 정보를 처리할 수 있는 가장 중요한 감각기능 중 하나이다. 하지만 청각장애인들은 언어 습득기 이전에 청각의 손상으로 의사소통을 가능하게 하는 언어의 습득, 발전 및 유지에 어려움을 겪게 된다. 청각장애인 자신의 말을 이해하려면 발성을 다시 청각으로 받아 지각을 하며 보완해야 한다. 이런 단계를 거치지 못해 자신이 올바른 발음을 하고 있는지 알 수가 없다. 결국 정상인과 다른 발음을 익히게 되어 부정확한 발음으로 의사소통에 큰 어려움이 있다. 이처럼 청각적 피드백 효과는 발음 명료도를 높이는데 아주 중요한 역할을 한다. 현대에는 이러한 효과를 청각장애인에게 보완하기 위해 많은 교육적 서비스가 제공되고 있다. 또한 공학 및 의학기술의 발달로 보청기나 인공와우 같은 다양한 의사소통에 도움이 되는 보조 장비가 개발되었으며 이 모든 것이 복합적으로 사용되고 있는 실정이다. 그렇지만 이런 장치와 특수학교에서 배우는 것 이외에 시각적인 피드백 효과를 이용하여 혼자서도 학습하고 자신의 발음을 체크할 수 있는 청각장애인 발음훈련기기가 필요하다.

청각장애인이 음성을 발성할 때 성도 모양은 건청인과 다른 모양으로 나타나는데, 이 특징으로 건청인과 흡사한 성도 형태를 갖추기 위해 조음기관을 교정해 나가는 청각장애인 발음 훈련기기가 연구되었다. 정상 발음의 성도 모양과 훈련자가 발음한 성도 형태를 동시에 표시하여 교정이 필요한 부분을 시각적으로 나타낸 것이다. 하지만 기본 주파수가 일정하지 않은 장애인의 경우 성대와 코에 진동 센서를 부착하여 데이터를 추출하기 때문에 특별한 하드웨어 장치가 필요하였다 [1][2][3]. 이와 같은 발음 훈련기기는 시각적 피드백으로 청각장애인의 발성 개선에 좋은 효과를 나타내는 연구[4]와 상통한다.

성도 면적 추정은 초기에 Ladefoged와 Harshmane 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 발음 교정을 시각적 피드백 효과를 이용하기 위하여 성도 면적을 추정 하고, 윈도우 크기, 모델 차수 등의 영향을 분석하였다. 성도 모델링은 기존 성인에 대한 수식을 한국 초중학생에 맞게 수정하였고, 추정된 건청아동과 청각장애아동의 성도 모양으로 높낮이 차이를 비교하여 조음기관에서 입술 높이 및 혀의 위치 등의 교정부위를 알아낸다.

본 논문의 구성은 2장에서 Burg 알고리즘으로 반사계수와 성도 면적을 구하는 방법을 소개하고, 3장에서 건청아동 및 청각장애아동의 음성 데이터베이스를 소개 하며, 아동의 성도면적을 추정하여 건청아동과 청각장애아동의 성도면적 결과를

비교 분석하며, 4장에서는 결론을 맺는다.

2. 성도 면적 추정

2.1. Burg 알고리즘

Burg 알고리즘은 autoregressive (AR) 파라미터를 추출하기 위해서 Levinson recursion과 lattice 구조[12]를 기초로 하여 반복적인 방법으로 반사계수(reflection coefficient)를 구하는 방법이다. 음성신호를 $s(n), n = 0, 1, \dots, N-1$ 이라고 정의하면, 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 오차 $f_m(n)$ 과 $g_m(n)$ 은 (3)과 (4)와 같이 정의된다.

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

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

여기서, $a_m(k)$ 는 m 차 필터의 k 번째 예측계수이고, $0 \leq k \leq m-1$ 과 $m = 1, 2, \dots, M$ 으로 주어지며, 그것의 Least square error는 (5)와 같이 정의한다.

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

이 오차는 예측계수를 선택함으로써 최소화되었고, (6)에서 주어진 Levinson-Durbin recursion을 필요로 하고 있다.

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

$K_m = a_m(m)$ 은 lattice filter의 m 차 반사계수이다. (6)은 $f_m(n)$ 과 $g_m(n)$ 을 치환한 것이고, 그 결과는 (7)에서부터 주어진 전향과 후향 예측에러를 한 쌍의 order-recursive equation으로 나타낸다.

$$\begin{aligned} f_m(n) &= f_{m-1}(n) + K_m g_{m-1}(n-1), & m = 1, 2, \dots, M \\ g_m(n) &= K_m^* f_{m-1}(n) + g_{m-1}(n-1), & m = 1, 2, \dots, M \end{aligned} \quad (7)$$

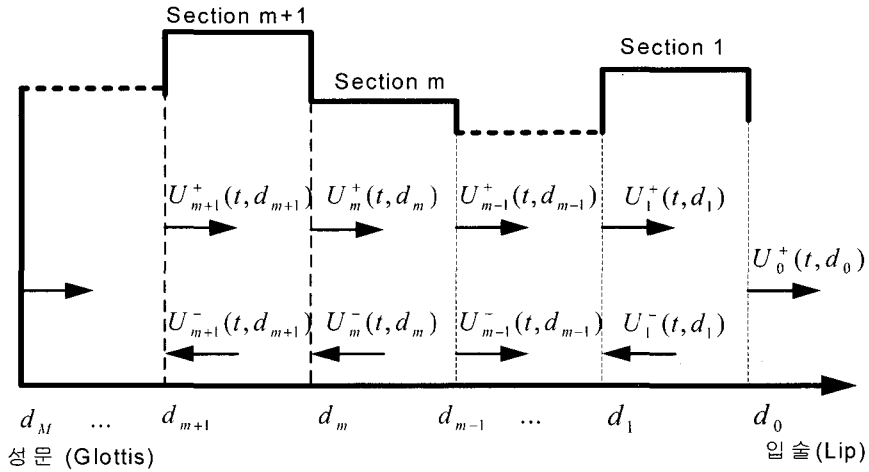
만약 (6)을 (7)식으로 치환하고, 반사계수 K_m 에 관하여 ε_m 에 대하여 최소화를 한다면, 식 (8)과 같은 반사계수를 얻을 수 있다. 여기서 *는 공액복소수를 의미한다.

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

2.2. 성도의 음향 튜브 모델

음성 신호로부터 성도의 단면적을 추정하기 위해서는 성도를 구간마다 면적이 다른 음향 튜브 모델을 연속적으로 나열한 비균일 모델을 이용하며, <그림 1>과 같다. $U_m^+(t, d_m)$ 은 구간 m 상태의 성문(glottis)에서 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 $F_s = M_c/2l$ 의 조건은 만족해야 한다. 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 성문에서 면적을 $S_{M+1} = 1$ 로 경계조건이 주어질 때, 추출된 반사계수로부터 성문에서 입술까지의 성도 면적을 (9)와 같이 구할 수 있다[6].

$$S_m = S_{m+1} \frac{1 + K_m}{1 - K_m}, \quad m = M, M - 1, \dots, 1 \tag{9}$$

K_m 은 반사계수, S_m 은 구간 m 에서의 성도 면적이다.

2.3. 아동 성도면적 추정

Burg 알고리즘으로 반사계수를 구하고, (9)를 이용하여 성도면적 비를 구한다. Ladefoged는 영어 모음의 입술 사이 거리를 구하기 위한 포먼트 주파수들의 관계식을 다음과 같이 정의하였다[10].

$$\begin{aligned} x_{M+1} &= C_1 F_2 + C_2 F_2 F_3 + C_3 (F_1 / F_2) + C_4 \\ C_1 &= 0.300 * 10^{-3} \\ C_2 &= -0.343 * 10^{-6} \\ C_3 &= 4.143 \\ C_4 &= -2.865 \end{aligned} \tag{10}$$

여기서 C_1, C_2, C_3, C_4 는 실험 데이터로부터 구한 값이다.

성인의 영어 모음을 초등학생 한국어 모음에 맞게 성인과 아동의 성도 길이에 따라 포먼트비의 차이 값을 고려하여 (11)과 같이 새로운 계수를 결정하였다. 포먼트비의 차이 값은 F1이 1.3배, F2는 1.4배, F3은 1.4배로 조사되었고, C_1 에 상수 값을 더하여 X_1 이 양수가 되도록 조정하였다.

$$X_1 = C_1 / 1.3 * F_2 + C_2 / 1.96 * F_2 * F_3 + C_3 * (1.3/1.4) * F_1 / F_2 + (C_4 + 2.8) \quad (11)$$

입술부터 성문까지 면적에 각각 대응하는 새로운 성도 면적비는 (9)에서 구한 면적비와 (11)의 입술 높이를 이용하여 (12)와 같이 비례식을 정의하였다.

$$X_m = S_m * X_{m-1} / S_{m-1}, \quad m = 2, 3, \dots, M \quad (12)$$

3. 실험 결과

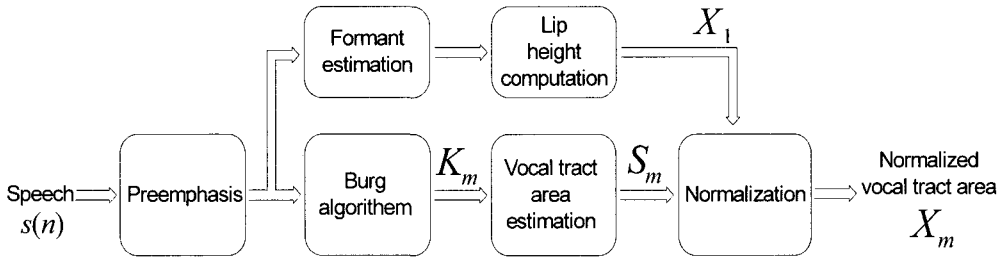
3.1. 음성 데이터베이스

음성 데이터베이스 대상자는 전주에 있는 특수학교에서 초등학생과 중학생으로 구성된 10명의 청각장애아동과 10명의 건청아동으로 구성되며, 남녀 각각 5명을 조용한 사무실에서 녹음하였다. 순음 청력이 90dB 이상인 장애아동이 보청기 착용 상태에서 단모음 /아/(a), /에/(e), /이/(i), /오/(o), /우/(u) 5개와 장애아 발음평가에 이용되는 40단어[4]를 발성하였다. 비교 평가를 위하여 건청아동도 같은 내용을 발성하였다. Shure사의 SM12A 헤드셋 마이크를 사용하여 녹음하였으며, 발성화자의 입에서 3cm 가량을 유지하였고, 샘플링 주파수 16kHz와 16비트 모노로 설정하였다.

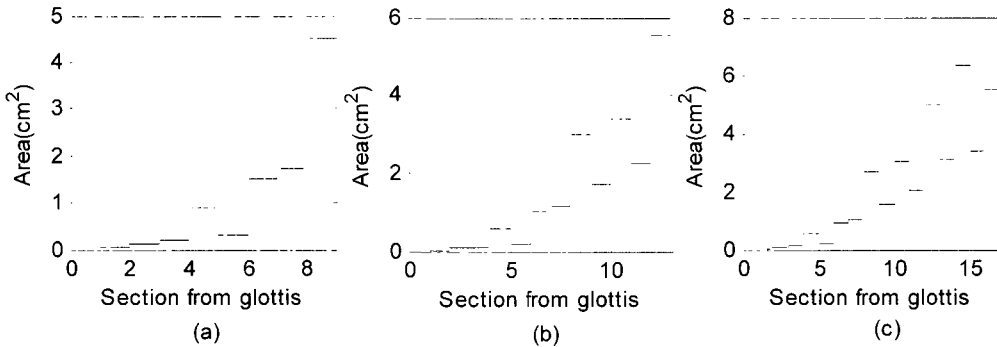
3.2. 건청아동 음성의 성도면적 추정 성능

<그림 2>와 같이 마이크로폰으로 입력된 음성 신호는 위너 필터와 계수값이 1.0인 프리엠퍼시스(preemphasis) 거친 후 해밍(Hamming) 윈도우를 씌웠고, Burg 알고리즘으로 반사계수를 구하여 성도 면적을 추정하게 된다. 다양한 샘플링 주파수, 윈도우 크기, 차수에 대해 성도 면적 추정 실험을 수행한 결과, 샘플링 주파수는 10kHz, 윈도우 크기는 20ms, 차수는 8정도가 가장 좋은 성도 모양을 나타내었다. 모델 차수를 8로 사용한 이유는 4kHz구간에 대개 4개의 포먼트가 위치하고 있는데 각 포먼트는 2개의 극점(pole)과 연관되기 때문이다. 그림 3과 같이 실험에

의한 최적의 차수를 확인할 수도 있고, 고차가 되는 경우에는 성도 면적 변화가 너무 세분화되어 급작스러운 피크(spurious peak)가 발생하는 경향이 있다.

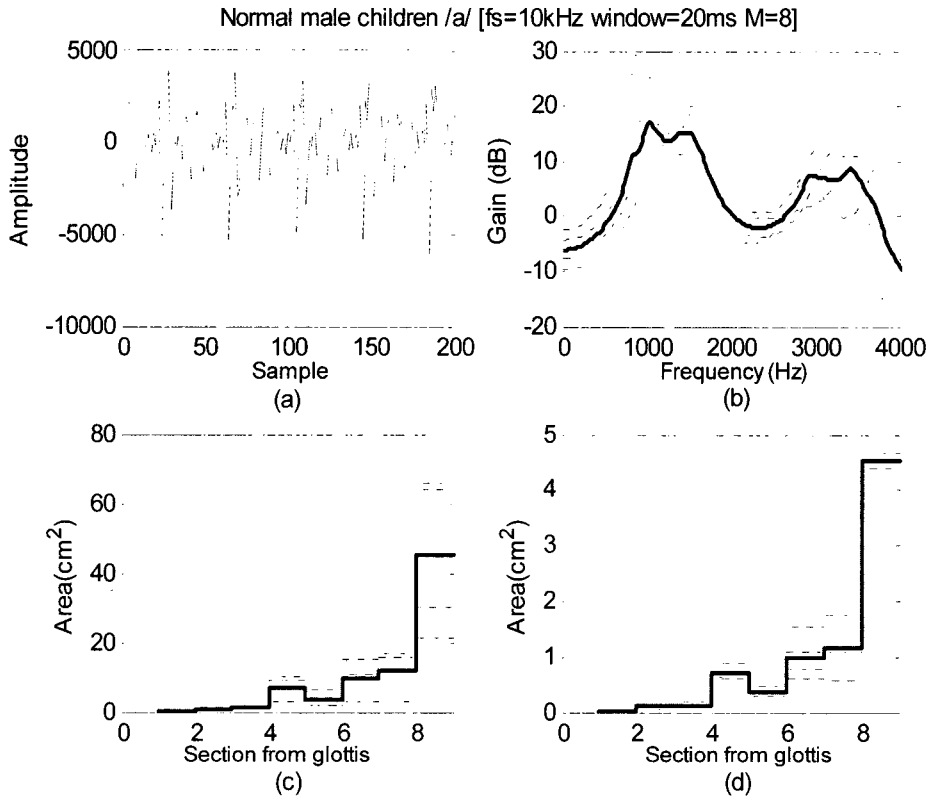


<그림 2> 성도 면적 추정 블록도.

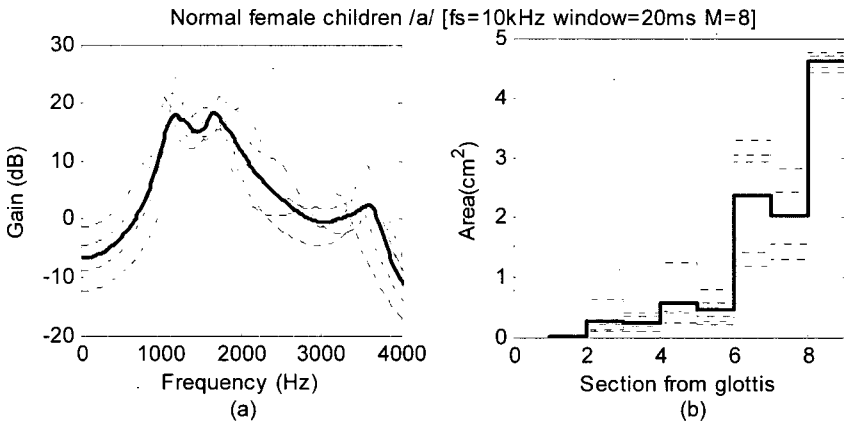


<그림 3> LPC 모델 차수에 따라 변하는 성도 면적, (a)모델차수 8차, (b)모델차수 12차, (c)모델차수 16차.

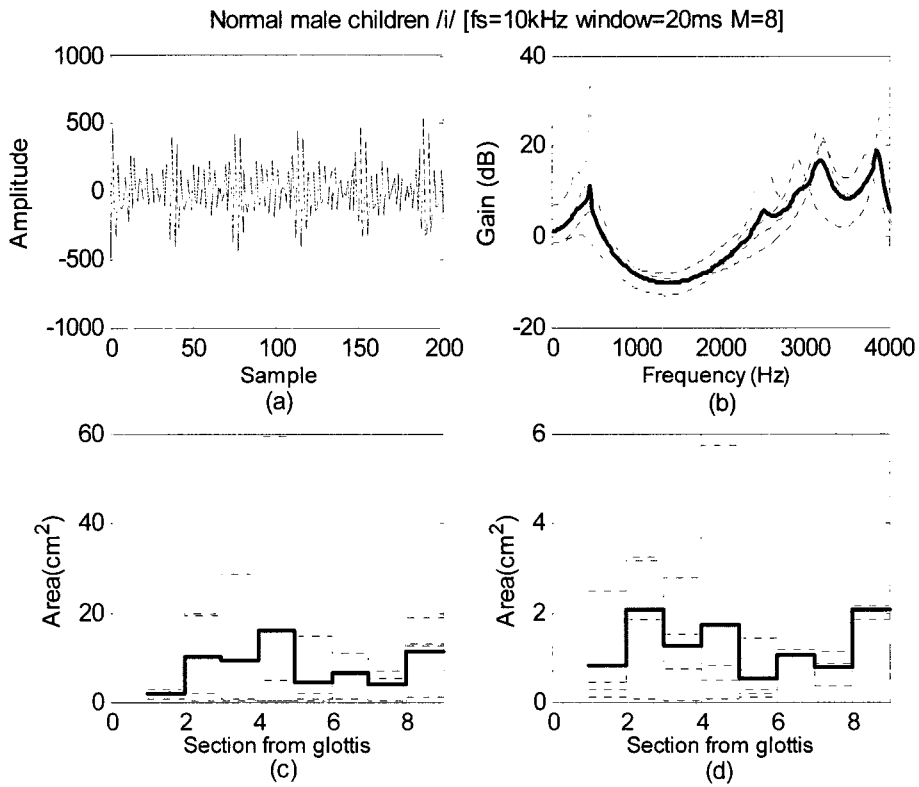
<그림 4>와 <그림 5>는 건청남자아동과 건청여자아동 각각 5명에 대한 한국어 모음 /아/에 대해 나타내었고, 점선 부분은 화자들의 출력 값이고 진한 실선은 평균값을 그린 것이다. (a)는 입력 파형, (b)는 linear predictive coding(LPC) 스펙트럼, (c)는 성도 면적비, (d)는 정규화된 성도 면적비를 나타낸다. 건청여자아동은 (a)는 LPC 스펙트럼, (b)는 정규화된 성도면적이다. <그림 6>과 <그림 7>은 건청아동의 /이/에 대해 성도 면적을 보여준다. 정규화된 성도 면적을 성도면적비와 비교해 보면 입술부분에서 두드러지게 정규화된 것을 알 수가 있다.



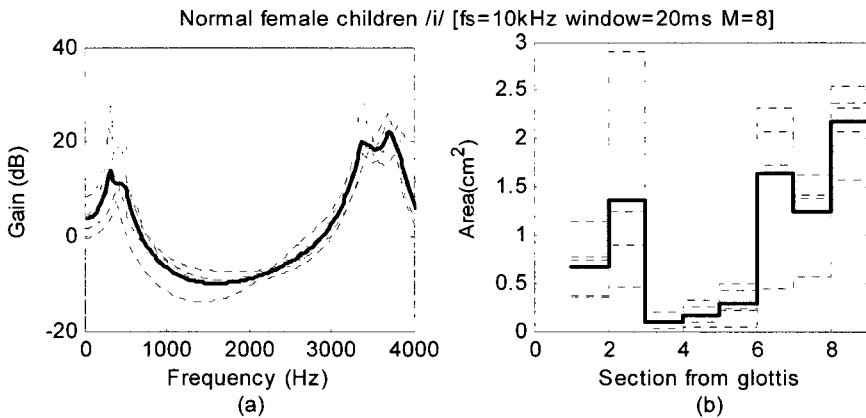
<그림 4> 건청남자아동의 모음 /아/에 대해 (a)입력 파형, (b)LPC 스펙트럼, (c)성도 면적비 (d)정규화된 성도면적.



<그림 5> 건청여자아동의 모음 /아/에 대해 (a)LPC 스펙트럼, (b)정규화된 성도면적.

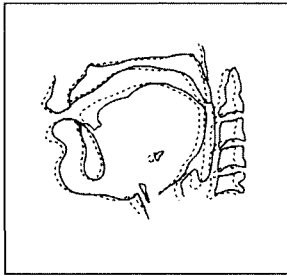


<그림 6> 건청남자아동의 모음 /이/에 대해 (a)입력 파형, (b)LPC 스펙트럼, (c)성도면적비, (d)정규화된 성도면적.



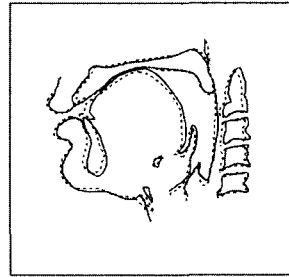
<그림 7> 건청여자아동의 모음 /이/에 대해 (a)LPC 스펙트럼, (b)정규화된 성도면적.

<그림 8>과 <그림 9>는 일본어 모음 /아/와 /이/를 자기공명영상(MRI)으로 직접 촬영한 것[11]으로서, 한국인과 같은 동양인으로 비슷한 조음기관을 나타내기 때문에 본 연구 결과와 비교 하기위해 첨부한 그림이다. 이와 같이 건청아동의 성도 면적 모양과 MRI로 추출한 조음기관 그림에서 성도 내에 공명구간의 면적비가 비슷한 추세로 되어 있는 것을 확인할 수 있다.



/a/

<그림 8> MRI 사진/a/.

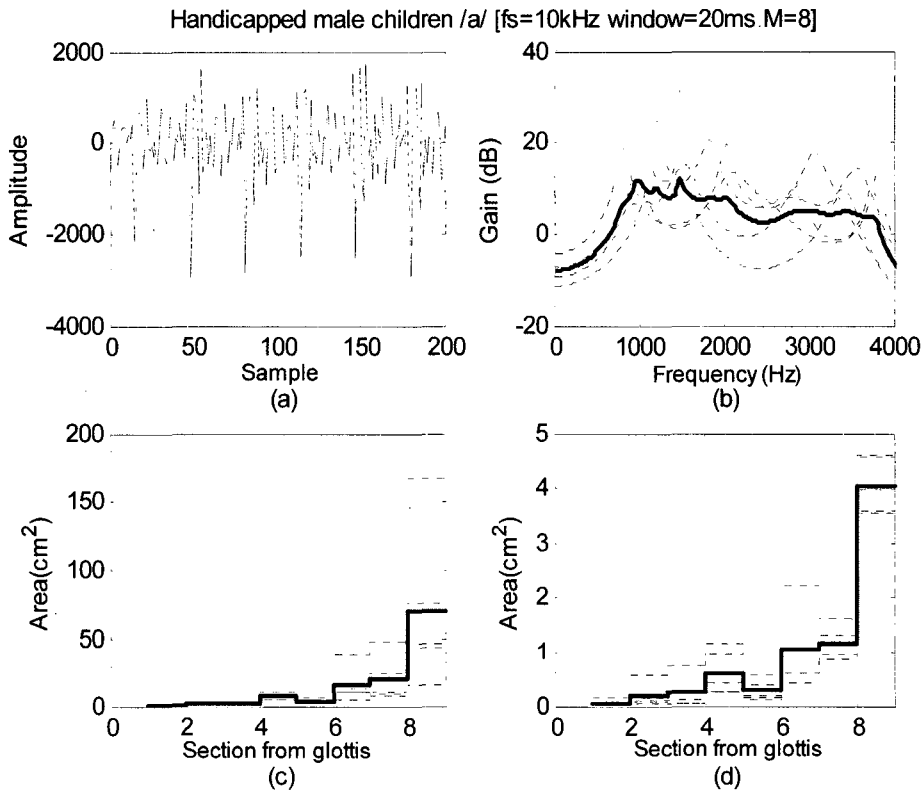


/i/

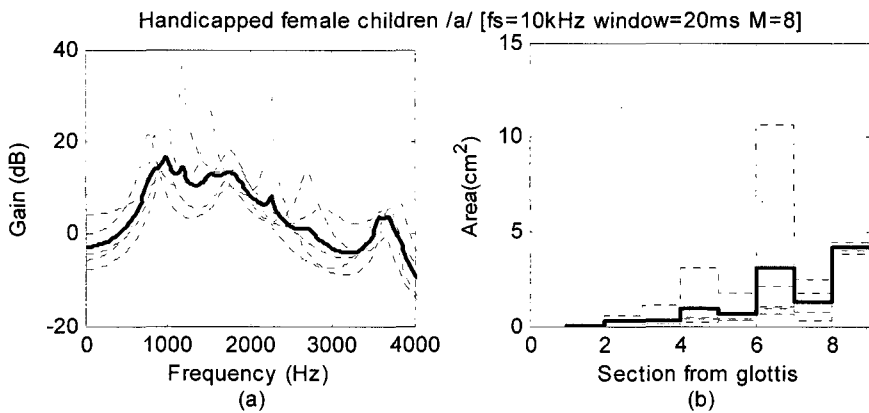
<그림 9> MRI 사진/i/.

3.3. 장애아동 음성의 성도면적 추정 성능

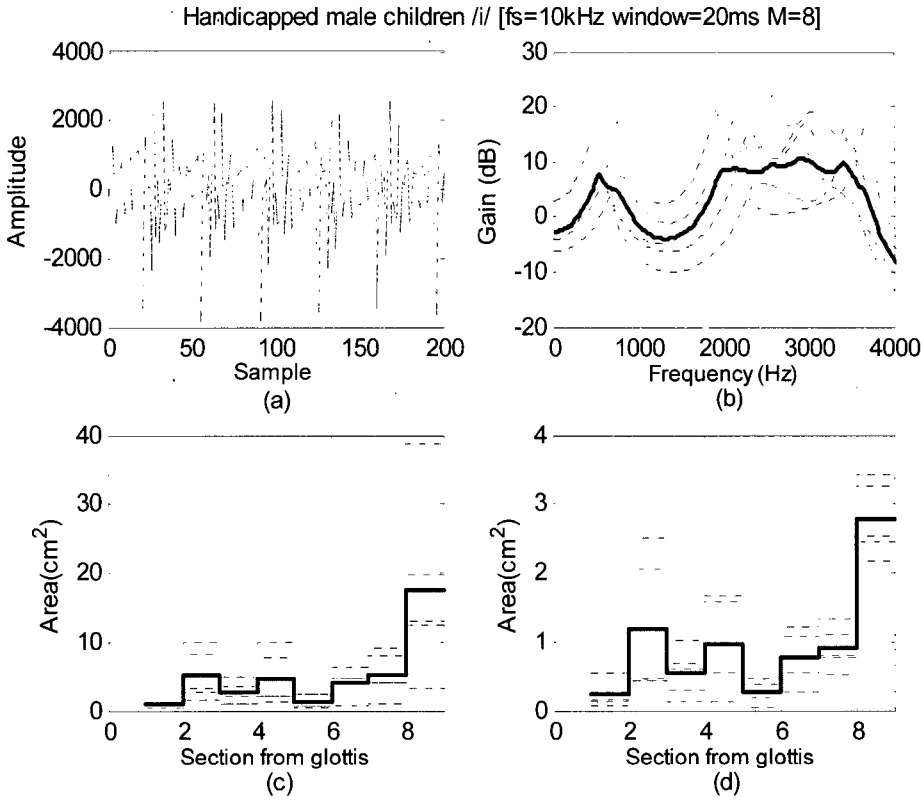
<그림 10>과 <그림 11>은 청각장애 남자아동과 청각장애 여자아동의 각각 5명의 모음 /아/로서 (a)는 입력 파형, (b)는 LPC 스펙트럼, (c)는 성도 면적비, (d)는 정규화한 성도 면적을 나타내고, <그림 12>와 <그림 13>은 청각장애 아동의 모음 /이/를 보여준다. 건청아동의 성도 모델 특성은 비교적 잘 나타내어졌지만, 장애아동의 경우는 분석결과 포먼트의 위치가 부정확해 상대적으로 성도 모델 추정에 어려움이 있었다. 청각장애아동의 성도 면적 추정은 기존 방법으로 얻어낸 성도 면적비를 어린아이에 맞게 계수값을 적용 후 정규화하여 성도 면적 모양을 나타내었다. 추정된 성도 면적을 건청아동의 성도 면적과 비교되는 부분, 즉 조음기관의 교정이 필요한 위치를 모음 /아/와 /이/에서 알 수 있다. /아/의 성도 모양에서는 혀 앞부분이 위쪽에 위치된 것으로 공명이 부족 한 것으로 혀 앞부분을 낮춰서 발음해야 하는 것으로 나타났고, /이/의 모양에서는 혀 뒤쪽에 있는 인두 부분의 면적이 너무 좁게 발음되었다는 것을 알 수 있다. 이와 같이 추출된 성도 면적비의 차이점으로 교정이 필요한 조음기관의 위치를 인지할 수 있다.



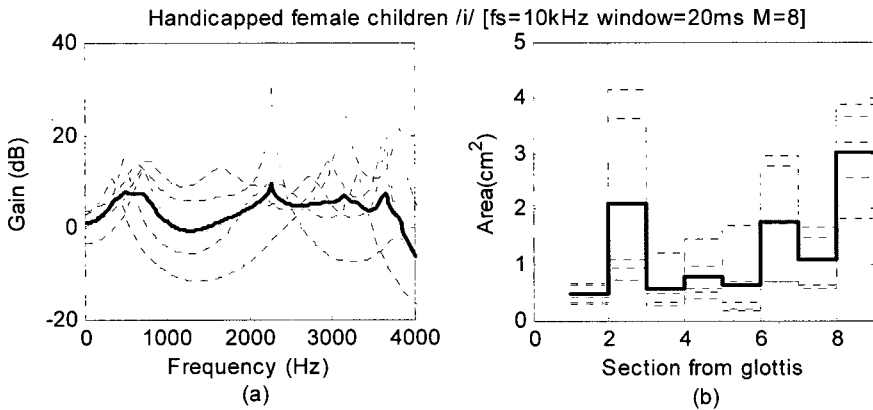
<그림 10> 청각장애 남자아동의 모음 /아/에 대해 (a)입력 파형, (b)LPC 스펙트럼, (c)성도면적비, (d)정규화된 성도면적.



<그림 11> 청각장애 여자아동의 모음 /아/에 대해 (a)LPC 스펙트럼, (b)정규화된 성도면적



<그림 12> 청각장애 남자아동의 모음 /i/에 대해 (a)입력 파형, (b)LPC 스펙트럼, (c)성도 면적비, (d)정규화된 성도면적.



<그림 13> 청각장애 여자아동의 모음 /i/에 대해 (a)LPC 스펙트럼, (b)정규화된 성도면적

4. 결 론

본 연구에서는 청각장애아동 발음 훈련기기를 개발하는 과정의 핵심 부분으로서 건청아동과 청각장애아동의 성도를 모델링하여 비교 분석하였고, 기존에 연구되었던 성인의 성도 모델 추출 과정에서 적용되던 입술 높이 추정식을 한국 어린이에 맞도록 새로운 비례식으로 정규화하여 청각장애아동 및 건청아동의 성도를 모델링하여 교정이 필요한 조음기관의 위치를 알아내었다. 추출된 성도 모델은 세밀하게는 차이가 있지만, 기존에 X-ray나 MRI로 연구된 조음기관 사진과 비교해보면 비슷한 추세를 유지하고 있다. 실험 결과에 의하면, 청각장애아동의 /아/는 건청아동과 비슷한 모양으로 출력되었지만 혀끝부분을 약간 낮춰서 발음해야 하는 것으로 나타났고, /이/의 경우는 혀의 뒤쪽부분에 면적이 좁게 나타나서 건청아동과는 다르게 발음된 것을 알게 되었다. 이와 같은 청각장애아동의 추출된 성도면적 결과를 건청아동의 성도 면적과 비교하여 교정이 필요한 조음기관의 위치를 알 수 있었다.

이러한 성도 면적 추출 시스템을 위해서 보완되어야 할 점은 청각장애아동의 기본주파수를 보다 정확하게 추출하기 위한 방법이 필요하며, 더 많은 음성데이터베이스로 신뢰할 수 있는 청각장애아동의 성도 모델 특성을 특징지어야 할 것이다. 앞으로 PDA같은 휴대용기기에 음성만으로 정확한 성도를 모델링하여 시각적으로 훈련할 수 있는 기기를 개발하여 공간적인 제약에서 벗어나 언제 어디서나 발음을 훈련할 수 있는 프로그램을 적용하고, 더 나아가 청각장애인과 일반인이 간단한 의사소통을 할 수 있는 음성인식 프로그램을 구현할 예정이다. 이와 같은 프로그램은 특수학교 및 여러 교육시스템과 연동하여 청각장애아동들이 언어 습득하는데 보조 역할로서 많은 도움을 줄 것으로 기대한다.

참 고 문 헌

- [1] 김동준, 윤태성, 박상희, "청각 장애자용 발음 훈련 기기의 개발," 한국자동제어학술회의 논문집, Vol. 1 of 2, pp. 746-751, 1991.
- [2] 송제혁, 김동준, 박상희, "성도 면적 함수를 이용한 음성 인식에 관한 연구," 대한의용생체공학회논문집, Vol. 16, No. 3, pp. 345-352, 1995.
- [3] 박상희, 김동준, 이제혁, 윤태성, "조음도를 이용한 발음 훈련 기기의 개발," 대한전기학회논문집, Vol. 41, No. 2, pp. 209-216, 1992.
- [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.
- [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 et al., "Estimating the vocal-tract area function and the derivative of the glottal wave from a speech signal," *Proc. EUROSPEECH*, Geneva, Switzerland, 2003.
- [8] H. Deng, R. K. Ward et al., "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. et al., "Effects of glottal and lip boundary conditions on vocal-tract area function estimates from speech signals," *Proc. ICASSP '05*, pp. 901-904, Philadelphia, PA, Mar. 2005.
- [10] P. Ladefoged, R. Harshman, et al., "Generating vocal tract shapes from formant frequencies," *Journal of 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.
- [12] J. Makhoul, "Stable and efficient lattice methods for linear prediction," *IEEE Trans. Acoustics, Speech, and Signal Processing*, Vol. 5, pp. 423-428, Oct. 1977.

접수일자: 2005년 11월 15일

게재결정: 2005년 12월 20일

▶ 김세환(Se-Hwan Kim)

주소: 361-763 충북 청주시 흥덕구 개신동 12번지

소속: 충북대학교 제어계측공학과

전화: (043) 261-3374

E-mail: shkim@chungbuk.ac.kr

▶ 김남(Nam Kim)

주소: 361-763 충북 청주시 흥덕구 개신동 12번지

소속: 충북대학교 전기전자컴퓨터공학부

전화: (043) 261-2482

E-mail: nkim@chungbuk.ac.kr

▶ 권오욱(Oh-Wook Kwon) : 교신저자

주소: 361-763 충북 청주시 흥덕구 개신동 12번지

소속: 충북대학교 전기전자컴퓨터공학부

전화: (043) 261-3374

E-mail: owkwon@chungbuk.ac.kr