

# 청각장애자용 발음훈련기기 개발에 관한 연구

안 상필, 이 재혁, \* 윤 태성, 박 상희  
연세대학교 전기공학과, \* 창원대학교 전기공학과

## A study on speech training aids for Deafs

Sang-Pil Ahn, Jae-Hyuk Lee, \*Tae-Sung Yoon, Sang-Hui Park  
Dept. of Electrical Engineering, Yonsei University  
\* Dept. of Electrical Engineering, Changwon University

### ABSTRACT

Deafs cannot speak straight voice as normal people in lack of feedback of their pronunciation, therefore speech training is required.

In this study, fundamental frequency, intensity, formant frequencies, vocal tract graphic and vocal tract area function, extracted from speech signal, are used as feature parameter. AR model, whose coefficients are extracted using inverse filtering, is used as speech generation model. In connection between vocal tract graphic and speech parameter, articulation distances and articulation distance functions in selected 15-intervals are determined by extracted vocal tract areas and formant frequencies.

### 1. 서 론

청각장애인들은 자신의 발음에 대한 feedback의 결여로 정상인들처럼 발음을 스스로 조절하면서 올바른 음성을 생성시킬 수 없다. 따라서 특수교육을 통해, 즉 교사의 입모양을 보거나 조음기관의 움직임을 촉각으로 감지하거나 발음교정기기를 통하여 발음을 훈련받게 된다.

인간의 발음생성과정을 음성신호로부터 역으로 추정하여 조음도의 형태로 표현하는 것은 음성훈련에서 매우 중요하다. 현재까지 외국에서 개발된 장애자용 음성훈련기기들은 음성의 단일파라미터(F0, /S/음, nasality)만을 측정할 수 있거나 또는 음성신호를 주파수분석하여 막대그래프 또는 color로 표시하는 형태였다. 음성의 단일파라미터 측정 형태의 훈련기기는 성대진동이나 비강의 개폐 또는 /S/ 등의 특징음에 대해서만 훈련이 가능하여 일반음성의 발음훈련에는 쓸 수 없는 단점이 있으며, 음성의 주파수별 막대그래프 표현형태의 훈련기기의 경우에는 화면에 나타난 주파수별 막대성분만 보고서는 훈련당사자가 자신의 조음기관중 어느 곳을 어떤 정도로 교정해야 하는지 알기 어려우므로 훈련효과가 많지 않았다. 따라서 발음기관의 단면도를 나타내고 그 위에 표준발음의 조음형태와 훈련자가 발음한 조음형태를 동시에 표시해 교정이 필요한 부위를 드러나게 하는 조음도그래프의 개발은 장애인의 발음훈련에 획기적인 진전을 가져오게 될 것이다.

### 2. 음성 분석 및 특징매개변수 추출 기술

음성신호를 분석하여 추출하는 특징매개변수로서 F0(기본주파수), Intensity, 형성음주파수(formant frequency), 그리고 조음도 생성 및 조음거리계산을 위한 성도면적함수 등을 선정하였다.

### 가. F0(기본주파수)

F0는 가장 기본적인 특징매개변수로서 허파로부터 발생한 공기의 흐름이 성대를 거치면서 Vocal Cord의 주기적인 진동에 의해 형성된 기본주파수이다. F0의 양에 의해 유·무성음이 구분되는데, 유성음은 Vocal Cord의 진동에 의해 생기는 소리고, 무성음은 Vocal Cord가 열린 채로 진동없이 공기를 통과시켜 결과적으로 F0가 정의되지 않은 소리이다.

자신의 발음에 대한 feedback이 결여되어 있는 청각장애자들의 경우, F0를 제어할 수 있는 조절능력이 현저히 저하되어 F0가 정상발음 보다 높은 경우가 많으며 무엇보다도 상당히 불규칙적이다.

### 나. Intensity

장애인들의 발음에 있어 또 하나의 심각한 오류는 Loudness의 조절이다. 장애인 발음의 경우 Loudness는 F0와 밀접한 연관을 갖게 되는데, F0를 조절하려는 시도가 Loudness에 영향을 주어 Loudness가 지나치게 크게 되거나 아주 불규칙해져 정상적인 음성과는 다르게 된다.

본 연구에서는 원래 음향정신적 개념인 Loudness에 대응되는 음향학적인 변수로서 Intensity를 선정하였다. Intensity는 음성에너지의 제곱에 비례하는 양으로서 F0와 마찬가지로 소구간을 택하여 시간합수인 객적형태를 되계한다.

### 다. 형성음주파수

형성음주파수는 성문을 거친 공기의 흐름이 조음체(혀, 입술, 턱)의 형태에 따라 성도내에서 발생하는 공진의 공진주파수이다. 17cm의 길이를 갖는 성도의 경우, 3KHz 내의 주파수 범위에서 3~4개, 그리고 5 KHz 이내의 주파수 범위에 4~5개의 형성음주파수를 갖는다고 알려져 있다. 또한 연속적인 발음에서 조음체들의 위치변화에 따라 형성음주파수도 시간에 따라 변화한다. 형성음주파수를 추출하는데 있어서, LPC 분석방식은 종종 서로 상이한 대역폭을 갖고있어서 위치가 인접한 두 형성음주파수가 하나로 중첩되어 나타나는 심각한 문제점이 있다. 따라서 정상적인 방법의 peak-picking을 취하면 주요형성음중 일부가 유실될 수 있다. 이러한 단점을 보완하고 원래의 형성음주파수를 찾기위해 1차 미분 및 2차미분을 통한 peak-picking방법을 이용하였다.

우선 음성발생 모델을 하나의 R-L-C 공진필터로 단순화시키면 그 전달함수의 크기는

$$H = 1 / [(-Fo^2 + F^2 + B^2)^2 + 4 F^2 B^2]$$

이며, Fo는 공진주파수, F는 입력주파수, B는 -3dB 대역폭이다.

이에 log 함수를 취해 1차 및 2차미분을 취하면, log magnitude  $L_i$ 가 최대가 되는 조건은  $dL_i/dF=0$  이며,  $dL_i^2/dF^2$ 가 극한치를 갖는 때는 본모가 최소를 때로서 두 경우 모두  $F^2=F_0^2-B^2$ 을 갖으므로 2차미분의 minima는 1차미분의 maxima와 같은 위치에서 일어난다. 또한 log spectrum에서 중첩되어 언덕형태로 나타나는 두 형성음 주파수는 각기 negative minima로 나타나게 되어 원래의 형성음주파수의 위치를 찾아낼 수 있게 된다.

실제 spectrum sequence  $A_n$ 에 대한 1,2차미분식은 다음과 같다.

$$\begin{aligned} D_n' &= A_{n-1} - A_n \\ D_n'' &= A_{n+1} + A_{n-1} - 2A_n \\ B_n'' &= B_0'' (D_n'' / |B_0''|) \end{aligned}$$

모음 /오/에 대한 log spectrum 및 1,2차미분된 spectrum은 그림 2-1과 같다.

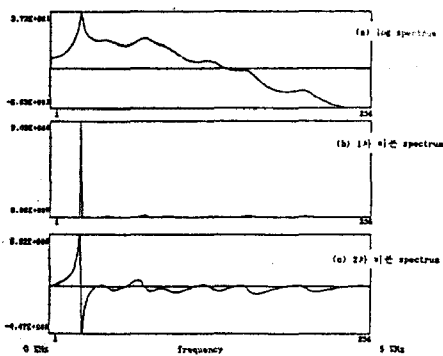


그림 2-1 모음 /오/에 대한 log 1,2차 spectrum

본 연구에서는 AR모델의 계수가 15차인 점에 근거하여 순서대로 그리고 크기별로 5개의 peak를 picking하여 각기 F1, F2, F3, F4, F5를 할당하는 방식을 택하였다. 그림 2-2는 이와 같은 방식으로 F1, F2, F3의 시간적 변화를 추적한 것이다. 그림 2-3은 형성음주파수 추출알고리즘의 블록선도이다.

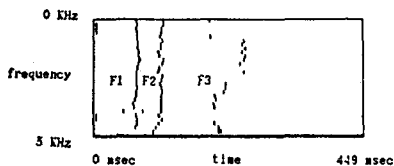


그림 2-2 /아/ F1, F2, F3 Spectrogram

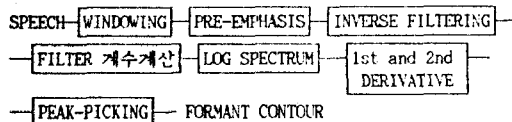


그림 2-3 형성음주파수 추출알고리즘 블록선도

### 3. 조음도그래픽의 설계와 개발

조음도는 발생할 때 각 조음기관의 모습을 vocal cord에서 입술 또는 코까지, 즉 성도(vocal tract)를 쫓아 단면도의 형태로 나타낸 것이다. 음성신호로부터 변별적 파라미터들을 추출하여 각 조음기관의 조음형태를 그래픽으로 복원하기 위해서는 우선 각 조음기관의 운동법위를 정의해야 하는데 이론적으로는 vocal tract의 아무 곳에서도 공기의 흐름을 폐쇄하거나 방해할 수 있는 무한한 가능성이 있지만 인간의 신경과 근육조직이 아주 세밀한 차이를 늘 구별해서

행동할 수 있게 할 수 없을 뿐만 아니라 또 허가 1mm의 조음위치의 차이를 착오없이 수행할 수 있다하더라도 청각적으로는 아무런 차이를 나타내 주지 않으므로 실제로 언어발음에서 쓰이는 조음위치는 제한된다.

조음기관의 운동을 표시하기 위해 다음과 같은 원칙에 입각하여 운동축을 설정하였다.

- 1) 최단거리 법칙에 따른다.
- 2) 성도를 15개 등간격으로 나눈다.
- 3) X선으로 얻은 조음단면도의 데이터에 일치시킨다.
- 4) 윗 턱에 설정된 조음점은 고정되어 있고 아래 턱에 연결된 조음체는 가변된다.

본 연구에서는 기본모델로 삼은 15개 등간격 pipe의 vocal tract model은 성인남자의 17cm vocal tract를 기본전제로 한 것이므로 여기서도 vocal tract전체 길이를 17cm로 잡고 각 section을 구분하였다.

X선 data는 Harshman 등의 PARAFAC 분석에 쓰인 자료에 근거하였다. 성도의 기본적인 틀은 각 화자의 차이를 평균한 중간모습으로 취했으며 각 모음에 대한 평균 조음거리로 grid line을 설정하여 조음점으로부터 조음체까지의 거리 및 조음체로부터 grid line까지의 거리를 분석하였다. 그림 3-1은 Harshman 등의 17개 등간격 section의 모습이다.

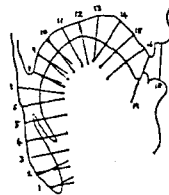


그림 3-1 Harshman 등의 X선 data의 조음축

이상의 원칙에 근거하여 성도길이를 입술과 치아에 2개 section, 치아에서 설근까지 (Velum까지) 11개, Velum에서 성문까지 2개로 나누어 총 15개의 section으로 나누었다. 각 축의 기울기는 조음점에 수직이 되도록 하였고 앞뒤축의 기울기와 되도록 평행을 유지하도록 하였다.

그림 3-2는 완성된 화면을 나타낸 그림이다.

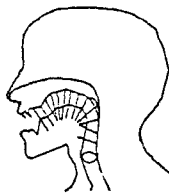


그림 3-2 완성된 조음도그래픽

### 4. 조음도와 음성파라미터의 연결

가. 조음거리의 계산

설정된 15개 구간에 대한 조음거리는 성도면적과 형성음 위치정보로 결정하였다. LP분석으로부터 성도모델의 전달함수  $A_n(z)$  ( $n$ 은 section No.)을 결정하였을때

$$A_n(z) = \frac{\text{forward volume velocity}}{\text{전파된 volume velocity}}$$

이때 해당 section  $n$ 에서의 면적  $A_n$ 은 다음과 같다.

$$A_n = A_{n-1} \frac{1 + C_n}{1 - C_n}$$

$f_n$ 은 그 이전 section  $f_{n-1}$ 과의 상대비율로 정해지므로 성도모델의 마지막 단이 입술에서 면적 1로 강제조건이 주어질 때, 입술에서 성문까지의 면적이 역으로 추산될 수 있다. 또한 입술사이의 거리는 형성음주파수로부터 다음과 같이 결정된다.

$$X_1 = C_1 F_2 + C_2 F_2 F_3 + C_3 F_1 / F_2 + C_4$$

$$C_1 = 0.300 \cdot 10^{-3}$$

$$C_2 = -0.343 \cdot 10^{-6}$$

$$C_3 = 4.143$$

$$C_4 = -0.174$$

위, 아래 치아 사이의 거리는  $X_1$ 과  $X_2$  사이의 거리의 평균치로 정한다.

$$X_3 = 1/2 (X_1 + X_2)$$

이 이외에도 입술의 rounding degree, 턱뼈의 height 등을 형성음 정보로부터 추정한다. 그림 4-1은 각 조음거리별 X-선 Data와 비교한 것이다. 굵은 선이 X-선 Data이고 가는 선이 추정된 조음거리로서 거의 일치됨을 알 수 있다.

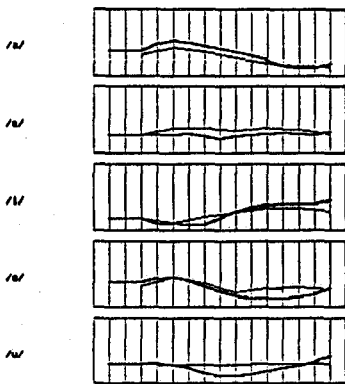


그림 4-1 X-선 data 와 추정된 조음거리

나. 조음체의 이동

추정된 조음거리별 각각의 속위에 mapping함에 있어 조음점을 고정점으로 보았다. 즉 윗턱뼈는 고정되어 윗치아, 치경, 경구개, 연구개는 움직이지 않다고 보고 고정좌표를 할당하였다. 또한 설근에서 성문까지의 성도의 윗턱, 즉 연구개에서 성문에 이르는 목의 뒷부분 역시 고정되었다고 가정하였다. 따라서 각 속위의 고정된 조음점에서부터 시작하여 해당조음거리에 해당하는 점이 조음체의 좌표가 된다.

또한 각 모음에 따라 그리고 각 자음에 따라 조음점의 위치는 서로 다른 가중치를 갖게된다. 표 4-1은 각 음소에 따른 조음점의 가중치이다.

표 4-1 각 음소에 따른 조음점의 가중치

음 소	가 중 치
/i/ /e/	폐모음으로서 F2에 의해 j번째 조음거리는 1+번째로 이동
/u/ /o/	Rounding degree = 1.
/u / ɜ / ɔ /	양순음으로 $X_1 = 0$ .
/ɛ / /ɛ / /ɛ /	치경음으로 $X_3 = 0$ .
/ɜ / /ɜ /	구개치경음으로 $X_5 = 0$ .
/ɔ / /ɔ / /ɔ /	연구개음으로 $X_6$ 이 factor 0.5
/ɔ / /ɔ / /o /	비음으로 nasal degree = 1.

표 4-1의 가중치와 조음거리 계산에 따라 작성된 양순음, 치경음, 구개치경음, 연구개음에 대한 조음도는 그림 4-2와 같다. 그림 4-3과 그림 4-4는 성도면적 계산 및 조음도 작성에 대한 블록선도이다.

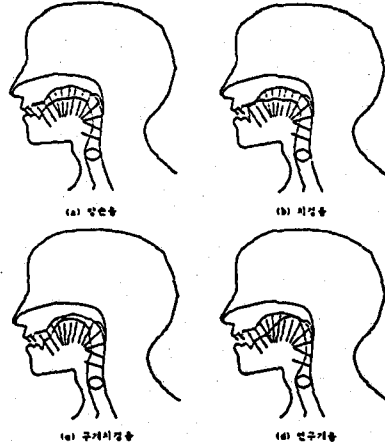


그림 4-2 양순음, 치경음, 구개치경음, 연구개음의 조음도

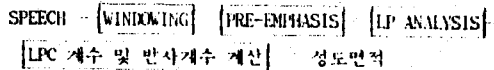


그림 4-3 성도면적 계산블록선도

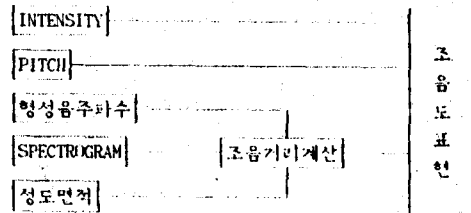


그림 4-4 조음도 블록선도

5. 결 본

종전에는 별도의 훈련기기로 훈련해야 했던 F0제어, Intensity제어 및 조음훈련 등을 하나로 통합하여 한 화면에서 처리함으로써 종전 훈련방식의 결점, 즉 한 가지 성분에 대한 훈련에 집중하는 동안 나머지 음성성분에 대한 제어능력이 저하되거나 등한시되는 단점을 해결하였다. 따라서 적절히 고안된 훈련프로그램이 수반될 경우 장애인의 발음훈련에 대한 종합적인 훈련환경을 조성할 수 있을 것이다.

본 연구는 상공부의 지원을 얻어 수행된 것임을 밝힘니다.

참고 문헌

[1] R. G. Crichton and F. Fallside, "Linear prediction model of speech production with applications to

- deaf speech training", Proc. Inst. Elec. Eng. Control & Science, Vol.121, pp.865-873, Aug.1974
- [2] Hisashi Wakita, "Direct Estimation of the Vocal Tract Shape by Inverse Filtering of Acoustic Speech Waveforms", IEEE TRANS. ON AUDIO ELECTRO-ACOUSTICS, VOL.AU-21, NO.5, OCTOBER 1973
- [3] Randall L. Christensen, William J. Strong, and E. Paul Palmer, "A Comparison of Three Methods of Extracting Resonance Information from Predictor-Coefficient Coded Speech", IEEE TRANS. ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, VOL. ASSP-24, NO.1, FEBRUARY 1976
- [4] J. Schroter, J. N. Larar, and M. M. Sondhi, "Speech Parameter Estimation Using a Vocal Tract / Cord Model", ICASSP-1987
- [5] Richard Harshman, Peter Ladefoged, and Louis Goldstein, "Factor analysis of tongue shapes", J. Acoust. Soc. Am., Vol.62, No.3, September 1977