

청각 장애자용 발음 훈련 기기의 개발

김 동준, 윤 태성, 박 상희
연세대학교 전기공학과, * 창원대학교 전기공학과

Speech Training Aids for Deafs

Dong-Jun Kim, Tae-Sung Yoon, Sang-Hui Park

Dept. of Electrical Engineering, Yonsei University
* Dept. of Electrical Engineering, Changwon University

ABSTRACT

Deafs train articulation by observing mouth of a tutor, sensing tactually the motions of the vocal organs, or using speech training aids. Present speech training aids for deafs can measure only single speech parameter, or display only frequency spectra in histogram or pseudo-color.

In this study, a speech training aids that can display subject's articulation in the form of a cross section of the vocal organs and other speech parameters together in a single system is aimed to develop and this system makes a subject to know where to correct.

For our objective, first, speech production mechanism is assumed to be AR model in order to estimate articulatory motions of the vocal tract from speech signal. Next, a vocal tract profile model using LPC analysis is made up. And using this model, articulatory motions for Korean vowels are estimated and displayed in the vocal tract profile graphics.

1. 서 론

청각장애자는 자신의 발음을 귀로 직접 들을 수 없으므로 정상인처럼 발음을 스스로 조절하면서 올바른 음성을 생성시킬 수 없다. 따라서 특수 교육을 통하여 발음을 훈련받아야 한다.

현재까지 국내에서는 아직 개발된 사례가 없고, 외국에서 개발된 청각장애자용 음성 훈련 기기들은 음성의 단일 매개변수(기본 주파수, /S/음, 비음성(nasality))만을 측정할 수 있거나, 또는 음성 신호를 주파수 분석하여 막대 그래프로 나타내거나 색을 달리하여 표시하는 형태였다. 음성의 단일 매개변수 측정 형태의 훈련 기기는 특정음에 대해서만 훈련이 가능하여 일반 음성의 발음 훈련에는 쓸 수 없는 단점이 있으며, 음성의 주파수별 막대 그래프 표현 형태의 훈련 기기의 경우에는 훈련 당사자가 주파수별 막대 성분만 보고서는 자신의 조음 기관중 어느 곳을 어떤 정도로 교정해야 하는지 알아내기가 어려우므로 훈련 효과가 크지 않았다.

음성 훈련 기기의 개발에 있어서, 인간의 발음 생성 과정을 음성 신호로부터 역으로 추경하여 조음도의 형태로 표현하는 것은 매우 중요하다.

본 연구에서는 청력의 상실로 인한 발음 기관 조절 능력의 저하로 정상적인 대화 소통이 어려운 청각장애자들을 대상으로 발음된 음성을 분석하여 발음 기관의 단면도를 나타

내고, 그 위에 표준 발음의 조음 형태와 훈련자가 발음한 조음 형태를 동시에 표시하여 교정이 필요한 부위를 드러나게 하는 조음도 그래픽을 개발한다. 또한 음성 발음에 필요한 여러 가지 매개변수들을 동시에 그래픽상에 표현하여 장애자가 표준 발음과 자신의 발음과의 차이를 쉽게 식별하여 음성 발음을 훈련할 수 있는 발음 훈련 기기를 개발한다.

2. 조음도 그래픽

(1) 조음도 그래픽의 설계와 개발

조음도는 발성할 때 각 조음기관의 모습을 성대(vocal cords)에서 입술 또는 코까지, 즉 성도(vocal tract)를 쫓아 단면도의 형태로 나타낸 것이다. 음성신호로부터 변별적 매개변수들을 추출하여 각 조음기관의 조음형태를 그래픽으로 복원하기 위해서는 우선 각 조음기관의 운동범위를 정의해야 하는데 이론적으로는 성도의 아무 곳에서도 공기의 흐름을 폐쇄하거나 방해할 수 있는 무한한 가능성이 있지만 인간의 신경과 근육조직이 아주 세밀한 차이를 구별해서 행동할 수 있게 할 수 없으며, 비록 혀가 1[mm]의 조음위치의 차이를 착오없이 수행할 수 있다하더라도 청각적으로는 아무런 차이를 나타내 주지 않으므로 실제로 언어발음에서 쓰이는 조음위치는 제한된다.

조음기관의 운동을 표시하기 위하여 다음과 같은 원칙에 입각하여 운동축을 설정하였다.

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

본 연구에서는 기본모델로 삼은 15개 등간격 파이프의 성도 모델은 성인남자의 17[cm] 성도를 기본 전체로 한 것이므로 여기서도 성도 전체 길이를 17[cm]로 잡고 각 구간(section)을 구분하였다.

X-선 데이터는 Harshman 등의 PARAFAC 분석에 쓰인 자료에 근거하였다. 성도의 기본적인 틀은 각 화자의 차이를 평균한 중간모습으로 취했으며 각 모음에 대한 평균 조음거리로 그리드 선(grid line)을 설정하여 조음점으로부터 조음

체까지의 거리 및 조음체로부터 그리드 선까지의 거리를 분석하였다.

이상의 원칙에 근거하여 성도길이를 입술과 치아에 2개 구간, 치아에서 설근(velum)까지 11개, 설근에서 성문까지 2개로 나누어 총 15개의 구간으로 나누었다. 각 축의 기울기는 조음점에 수직이 되도록 하였고 앞뒤축의 기울기와 가능한 한 평행을 유지하도록 하였다.

그림 2.1은 완성된 화면을 나타낸 그림이다.

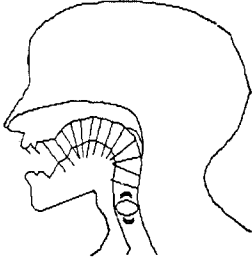


그림 2.1 완성된 조음도 그래픽
Fig. 2.1 Completed vocal tract profile graphics

(2) 조음도와 음성 매개변수의 연결

가. 조음거리의 계산

설정된 15개 구간에 대한 조음거리는 성도면적과 형성음 위치정보로 결정하였다. LP 분석(linear predictive analysis)으로부터 성도모델의 전달함수 $A_n(z)$ (n 은 구간 번호)을 결정하였을 때

$$A_n(z) = \frac{\text{전방향 부피 속도(forward volume velocity)}}{\text{전파된 부피 속도(propagated volume velocity)}}$$

이 때 해당 구간 n 에서의 면적 ϕ_n 은 다음과 같다.

$$\phi_n = \phi_{n-1} \frac{1 + C_n}{1 - C_n}$$

ϕ_n 은 그 이전 구간 ϕ_{n-1} 과의 상대비율로 정해지므로 성도모델의 마지막 단이 입술에서 면적 1로 경계조건이 주어질 때, 입술에서 성문까지의 면적이 역으로 추산될 수 있다. 또한 입술사이의 거리는 형성음 주파수로부터 다음과 같이 결정된다.

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

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

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

$$C_3 = 4.143$$

$$C_4 = -0.174$$

위, 아래 치아 사이의 거리는 X_1 과 X_3 사이의 거리의 평균치로 정한다.

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

그림 2.2는 형성음주파수 추출알고리즘의 블록 선도이다. 이외에도 입술의 둥근 정도(rounding degree), 턱뼈의

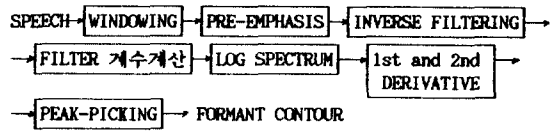


그림 2.2 형성음주파수 추출알고리즘 블록선도
Fig. 2.2 Block diagram of formants extraction

높이 등을 형성음 정보로부터 추정한다.

나. 조음체의 이동

추정된 조음거리를 각각의 축위에 사상함에 있어 조음점을 고정점으로 보았다. 즉 윗턱뼈는 고정되어 윗치아, 치경, 경구개, 연구개는 움직임이 없다고 가정할 수 있으므로 고정좌표를 할당하였다. 또한 설근에서 성문까지의 성도의 윗벽, 즉 연구개에서 성문에 이르는 목의 윗부분 역시 고정되었다고 가정하였다. 따라서 각 축위의 고정된 조음점에서부터 시작하여 각 조음거리에 해당하는 점이 조음체의 좌표가 된다.

또한 각 모음에 따라 그리고 각 자음에 따라 조음점의 위치는 서로 다른 가중치를 갖게 된다. 표 2.1은 모음에 따른 조음점의 가중치이다.

표 2.1 각 모음에 따른 조음점의 가중치
Table 2.1 Articulatory weights for each vowel

모 음	가 중 치
/이/ /에/	폐모음으로서 F2에 의해 j번째 조음거리는 j+1번째로 이동
/우/ /오/	Rounding degree = 1.

그림 2.3과 그림 2.4는 성도면적 계산 및 조음도 작성에 대한 블록 선도이다.

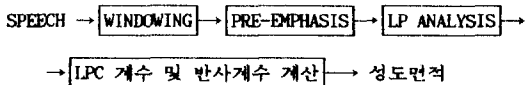


그림 2.3 성도면적 계산 블록 선도
Fig. 2.3 Block diagram of vocal tract area extraction

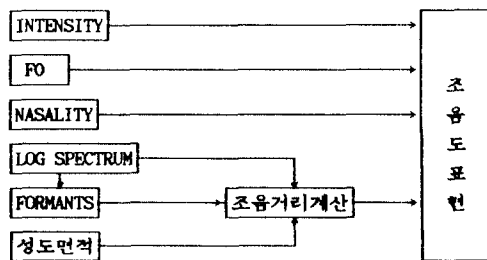


그림 2.4 조음도 블록선도
Fig. 2.4 Block diagram of vocal tract profile graphics

3. 하드웨어 구현

(1) 데이터 추출 장치의 구현

인간이 음성을 발생시킬 때 입을 통하여 음성 신호가 방사되는 동시에 조음 기관의 각 부분들이 계각기 각자의 기능을 수행하여 줌으로써 완전한 음성이 표현된다. 따라서 조음 기관의 각 부분들의 정보를 충분히 반영할 때 보다 정확히 조음 운동의 추정이 가능하다.

본 연구에서는 조음 운동시 입을 통하여 발산되는 음성 신호와 이 때 조음 기관의 외부에서 가장 두드러진 떨림을 나타내는 성대와 코에서 특징 매개변수를 추출하는 데이터 추출 장치를 구현하였다.

음성 신호 추출을 위한 블록 선도는 그림 3.1과 같다.

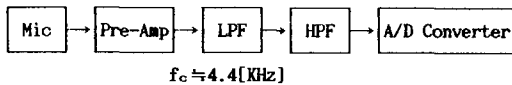


그림 3.1 음성 신호 추출을 위한 블록 선도
Fig. 3.1 Block diagram of speech signal detector

마이크를 통하여 입력된 음성 신호는 전체 이득 약 350으로 증폭되며, 이 때 약 4.4[KHz]의 차단 주파수(cutoff frequency ; f_c)를 갖는 저역 통과 필터(low pass filter ; LPF)를 통과하게 된다. 저역 통과 필터에서의 출력 신호는 직류 성분 오프셋(offset)을 차단하기 위한 $f_c \approx 1.6$ [Hz]의 고역 통과 필터(high pass filter ; HPF)를 거쳐서 TMS 보드의 A/D 변환부로 들어가게 된다.

그리고 음성 신호에서 간접적으로 소프트웨어에 의하여 기본 주파수를 검출하는 경우, 기본 주파수가 일정하지 않은 장애자의 경우에는 특히 검출이 어려울 것이다. 따라서 본 연구에서는 목 부분에서 성대의 떨림이 가장 잘 검출되는 부위에 진동 센서를 부착하여 이를 통하여 진동이 전기적 신호로 바뀐 데이터를 하드웨어에 의하여 적절히 변형하여 기본 주파수를 검출하고자 한다.

기본 주파수 추출을 위한 블록 선도는 그림 3.2와 같다.

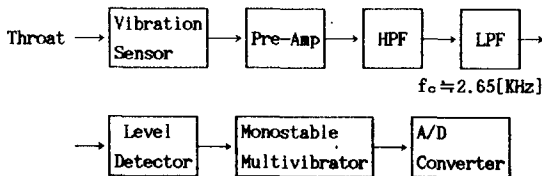


그림 3.2 기본 주파수 추출을 위한 블록 선도
Fig. 3.2 Block diagram of FO detector

목의 진동 신호는 전체 이득 약 11.9로 증폭되며, 직류 성분 제거용 고역 통과 필터를 거쳐서 $f_c \approx 2.65$ [KHz]인 저역 통과 필터를 통과하게 된다. 출력된 신호는 레벨 검출기(level detector)로 입력되어 어떤 레벨 이상에 대하여 펄스 파형을 출력하게 되며, 최종적으로 단안정 멀티바이브레

이터(monostable multivibrator)의 출력을 A/D 변환하게 된다.

비음은 모음처럼 주기적인 파형을 갖으며 그 진폭은 다른 자음보다는 크지만 모음보다는 작다. 비음을 추출하기 위하여 기본 주파수 추출부와 마찬가지로 코의 옆 부분에 진동 센서를 부착하여 진동의 포락선(envelope)을 추출하고자 한다.

비음 추출을 위한 블록 선도는 그림 3.3과 같다. 비음이 발음될 때 코의 진동이 진동 센서에 의해 전기적 신호로 바뀌고, 전체 이득 약 357로 증폭되며 직류 성분 제거를 위한 고역 통과 필터를 거쳐 $f_c \approx 2.5$ [KHz]인 저역 통과 필터를 통과한다. 그리고 출력된 신호는 정류기를 거쳐 다시 $f_c \approx 10$ [Hz]의 저역 통과 필터를 거치면 비음의 포락선이 출력되어 A/D 변환된다.

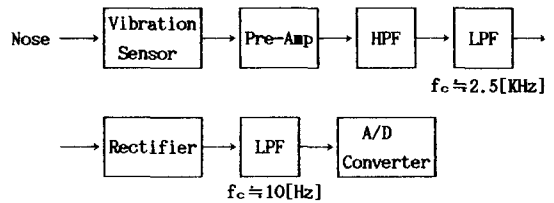


그림 3.3 비음 추출을 위한 블록 선도
Fig. 3.3 Block diagram of nasality detector

(2) 매개변수의 고속 추출 장치 구현

음성의 매개변수 중에서 기본 주파수와 비음성은 음성 처리 장치에서 직접 추출이 가능하나, 음의 세기, 주파수의 대수 스펙트럼, 및 조음도 그래픽에 필요한 매개변수들은 많은 연산이 필요하다. 따라서 PC에서는 많은 시간이 소요되므로 이에 따르는 문제점, 즉, 자신이 방금 발음한 음성 에 대한 경험을 연산이 끝나기를 기다리는 동안 잊어 버릴 수 있으므로 고속 연산이 필요하게 된다. 본 연구에서는 이러한 고속 연산이 가능한 TMS 칩(chip)을 이용하여 보다 빠른 연산을 가능하게 하였다.

【 TMS320C25 】

- 명령어 수행 시간(instruction cycle time) : 100 [nsec]
 - 해상도(resolution) : 16 [bit]
 - 구동 언어 : C, TMS 어셈블리어
- 고속 연산을 요할 때에는 어셈블리어로 작성

(3) 전체 시스템의 결합

부분적으로 구현된 모든 장치들은 그림 3.4와 같이 결합되었다.

여기서, 데이터 추출장치는 한 보드 위에서 제작되었고, 음성 신호는 TMS의 A/D부로 들어가고 나머지는 다른 A/D보드를 이용하였다. TMS에 의하여 실시간에 가까운 연산이 이루어지며 이 모든 장치는 PC에서 구동하게 된다.

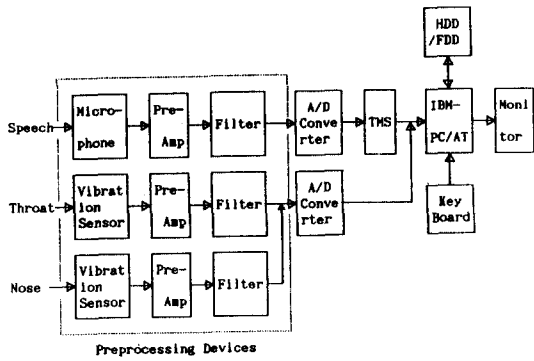


그림 3.4 전체 시스템의 블록 선도
Fig. 3.4 Block diagram of total system

4. 실험 및 결과 고찰

본 연구에서는 조음도 그래픽에 추가로 윈도우를 할당하여 음의 세기, 기본 주파수, 비음성, 대수 스펙트럼 등의 4개의 음성 매개변수를 동시에 나타낸다.

(1) 조음도 그래픽의 실험

4명의 남성 화자가 발음한 잡음이 적은 5개의 한국어 단모음 /아/, /에/, /이/, /오/, /우/에 대하여 성도 면적을 추출하여 그림 4.1에 그래프로 나타내었다.

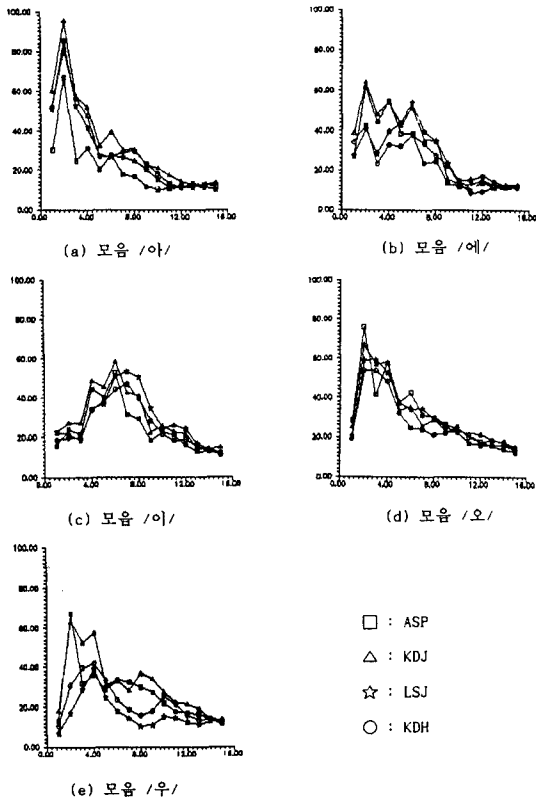


그림 4.1 5개 한국어 단모음에 대한 성도 면적
Fig. 4.1 Estimated vocal tract areas for 5 Korean vowels

추출된 성도면적과 형성음 주파수로부터 조음거리를 계산하게 된다. 그림 4.2는 4명의 화자의 5개 모음에 대한 성도 면적을 이용하여 조음거리를 계산한 결과와 Harshman 등의 의한 X-선 데이터와의 비교를 나타낸 것이다.

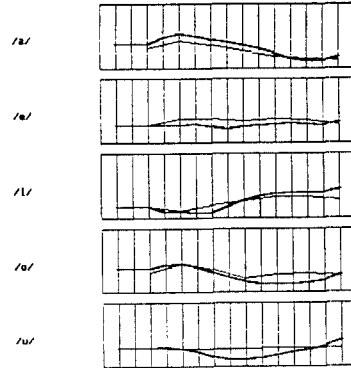


그림 4.2 추정된 조음 거리와 X-선 데이터와의 비교
Fig. 4.2 Comparison of estimated articulatory distances and X-ray data

(2) 기타 음성 매개변수의 실험

그림 4.3은 모음 /아/의 원신호에 대하여 추출된 음의 세기, 기본 주파수를 나타내며, 그림 4.4는 비음을 포함한 /마/를 발음했을 때의 추출된 궤적과 /ㅁ/부분과 /ㅌ/부분의 대수 스펙트럼을 나타낸다.

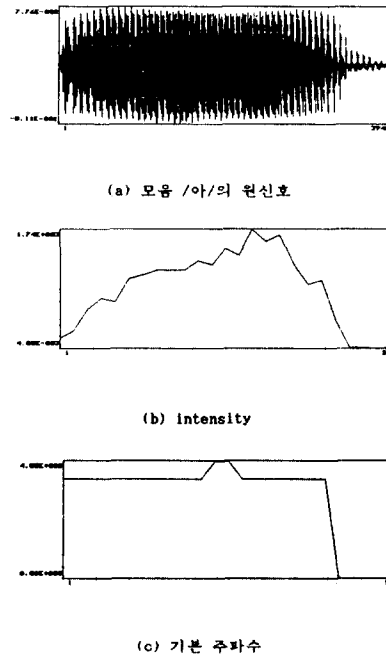


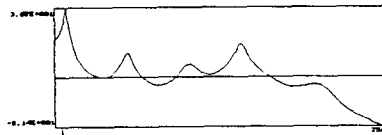
그림 4.3 모음 /아/의 추출된 음의 세기와 기본 주파수
Fig. 4.3 Intensity and F0 for Korean vowel /아/



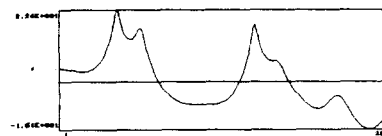
(a) /마/의 원신호



(b) 추출된 비음성



(c) /ㅁ / 부분의 스펙트럼



(d) /ㅏ / 부분의 스펙트럼

그림 4.4 /마/에 대한 비음성과 대수 스펙트럼
Fig. 4.4 Nasality and log spectrum for /ㅁㅏ/

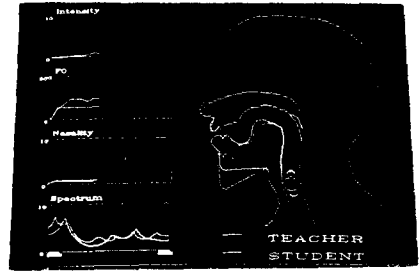
(3) 전체 실험

전체적으로 결합된 시스템에서 목에 센서를 부착할 수 있는 목띠를 제작하여 이용하였고, 코의 센서는 혼련자가 직접 코의 펄프가 잘 검출되는 부위에 부착하고 마이크에 발음하게 된다. 먼저 교사가 정상적인 표준 발음을 하여 입력시키고 이어서 혼련자가 따라 발음한 후, 저장된 교사의 표준 발음과 비교 검토하면서 반복적으로 훈련을 수행한다.

청각 장애자에게 직접 실험하기 전에 예비 실험으로서 여러 가지 실험을 수행하였으며, 그 결과 중의 일부를 그림 4.5에서 나타내었다.



(a) /ㅏ/의 정상 발음



(b) /오/ 와 /ㅏ/

그림 4.5 /ㅏ/와 /오/에 대한 화면 출력 (학생이 정상 /ㅏ/와 중성 모음 /ㅓ/를 발음했을 때)

5. 결론

본 연구에서는 청력의 상실로 인한 발음 기관 조절 능력의 저하로 정상적인 대화 소통이 어려운 청각장애자들을 대상으로 하여 발음된 음성을 분석하여 발음 기관의 단면도를 나타내고 그 위에 표준 발음의 조음 형태와 혼련자가 발음한 조음 형태를 동시에 표시하여 교정이 필요한 부위를 드러나게 하는 조음도 그래픽을 개발하며, 또한 음성 발음에 필요한 여러 가지 매개변수들을 그래픽상에 표현하여 장애자가 표준 발음과 자신의 발음의 매개변수 상의 차이를 쉽게 식별하여 음성 발음을 훈련할 수 있는 발음 훈련 기기를 개발하였다.

얻어진 결과는 다음과 같다.

- ① 모음의 경우로 한정되지만 일선 교사의 요망인 발음 기관 내부 표현을 이용함으로써 혼련자가 자신의 발음 기관 중에서 교정이 필요한 부위를 한눈에 알아볼 수 있다.
- ② 표준 발음과 혼련자의 발음을 한 화면에서 색을 달리하여 나타내며, 특히 일치된 부분을 눌러감으로써 혼련이 효율적으로 이루어질 수 있다.
- ③ 발음에 필요한 여러 특징 매개변수들(기본 주파수, 음의 세기, 비음성, 주파수 스펙트럼 등)을 조음도 그래픽과 함께 하나의 화면에 나타냄으로써, 발음 상태를 종합적으로 관찰할수 있는 환경을 제공할 수 있다.
- ④ 형성음 주파수 추출에 있어서 1차 미분과 2차 미분을 같이 사용함으로써 형성음의 첨두치가 중첩되는 현상과 이동되는 문제점을 해결하여 정확한 결과를 얻을 수 있었다.
- ⑤ 조음도에서 조음점을 윗턱으로 설정하여 고정시키고, 조음체를 아랫턱으로 설정하여 변화가 가능하게 하여 성도 면적에 따른 조음 운동의 변화를 쉽게 관찰할 수 있었다.
- ⑥ 한국어 단모음에 의한 조음 운동을 추정하여 X-선의 성도 면적 데이터와 비교하여 본 결과 상당한 신뢰성을 확인하였다.
- ⑦ 특징 매개변수들 그래픽으로 표현하기 위한 연산을 고속 연산이 가능한 TMS 보드를 이용함으로써 빠른 결과를 볼 수 있으므로 혼련자가 발음하고 난 후 발음의 방법을 잊어버리기 전에 화면 상에서 비교 검토할 수 있다.

본 연구는 상공부의 공업 기반 기술 개발 사업의 기술 개발 결과임을 밝힙니다.

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. Schroeter, 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
- [6] 안 상필, 이 재혁, 박 상희, 윤 태성, "청각장애자용 발음 훈련 기기 개발에 관한 연구," 대한전기학회 학술대회, 1990.7.
- [7] 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.
- [8] T. Chiba and M. Kajiyama : The Vowel, Its Nature and Structure, Kaiseikan, 1941.
- [9] J. D. Markel and A. H. Gray, Jr. : Linear Prediction of Speech, Springer-Verlag, 1980.
- [10] 김용목 편역 : 최신 아날로그 IC 회로기술, 세화, 1987.
- [11] Digital Signal Processing Applications with the TMS320 Family, Texas Instruments, 1986.