醫工學會誌 : 第 13 卷, 第 4 號, 1992
J. of KOSOMBE : Vol. 13, No. 4

研 究 論 文
92 - 13 - 4 - 02

청각 장애인용 통합형 발음 훈련 기기의 개발

박상희 · 김동준 · 이명국 · 송제혁

=Abstract=

Development of Integrated Speech Training Aids for Hearing Impaired

Sang-Hui Park, Dong-Jun Kim, Myung-Kuk Lee, and Jei-Hyuk Song

In this study, a speech training aids that can do real-time display of vocal tract shape and other speech parameters together in a single system is implemented and self-training program for this system is developed. To estimate vocal tract shape, speech production process is assumed to be AR model. Through LPC analysis, vocal tract shape, intensity, and log spectrum are calculated. And, fundamental frequency and nasality are measured using vibration sensors.

서 론

청각 장애인의 발음 훈련을 보조해 주기 위하여 많은 발음 훈련 기기들이 개발되어 왔다[1]. 현재까지 국내에서는 아직 개발된 사례가 없고, 외국에서는 음성의 단일 매개변수만을 측정할 수 있거나, 음성 신호를 주파수로 분석하여 막대 그래프로 나타내거나 색을 달리하여 표시하는 형태였다. 이러한 단일 매개변수 형태의 훈련 기기는 특정음에 대해서만 훈련이 가능하므로 훈련할 수 있는 발음이 한정되는 단점이 있으며, 음성의 주파수별 막대 그래프 표현 형태의 훈련 기기의 경우에는 훈련자가 주파수 성분만 관찰하면서 자신의 조음 기관중 어느 곳을 어떤 정도로 교정해야 하는지 알아내기가 어려우므로 훈련 효과가 크지 않았다. 따라서, 인간의 발음 생성 과정을 음성 신호로부터 역으로 추정하여 조음도의 형태로 표현하는 것은 매우 중요하다.

본 연구에서는 청력의 상실로 인한 발음 기관 조절 능력의 저하로 정상적인 대화 소통이 어려운 청각장애인들을 대상으로 발음된 음성을 분석하여 발음 기관의 단면도를 나타내고, 그 위에 표준 발음의 조음 형태와 훈련자가 발음한 조음 형태를 동시에 표시하여 교정이 필요한 부위를 알 수 있는 조음도 그래픽과 음성 발음에 필요한 여러가지 매개변수들을 동시에 한 시스템의 그래픽상에 표현하여 장애인의 표준 발음과 자신의 발음과의 차이를 쉽게 식별할 수 있고, 또한 체계적으로 개발된 훈련 프로그램에 따라 훈련자가 스스로 발음을 훈련할 수 있는 발음 훈련 기기를 개발하고자 한다.

<접수 : 1992년 11월 28일>
연세대학교 전기공학과
Dept. of Electrical Engineering, Yonsei University

2. 음성 매개변수의 추출

2. 1 조음도 그래픽

인간의 성도 형태를 추정하기 위하여 직접적인 방법으로 X-선 기법이 이용되었으나, 방법의 어려움과 과다 노출의 위험성 때문에 간접적인 방법이

많이 개발되었다[2][3][4].

본 연구에서는 조음도에서 조음기관의 운동을 표시하기 위하여 최단거리 법칙을 따르고, 성도는 15개 등간격으로 나누고, X-선으로 얻은 조음단면도의 데이타에 일치시켰으며, 성인남자의 평균 성도 길이가 17[cm]이므로 성도 전체 길이를 17[cm]로 잡고 각 구간을 구분하였다[11].

그림 1은 설계가 완성된 화면을 나타낸 그림이다.

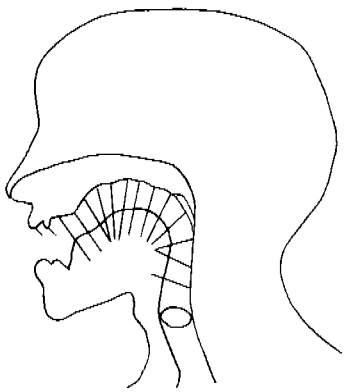


그림 1 설계된 조음도 그래픽
Fig. 1 Designed vocal tract profile graphics

구분된 구간 n에서의 성도면적 Φ_n은 다음과 같다[3].

$$\Phi_n=\Phi_{n-1}\frac{1+r_n}{1-r_n} \quad (1)$$

여기서, r_n은 반사계수이다.

Φ_n은 그 이전 구간 Φ_{n-1}과의 상대적인 비율로 정해지므로 성도모델의 마지막 단인 입술에서 면적 1로 경계조건이 주어질 때, 입술에서 성문까지의 면적이 역으로 계산될 수 있다. 또한 입술 사이의 거리는 형성음 주파수로부터 다음과 같이 결정된다[4].

$$X_1=C_1F_2+C_2F_2F_3+C_3F_1/F_2+C_4 \quad (2)$$

$$C_1=\ 0.300\times10^{-3}$$
$$C_2=-0.343\times10^{-6}$$
$$C_3=\ 4.143$$
$$C_4=-0.174$$

여기서, C_1~C_4는 Harshman 등의 PARAFAC 분석 결과를 추출된 성도 면적과 그래픽 표현을 위하여 수정한 실험치이다.

위, 아래 치아 사이의 거리는 X_1과 X_3의 평균치로 정한다.

$$X_2=\frac{X_1+X_3}{2} \quad (3)$$

그림 2는 성도면적과 형성음주파수 추출의 블록 선도이다.

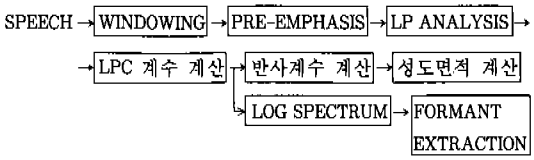


그림 2 성도면적과 형성음 주파수 계산 블록 선도
Fig. 2 Block diagram calculation of vocal tract area and formants

2.2 기타 매개변수

조음도 그래픽 이외에 추출된 특징 매개변수들은 기본 주파수(fundamental frequency), 음의 세기(intensity), 비음성(nasality), 그리고 주파수 스펙트럼(frequency spectrum) 등이다.

기본 주파수는 가장 기본적인 특징 매개변수로서, 청각장애인들의 경우, 정상 발음보다 높은 경우가 많으며, 상당히 불규칙적이다.

장애인들의 발음에 있어 또 하나의 심각한 오류는 음의 세기의 조절이다. 장애인 발음의 경우 음의 세기는 기본 주파수와 밀접한 연관을 갖게 되는데, 기본 주파수를 조절하려는 시도가 음의 세기에 영향을 주어 음의 세기가 지나치게 크게 되거나 아주 불규칙해져 정상적인 음성과는 다른 양상을 띄게 된다. 본 연구에서 음의 세기는 인간의 귀가 음의 세기의 대수로 인지하는 특성을 고려하여 진폭의 제곱에 로그를 취한 값으로 구한다[5].

비음성은 하드웨어에 의하여 코의 진동의 포락선(envelope)을 구하여 나타내게 된다.

주파수 스펙트럼은 선형 예측법(linear prediction method)에 의하여 선형 예측 계수를 구하고, 이 값들을 음성 신호에 대한 자기 회귀(auto-regressive; AR) 모델에 대입함으로써 얻어진다[8].

그림 3은 지금까지 구해진 모든 특징 매개변수들의 화면 디스플레이 과정에 대한 블록 선도이다.

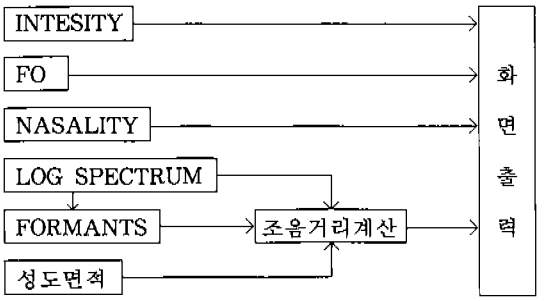


그림 3 전체 매개변수의 화면 출력 블록 선도
Fig. 3 Block diagram of total parameters display

3. 하드웨어 구현

본 연구에서는 음성 신호와 성대와 코에서 특징 매개변수를 추출하는 데이타 추출 장치를 구현하였다.

음성의 매개변수 중에서 기본 주파수와 비음은 하드웨어에 의하여 실시간(real time)으로 직접 추출이 가능하나, 음의 세기, 주파수 스펙트럼, 및 조음도 그래픽에 필요한 매개변수들은 많은 연산이 필요하므로 고속연산 장치가 필요하다. 본 연구에서는 이러한 고속 연산이 가능한 TMS 보드를 이용하였고, 부분적으로 구현된 모든 장치들은 그림 4와 같이 결합되었다[11].

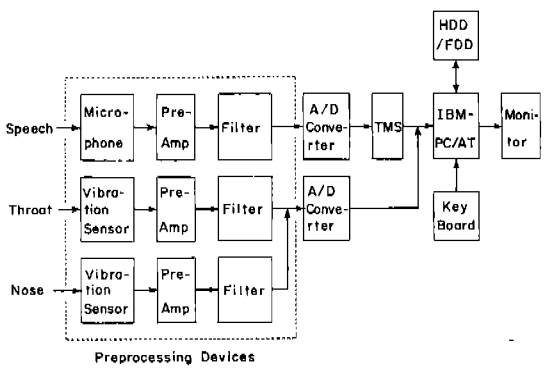


그림 4 전체 시스템의 블록 선도
Fig. 4 Block diagram of total system

4. 자율 훈련 프로그램의 개발

청각 장애인을 위한 기존의 대부분의 발음 훈련 기기들은 적절한 훈련 프로그램의 결여로 인하여 훈련자 자신이 스스로 체계적으로 훈련하는 것이 불가능하다. 따라서 교사의 세심한 지도가 수반되어야 한다.

본 연구에서는 추출된 여러가지 매개변수를 이용하여 훈련자 스스로 쉽게 알아 볼 수 있는 조음 방법 상의 차이로 나타내어 이러한 조음 방법의 교정 명령에 의하여 발음을 교정하여 나갈 수 있는 훈련 프로그램을 개발하였다.

4.1 모음에 대한 훈련 프로그램

발음이 정상 상태일 때의 매개변수들을 실험적으로 정확하게 추출하여 기준값으로 삼고, 이 기준과 훈련자의 발음을 비교하게 된다. 이때 여자와 어린이의 기본 주파수의 기준값은 같다고 가정하고, 남자의 기준 기본 주파수를 구한 뒤, 남자와 여자의 기본 주파수 차이를 고려하여 여자의 기본 주파수 기준값을 설정하였다.

4.1.1 단모음

9개의 한국어 단모음 /아/, /어/, /오/, /우/, /으/, /이/, /에/, /애/, /외/에 대하여 훈련 프로그램을 개발하였다. 훈련프로그램의 판단과 교정 명령은 다음의 원리에 의하여 이루어졌다.

① **음의 세기**

훈련자의 음의 세기가 허용 범위안에 들어올 때는 정상이며, 허용범위에서 벗어나면 각각 해당하는 교정 명령이 출력된다. 이때 허용 범위는 실험적으로 결정된다. 이를 수식으로 나타내기 위하여 기준 음의 세기를 I_{ref}, 훈련자의 음의 세기를 I_{st}, 허용 범위를 θ_1라 하면, 다음과 같다.

구간 1 : $I_{st}-I_{ref} > \theta_1$⇨"소리를 작게 발음하시오"
구간 2 : $|I_{st}-I_{ref}| \le \theta_1$⇨"잘 했습니다."
구간 3 : $I_{st}-I_{ref} < \theta_1$⇨"소리를 크게 발음하시오."

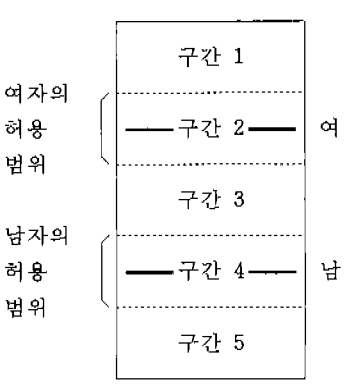
② **기본 주파수**

기본 주파수의 경우에는 남자와 여자의 허용 범위에 대한 판단을 수행한 후 이에 대한 교정 명령이 이루어지며, 이를 위하여 그림 5와 같이 5개 구간으로 나누어져 판단을 하게 된다.

③ **대수 스펙트럼**

훈련하고자 하는 모음의 기준 대수 스펙트럼과 훈련자의 대수 스펙트럼간의 제곱 오차를 구하여 판단하게 된다. 즉, 기준 대수 스펙트럼을 S_{ref}, 훈련자의 대수 스펙트럼을 S_{st}라 하고, 허용 오차를 θ_s라 할때

$$\begin{cases} \sum_{i=1}^{128} (S_{ref}-S_{st})^2 \le \theta_s ⇨ \text{"잘 했습니다."} \\ \text{그외} ⇨ \text{"틀렸습니다."} \end{cases}$$



구간 1 ⇨ 남 : "낮게 발음하시오."
여 : "약간 낮게 발음하시오."
구간 2 ⇨ 남 : "낮게 발음하시오."
여 : "잘 했습니다."
구간 3 ⇨ 남 : "약간 낮게 발음하시오."
여 : "약간 높게 발음하시오."
구간 4 ⇨ 남 : "잘 했습니다."
여 : "높게 발음하시오."
구간 5 ⇨ 남 : "약간 높게 발음하시오."
여 : "높게 발음하시오."

그림 5 기본 주파수에 대한 판단 기준
Fig. 5 Decision criteria for fundamental frequency

④ **조음도 그래픽**

조음도 그래픽에서도 기준 조음도 그래픽과 훈련자의 조음도 그래픽을 비교한다. 기준 턱의 높이를 H_{ref}, 훈련자의 턱의 높이를 H_{st}라 하고 허용 오차를 θ_H라 할 때

$|H_{st}-H_{ref}| \leq \theta_H$⇨"잘 했습니다."
$H_{st}-H_{ref} > \theta_H$⇨"턱을 올리고 발음 하시오."
$H_{st}-H_{ref} \leq \theta_H$⇨"턱을 낮추고 발음 하시오."

4.1.2 복모음

복모음(complex vowel)에 대한 훈련 프로그램은 /야/, /여/, /요/, /유/, /예/, /애/, /와/, /워/, /웨/, /왜/, /위/, 그리고 /의/에 대하여 개발하였으며, 음의 세기와 기본 주파수는 단모음과 같고, 대수 스펙트럼과 조음도 그래픽은 앞과 뒤의 모음에 대하여 각각 판단 및 교정명령이 이루어진다.

4.2. 자음에 대한 훈련 프로그램

자음의 경우 음파의 진행이 막혀진 상태에서 발생하므로 정확한 성도 면적의 추출이 불가능하다. 따라서, 본 연구에서는 추출된 모든 매개변수들을 이용하여 음성이 음향학적 특성을 고려하여 자음

표 1 자음의 판단 기준
Table 1. Decision criteria for consonants

음 성 분 류	음 소	판 단 기 준
1. 순 음 (Labial)	ㅁ	inten +, nasal +
	ㅂ	inten +, nasal −
	ㅍ	inten −, nasal −
2. 치경음 (Alveolar)	ㄴ	inten +, nasal +
	ㄷ	inten +, nasal −
	ㅌ	inten −, nasal −
	ㄹ	sepctrum peak가 1953～2540 Hz에서 양호
	ㅅ	spectrum peak가 4199～4785 Hz에서 양호
3. 구개치경음 (Plato-alveolar)	ㅈ	inten +, nasal −
	ㅊ	inten −, nasal −
4. 연구개음 (Velar)	ㄱ	inten +, nasal −
	ㅋ	inten −, nasal −

의 훈련 프로그램을 개발하였다.

먼저 표준 조음 위치를 나타내는 조음도를 화면에 디스플레이하고, 조음위치는 제대로 되었다고 가정하고 그 외의 매개 변수에 대하여 표 1의 판단 기준에 의하여 판단이 이루어진다. 여기서, inten-는 발음의 시작 부분에서 상대적으로 작은 값이 나타나는지를 판단하고, inten+는 그렇지 않은 경우를 판단한다. nasal+는 비음 성분이 특정 문턱치보다 큰지를 판단하고, nasal—는 반대의 경우를 판단한다.

5. 실 험

5.1. 시뮬레이션 결과

그림 6은 4명의 화자의 5개 모음에 대한 성도 면적을 이용하여 조음거리를 계산한 결과와 Harshman 등에 의한 X-선 데이타와의 비교를 나타낸 것이다. 그림 6에서 굵은 선이 X-선의 조음거리이고, 가는 선이 추출된 데이타이다.

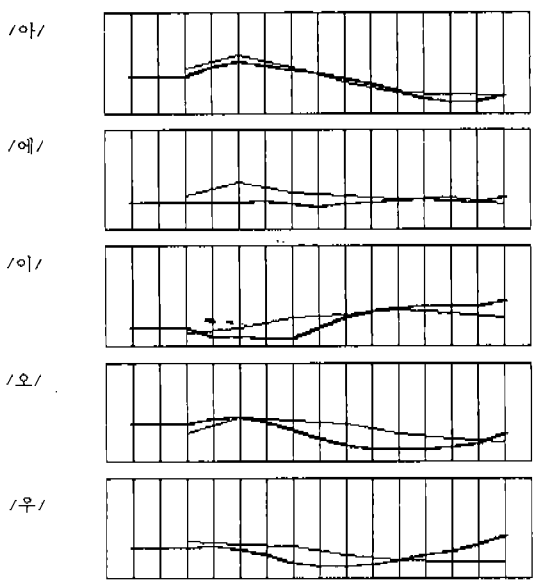


그림 6 추정된 조음거리와 X-선 데이타의 비교

Fig. 6 Comparison of estimated articulatory distances and X-ray data

그림 7은 모음 /아/의 원신호에 대하여 추출된 음의 세기, 기본 주파수를 나타내며, 그림 8은 비음을 포함한 /마/를 발음했을 때의 추출된 비음 궤적과 /ㅁ/부분과 /ㅏ/부분의 대수 스펙트럼을 나타낸다.

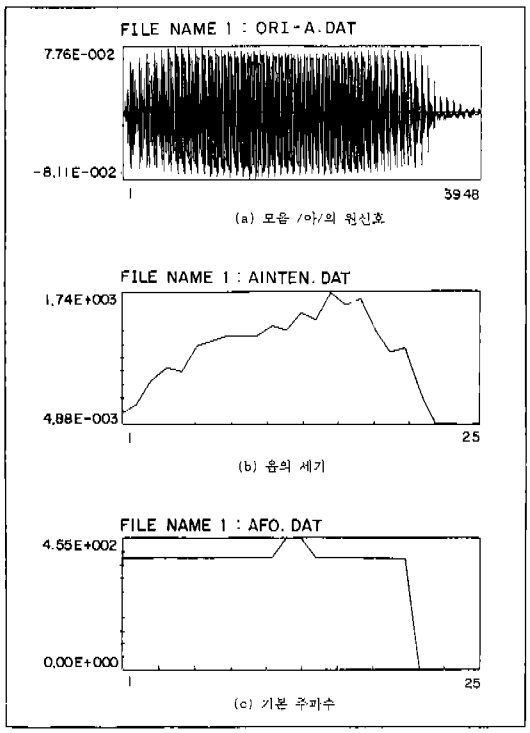


그림 7 모음 /아/의 추출된 음의 세기와 기본 주파수

Fig. 7 Intensity and fundamental frequency for Korean vowel /아/

5.2. 임상 실험

임상 실험의 대상은 13~14세의 비교적 발음 능력이 약한 남자 어린이를 선택하였고, 일주일에 4회 1인당 약 20분씩 훈련을 하게 하였다. 복모음과 자음 중에서 잘 되지 않는 /야/와 /표/을 선택하여 2인의 어린이를 대상으로 훈련을 수행하였다.

그림 9는 복모음 /야/에 대한 훈련의 초기과정에서 발음 도중에 화면상에서의 매개변수 디스플레이를 나타낸 것이다.

그림 9에서 표준 데이타는 하늘색으로 나타나며 훈련자의 현재 발음 상태는 보라색으로 나타난다. 음의 세기, 기본 주파수, 그리고 비음을 나타내는 윈도우에서 각 매개변수의 정도에 따라 막대기가

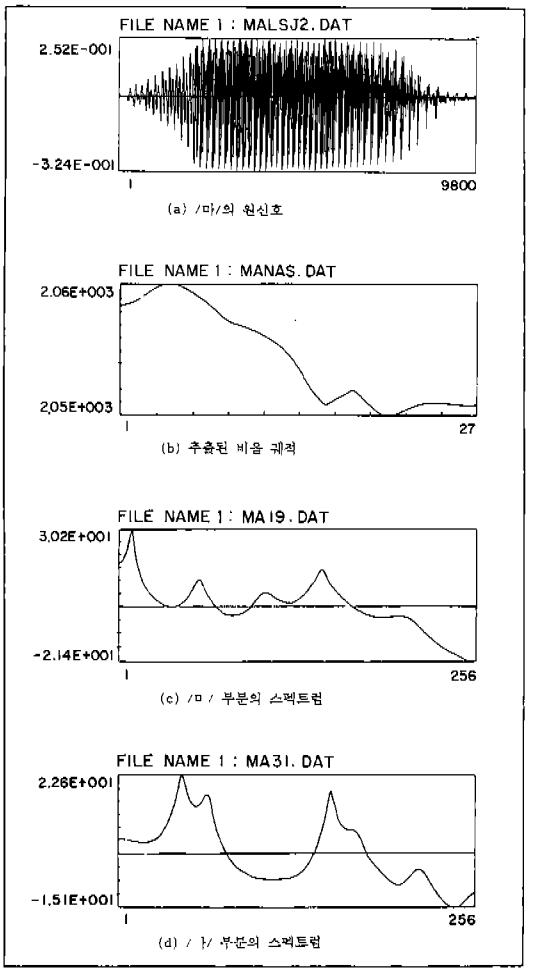


그림 8 /마/에 대한 비음과 대수 스펙트럼
Fig. 8 Nasality and log spectrum for /마/

움직이게 되어 있다. 윈도우의 좌측에 정상적인 기준값을 나타내었다. 스펙트럼과 조음도에서 하늘색이 두개인 것은 복모음의 앞의 성분과 뒤의 성분을 나타내며 훈련자는 이를 차례로 쫓아가도록 훈련하여야 한다.

그림 10에서는 CV 음절 /파/에 대한 훈련의 초기 상태 매개변수 디스플레이를 나타낸다. 그림에서 알 수 있듯이 음의 세기, 기본 주파수, 비음 등의 매개변수는 전체적인 궤적을 나타내게 되어 있다. 이는 모음에서와 같이 움직이는 막대기 형태로 표현하면 자음 부분에서 순간적인 변화를 관찰하

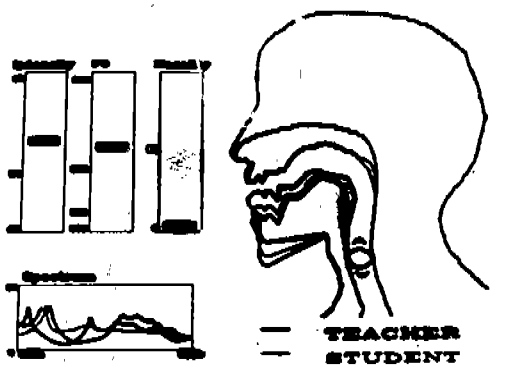


그림 9 복모음 /야/에 대한 발음도중의 화면 출력
Fig. 9 Monitor display in articulating diphthong /야/

기가 어렵기 때문에 이와 같이 디스플레이하기로 하였다.

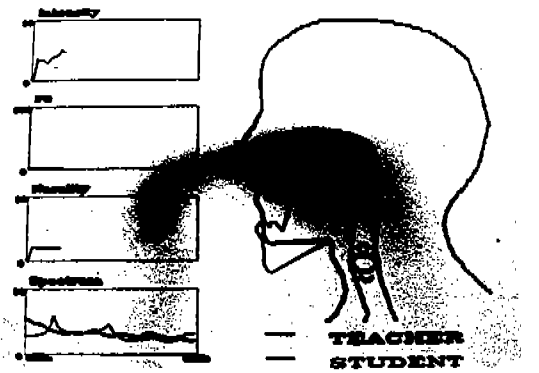


그림 10 CV 음절 /파/의 발음 후 화면 출력
Fig. 10 Monitor display after articulating CV syllable /파/

훈련 결과를 표 2와 3에 나타내었다.

표 3 CV 음절 /파/에 대한 훈련 결과 (단위:%)
Table 3 Progress of training for CV syllable /파/

훈련일	일 치 율	
	P P S	K S W
1일	10	10
2일	0	10
3일	30	10
4일	50	10
5일	70	70

표 2 복모음 /야/에 대한 훈련 결과 (단위:%)

Table 2 Progress of training for diphthong /야/

피검자	훈련일	음의세기	기본주파수	조음도		스펙트럼	
				이	아	이	아
P P S	1일	25	50	0	25	0	0
	2일	50	0	90	30	90	30
	3일	22	0	78	56	56	22
	4일	44	36	64	44	0	32
	5일	71	64	100	43	64	50
	6일	90	50	0	30	100	50
	7일	87	53	0	40	100	80
	8일	95	55	25	90	95	100
	9일	75	44	100	44	100	95
	10일	75	45	10	25	100	100
	11일	90	50	10	35	95	100
K S W	1일	0	20	80	20	0	0
	2일	70	30	100	30	100	40
	3일	44	0	0	78	78	11
	4일	76	84	100	24	48	12
	5일	86	57	100	0	100	79
	6일	100	36	55	18	100	18
	7일	80	53	0	100	93	20
	8일	63	37	16	21	100	100
	9일	88	50	100	13	100	100
	10일	70	95	100	50	100	100
	11일	80	60	50	45	95	100

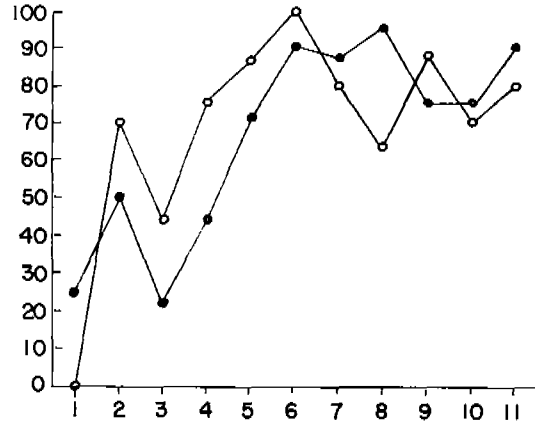
그림 11은 복모음 /야/에 대하여 각 매개변수별로 피검자들의 훈련 결과를 그래프로 나타낸 것이다.

6. 결 론

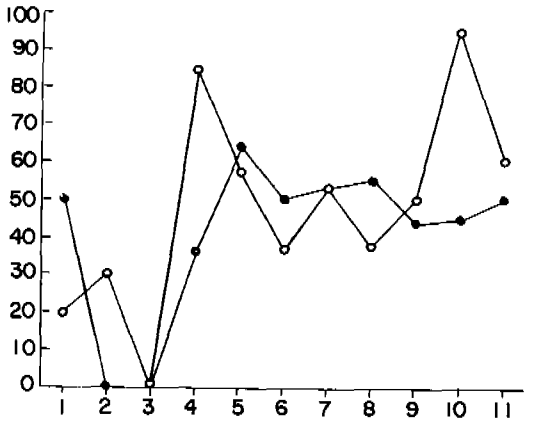
본 연구에서는 청각장애인들의 발음된 음성을 분석하여 표준 발음과 훈련자가 발음한 조음 형태를 동시에 실시간으로 표시하고, 또한 음성 발음에 필요한 여러 가지 매개변수들을 동시에 그래픽상에 표현하여 장애인의 표준발음과 자신의 발음과의 차이를 쉽게 식별하고, 체계적인 훈련 프로그램에 의하여 스스로 음성 발음을 훈련할 수 있는 통합형 발음 훈련 기기를 개발하였다.

발음 훈련 기기의 개발 과정에서 얻어진 결론은 다음과 같다.

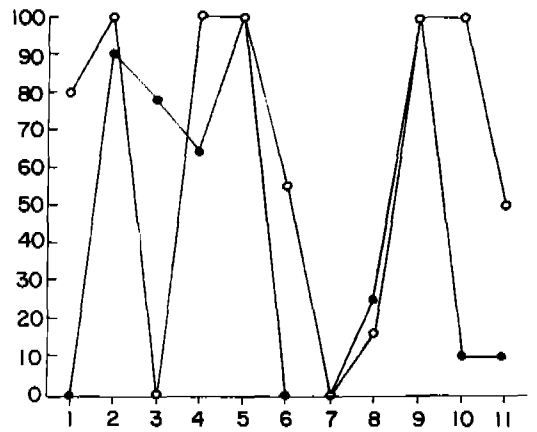
1. 일선 교사의 요망인 발음 기관 내부 표현을 이룸으로써 훈련자가 자신의 발음 기관 중에서 교정이 필요한 부위를 한눈에 알아볼 수 있었다.
2. 표준 발음과 훈련자의 발음을 한 화면에서 색을 달리하여 나타내므로, 일치된 부분을 늘려감으로써 훈련이 효율적으로 이루어질 수 있었다.
3. 발음에 필요한 여러 특징 매개변수들을 조음도 그래픽과 함께 하나의 화면에 나타냄으로써, 발음 상태를 종합적으로 관찰할 수 있는 환경을 제공하였다.
4. TMS 보드를 이용함으로써 모음의 경우 실시간으로 발음 상태를 볼 수 있으므로 훈련자가 발음의 상태를 눈으로 보면서 수시로 교정해 갈 수 있었다.
5. 임상 실험에서 모든 매개변수들이 골고루 향상



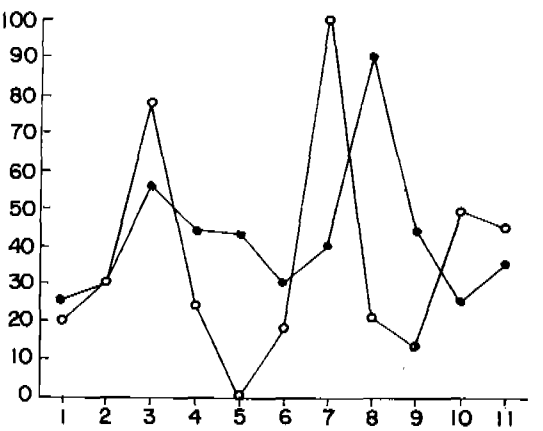
(a) 음의 세기의 변화



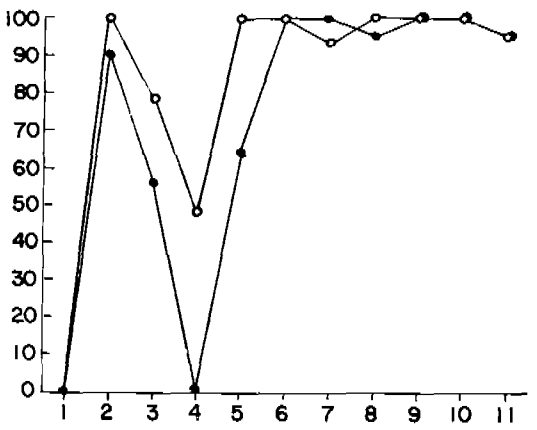
(b) 기본 주파수의 변화



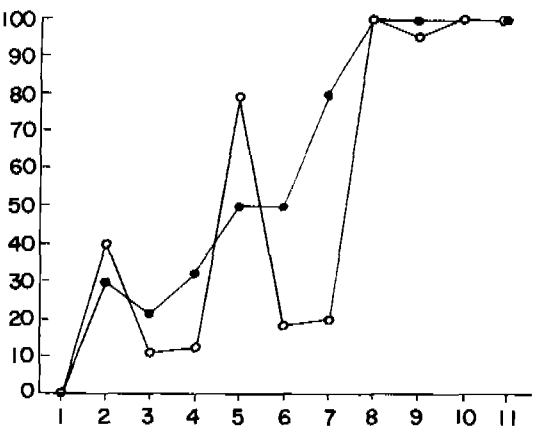
(c) /이/부분 조음도의 변화



(d) /아/부분 조음도의 변화



(e) /이/부분 스펙트럼의 변화



(f) /아/부분 스펙트럼의 변화

그림 11 각 매개변수별 훈련 결과 변화

Fig. 11 Progress of training each parameter

되었다. 즉, 한가지 성분에 대한 훈련에 집중하는 동안 나머지 다른 음성 성분들에 대한 조절 능력이 저하되거나 둔한시되는 단점을 해결하였다.

6. 비교적 발음 능력이 약한 청각 장애 어린이들을 대상으로 복모음 /야/와 자음/ㅍ/을 훈련한 결과, 약 5~6일만에 듣기에 자연스러운 정도의 발음 상태에 이를 수 있었다.

참 고 문 헌

1) 최 참도, "청각 장애아를 위한 음성 언어의 시각 표시 장치 개발에 대한 검토," 난청과 언어 장애 연구, 제5권 제2호, 1982.

2) T. chiba and M. Kajiyama: The Vowel, Its Nature and Structure, Kaiseikan, 1941.

3) R. Harshman, P. Ladefoged, and L. Goldstein, "Factor analysis of tongue shapes", J. Acoust. Soc. Am., Vol. 62, No. 3, September 1977.

4) 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.

5) 윤태성, "청각모델을 이용한 한국어 단음의 인식에 관한 연구," 박사학위논문, 연세대학교 대학원, 1988.

6) P. Ladefoged, R. Harshman, L. Goldstein, and L. Rice, "Generating Vocal Tract Shpaes from Formant Frequencies," J. Acoust, Soc. Am., Vol. 64(4), Oct. 1978.

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) J. D. Markel and A. H. Gray, Jr.: Linear Prediction of Speech, Springer-Verlag, 1980.

9) 김응묵 편역: 최신 아날로그 IC 회로기술, 세화, 1987.

10) Digital Signal Processing Applications with the TMS320 Family, Texas Instruments, 1986.

11) 안상필, 이재혁, 윤태성, 박상희, "청각장애자용 발음 훈련 기기 개발에 관한 연구," 대한전기학회 학술대회, 1990. 7.

12) Dong-Jun Kim, Sang-Pil Ahn, Young-Soo Song, jae-Hyuk Lee, Tae-Sung Yoon, Sang-Hui Park, "A Speech(vowels) Articulatory Motion Estimating System," 16th International Conference on Medical and Biological Engineering, Vol. 29, pp 92, July 1991.

13) 박상희, 김동준, 이재혁, 윤태성, "조음도를 이용한 발음훈련기기의 개발," 전기학회 논문지 41권 2호. pp. 209-216, 1992. 2.

14) 황희영: 한국어 음운개설, 이우 출판사, 1986.