

# 포먼트 통계치를 이용한 발음교정 지시 방법에 관하여

박일서, 조철우

창원대학교 제어계측공학과 음성 및 음향 신호처리 실험실

## Teaching Method of Correct Pronunciation from Formant Statistics

Il-Suh Bak, Cheol-Woo Jo

SASPL, Changwon National University

e-mail : cwjo@sarim.changwon.ac.kr, ilsuh@korea.com

### Abstract

In this paper, we tried to develop a vowel training assistant method using vowel formant statistics. Formant statistics were obtained from PBW set consists of 452 words from 8 persons. Then, we calculated distance from input formants to each center of vowel formant space. Based on the distance, directions to correct the speaker's manner of articulation, i.e. position of jaw and tongue.

### 1. 서론

정보화 시대에 있어서 올바른 발음을 통한 의사소통의 중요성은 날로 증가하고 있다. 그러나, 발성 장애가 있는 사람의 발성을 교정하는 과정은 극소수의 언어 치료 전문가에 의해서만 수행되고 있으며 전문가가 되기 위해서는 많은 시간과 훈련이 필요하다.

그래서, 장애음성 교정을 위한 훈련 방법으로 교정용 소프트웨어를 이용한 방법이 여러 가지 측면에서 시도되고 있다. 소프트웨어에 의한 훈련방법은 시각적인 피드백이 가능하며, 교정대상자에게 구체적인 목표치와 동기를 부여할 수 있기 때문에 효과적인 방법으로 다양한 측면에서 개발되고 있다.[1],[2],[3]

본 논문에서는 음성 처리 기술을 이용한 자동 언어장애 교정 및 훈련을 위한 보조장치의 개발을 목적으로

미리 수집된 음성 샘플의 포먼트를 분석하여 통계분포를 구하고, 데이터베이스에서 분석된 정상 발음의 포먼트 분포의 중심값(평균)과 입력 음성의 포먼트 값을 비교하여 입력된 발음의 교정을 위한 혀와 턱을 위치에 대한 교정 정보를 구하여 발성자의 발화 태도를 수정할 수 있는 방법을 제시하였다.

### II. 제안된 방법의 개요

본 논문에서는 장애 음성을 훈련하기 위해서 장애음성을 가지고 있는 환자가 발화한 음성으로부터 파라미터를 분석하고 분석된 특징을 토대로 발화하고자하는 음성으로 유도해 줄 수 있는 방법을 제안하고자 한다.

일반적으로 모음을 발성할 때, 혀의 위치와 높이에 따른 모음의 분포도는 그림 1과 같다. 이 분포는 포먼트 분포와 유사한 분포를 가지는데 이 특성을 이용하였다.[4],[5]

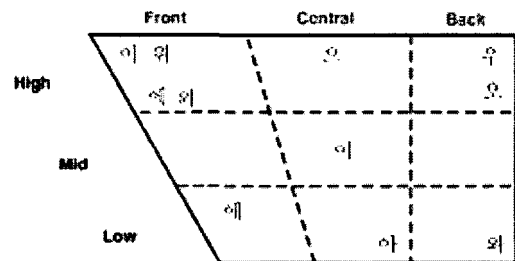


그림 1. 혀에 따른 모음의 조음 위치

제안된 방법에 의한 교정원리는 그림 2와 같다. 먼저

각 모음의 포먼트 분포도를 구한다. 통계치가 일반성을 갖게 하기 위하여 우선 충분한 양의 데이터베이스가 확보되어야 한다. 또한 데이터베이스는 성별, 연령별에 따라 독립적으로 데이터가 확보되어야 한다. 이렇게 구한 통계분포를 기준으로 각 표준모음의 중심점이 추출된다. 그림 1에서는 o로 표시되어 있다.

그 다음 마지의 음성이 입력되면 역시 포먼트 값을 계산하고, 이 값에 의해 각 표준모음과 입력음성의 포먼트 중심값과의 거리를 계산한다. 이때 계산된 거리가 가장 작은 것부터 나열해 보면 입력음성과 표준모음간의 유사도 순위를 계산할 수 있다.

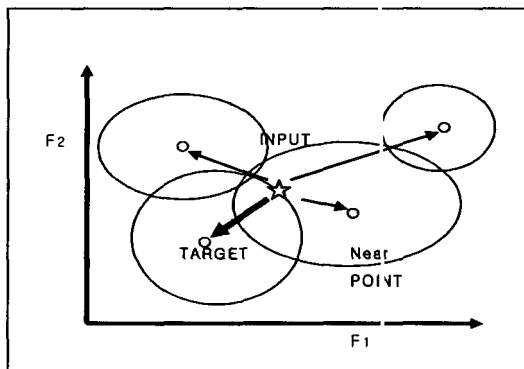


그림 2. 입력음성에 대한 목표음성의 추정

만약 거리가 최소인 음성이 목표음성과 일치하지 않을 경우 F1과 F2의 차를 계산하고 이를 바탕으로 조음방법을 조절하도록 유도하여 입력된 음성이 목표값에 근사하도록 피드백을 시켜준다. 일반적으로 F1, F2값은 각 모음에서 구강내에서의 혀의 위치와 밀접한 상관관계가 있다. 포먼트의 F1에 해당하는 값은 혀의 위치 조절을 통해서 F2에 해당하는 값은 턱의 위치와 혀의 위치를 이용하여 조절이 가능하기 때문이다. 즉, 혀의 위치가 낮을 수록 F1가 높아지고, 낮을수록 F1가 높아지는 경향이 있다. F2은 전설모음의 경우 값이 커지고, 후설모음일수록 값이 작아진다. 턱의 위치는 혀의 위치가 낮을수록, F1의 값이 적을수록 커진다. 이러한 현상을 바탕으로 현재의 발성과 목표발성간의 조음방법의 차이를 추정하고 개선법을 제시해 줄 수 있다. 이런 방법을 사용할 경우 각 화자의 특성, 성별, 연령 등에 따라 포먼트의 분포가 다르기 때문에 포먼트 공간의 정규화 과정이 필요하나 본 실험에서는 별도의 정규화 과정을 거치지 않고 20-30대 남성이나운서의 목소리로 제한하여 실험하였다.[4]

수집된 음성을 분석하기 위하여 선형예측계수(Linear Predictive Coefficient)를 이용해서 각 모음마다 성도의 포먼트를 분석하였다.

시스템의 전달 함수  $H(z) = \frac{1}{A(z)}$  와  $A(z)$ 의 근을 갖는

전극 시스템은  $H(z)$ 의 극점을 표현한다. 그러므로 선형 예측계수 ( $a_i$ )가 분석에 의하여 구해질 때, 다음의 복소수 방정식의 극점들을 결정한다.

$$z^p + a_1 z^{p-1} + a_2 z^{p-2} + \dots + a_{p-1} z + a_p = 0 \quad (1)$$

식 (1)은 실수 계수 값을 가지는 P차의 방정식이며 일반적으로 p/2개의 공액 복소수를 가진다. 만약 한쌍의 근이 각각  $z_i = r_i e^{-j\theta}$ ,  $z_i = r_i e^{j\theta}$  라면 포먼트 주파수는 다음과 같다.

$$f_i = \frac{w_i}{2\pi} = \frac{1}{2\pi} \frac{1}{T} \arg(z_i) \quad (2)$$

여기에서 표본화 주기는 T이며, 분석차수인 p가 0에서  $\frac{1}{2T}$ 의 주파수 범위에 있는 포먼트 개수의 두배라면 위 식의 근은 포먼트들과 일치하게 된다. 만약 p가 포먼트 개수보다 크다면 그 근은 포먼트 뿐만 아니라, 스펙트럼의 작은 봉우리까지 나타내게 된다. 이 실수 극점은 스펙트럼 포락의 기울기를 나타낸다. 본 논문에서는 12차의 LPC 계수의 이용하였다.

각 모음의 포먼트 통계 값과 입력된 음성의 포먼트간의 거리는 유클리디안 방정식을 이용하였다.

### III. 실험 및 결과

우선 발성자의 음성을 구분하는 기준을 정하기 위하여 정상 음성을 가지고 있는 직업인인 20-30대의 남자 아나운서 8명으로부터 녹음된 452개의 고빈도의 고립 단어나 문장을 발음한 SITEC의 PBW DB중의 일부를 사용하였다. 사용된 데이터베이스는 HMM을 이용한 자동 세그멘테이션에 의해 각 음소를 분할한 뒤 모음부분만을 추출하여 사용하였다.

이 DB의 녹음 조건은 방음 부스에서 Senheizer HMD224X를 사용하여 녹음했으며, 디지털 오디오 테이프에 저장된 뒤 A/D는 PC환경에서 실시하였으며, AD/DA Module은 KAY CSL 4300B를 사용하였다. 그리고 16 kHz로 샘플링하고 16 Bits로 양자화 되어 있다.

위의 데이터베이스를 선형 예측 계수를 이용해서 포먼트를 구한 결과를 각 음소 정보에 따른 평균값과 분산값은 표 .1 과 같다.

표.1 모음별 포먼트의 평균과 분산

조음성분		F1		F2		빈도수
영문	한글	평균	분산	평균	분산	
aec	애	468	89	1895	157	1441
axc	ㅏ	662	115	1440	168	1357
coc	ㅑ	502	90	1193	218	1090
euc	ㅡ	409	144	1604	309	829
euic	ㅣ	337	42	2166	321	336
ixc	ㅣ	347	112	2174	234	1198
jac	ㅓ	665	93	1496	175	266
jec	ㅕ	404	72	2047	177	105
jeoc	ㅛ	493	79	1449	237	609
joc	ㅜ	395	64	1375	265	277
juc	ㅠ	334	73	1850	263	217
oxc	ㅗ	388	75	1087	261	932
uxc	ㅜ	372	151	1360	401	708
wac	ㅘ	651	103	1292	191	343
wec	ㅙ	445	70	1799	162	468
wic	ㅚ	332	59	2108	195	259
woc	ㅜ	465	51	1062	153	217

이때 대표음성 '아', '어', '오', '우', '이', '으', '에'에 대한 8명의 남자 아나운서들의 포먼트 분포는 그림.3와 같다.

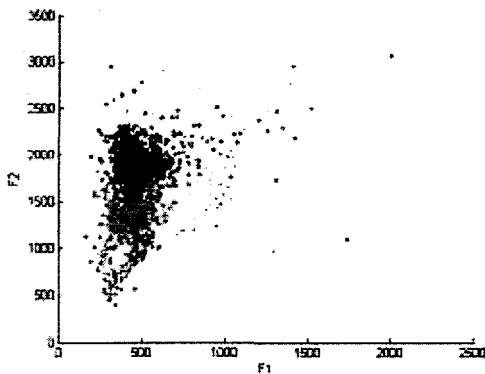


그림3. 발성 데이터베이스의 포먼트 분포

여기에서 발화된 음성이 일정한 포먼트 분포를 가짐을 확인할 수 있었다. 또 각각의 F1과 F2의 분산과 평균값의 간격을 이용하여 훈련기에서 모음을 판단할 수 있는 구간을 정할 수 있다. 즉, F1의 경우 '오'와 '우'의 간격이 가장 작고 그 크기가 10이므로 F1의 동일한 모음으로 판단하는 구간의 최고 값은 5로 할당할 수 있다. 마찬가지로 F2는 '우'와 '아' 구간으로 40의 최고 구간 값을 가진다.

이러한 포먼트 기준을 이용하여 발화된 음성 정보를

분석해보기 위하여 임의의 대상에 대한 음성 데이터를 채집하였다. 사용된 음성은 일반을 대상으로 '어' 발음에 대하여 방음실에서 DAT 녹음기를 이용하여 녹음하였다 16khz의 샘플링 주파수를 가지며 16비트로 양자화 하였다.

녹음된 음성으로부터 포먼트를 구하고, 구해진 포먼트와 구해진 기준값과의 거리를 환산하였다. 이때, 가장 적은 거리값을 가지는 음성을 선택하였고, 동일한 음성에 대해서는 편차값의 차이로 그 음성을 추론하였다. 추론한 결과를 이용하여 혀의 움직임을 보정하였는데 혀의 높낮이는 F1의 양수일 경우 혀를 낮게, 음수일 경우 혀를 높게 유도하였다. 이때 크기는 F1의 절대값이 20보다 작으면 '조금', 20보다 크면 '많이' 움직이도록 하였다. 혀의 위치는 F2의 크기가 양수일 경우 혀를 앞으로, 음수일 경우 뒤로 유도하였고, 움직이는 크기는 F2의 절대값이 200보다 작으면 '조금', 200보다 많으면 '많이' 움직이게 하였다.

다음은 '어' 모음에 대하여 동일 조음으로 판단한 경우와 유사한 다른 조음으로 판단한 경우에 훈련기의 내용과 각 조음간의 크기를 표로 나타낸 것이다

우선, 동일한 조음으로 판단한 경우 표.2에서 분석을 위해 사용된 음성 '어'와 각각기준음성의 F1, F2 값의 차이값과 이 차이를 거리값으로 환산한 내용을 확인할 수 있다

표.2 '어'와 동일 조음으로 판단한 경우의 조음별 크기

	F1	F2	Distance
acx	197	283	345
oxc	-77	-70	104
uxc	-93	203	223
euc	-56	447	450
ixc	-118	1017	1024
aec	3	738	738

사용된 훈련기의 결과는 그림 4

발성된 음성과 통계적 기준치가 유사함을 확인할 수 있고, 좀 더 정확한 발음을 위해 혀를 아주 조금 아래쪽으로 유도하는 내용을 확인할 수 있다.

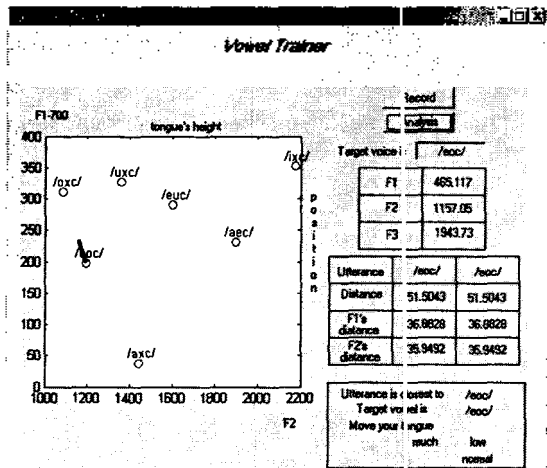


그림4. '어' 와 동일 조음으로 판단한 경우

다음으로 '어' 와 유사한 다른 조음 '어' 로 판단한 경우는 다음과 같다.

표.3과 그림.5에서 각각의 조음간의 간격의 차이의 훈련기의 결과를 확인 할 수 있다.

표.3 '어' 와 유사 위치의 다른 조음으로 판단한 경우의 조음별 크기

	F1	F2	Distance
acx	264	573	630
eax	104	326	342
uax	-26	493	493
euc	11	737	737
ixc	-51	1307	1308
aec	70	1028	1030

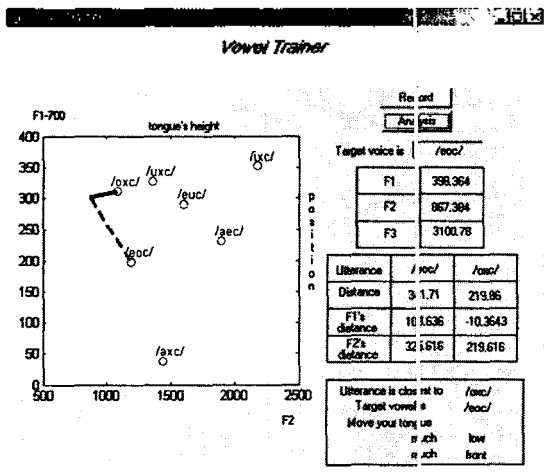


그림5. '어' 를 유사 위치의 다른 조음으로 판단한 경우

이 경우에는 혀를 더 안쪽으로 유도하면 정상적인 발음을 유도 할 수 있을 것이다.

## VI. 결론

본 논문에서는 정상발화음성의 포먼트 값의 통계적 특성을 구하고, 포먼트 공간과 조음공간의 유사도를 이용하여 발화된 음성의 목표음성과의 조음유사도를 판단하여 장애음성을 교정하는 한가지 방법을 제안하였다.

실험에 사용된 음성에서는 입력값에서 기준 점까지의 거리가 최소인 지점이 목표값으로 선택 되거나 최소값에 근사한 값을 가지는 것을 확인 할 수 있었다.

그러나 포먼트의 조음 분포는 연령이나 성별에 따라 차이가 있으므로 포먼트 값의 통계적 분포를 구하는 과정에서의 성별, 연령별 차이에 따른 정규화 과정이 필요하고, 발성 조음과 통계치 사이의 관계를 더 정확히 파악하기 위해서 조음의 분포를 평균값과 함께 GMM과 같은 통계적인 방법을 이용하여 각 조음의 허용 구간을 더 명확히 하는 작업이 필요하다.

## 참고문헌

1. 조철우, 박일서, 정은태, '포먼트 통계치를 이용한 장애모음발음 훈련 보조 방법에 관한 연구', 2003년도 신호처리소사이어티 추계학술발표대회, pp 325-328, 2003
2. ALAN D. BLAIR, JOHN INGRAM 'Learning to Predict the Phonological Structure of English Loanwords in Japanese', Applied Intelligence 19, pp 101-108, 2003
3. K. VICSI, P. ROACH, 'A Multimedia, Multilingual Teaching and Training System for Children with Speech Disorders', INTERNATIONAL JOURNAL OF SPEECH TECHNOLOGY 3, pp289-300, 2000
4. Blamey P.J, Sarant J.Z, Paatsch L.E, 'Effects of Articulation Training on the Production of Trained and Untrained Phonemes in Conversations and Formal Tests', Journal of Deaf Studies and Deaf Education vol. 6, no.1, pp. 32-42, 2001
5. 정일진, '표준어 단순 모음의 세대간 차이에 대한 실험음성학적 분석 연구', 말소리 33-34, pp 111 - 138, 1997