

# 성도 자기공명 영상과 음향정보(F1/F2)를 이용한 한국어 단모음 [이, 에, 아, 오, 우, 으] 판별\*

성철재(충남대), 박종원(충남대), 김귀룡(충북대)

## <차 례>

1. 머리말
2. 성도 자기 공명 영상 수집
3. 성도 분할 및 3차원 복원
  - 3.1. 성도의 분할 작업
  - 3.2. 성도의 3차원 복원
4. 성도 단면적 및 체적 계산
  - 4.1. 성도 단면적 계산
  - 4.2. 성도 체적 계산
5. 성도체적의 모음판별 파라미터 응용
6. 모음판별
  - 6.1. F1, F2 값 측정
  - 6.2. 모음 판별 분석
7. 맺음말

## <Abstract>

### A Vowel Discrimination of Korean Monophthongs [i, e, a, o, u, w] Using Vocal Tract Magnetic Resonance Image and F1/F2

Cheoljae Seong, Jongwon Park, Guiryong Kim

We present a new method of measuring the volume and cross-sectional area of the vocal tract from magnetic resonance images. The vocal tract was divided by the 2 constriction points on the horizontal and vertical planes. The ratios of the volumes of the segment vocal tracts to that of the entire vocal tract play a crucial role in discriminating Korean monophthongs in that vowels were successfully discriminated by the ratios. The discriminant analysis also demonstrated that the acoustic parameters F1 and F2, in addition to the segment volumes, serve as significant parameters in discriminating Korean monophthongs.

\* Keywords: Volume, Cross-sectional area, Vocal tract, 3D MRI, Pixel space, Constriction points

\* 본 논문은 한국학술진흥재단의 인문사회 기초분야 연구지원(KRF-2002-074-AM 1021)의 도움으로 이루어졌다.

## 1. 머리말

성도(vocal tract)의 3차원 모델링 작업은 좀 더 정확한 조음-음향 모델을 얻기 위한 노력의 일환으로 발달되어 왔다. 60 - 70년대 X-ray 영상을 이용한 [1]의 모델과 [2]에 소개된 성도 공명 주파수를 계산하기 위한 면적함수 등은 2차원적 평면에서 성도를 튜브 형태로 모델링한 뒤, 성도를 지나는 중앙선을 기초로 하여 이루어졌다.

[3], [4], [5], [6], [7] 등으로 이어지는 자기공명 영상(MRI)에 바탕한 연구들은, 2차원 중앙 시상면(midsagittal plane)에 국한되었던 지난 연구들의 한계를 넘어서, 직접적으로 3차원 영상을 이용하여 면적 함수를 구할 수 있는 단계로 나아가는 모습을 보여준다. 조음 합성 시스템 구축을 지향하는 이와 같은 일련의 논문들은, 성도 전달함수를 구하기 위하여 다각형 모델과 같은 여러 가지 방법으로 성도를 모델링한 뒤 면적 함수를 구하는 방식으로 연구를 진행하고 있다[8].

한국어의 경우, [9]와 [10]에서 이러한 시도를 찾아볼 수 있다. [9]는 2차원 성도 모델을 MATLAB으로 구현하고 이에 따른 합성음을 만드는 시도와 함께, 자기공명 각 단면 영상으로부터 구한 단면적을 좌표 변환하여 3차원으로 구현하는 과정을 담고 있다. [10]은 자기공명 영상의 정중 시상 단면을 구한 뒤, 성도의 윤곽을 수작업으로 얻고 이를 바탕으로 소프트웨어를 활용하여 성도 단면적과 체적을 구하는 과정을 소개하고 있다. 이렇게 구한 성도 단면적값을 Formfrek 프로그램에 입력하여 포먼트 주파수를 얻고 모음합성을 한 뒤 청취실험을 실시하였다.

지금까지 언급한 대개의 연구들은, 자기공명영상을 통해 성도의 정중 시상 단면 영상을 얻고 그로부터 성도 중앙선을 확정하고 적절한 모델링을 통하여 단면적을 추정하는 공통의 과정을 보여주고 있다. 본 연구에서는, 그러나, 이러한 접근 방법과는 다르게 현대의 발달된 영상처리 기법을 직접적으로 성도모델링에 적용하고자 한다. 좀 더 정확하게는 분절된 성도 자기공명 영상 슬라이스의 픽셀 정보만으로 성도의 단면적과 체적을 계산해보려는 시도를 담고 있다.

이 논문은, 이러한 전체를 바닥에 깔고, 크게 두 가지 주제로 진행될 것이다. 첫번째는 자기공명 영상 슬라이스의 픽셀 정보를 활용하여 한국어 모음을 조음적 관점에서 기술할 수 있는 해부학적 조음 파라미터를 얻는 것이다. 좀 더 구체적으로, 능동 조음체와 수동 조음체가 이루는 전통적 관점에서의 좁힘점을, 슬라이스의 픽셀 정보에 기반한 성도 면적 추적 기법을 활용하여 성도의 수평면과 수직면에서 각각 하나씩 총 2개를 구하고, 이를 중심으로 얻을 수 있는 총 3개 튜브에 대한 체적 비율값을 구하고자 하는 조음 파라미터값으로 삼는다. 두 번째는, 음성 정보를 통해 얻은 음향 파라미터(F1, F2 정보)를 조음 파라미터와 함께 복합적으로 활용하여 한국어 단모음을 기술하고 판별하는 일련의 과정을 시험적으로 제안하는 것이다.

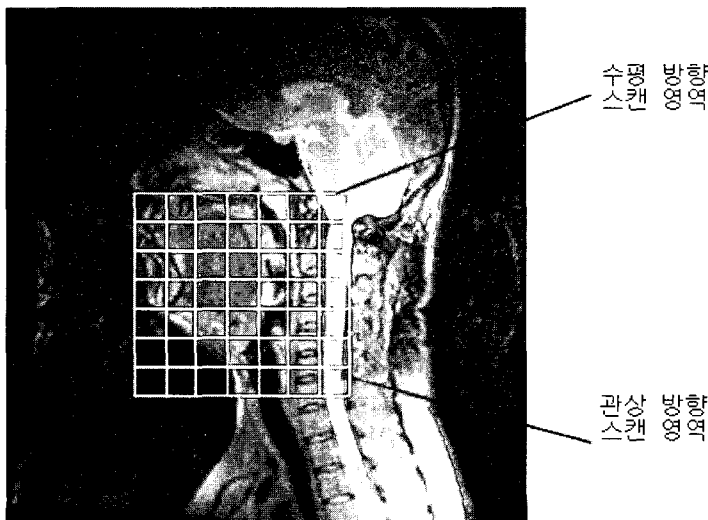
이러한 작업은 한국어 조음 합성기를 구현하기 위한 가장 기초적인 작업의 일환임과 동시에, 세계 언어들 간의 미묘한 조음적 차이를 해부학적 차원에서 기술할 수 있는 새로운 분류 기준을 세우는 작업에 도움이 될 수 있을 것으로 생각한다.

## 2. 성도 자기 공명 영상 수집

성도 자기 공명 영상은 20대 남성을 대상으로 하여 얻었다. 키 172 cm, 몸무게 74 Kg의 성악가로 치아 보철에 의한 영상간섭이 없는 표준어 화자이다. 촬영은 충남대학 병원에서 이루어졌는데, 한국어 단모음 6개 [이, 에, 아, 오, 우, 으]를 각각 25초 내외로 입모양을 최대한 동일하게 유지하며 발성하게 한 후, 수평(Horizontal) 방향으로 60장, 관상(Coronal) 방향으로 36장이 촬영되었다. <표 1>에 상세한 파라미터값이 나타나 있다. <그림 1>은 수평 방향 슬라이스와 관상 방향 슬라이스의 대략적인 촬영 범위를 보여주고 있다.

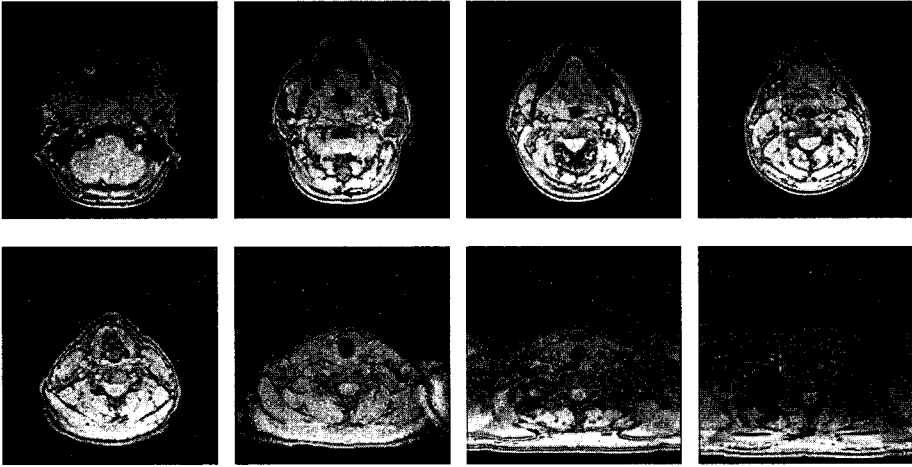
<표 1> 자기공명 영상 촬영을 위한 파라미터

파라미터 명	파라미터 수치
촬영 방향	수평방향 : 60장 관상방향 : 36장
이미지 두께/ 간격	3mm/ 3mm
이미지 크기	256 픽셀 x 256 픽셀
픽셀의 크기	수평방향 : 0.987500mm x 0.987500mm 관상방향 : 0.703135mm x 0.703125mm

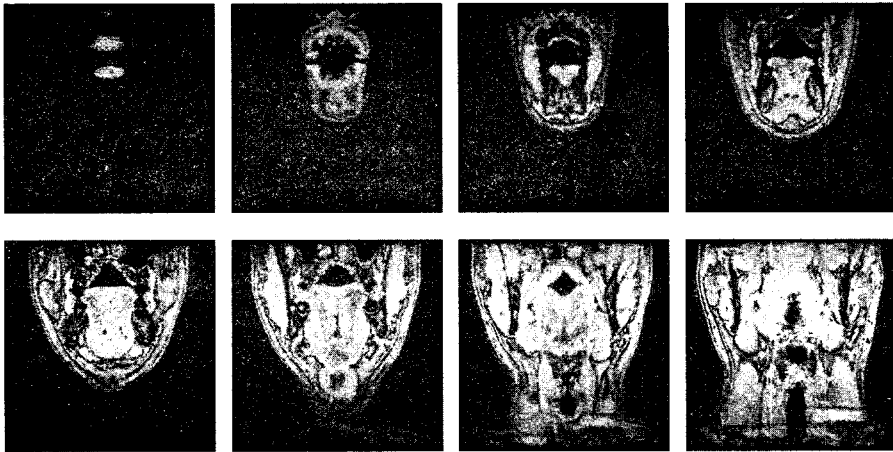


<그림 1> 시상면(sagittal plane)에서 바라본 수평방향 슬라이스와 관상방향 슬라이스의 촬영범위

다음 <그림 2>와 <그림 3>은 본 연구에서 사용된 자기공명 영상 중 모습 [아]에 대한 각 방향별(수평, 관상) 영상 자료 일부를 보여주고 있다.



<그림 2> [아] 모습, 수평 방향의 자기공명 영상



<그림 3> [아] 모습, 관상 방향의 자기공명 영상

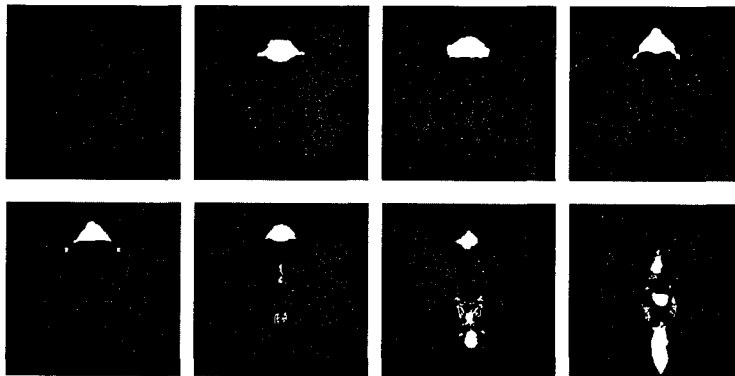
### 3. 성도 분할 및 3차원 복원

#### 3.1. 성도의 분할 작업

본 연구에 사용된 자기공명 영상은 1차적으로 병원의 의료 영상 저장 전달 시스템(PACS)에서 사용되는 다이콤(DICOM) 파일 포맷으로 저장된다. 다이콤 파일은

환자의 신상 정보 및 촬영에 관련된 각종 파라미터와 영상 자료가 모두 포함된 형태의 파일 포맷으로서, 본 연구팀은 이들 파일을 영상처리에 이용하기 위하여 영상 자료 부분만을 재처리하여 성도부분의 분할(segmentation) 작업을 수행하였다. 성도 분할을 위한 첫 번째 단계로, 다이콤 파일로부터 영상 자료 부분만을 획득한 후 이를 흑백 영상을 위한 무 손실 압축 파일 포맷인 PGM 파일 포맷으로 변환하였으며, 두 번째 단계로, 변환된 PGM 파일을 대상으로 일반적인 영상 편집 응용 프로그램과 도구(Tablet)를 활용하여 수동 방식으로 한 슬라이스씩 분할작업을 수행하였다[11, 12, 13, 14, 15]. 최종적으로 얻은 영상은 흰색 픽셀과 검은색 픽셀로만 이루어진 2진 영상 자료로 만들어진다. <그림 4>는 분할작업이 완료된 결과 영상 중 [아] 모음의 관상 방향 원본 영상 처리 결과를 보여주고 있다.

수동 방식의 분할 작업은, 짐작할 수 있듯이, 사람의 노력과 시간을 많이 들여야 하는 작업이라 자동 분할 알고리즘을 만들어내어야지만 경제적인 성과를 올릴 수 있게 된다. 성도의 분할과 관련해서, [5]에서 다루고 있는 3차원 영역 증식(region growing) 방법이 일반적으로 유용하게 사용되고 있다. 그러나, 자기공명영상에서는 치아와 경구개 부분의 분할이 힘들고 3차원 증식만으로는 공기층에 대한 깨끗한 분할이 힘든 까닭에 [7]에서는 치아만을 위한 CT 영상을 따로 찍어서, 이를 자기 공명 영상에 포함시켜 처리하는 방식을 쓰고 있다.



<그림 4> [아] 모음, 관상 방향의 분할 결과 영상

### 3.2. 성도의 3차원 복원

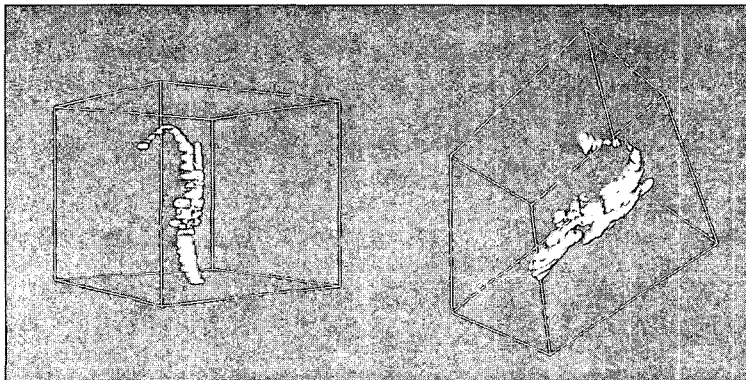
성도의 3차원 복원은, 분할작업 결과로 얻은 영상을 이용하여 이루어지는데 이를 위해 보간 작업이 필요하다. 보간은 분할 작업된 영상을 3차원으로 재구성하기 위한 첫 단계로서, 각각 슬라이스 사이의 연속된 변화를 표현해 주기 위해 임의의 슬라이스를 삽입하는 작업이다. 3차원 재구성을 위한 응용프로그램 중의 하나인 ImagJ를 이용하여 작업을 수행하였다. 작업 완료 후 총 슬라이스 수의 변화는 다

음 <표 2>와 같다.

<표 2> 보간(Interpolation) 작업에 의한 슬라이스 수의 변화

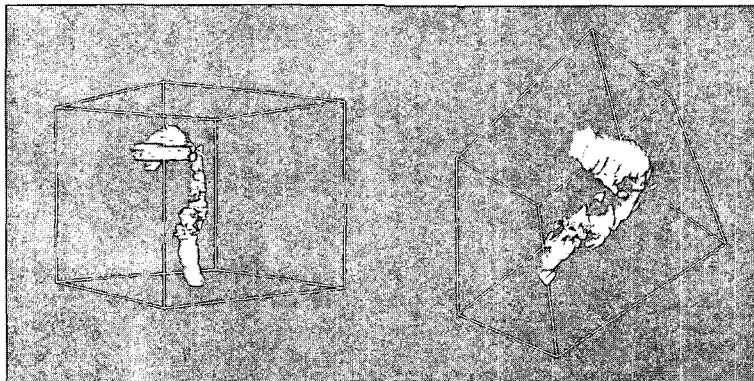
촬영 방향	처리 전 슬라이스 수	처리 후 슬라이스 수
Horizontal	60	378
Coronal	36	299

3차원 복원 작업은 마이크로소프트사의 Visual C++ 6.0과 OpenGL 라이브러리를 이용하여 수행하였으며, 작업의 결과 3차원으로 재구성된 성도의 동영상 파일(상영시간 약 21초)이 제작되었다. <그림 5, 6, 7>은 [이, 아, 우] 모음 동영상 결과물의 수평, 관상 방향의 영상을 보여주고 있다. 복원된 3차원 영상은 모음에 따라 구강과 인두강에서 좁혀지고 확장되는 모습을 잘 보여주고 있다.<sup>2)</sup>



<그림 5> [이] 수평 방향 영상

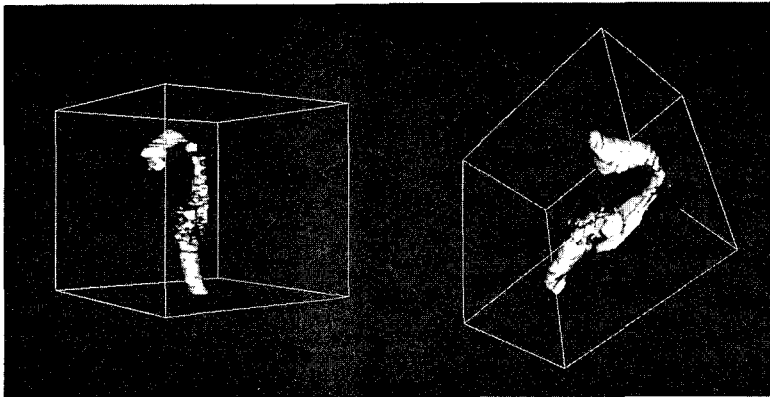
[이] 관상 방향 영상



<그림 6> [아] 수평 방향 영상

[아] 관상 방향 영상

2) 3차원 영상은 VTK(Visualization Toolkit)로 만들었는데, 이 도구는 주로 시각적 효과의 극대화에 초점이 맞추어져 있다. VTK는 광범위한 영상 표현, texture mapping, 특수 효과 및 강력한 시각효과만을 보여줄 뿐 3차원 상태에서 데이터의 측정是不可能的하다.



<그림 7> [우] 수평 방향 영상

[우] 관상 방향 영상

#### 4. 성도 단면적 및 체적 계산

##### 4.1. 성도 단면적 계산

복원된 3차원 이미지에서 실질적인 계산이 불가능하므로, 성도의 단면적과 체적은 자기공명 영상 한 슬라이스에 담겨있는 분할된 성도 이미지 부분 픽셀의 숫자를 활용하여 구하였다. 머리말에서 언급한 다각 모델링의 경우[6, 8], 근사치 접근이라는 한계가 이미 노정되어 있기 때문에 본 논문에서 활용하고 있는 픽셀크기와 한 슬라이스 내에서의 픽셀 숫자를 이용한 단면적 계산법이 좀 더 실측치에 가깝다고 평가한다. 성도의 체적 또한 픽셀을 이용하여 구한 단면적 값에 (높이에 해당하는) 전체 슬라이스의 수와 슬라이스간의 간격을 응용하면 고전적인 면적 합수를 고려하지 않고도 어렵지 않게 구할 수 있다. 이미 이 방법은[11, 12, 13, 14, 15]를 통하여 충분히 검증되었다. 각 모음별 성도의 단면적은 다음 식 (1)의 방식으로 구할 수 있다.

$$A = N \cdot V_p \cdot H_p \quad (1)$$

A: 단면적

N: 한 슬라이스에서 성도에 해당하는 영역의 픽셀 수

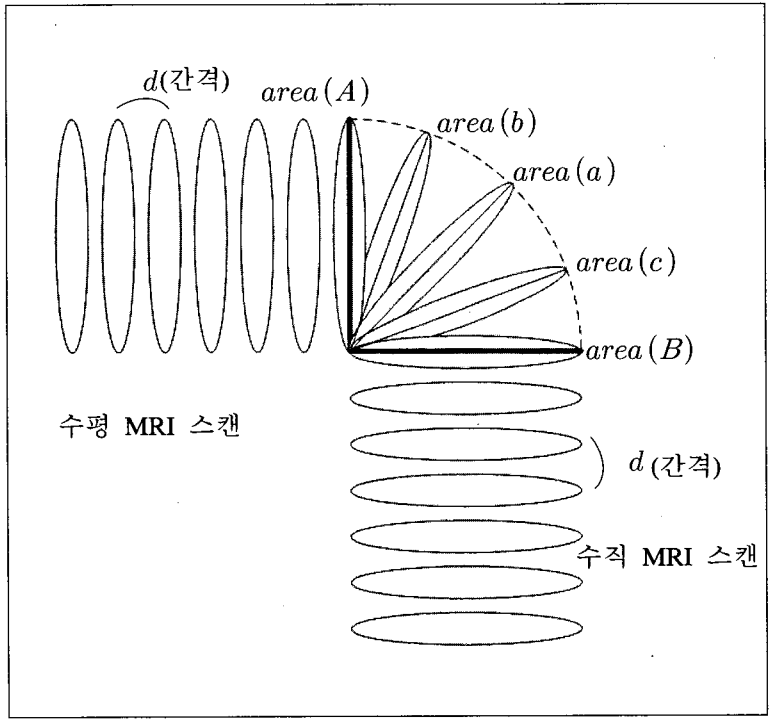
$V_p$ : 수직 픽셀 영역,  $H_p$ : 수평 픽셀 영역

고전적 관점에서 성도의 단면적 값은 머리말에서도 밝혔듯이 성도 전달함수를 구하는 데 있어 중요한 역할을 수행한다. 성도의 가운데를 통과하는 중앙선을 여러 계산 방식에 의해 추정하고, 성문에서부터 이 선을 따라 성도의 단면적 값을 면적합수 공식에 의해 구하게 되면, 공명 주파수로 특징지을 수 있는 성도 전달

함수를 얻게 된다[6, 7]. 즉, 필터의 특성을 구체적으로 기술하기 위해서 반드시 알아야 하는 부분이 성도의 단면적 값 추정이라 할 수 있다. 장비와 기술의 발달로 성도의 단면적을 영상처리 기법을 통해 그렇게 큰 어려움 없이 구할 수 있음을 식 (1)을 통해 알 수 있다.

그러나, 본 연구에서 시도한 방식도 촬영 방향에 의한 단점이 잠재되어 있음을 시인하지 않을 수 없다. 수평방향, 관상방향, 두 방향만의 촬영이 이루어졌기 때문에, 성도가 '기역'자로 꺾어지는 연구개, 목젓 이후 연인두문(velo-pharyngeal port)에서 구-인두(oro-pharynx) 부분 연속된 부분의 단면적 값을 정확하게 추정할 수 없다는 것이다. 성도의 수평면은 수평면과 직각으로 만나는 관상방향의 슬라이스를 이용하면 되고 수직면은 수평 방향 슬라이스를 이용하면 되는데, 원호 모양으로 꺾어지는 부분은 이 면에 직각으로 작용하는 사각(oblique) 방향 영상이 없어 정확한 추정이 안되는 것이다. 이를 보완하기 위해 수직, 수평의 면적 사이값을 연속 추정하는 다음과 같은 수학적 보간 기법을 사용하였다.

<그림 8>에서 수평 MRI로 스캔한 마지막 슬라이스의 단면적을  $area(A)$ 라고 하고 수직 MRI로 스캔한 처음 슬라이스의 단면적을  $area(B)$ 라고 할 때, 중간에 있는 가상 슬라이스의 단면적을 다음 식 (2)와 같이 보간법에 의해 계산할 수 있다.



<그림 8> 보간법에 의한 중간 슬라이스의 단면적 계산 개념도.



$$\begin{aligned}
 area(a) &= \frac{1}{2} (area(A) + area(B)) \\
 area(b) &= \frac{1}{2} (area(A) + area(a)) = \frac{1}{2} \left( area(A) + \frac{1}{2} (area(A) + area(B)) \right) \\
 area(c) &= \frac{1}{2} (area(B) + area(a)) = \frac{1}{2} \left( area(B) + \frac{1}{2} (area(A) + area(B)) \right)
 \end{aligned} \tag{2}$$

위의 식은 다음 식 (3)과 같이 정리될 수 있다.

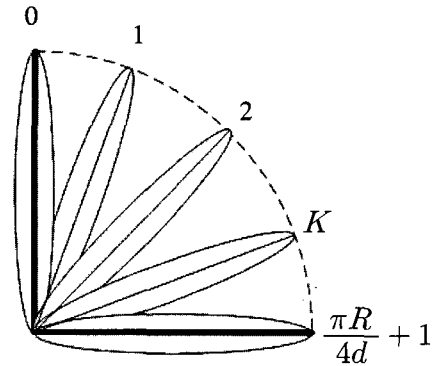
$$\begin{aligned}
 area(a) &= \frac{1}{2} (area(A) + area(B)) \\
 area(b) &= \frac{3}{4} (area(A) + \frac{1}{4} area(B)) \\
 area(c) &= \frac{1}{4} (area(A) + \frac{3}{4} area(B))
 \end{aligned} \tag{3}$$

이 경우, 슬라이스를 원이라고 가정하고, 다음과 같이  $area(A)$  와  $area(B)$  중에서 큰 면적을 가진 슬라이스의 직경을  $R$ 이라고 하면  $R$ 은 다음 식 (4)와 같다.

$$R = \sqrt{\max(area(A), area(B)) / \pi} * 2 \tag{4}$$

그러므로,  $area(A)$ 와  $area(B)$ 의 사이에 같은 간격  $d$ 로 채워 넣어야하는 슬라이스의 수는  $\frac{\pi R}{4d}$  (1/4 원둘레를 동일 간격(d)으로 나눈 것)이고 그 수만큼 보간법으로 슬라이스들의 단면적을 구한다. 따라서, 그림 9에서의  $K$ 번째 위치 슬라이스의 단면적은 다음 식 (5)와 같다.

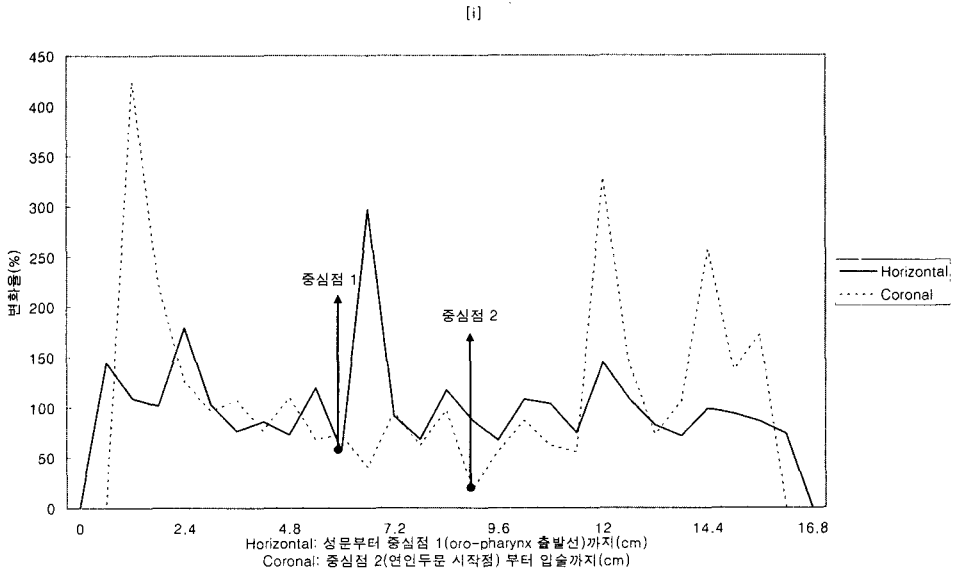
$$K\text{번째 Slice의 단면적} = area(A) \times \frac{\frac{\pi R}{4d} + 1 - K}{\frac{\pi R}{4d} + 1} + area(B) \times \frac{K}{\frac{\pi R}{4d} + 1} \tag{5}$$



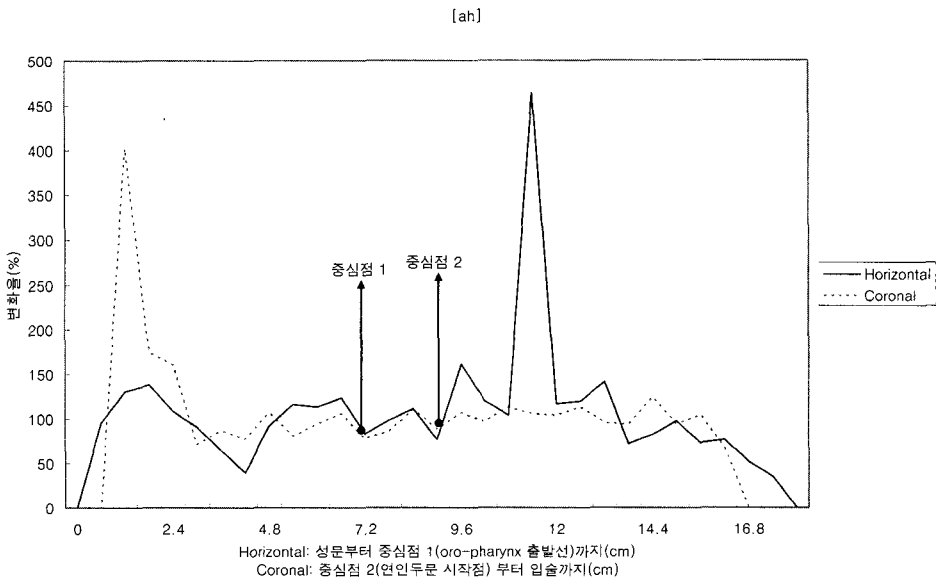
<그림 9> K번째 위치의 슬라이스

이제 남은 문제는, 성대(성문)에서 출발하여 입술 쪽으로 진행하다가 성도의 수직면이 끝나고 곡면(그림 8의 원호 부분) 시작 부위를 나타내는  $area(B)$  슬라이스와, 원호 구간이 끝나고 구강 쪽 수평면 시작을 알리는  $area(A)$  슬라이스를 확정짓는 것이다. 이를 위하여 수평, 관상 각 슬라이스의 방향별 단면적 변화율<sup>3)</sup>을 추적하였다. 수평 방향 슬라이스의 경우는, 성대에서부터 출발하여 성도의 가운데 지점 이전 부근의 면적 변화율이 높아지기 직전 최소값 지점을  $area(B)$  슬라이스로 확정하고, 관상 방향 슬라이스의 경우는 인두벽에서 출발하여 성도의 가운데 지점 부근을 지나 면적 변화율이 최소값으로 떨어지는 지점을  $area(A)$  슬라이스로 확정하였다. <그림 10>과 <그림 11>은 [이, 아] 모음에 대해 두 지점을 확정해 본 내용이다. 그림에서 중심점 1로 표현된 것은  $area(B)$  슬라이스이고, 중심점 2로 지적한 지점은  $area(A)$  슬라이스로 간주하면 된다.

3) 변화율 =  $\left(\frac{\text{current slice}}{\text{past slice}}\right) \times 100$



<그림 10> [이] 모음의  $area(B)$ (중심점 1)와  $area(A)$ (중심점 2) 확정 지점

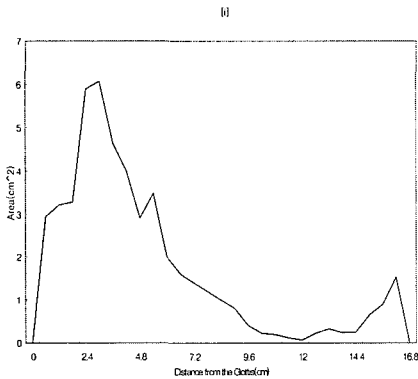


<그림 11> [아] 모음의  $area(B)$ (중심점 1)와  $area(A)$ (중심점 2) 확정 지점

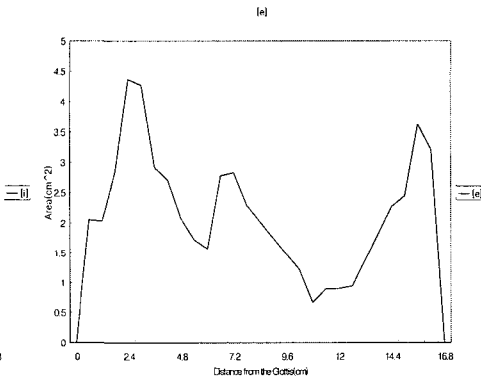
<그림 10>과 <그림 11>에서 실선 부분은 수평방향 슬라이스의 면적 변화율을, 점선으로 표시된 부분은 관상방향 슬라이스의 면적 변화율을 보여준다. 수평방향 슬라이스에서 찾은 단면적은 성문에서부터 출발하여 후-인두(laryngo-pharynx)가 끝

나고 구-인두(oro-pharynx)가 출발하는 선 즉 <그림 10과 11>의 중심점 1까지만 유효하며, 그 이후 값은 의미가 없다. 관상방향 슬라이스에서 찾은 단면적 값은 대략 연인두문 시작점 부근 혹은 목젓과 연구개 후단부 부근(<그림 10과 11>의 중심점 2 부분)에서 입술까지의 결과만 유효하며 나머지는 의미가 없다.

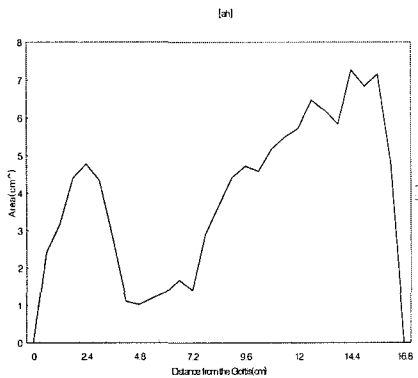
성대에서 입술까지의 단면적은 결국, 세 부분 계산 결과의 조합으로 완성된다. 첫 번째는 성대에서부터 중심점 1까지의 수평방향 슬라이스 단면적 값을 이용한 성도 수직면 부분이고, 두 번째는 중심점 2에서 입술까지의 관상방향 슬라이스 단면적 값을 이용한 성도 수평면 부분이다. 세 번째는 중심점 1과 2 사이의 유사-원호로 모델링된 부분의 수학적 추정치이다. 이렇게 하여 추정된 성도 전체의 단면적 값을, 각 모음별로 <그림 12~17>에 제시한다. x-축은 입술로부터의 거리를 cm로 나타내고, y-축은 각 지점에서 측정된 성도 단면적 값을 cm<sup>2</sup>로 나타낸다.



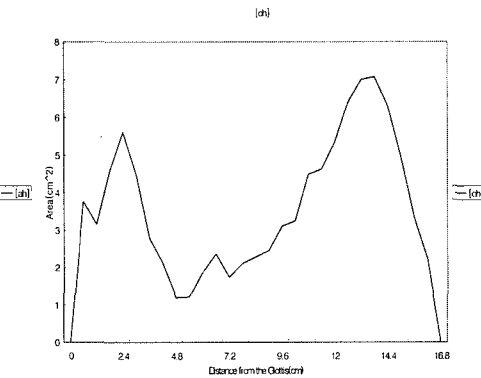
<그림 12> [이]의 성도 단면적



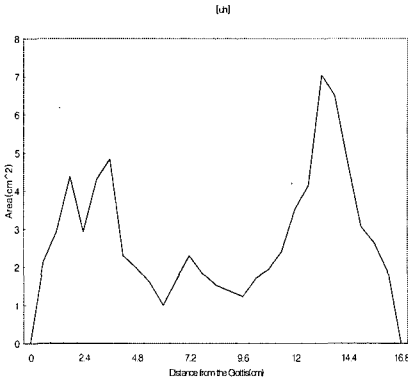
<그림 13> [에]의 성도 단면적



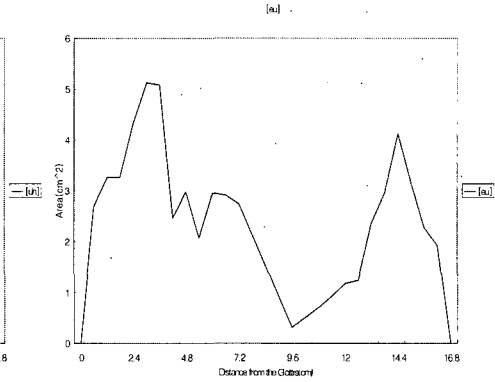
<그림 14> [아]의 성도 단면적



<그림 15> [오]의 성도 단면적



<그림 16> [우]의 성도 단면적



<그림 17> [오]의 성도 단면적

#### 4.2. 성도 체적 계산

이미 언급했듯이, 성도의 체적은 성도로 분절된 슬라이스의 픽셀 정보에서 얻은 단면적 값에 전체 슬라이스의 숫자와 슬라이스간 간격을 곱해주면 얻어질 수 있다. 이를 공식화시키면 식 (6)과 같이 표현될 수 있다.

$$V_i = \begin{cases} p_i \cdot s_i \cdot t_i \cdot \frac{d_i}{2} & , i = 1, N \\ p_i \cdot s_i \cdot t_i \cdot d_i & , \text{이외의 경우} \end{cases} \quad (6)$$

$V_i$ :  $i$  번째 슬라이스의 성도체적

$N$ : 전체 슬라이스의 수

$p_i$ :  $i$  번째 슬라이스에서 성도에 해당하는 영역의 픽셀 수

$s_i$ :  $i$  번째 슬라이스에서 한 픽셀의 실제 크기

$t_i$ :  $i$  번째 슬라이스에서 한 슬라이스의 두께

$d_i$ :  $i$  번째 슬라이스에서 슬라이스간의 간격

이렇게 해서 나온 각 모음별 성도 체적값은 다음 <표 3>과 같다.

<표 3> 각 모음별 성도 체적값(ml)

모음	체적(ml)
이	53.8
에	45.8
아	98.2
오	77.2
우	62.2
으	55.2

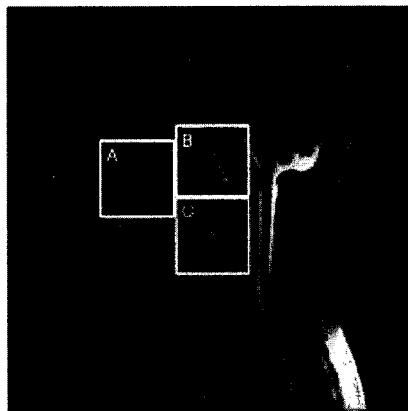
## 5. 성도 체적의 모음 판별 파라미터 응용

모음을 조음할 때 혀의 최고점보다는 좁힘점의 위치와 그 간극이 모음의 음가를 결정하는 데 있어서 더 중요하게 작용한다는 조음음성학적 지식에 바탕하여 ‘좁힘점’을 모음 기술에 활용할 수 있는 방법을 찾아보고자 한다[16]. 성도 내에서 각 모음이 보여주는 좁힘점의 위치는 각 모음의 조음자리 특성과 어느 정도 연관 관계를 가진다. [이, 에, 애] 등의 전설 모음 그룹은 성도의 앞쪽, 즉 경구개 근처에서 좁힘점이 형성되고, [으, 우] 모음은 연구개 부위에서, [어, 아, 오] 등의 모음은 후인두에서 좁힘점이 만들어진다고 한다[17].

좁힘점의 위치가 모음에 따라 위와 같이 각기 다르며 그 위치도 다르게 형성되므로, 이를 모음 판별 작업에 유용한 파라미터로 활용하기 위해 각 모음별 좁힘점을 성도 수평면과 수직면으로 분할하여 2개의 값을 구하였다. 체적의 관점에서 각 좁힘점을 기준으로 세 개의 튜브를 구분 짓는 결과를 낳게 된다. 각 영역의 체적을 구강 앞쪽 부분부터 차례로 체적 A, 체적 B, 체적 C로 명명하겠다. <표 4>에 각 영역의 정의가 나와 있다. <그림 18>은 이 세 영역을 명시적으로 잘 보여주고 있다.

<표 4> 성도 수평면과 수직면의 각 좁힘점을 중심으로 구분된 세 튜브의 체적 정의

- |                                     |
|-------------------------------------|
| A: 성도 수평면 좁힘점과 입술 사이이 성도 체적         |
| B: 성도 수평면 좁힘점과 성도 수직면 좁힘점 사이의 성도 체적 |
| C: 성도 수직면 좁힘점과 성문 사이의 성도 체적         |



<그림 18> 성도 수평/수직 좁힘점을 중심으로 구분한 3 영역(A, B, C)

각 모음의 성도 수평면과 수직면 좁힘점은 4장의 내용을 최대한 응용하면 쉽게 얻어진다. <그림 10과 11>에서 수평 방향 슬라이스를 통해 얻은 단면적 값은

[성문에서 중심점 1까지] 유효하다고 했으며, 관상 방향 슬라이스를 이용하여 얻은 단면적 값은 [중심점 2에서 입술까지] 유효하다고 하였다. 성도의 수평면(구강)과 수직면(인두강)에서 각각 좁힘점을 하나씩 구하는 것이므로, 수평방향 슬라이스 단면적 유효 구간에서 최소값 하나를 찾고, 관상방향 슬라이스 단면적 유효구간에서 최소값 하나를 찾으면 총 2개의 성도 최소 단면적값 즉 좁힘점을 구하게 된다. <표 5>는 [아] 모음을 대상으로 좁힘점 2개와 <표 4>의 A, B, C 세 튜브를 결정짓는 과정을 담고 있다. 성문에서 출발하여 4.8cm 부근의 단면적이  $1.03 \text{ cm}^2$ 로서 성도 수직면 즉 인두강의 최소 단면적 지점을 나타내고 있으며 이는 일반적으로 알려져 있는 [아] 모음 좁힘점과 일치하는 곳이 된다.

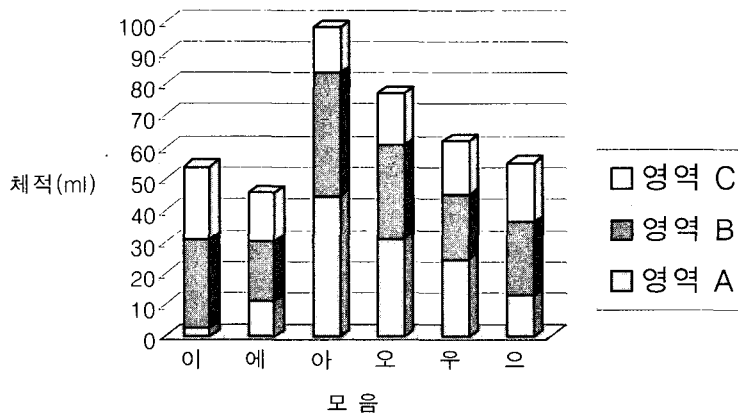
<표 5> [아] 모음의 중심점 1/2, 성도 수평/ 수직면 좁힘점, 성도 A/B/C 영역 확정

성문-입술	단면적	중심점	좁힘점	영역
0	0			C영역
0.6	2.4			
1.2	3.2			
1.8	4.4			
2.4	4.8			
3	4.3			
3.6	2.8			
4.2	1.1			
4.8	1.0		수직면 좁힘점	
5.4	1.2			B영역
6	1.4			
6.6	1.7			
7.2	1.4	중심점1		
7.8	2.9			
8.4	3.7			A영역
9	4.4	중심점2	수평면 좁힘점	
9.6	4.7			
10.2	4.6			
10.8	5.2			
11.4	5.5			
12	5.7			
12.6	6.5			
13.2	6.2			
13.8	5.8			
14.4	7.3			
15	6.8			
15.6	7.2			
16.2	4.8			
16.8	0			

좁힘점 두 군대를 중심으로 구한 A, B, C 영역의 체적과 비율 값이 <표 6>에 제시되어 있다. <그림 19>는 이 내용을 수직 누적 막대 그래프로 표현해본 것이다.

<표 6> 성도의 영역 분할에 따른 체적(ml)/ ( )는 비율값

모음	영역 A	영역 B	영역 C	전체 영역
이	2.5(4.6)	28.3(52.5)	23.1(42.9)	53.9(100)
에	10.9(23.8)	19.0(41.4)	15.9(34.8)	45.8(100)
아	44.8(45.6)	38.9(39.7)	14.4(14.7)	98.1(100)
오	31.0(40.2)	29.6(38.3)	16.6(21.5)	77.2(100)
우	24.5(39.4)	20.7(33.2)	17.0(27.4)	62.2(100)
으	13.0(23.5)	23.5(42.5)	18.8(34)	55.3(100)



<그림 19> 각 모음의 성도 영역 분할에 따른 체적값

각 영역로 부여된 비율값은 음향공간의 F1, F2 값과 마찬가지로 그 모음 고유의 특성을 나타내주는 파라미터로 설정할 수 있다. 어떤 언어의 모음이든 자기공명 영상 정보만 있다면 그 조음적 특성을 상호 비교 분석할 수 있는 파라미터라 할 만 하다. 이러한 비율값을 응용하여 한국어 6개 단모음을 성도의 체적 파라미터로 기술하면, 각 모음별 성도 전체 체적값과 비율값을 곱하여 단순하게 표현할 수 있을 것이다. [이] 모음을 예로 들면, 세 부분의 체적이 이루는 비율값이 4.6:52.5:42.9이므로, 전체 체적인 53.9(ml)을 이 세 영역 비율값으로 표현해 본다면  $(0.05 \times 53.9) + (0.53 \times 53.9) + (0.43 \times 53.9)$ 가 될 것이다. <표 7>은 각 모음에 대해서 이러한 표현을 일반화해 본 것이다.



<표 7> 각 모음의 좁힘점으로 구분된 성도 체적 비율 일반화 표현

모음	성도 체적 비율 일반화 표현
[이]	$0.05V[이]+0.53V[이]+0.43V[이]$
[에]	$0.24V[에]+0.41V[에]+0.35V[에]$
[아]	$0.46V[아]+0.4V[아]+0.15V[아]$
[오]	$0.4V[오]+0.38V[오]+0.22V[오]$
[우]	$0.39V[우]+0.33V[우]+0.27V[우]$
[으]	$0.24V[으]+0.43V[으]+0.34[으]$

(각 계수는 순서대로, 성도 내 A, B, C 체적 비율값, V[\*]: 각 모음별 성도 전체의 체적)

## 6. 모음 판별

이 연구의 결과와 함께, 음향 파라미터인 모음의 F1, F2를 첨가하여 총 5개의 파라미터를 대상으로, 통계학적 판별분석(discriminant analysis)을 수행해 보았다. 5개의 파라미터를 정리하면, 성도를 세부분으로 구분하여 구분한 A, B, C 각 영역의 체적값 3개, 그리고 F1, F2 값이다. 각 모음에 대한 체적 변수 측정결과가 하나의 자료 밖에 없으므로, 원 값을 기준으로  $\pm 5\%$ (F1, F2는 Hz, 각 영역의 체적은 ml)씩 차례로 더하고 빼서 하나의 변수 당 총 20개의 토큰을 새롭게 생성하여 통계 작업을 하였다(최종적으로  $\pm 50\%$  까지 진행).

일반적으로 20개 정도의 데이터 샘플이 있어야 통계적으로 의미 있는 결과가 얻어진다고 하는데 주어진 영상 실험 자료는 1인의 1회 자료 밖에 없으므로 임의로 20개 정도의 자료를 만들 필요가 있었다. 이러한 임의성은 당연히 언어학적 중요도와 반비례하겠지만 추후, 실제의 자료를 얻어 동일한 통계과정을 수행할 경우를 예상하여 시험적으로 판별함수와 모음 재배치결과를 얻어 보는데 의의가 있다고 할 수 있다. 따라서 이 작업의 결과는 실제의 데이터를 이용하여 진행한 것과 많은 차이가 날 수도 있고, 그렇지 않을 수도 있다.

### 6.1. F1, F2 값 측정

피험자는 자기공명 영상 촬영에 참가한 동일 인물이다. 한국어 단모음 목록 중 [아, 에, 이, 오, 우, 으]를 무작위로 10회 반복하였다. 자기공명영상 촬영이 누운 상태에서 이루어졌으므로, 음성 자료 역시 누운 상태에서 얻었다. 자기공명 영상 촬영기 내에서는 강한 자기장과 소음으로 인해 동시 녹음이 불가능하다. 이런 경우 사용하는 특수한 마이크가 있다고 하나 구할 수 없었고, 내부 소음이 워낙 커서 처음부터 시도할 엄두를 내지 못했다. 평상시 직립의 자세에서 녹음한 모음 음

가와 누워서 발성한 음가 사이에 어느 정도의 편차가 있음은 예상할 수 있는 사실이다. 성도 체적과 좁힘점에 미묘한 차이가 생기기 때문이다.

녹음은 음악 및 영화 전문 스튜디오인 M&F에서 이루어졌는데, Digidesign 사의 소프트웨어, 컨트롤러, 컨버터와 애플 컴퓨터를 본체로 한 Protools system을 이용하여 Panasonic의 SV-3800 DAT로 옮겼다. 마이크는 Neumann의 컨덴서 마이크인 U87Ai를 사용하였다. 녹음 시 표본화율은 48 kHz로 하였다. 분석을 위한 컴퓨터 입력 시(A/D) 표본화율은 42.1 kHz로 낮추었으며 양자화는 16 bit로 하였다. 포먼트 주파수는 Praat 4.3.12를 이용하여 측정하였다. Praat 4.3.12에서 포먼트 측정을 위한 세팅 조건은 다음과 같이 이루어졌다.

<표 8> Praat 4.3.12에서의 포먼트 검출 조건

parameter	setting	기타
time step	10 msec	
maximum number of Formants	5 개	
maximum Formant frequency	5500 Hz	
window length	25 msec	Gaussian-like
pre-emphasis	50 Hz	a=0.97
dynamic-range	20 dB	
method	Burg	

[오, 우]와 같이 제 1, 2 포먼트가 가까이 있을 경우, 최대 포먼트 주파수 영역을 4,000 Hz로 하고 구하는 포먼트 개수를 4개로 하여 측정하였다. [이, 에, 아, 오, 우, 으]의 모음 안정구간(모음 신호 시작부에서 약 1/3 지점) 약 25 msec 정도의 구간을 잡아 그 평균치로 측정하였다. <표 9>에 전체 측정 결과와 평균값, 표준편차 값을 제시한다.

<표 9> [이, 에, 아, 오, 우, 으] 모음의 F1, F2, 평균, 표준편차(Hz)

		1	2	3	4	5	6	7	8	9	10	mean	s.d.
아	F1	661	686	720	652	645	620	702	715	750	695	685	39.8
	F2	1152	1194	1204	1500	1602	1187	1185	1207	1227	1154	1261	156.3
에	F1	448	518	452	443	460	503	511	495	451	515	480	31.3
	F2	2005	1998	1857	1758	1954	2114	2050	1657	1784	2001	1918	146.6
이	F1	277	244	203	282	245	227	265	274	285	291	259	28.7
	F2	2154	2546	2448	2540	2215	2003	1987	2067	2234	2167	2236	208.2
오	F1	362	377	398	402	392	421	405	387	411	385	394	17.2
	F2	746	758	784	725	695	710	745	708	796	743	741	32.7
우	F1	341	320	345	310	324	345	350	335	323	340	333	13.2
	F2	810	905	878	901	925	800	911	807	838	879	865	47.5
으	F1	384	354	305	370	335	374	355	390	308	312	349	31.9
	F2	1237	1140	1154	1355	1622	1158	1245	1130	1110	1204	1236	153.9

## 6.2. 모음 판별분석

각 모음별로 Fisher의 판별함수(discriminant function)와 정준(canonical) 판별 함수를 구하였고, 이 함수를 바탕으로 원래 모음 재배치를 해본 결과, 100 %의 결과를 얻어 내었다. [아] 모음을 예로 들어, 판별분석에 사용한  $\pm 5\%$ 씩 증감한 변수값을 제시하면 다음 <표 10>과 같다. 음향 파라미터인 F1, F2의 경우, 원래 10개의 측정값이 있었으므로, 그대로 사용할 수도 있었으나, 다른 파라미터와의 적용 방법을 동일하게 하기 위하여 똑같이 평균값으로부터  $\pm 5\%$  씩 증감하여 전체 변수를 구성하였다.

<표 10> 모음 [아]의 각 파라미터별  $\pm 5\%$ (F1, F2는 Hz, 각 영역의 체적은 ml) 증감 자료

모 음	F1(Hz)	F2(Hz)	vol_a(ml)	vol_b(ml)	vol_c(ml)	증감
아	342	631	22.4	19.5	7.2	-50%
아	377	694	24.7	21.4	7.9	-45%
아	411	757	26.9	23.4	8.7	-40%
아	445	820	29.1	25.3	9.4	-35%
아	479	883	31.4	27.3	10.1	-30%
아	514	946	33.6	29.2	10.8	-25%
아	548	1009	35.9	31.2	11.6	-20%
아	582	1072	38.1	33.1	12.3	-15%
아	616	1135	40.4	35.0	13.0	-10%
아	650	1198	42.6	36.9	13.7	-5%
아	685	1261	44.8	38.9	14.4	0%
아	719	1324	47.1	40.9	15.2	+5%
아	753	1387	49.3	42.8	15.9	+10%
아	787	1450	51.6	44.8	16.6	+15%
아	822	1513	53.8	46.7	17.3	+20%
아	856	1577	56.0	48.7	18.1	+25%
아	890	1640	58.3	50.6	18.8	+30%
아	924	1703	60.5	52.6	19.5	+35%
아	958	1766	62.8	54.5	20.2	+40%
아	993	1829	65.0	56.5	20.9	+45%
아	1027	1892	67.3	58.4	21.7	+50%

Fisher의 판별함수 결과는 다음과 같이 나타났다.

<표 11> 모음 [이, 에, 아, 오, 우, 으]에 대한 Fisher의 선형 판별 분석결과

	모 음					
	이	에	아	오	우	으
F1(Hz)	-2.59	.145	-.011	.004	-.251	.429
F2(Hz)	.059	.021	.019	-.034	.062	-.132
vol_a(ml)	1.610	.100	.486	-.552	6.146	-7.539
vol_b(ml)	1.608	-3.343	2.087	1.407	-6.222	3.662
vol_c(ml)	-2.229	.412	-4.357	1.624	3.686	4.288
(상수)	-32.909	-29.285	-30.046	-15.910	-29.389	-29.057

이에 의해 각 모음은 다음과 같은 판별함수로 표현될 수 있다.

$$F([\text{아}]) = -30.046 - 0.011[F1] + 0.019[F2] + 0.486[\text{vol\_a}] \\ + 2.087[\text{vol\_b}] - 4.367[\text{vol\_c}]$$

$$F([\text{에}]) = -29.285 + 0.145[F1] + 0.021[F2] + 0.1[\text{vol\_a}] \\ - 3.343[\text{vol\_b}] + 0.412[\text{vol\_c}]$$

$$F([\text{이}]) = -32.909 - 0.259[F1] + 0.059[F2] + 1.610[\text{vol\_a}] \\ + 1.608[\text{vol\_b}] - 2.229[\text{vol\_c}]$$

$$F([\text{오}]) = -15.910 + 0.004[F1] - 0.034[F2] - 0.552[\text{vol\_a}] \\ + 1.407[\text{vol\_b}] + 1.624[\text{vol\_c}]$$

$$F([\text{우}]) = -29.389 - 0.251[F1] + 0.062[F2] + 6.146[\text{vol\_a}] \\ - 6.222[\text{vol\_b}] + 3.686[\text{vol\_c}]$$

$$F([\text{으}]) = -29.057 + 0.429[F1] - 0.132[F2] - 7.539[\text{vol\_a}] \\ + 3.662[\text{vol\_b}] + 4.288[\text{vol\_c}]$$

각 모음 그룹의 데이터는 평균 벡터와 공분산 행렬로 정리된다. 각 그룹이 모집단 수준에서 동일한 공분산 행렬을 갖는다고 가정하면 임의의 관측 벡터가 여타의 다른 그룹 중심점(centroid)보다 특정 그룹의 중심점에 더 가까이 있다고 판단할 기준은 마할라노비스(Mahalanobis) 거리의 대소로 표현된다. 이를 위 여섯 개의 식으로 설명해보면, 각 분류함수에 독립변수 값 즉, 동일 데이터 열의 [F1], [F2], [vol\_A], [vol\_B], [vol\_C] 값을 대입하여 결과적으로 가장 큰 값(Mahalanobis distance)을 제시하는 모음의 집단으로 재대입 분류가 이루어진다. 예컨대, [아] 모음의 특정 데이터열에서 [F1], [F2], [vol\_A], [vol\_B], [vol\_C] 값을 여섯 개의 판별함수식에 집어넣게 되면, [아] 모음의 마할라노비스 거리값이 가장 크게 나타나게 된다.

<표 12>는 위 판별함수에 근거하여 재대입 분류를 한 결과, 100% 원래 모음으로의 재배치가 이루어졌음을 보여준다. 1회 측정 자료를 대상으로 ±5% 씩 20단계의 증감을 가한 인위적 데이터로 이루어진 통계이므로 재대입 분류의 결과 자체는 언어학적으로 중요한 의미를 가지지 못한다. 다만, 실측된 자료가 통계적으로 의미 있는 양이 확보되었다고 가정할 경우, 위와 같은 과정을 거쳐 판별 함수를 도출하고, 미지 확률만큼의 재대입 분류가 이루어질 수 있을 것이라는 점을 밝히기 위한 실험적 시도라고 이해해주시기 바란다.

<표 12> Fisher의 판별함수에 의한 재대입 분류 결과(괄호 속은 백분율(%))

		모음	예측 소속집단						전체
			1	2	3	4	5	6	
원래값	빈도(%)	1	21(100)	0	0	0	0	0	21(100)
		2	0	21(100)	0	0	0	0	21(100)
		3	0	0	21(100)	0	0	0	21(100)
		4	0	0	0	21(100)	0	0	21(100)
		5	0	0	0	0	21(100)	0	21(100)
		6	0	0	0	0	0	21(100)	21(100)
교차 유효값(a)	빈도(%)	1	21(100)	0	0	0	0	0	21(100)
		2	0	21(100)	0	0	0	0	21(100)
		3	0	0	21(100)	0	0	0	21(100)
		4	0	0	0	21(100)	0	0	21(100)
		5	0	0	0	0	21(100)	0	21(100)
		6	0	0	0	0	0	21(100)	21(100)

## 7. 맺음말

성도(필터)의 특성을 가장 잘 나타내는 음향변수는 성도의 공명특성인 F1, F2 값이라 볼 수 있다. 성도를 3차원으로 표현하여 각 모음에 따라 달라지는 좁힘점과 이에 따라 구분되는 세 가지 영역의 체적값을 비율값으로 나타내본 이번 연구의 접근법은 성도의 조음적 특성을 나타내는 새로운 물리적 접근법이라 할 만하다. 음향적 파라미터인 F1, F2 값, 조음적 파라미터인 체적값 등을 변수로 활용하여 설정된 모음 판별함수는, 다양한 모음의 음향적/조음적 특성을 동시에 기술하고 비교할 수 있게 하는 자질 특성이라 할 수 있다. 지금까지의 결과를 정리하면 다음과 같다.

첫째, 한국어 단모음 [이, 에, 아, 오, 우, 으]의 성도 체적을 자기 공명 영상을 이용한 3차원 영상 처리 기법으로 구해본 결과, 성도의 수평, 수직면에서의 좁힘점 두개를 중심으로 분할한 A:B:C 튜브 영역의 체적 비율값이 [이]는 4.6:52.5:42.9, [에]는 23.8:41.4:34.8, [아]는 45.6:39.7:14.7, [오]는 40.2:38.3:21.5, [우]는 39.4:33.2:27.4, [으]는 23.5:42.5:34로 나타났다.

둘째, 성문에서 입술까지 연속적으로 측정할 수 있는 성도 단면적의 관점에서, 최소 단면적 형성은 [이]는 성문에서 12 cm, [에]는 10.8 cm, [아]는 4.8 cm, [오]는 4.8 cm, [우]는 6 cm, [으]는 9.6 cm에서 이루어졌다.

셋째, 음향 파라미터인 F1, F2 값과 체적 파라미터인 성도 내 3 영역(A/B/C) 체적

값, 총 5개 파라미터의  $\pm 5\%$  씩 20단계의 증감 데이터를 이용한 모음 판별 함수 (Fisher's discriminant function)를 구하여 보았다. 이 함수를 이용하여, 전체 독립 변수를 재배치한 결과, 100 % 원래의 모음으로 재분류됨을 확인하였다.

넷째, 성도의 단면적과 체적을 구하는 데 있어, 자기공명 영상 슬라이스의 픽셀 영역과 수에 기반한 새로운 영상처리 기법을 적용하였다.

자기공명영상을 이용한 연구는 경제적, 시간적(주로 촬영 자체에 걸리는 시간)인 문제로 인해 난관을 겪게 된다. 실험내용에서, 한국어 단모음 목록 8개를 6개로 줄인 1차적 이유이다. 또한, 이 연구는 향후 본격적 연구에 앞선 파일럿 실험의 성격이 강하기에 모음사각도에서 단모음의 윤곽을 최대한 얻어낼 수 있는 [이, 예, 아, 어, 오, 우] 6개 모음만으로 실험 자료를 한정하여 진행하였다. 따라서 [애, 어] 모음은 이번 실험에서는 다루어지지 못했다. 후속 연구에서는 단모음 8개 모두를 대상으로 특징 분석과 기술을 시도할 예정이다. 향후, 이러한 방법론을 다양한 외국어에 적용하여, 각 나라의 모음 목록을 상호 비교 분석해보려고 하며, 얻어진 조음 파라미터를 심분 활용하여, 양질의 조음 합성기 개발에도 적용할 예정이다.

## 참 고 문 헌

- [1] G. Fant, *Acoustic Theory of Speech Production*, Mouton: den Haag, 1960.
- [2] J. Liljencrants and G. Fant, "Computer program for VT-resonance frequency calculations", *KTH STL-QPSR*, Vol. 4, pp.15-20, 1975.
- [3] T. Baer, J. Gore et al., "Analysis of Vocal Tract Shape and Dimension using Magnetic Resonance Imaging: Vowels", *JASA*, 90(2), pp.799-828, 1991.
- [4] S. Narayanan, A. Alwan, and K. Haker, "Towards Articulatory-Acoustic Models for Liquid Approximants Based on MRI and EPG Data", *JASA*, 101(2), pp.1064-1089, 1995.
- [5] I. Titze and B. Story, "Vocal Tract Area Functions from Magnetic Resonance Imaging", *JASA* 100(1), pp.537-554, 1996.
- [6] O. Engwall, "Modeling of the Vocal Tract in 3 Dimensions", *Proc. of Eurospeech 1999*, pp. 113-116, 1999.
- [7] J. Behrends & A. Wismüller, "A Segmentation and Analysis Method for MRI Data of the Human Vocal Tract", *FIPKM*, 37, pp.179-189, 2001.
- [8] A. Foldvik, U. Kristiansen et al., "3-Dimensional and Magnetic Resonance Imaging: A new Dimension in Phonetic Research", *Proc. of ICPHS 95*, pp.46-49, 1995.
- [9] 조철우, "MRI를 이용한 조음모델 시뮬레이터 구현에 관하여", *음성과학*, 2권, pp.45-55, 1997.
- [10] 양병근, "MRI에 의한 모음의 성도 단면적 측정 및 면적 변이에 따른 합성 연구", *음성*

과학, 4권 1호, pp.19-34, 1998.

- [11] Y. C. Sung, C. J. Song et al., "Tissue Segmentation and Volumetry on Brain MR Image of Coronal Section having the Partial Volume Artifact", *Proc. of 2000 IEEE International Symposium on Intelligent Signal Processing and Communication Systems*, Vol. I, pp.369-372, 2000.
- [12] Y. C. Sung, K. S. Han et al., "Threshold Estimation for Region Segmentation on MR image of Brain having the Partial Volume Artifact", *5th International Conference on Signal Processing Proceedings*, Vol.II, pp.1000-1009, 2000.
- [13] Y. C. Sung, C. J. Song et al., "Tissue Segmentation and Volumetry on Brain MR Image of Coronal Section having the Partial Volume Artifact", *Proceeding of 2000 IEEE International Symposium on Intelligent Signal Processing and Communication Systems*, Vol.I, pp.369-372, 2000.
- [14] Y. C. Sung, C. J. Song et al., "Fully Automatic Segmentation and Volumetry on Brain MRI of Coronal Section", *2000 International Conference on Circuits/Systems, Computers and Communications*, Vol.1, pp.441-445, 2000.
- [15] D. Y. Kim, J. W. Park, "Computerized quantification of carotid artery stenosis using MRI axial images", *Magenetic Resonance Imaging*, Vol 22/3, pp.353-359, 2004.
- [16] 이호영, 국어음성학, 서울: 태학사, 1996.
- [17] 김영송, 우리말 소리의 연구, 고천판, 서울: 과학사, 1981.

접수일자: 2005년 11월 17일

게재결정: 2005년 12월 20일

▶ 성철재(Cheol Jae Seong) : 교신저자

주소: 305-764 대전광역시 유성구 궁동 220번지 충남대학교

소속: 충남대학교 인문대학 언어학과

전화: 042) 8221-6395

E-mail: cjseong@cnu.ac.kr

▶ 박종원(Jong Won Park)

주소: 305-764 대전광역시 유성구 궁동 220번지 충남대학교

소속: 충남대학교 공과대학 정보통신공학과

전화: 042) 821-6863

E-mail: jwpark@cnu.ac.kr

▶ 김귀룡(Gui-Ryong Kim)

주소: 361-763 충북 청주시 흥덕구 개신동 산 48번지

소속: 충북대학교 인문대학 철학과

전화: 043) 261-2150

E-mail: tragon@chungbuk.ac.kr