

'ArtSim'을 이용한 모음의 조음점 추정에 관한 연구

김 대현, 조 철우(창원대학교 제어계측공학과)

<차 례>

1. 서론
2. 기본모음에 대한 이해
3. 조음합성시뮬레이터
4. 조음점(혀의 최고점) 추정을 위한 실험
 - 4.1 조음점의 설정방법
 - 4.2 혀 형태의 제한
 - 4.3 합성 및 청취실험
5. 분석 및 검토
6. 결론

<Abstract>

Estimation of Articulatory Characteristics of Vowels Using 'ArtSim'

Dae-Hyun Kim et al.

In this paper, articulatory simulator 'ArtSim' is used as a tool for the experiments to examine the articulatory characteristics of 6 different vowels. Each vowels are defined by some articulatory points from their vocal tract area functions and shapes of tongues.

Each points are varied systematically to synthesize vowels and the synthesized sound is evaluated by human listners. Finally distributions of each vowels within vowel space is obtained. From the experimental results it is verified that our articulatory simulator can be used effectively to investigate the articulatory characteristics of speech.

1. 서론

사람의 음성은 혀의 최고점의 높이와 위치, 그리고 입술의 모양에 따라 모음을 분류하고 모음의 음가를 나타내고 비교하는 데 많은 어려움이 있다. 입술의 모양은 눈으로 관찰할 수 있어서 어떤 모음이 원순 모음인지 비원순 모음인지 쉽게 판단할 수 있다. 그러나 혀의 위치는 눈으로 관찰할 수 없고, 모음을 조음할 때 구강 안에 막음이 나 마찰이 생길 정도의 좁힘이 형성되지 않기 때문에 혀의 위치를 감각으로도 느끼기 어렵다.

혀의 위치를 객관적으로 관찰할 수 있는 방법은 X-Ray, MRI나 CT기계 등을 이용하는 것이다. 이러한 기계를 이용하면 구강 안에서의 혀의 위치를 정확하게 파악할 수 있지만 여러 가지 문제점이 있다.

X-Ray는 방사선의 위험 때문에 많은 양을 조사하는 것이 안전하지 않고, 휴대할 수 없기 때문에 기동성이 떨어지며, 많은 양의 자료를 분석할 수 없다. MRI는 촬영에 소요되는 비용이 막대하므로 이러한 기기가 설치되어 있는 병원 등 기관의 도움이 없이는 사용이 불가능하다. 뿐만 아니라, 이러한 것을 이용해 혀 최고점의 높이와 위치를 정확하게 수치로 나타낸다 하더라도 이 수치가 모음의 음가에 대한 우리들의 청각인상(auditory impression)을 정확하게 반영하지 못한다. 구강의 크기가 사람마다 다르기 때문에 같은 모음을 발음하더라도 입천장과 혀의 최고점과의 거리가 사람마다 차이날 수밖에 없기 때문이다. 그리고 MRI를 이용한다고 하더라도 다양한 조음위치의 변화를 임의로 바꾸어가면서 음의 변화를 관찰하는 것은 어렵다. 현재 여러 가지 성도 형상을 이용한 조음합성에 관한 연구결과가 많이 나와 있지만 아직 그러한 조음합성기를 이용하여 음성의 구강 내에서의 공간적 분포를 측정해 본 경우는 없다.

본 연구에서는 성도의 단면을 입술과 혀의 변화에 따른 모음의 위치분포를 찾기 위해 본 연구실에서 개발한 'ArtSim'이란 조음합성 시뮬레이터를 이용해서 혀 최고점의 위치에 따른 입술과 혀의 형태를 여러 가지로 변화시켜 얻은 성도함수(단면적)를 구하고, 이를 이용해 합성된 음을 인지실험을 통해 분석하였다.

이러한 실험의 필요성은 음성합성을 할 경우 조음결합을 조음모델에 의해 구현하고자 한다면 동일한 음소의 경우도 다양한 조음위치가 설정될 필요가 있기 때문에 마치 조음삼각도 예서와 같은 조음점의 분포를 추정하여 차후 합성기의 제작에 필요한 조음점의 이동에 따른 정확한 성도파라미터의 추정에 이용할 수 있을 것이다. [1][2]

2. 기본 모음에 대한 이해

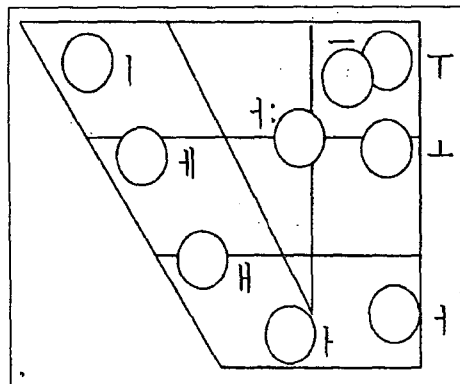
모음은 음가에 영향을 미치는 요인들을 기준으로 하여 분류된다. 모음의 음가를 결정하는 가장 중요한 요인은 혀의 위치와 입술의 모양이다. 턱의 움직임도 모음의 음가에 영향을 미치나 혀의 위치와 입술의 모양만큼 중요하게 기능하지는 않는다.

모음은 혀 최고점의 높이, 즉 혀의 최고점이 입천장으로부터 얼마나 떨어졌느냐에 따라 고모음(high vowel : ㅣ, ㅜ, ㅡ)과 저모음(low vowel : ㅏ, ㅓ)으로 나눈다. 고모음과 저모음 사이에 중모음을 설정하기도 하고, 중고모음(ㅘ, ㅙ)과 중저모음(ㅞ)을 설정하기도 한다.

모음은 또한 혀 최고점의 위치, 즉 혀의 최고점이 구강의 앞쪽에 위치하는가 뒤쪽에 위치하는가에 따라 전설모음(front vowel : ㅣ, ㅘ, ㅙ)과 후설모음(back vowel : ㅜ, ㅡ, ㅓ, ㅗ)으로 나눈다. 전설모음과 후설모음 사이에 중설모음(mid vowel : ㅚ)을 설정하기도 한다. 이러한 혀의 최고점은 모음을 분류하고, 모음을 발음할 때의 혀의 상대적인 위치를 기술하고, 모음의 발음을 훈련시키는 데에 효과적으로 사용된다.

그리고 모음은 입술의 모양에 따라 입술을 둥글게 오므리고 발음하는 원순모음(rounded vowel : ㅜ, ㅓ)과 입술을 둥글이지 않고 발음하는 비원순모음(unrounded vowel : ㅣ, ㅡ, ㅘ, ㅙ, ㅏ, ㅓ)으로 분류한다.

국어의 단순모음들의 음가는 <그림 1>과 같다. [1]



<그림 1> 국어의 단순모음 사각도

3. 조음합성 시뮬레이터

본 연구실에서 개발한 조음합성 시뮬레이터인 'ArtSim'을 이용하여 성도의 단면적을 얻어 모음을 합성하였다.

'ArtSim'은 먼저 MRI 이미지에서 얻은 그림을 이용해 먼저 전체적인 성도와 입술의 형태를 구성한다. 그런 후 합성하고자 하는 음에 대한 혀의 형태와 입술모양을 입력하는 부분을 선택하여 입술과 혀의 형태를 결정하는 혀뿌리, 혀끝, 혀의 최고점, 목젓 부근의 혀, 후두개, 후두개의 끝 부분을 차례로 입력하여 6개의 부분을 interpolation함으로써 혀의 형상을 구성하고 cutting 방법에 따라 성도의 단면적을 구한다. 이렇게 하여 구한 성도의 단면적을 이용해 합성음 및 스펙트럼 등을 구한다.

Cutting 방법은 세 가지가 있는데, 첫 번째는 하스킨스(Haskins) 연구실에 사용한 방법과 유사하며 두 번째 방법은 성도의 형태를 고려해 cutting한다. 그리고 세 번째 방법은 두 번째 방법에서 구한 성도의 중심을 기준으로 그 중심선과 수직으로 만나는 선을 이용하여 구하는 방법이다.[5]

조음합성을 위해 먼저 성도의 단면적과 디지털 필터 모델을 이용하여 전달함수를 구하였다. 그리고 이 전달함수를 원하는 주파수 범위로 resample하여 12차 all-pole 형태로 근사화 함으로써 파라미터를 구하였다. 이 파라미터를 이용해 실험용 단모음을 만들 때 피치와 크기의 변화를 주어 합성 음성을 만들었다.[3][4][7]

4. 조음점(혀의 최고점) 추정을 위한 실험

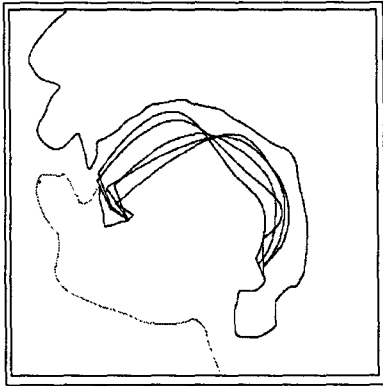
4.1 조음점의 설정방법

모음의 음가에 영향을 주는 것은 혀 최고점의 높이와 위치 그리고 입술의 모양이다. 여기서는 MRI 이미지에서 얻은 다섯 개의 모음(ㅏ, ㅓ, ㅗ, ㅜ, ㅡ)를 이용해 혀의 최고점의 위치를 정하였다. 그림 2에 다섯 개의 모음을 나타내었다. 조음합성 시뮬레이터인 'ArtSim'에서 입술과 혀를 입력하는 부분에 필요한 7개의 부분 중에서 혀의 뿌리 부분과 혀의 후두개 끝 부분은 고정시키고, 입술과 혀끝, 혀의 최고점, 목젓 부근의 혀, 후두개 부분은 변화를 주면서 음을 합성하였다. 여기서 혀의 최고점이 주요한 부분이므로 다른 부분보다 다양한 변화를 주었다. 또한 목젓 부근의 혀와 후두개 부분은 혀 최고점의 위치에 따라서 변하게 하였다.

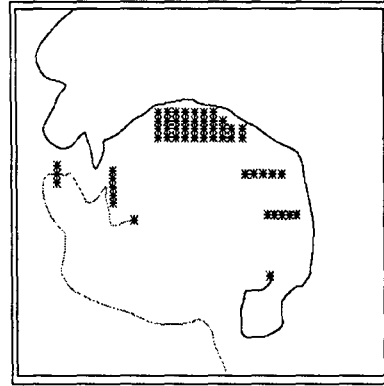
입술부분은 3개의 점을 이용해 변화를 주었으며, 혀끝은 MRI 이미지에서 얻은 5

개의 부분으로 하였고, 혀의 최고점은 35개의 점으로 나누어 다양한 변화를 주었다. 그리고 목젓부근의 혀와 후두개는 각각 5개의 점으로 혀의 형태 변화에 이용하였다.

<그림 3>에 입술과 혀 형태의 변화를 주는 모든 점을 나타내었다.



<그림 2> 5개 모음의 혀의 형태



<그림 3> 입술과 혀의 형태 변화점

4.2 혀의 형태의 제한

입술과 혀의 형태 변화점을 이용하여 혀의 형태를 변화시킬 때 실제의 혀에서는 나타나지 않는 혀의 형태가 나타나게 된다. 이런 경우를 제외시키기 위해 약간의 제한을 두었다.

혀의 최고점에 따라 혀끝과 목젓부근의 혀, 그리고 후두개를 변화시켰다. 혀의 최고점이 앞으로 갈 경우에는 목젓부근의 혀와 후두개가 다같이 인두벽에서 멀어지며, 뒤쪽으로 갈 경우에는 인두벽에 가까워진다.

또한, 혀의 최고점이 위쪽으로 올라갈 경우(경구개에 가까워질 경우)에도 목젓부근의 혀와 후두개는 인두벽에서 멀어진다. 이러한 제한 사항을 이용해 테이블 형태로 규칙화하여 혀의 형태를 변화시켜 가면서 270개 혀의 형태를 형성하였다.[6][8]

4.3 합성 및 청취실험

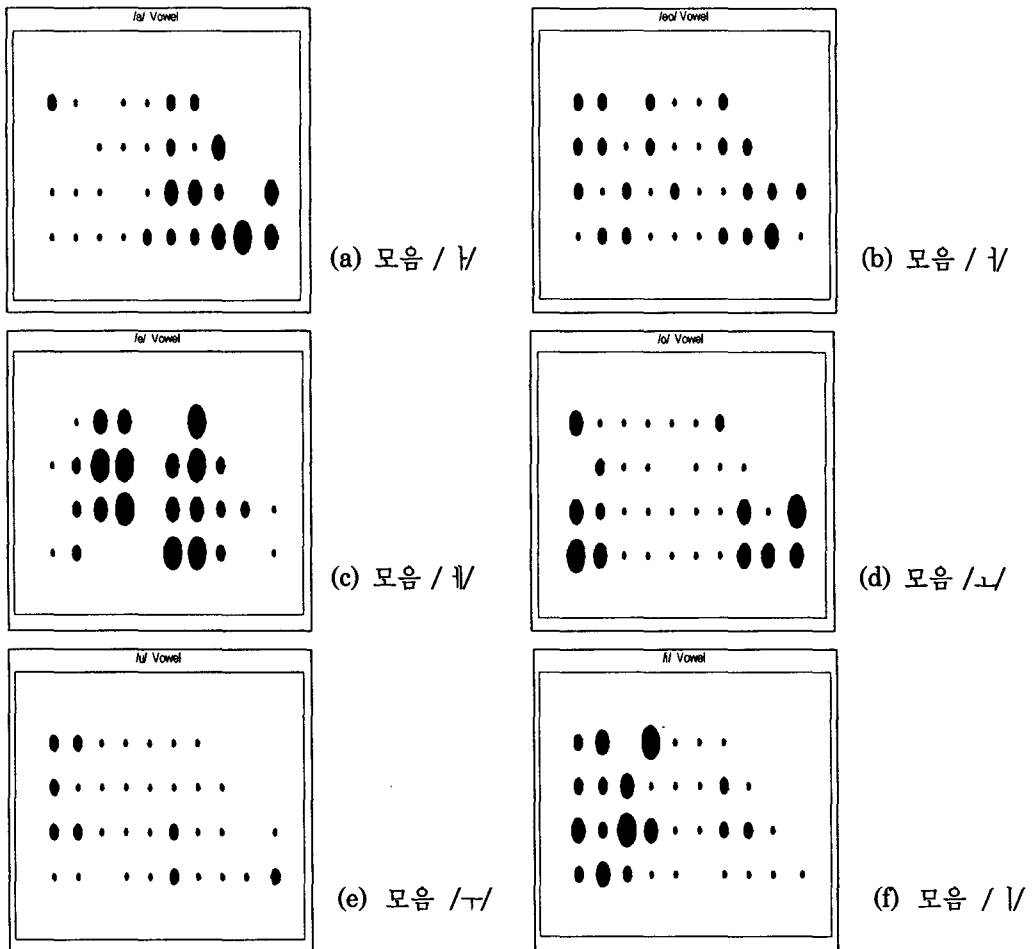
위에서 얻은 270개 정도의 혀의 형태를 가지고 조음합성 시뮬레이터인 'ArtSim'에 입술과 혀의 입력부분에 각각을 대입하여 성도의 단면적 및 합성음, 그리고 스펙트럼을 얻었다.

그 후 여기서 얻은 합성음을 이용해 청취실험을 하였다. 청취실험은 10명의 청취자를 선택하여 합성된 음을 각각 3번씩 들려주어 6개의 모음(ㅏ, ㅓ, ㅗ, ㅜ, ㅝ, ㅟ)과

‘그 외의 음’ 중에서 하나를 선택하게 하였다. 각 청취자는 헤드폰을 쓰고 독립적으로 실험을 행하였다.

5. 분석 및 검토

청취실험을 통해 얻은 6개의 모음(ㅏ, ㅑ, ㅓ, ㅕ, ㅗ, ㅣ)의 결과를 그림 4에 나타내었다. 그림에서 점으로 표시한 부분은 모음에 따른 혀의 최고점 분포를 나타내고 있으며, 왼쪽은 전설이고 오른쪽은 후설 부분을 의미한다.



<그림 4> 6개의 모음에 대한 청취실험 결과

먼저 그림 4 (a)의 모음 /ɪ/에 대한 결과에서는 분포가 뒷부분의 아래쪽에 많이 위치해 있다. /ɪ/음은 중설 저모음이며 혀의 최고점이 중설에 형성되며, 좁힘은 혀뿌리와 인두벽 사이에 형성된다. 그러나 그림에서는 약간의 차이가 있다. 그 이유는 좁힘점이 뒤쪽에 위치하므로 혀의 최고점의 위치가 뒤쪽으로 갈수록 후두개의 입력부분이 뒤쪽으로 위치하기 때문인 것 같다. 그리고 그림 4 (c)의 모음 /ɨ/에 대한 분포는 가운데 부분에 모여있다. /ɨ/음은 전설 중고모음으로 혀의 최고점이 전설에 형성되고, 좁힘점도 전설과 경구개 사이에 형성된다. 그러나 실험결과가 예측과 달리 나타난 것은 청취자가 전설 중저모음인 /ɨ/와 같이 인식한 것이라 판단된다. 그리고 그림 3 (d)의 모음 /ɪ/의 분포는 앞쪽에 위치해 있다. 모음 /ɪ/는 전설고모음으로 혀의 최고점은 전설에 형성되며, 좁힘점도 전설과 경구개 사이에 형성된다. 그림에서도 비슷하게 위치해 있다.

여기서 볼 수 있는 바와 같이 /ɪ/, /ɨ/, /ɪ/의 3개의 모음에 대한 인지도를 나타내는 그래프가 한 곳에 모여있다. 다른 모음들은 분산되어 있거나, /ɪ/와 같이 앞쪽과 뒤쪽에 나타나 있다. 이는 조음합성기에서 성도의 단면적을 얻어서 합성하기 때문에 정확한 합성음을 얻기 어려우며, 또한 본 실험에서는 혀의 최고점을 중심으로 모음을 구별하려고 하였으나, 좁힘점과 입술의 변화도 모음의 음가에 영향을 주기 때문인 것 같다.

그리고 청취실험을 위한 혀의 형태가 270개로 많아 청취실험을 할 때 앞부분에서는 다소 정확하게 인지하나, 뒷부분으로 갈수록 음이 단순하기 때문에 학습되어 비슷한 음으로 인식하는 경향이 있는 것 같다.

그리고 인지실험의 결과에서 현재 가정한 조음점의 위치 분포가 후설모음에 해당하는 위치에는 거의 나타나지 않았으므로 차후 재설정할 필요가 있다.

6. 결 론

본 연구에서 조음합성 시뮬레이터인 ‘ArtSim’을 이용하여 6가지 모음에 대한 성도의 단면 및 혀의 형상으로부터 조음위치를 설정하였고, 조음점들의 위치를 체계적으로 변화시켜 가면서 합성음을 생성하였으며, 각 합성된 음이 어떤 모음에 가까운지를 청취실험을 통해 확인하였다.

이번 실험으로 다소 오차는 있었으나, 분포가 대체적으로 한 곳으로 집중되어 있었다. 그러나 혀 최고점의 위치에 중점을 두어 실험을 하였으나 앞으로의 실험에서는

입술의 부분의 영향과 후두개 부분에서 일어나는 좁힘점의 영향도 고려한다면 보다 정확한 결과를 얻을 수 있을 것이다. 또한 혀의 최고점의 위치를 35개의 점으로 정하였으나, 보다 많은 점을 지정한다면 모음들간의 간섭이나 비슷한 모음으로의 변환을 위해서도 좋을 것이다.

향후의 연구에서는 정확한 모음의 음가를 생성하기 위해 계속 보완할 것이며, 자음의 경우에 대해서도 연구를 시도하고자 한다. 또한 이번 실험의 결과를 바탕으로 음으로부터 혀의 형태를 형성하는 부분을 조음합성 시뮬레이터에 부가하여 연속적인 모음의 변화를 실험할 예정이다.

감사의 글

본 논문에 사용된 MRI 영상을 제공해 주시고 많은 도움을 주신 부산대학교 이비인후과 왕수건 교수님과 관련된 여러 분께 감사의 말씀을 드립니다.

본 연구는 현재 수행중인 한국과학재단 핵심전문연구과제 “멀티미디어 환경에서의 정서음성의 모델링 및 합성에 관한 연구”의 수행결과의 일부입니다.

<참고문헌>

- [1] 이 호영(1996), 「국어 음성학」, 51-66, 107-115, 태학사.
- [2] Peter Ladefoged 저, 황 귀용 역(1994), 「음성학 입문」(*A course in Phonetics*), 3-19, 한신문화사.
- [3] G. Fant(1980), The relations between area function and the acoustic signal, *Journal of Phonetica*, 37, 55-86.
- [4] J. L. Flanagan(1972), *Speech analysis, synthesis and perception*, 2nd ed, Springer, Berlin,
- [5] B. Lindblom, J. Sundburg(1969), A quantitative model of vowel production and the distinctive features of Swedish vowels, 14-32, *Q. Prog. Status Rep., Speech Transm. Lab, Inst. Technol. Stockh*,
- [6] C. Yang and H. Kasuya(1994), Accurate measurement of vocal tract shapes from magnetic resonance imaging of child, female and male subject, *Technical Report of IEICE, SP94-6, 623-626 (in Japanese)*.
- [7] L. R. Rabiner and R. W. Schafer(1981), *Digital Processing of speech signals*, 33-107, Prentice-Hall, inc.
- [8] K. N. Stevens(1989), On the quantal nature of speech, *Journal of Phonetics*, 17, 3-45.

접수일자: 1998년 9월 2일

게재결정: 1998년 10월 30일

▶ 김대현

주소: 경남 창원시 사림동 9번지
소속: 창원대학교 제어계측공학과
전화: 0551) 279-7559
e-mail: midas03@hanmail.net

▶ 조철우

주소: 경남 창원시 사림동 9번지
소속: 창원대학교 제어계측공학과
전화: 0551) 279-7552
e-mail: cwjo@sarim.changwon.ac.kr