

MRI를 이용한 조음모델시뮬레이터 구현에 관하여

On the Implementation of Articulatory Speech Simulator Using MRI

조철우*
(Jo, Cheol-Woo)

ABSTRACT

This paper describes the procedure of implementing an articulatory speech simulator, in order to model the human articulatory organs and to synthesize speech from this model after. Images required to construct the vocal tract model were obtained from MRI, they were then used to construct 2D and 3D vocal tract shapes.

In this paper 3D vocal tract shapes were constructed by spatially concatenating and interpolating sectional MRI images. 2D vocal tract shapes were constructed and analyzed automatically into a digital filter model. Following this speech sounds corresponding to the model were then synthesized from the filter. All procedures in this study were using MATLAB.

Keywords : Articulatory, MRI, Simulator, Vocal tract, digital filter

1. 서 론

사람의 음성은 성도의 형태를 특정한 모습으로 한 뒤 성대를 울려서 소리를 낸다. 3차원의 성도의 형태 관찰은 음성 생성의 조음모델이나 음향적인 처리에 유용하게 이용될 수 있다.

지금까지 성도의 형태를 구하는데 X-ray, CT, MRI 등의 사진을 이용하여 왔다. 그러나 이러한 사진을 바탕으로 한 혀의 모습이 변할 경우 발생하는 음향적 특성을 조사하기는 어려운데 이는 건강상, 비용상의 문제 때문이다. 신체의 횡단면이나 두개골의 부분을 인체에 영향을 주지 않고 얻을 수 있는 MRI 이미지는 의학적인 실험에 광범하게 이용된다. MRI를 이용할 경우 치아 등 식별이 어려운 부분이 있기는 하지만 대체로 성도의 형태를 추정할 수 있다.[1][2][3][4]

본 연구에서는 MRI를 이용하여 성도의 형태에 따라 얻은 이미지를 얻어, 먼저 2차원 성도 모델을 추출하는 방법을 MATLAB으로 구현하고 그에 따른 합성음을 만들어내도록 하였으며, MRI의 각 단면 영상으로부터 단면적을 구하고 이를 3차원으로 영상화하기 위해 소프트웨어 도구를 구현하고 실제 자료를 이용하여 3차원 영상을 구현하였다.

* 창원대학교 공과대학 제어계측공학과

2. 2차원 성도모델 시뮬레이터(ArtSim)

2차원 성도모델은 머리, 중안의 종단면을 기준으로 획득한 영상을 이용하여 성도의 형태를 추정하는 것으로 기존에는 X-ray 영상을 많이 이용하였고 최근에는 MRI 영상 등을 이용할 수도 있다. 이러한 성도모델을 구현하기 위해서는 MRI나 X-Ray 영상을 디지털화하여 성도의 형태를 추정하는 과정이 필요한데 종래에는 수작업을 통해 성도의 형태를 얻어왔고 따라서 영상의 입력과정이나 성도의 음향출력을 추정하는 과정이 상당히 복잡하였다. 본 연구에서는 영상의 입력과정, 성도의 음향응답계산과정을 쉽게 하기 위하여 MATLAB의 GUI 기능을 이용하여 ArtSim (Articulatory Speech Simulator) 이라는 시뮬레이터를 작성하였다. MRI를 통해 구해진 영상은 디지털 카메라를 통해 입력된 뒤, 먼저 별도의 영상처리 프로그램에 의해 성도의 각 부분이 강조되도록 영상처리된다. 처리된 영상을 통해서 조음기관의 각 부분을 입력한다. 조음기관의 입력과정은 다음과 같이 세 부분으로 나누어진다. 첫 번째는 코, 경구개, 후두부를 포함하는 변하지 않는 부분, 두 번째는 입술과 턱 부분, 세 번째는 성도의 형태를 결정하는 데 가장 큰 역할을 하는 혀의 부분으로 나누어 입력된다. 이러한 입력과정에서 MATLAB의 spline 함수를 사용한다. 성도의 형상은 상당히 복잡하므로 곡선의 각 부분을 모두 입력하기가 힘들기 때문에 중요한 부분만 마우스와 그래픽 커서를 통해 입력하고 나머지 부분은 spline 곡선을 통해 근사화하여 전체적인 형상을 결정하였다. 입력된 각 점들은 spline 곡선을 통해 근사화되고 다시 표본화되어 일정한 개수의 점들로 재구성되게 된다.

2.1 디지털 필터 모델

조음합성에 이용한 디지털 필터모델은 일반적으로 이용되고 있는 Lossless tube model과 동등한 모델을 이용하였다.

이 모델에서의 전달함수를 표현하면,

Glottal 반사계수가

$$r_G = \left(\frac{Z_G - \frac{\rho c}{A_1}}{Z_G + \frac{\rho c}{A_1}} \right) \text{ 이고, } r_k = \left(\frac{A_{k+1} - A_k}{A_{k+1} + A_k} \right) \text{ 일 때}$$

$$V(z) = \frac{U_{lips}(z)}{U_{glottis}(z)} = \frac{0.5(1+r_G) \prod_{k=1}^N (1+r_k) z^{N/2}}{D(z)} \text{ 이며, } D(z) \text{는 } z^{-1} \text{를 가지는 다}$$

항식으로 이루어진 행렬이다. 여기서 N은 tube의 갯수이다.

$$D(z) = [1, -r_G] \begin{bmatrix} 1 & -r_1 \\ -r_1 z^{-1} & z^{-1} \end{bmatrix} \cdots \begin{bmatrix} 1 & -r_N \\ -r_N z^{-1} & z^{-1} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{과 같이 주어지고, 다시 나}$$

타내면

$$D(z) = 1 - \sum_{k=1}^N \alpha_k z^{-k} \text{ 와 같이 나타낼 수 있다.}$$

$r_G = 1 (Z_G = \infty)$ 인 경우에, Lossless model에서 $D(z)$ 를 구할 수 있다.

그리고 Section갯수의 선택은 음성신호를 표현하기 위해 선택된 Sampling rate에 의존한다. Lossless모델의 주파수응답은 주기적이고, Sampling rate가 T 일 때 $|F| < 1/(2T)$ 의 주파수 범위를 가진다. 또한 τ 가 하나의 단일 section의 전달시간일 때, $T = 2\tau$ 가 된다.

2.2 단면 Cutting 방법과 입술 및 혀의 입력방법

이 실험에서 사용한 단면의 Cutting 방법은 세 가지를 이용하였다. 첫 번째 방법은 하스킨스(Haskins) 연구실에 사용한 방법을 이용하였다. 그림 1이 첫 번째 방법의 예이다. 두 번째 방법은 성도의 형태를 고려한 Cutting 방법으로 그림 2가 그것이다. 마지막으로 세 번째 방법은 두 번째 방법에서 구한 성도의 중심을 기준으로 그 중심선과 수직으로 만나는 선을 이용하여 구하는 방법이다. 그림 3이 세 번째 방법을 이용한 방법이다.

그림 1. 방법1

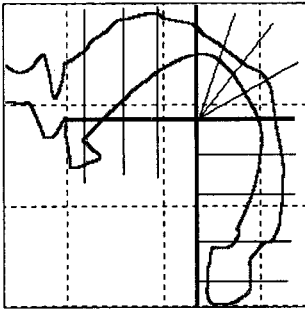


그림 2. 방법2

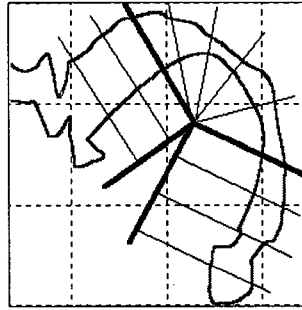
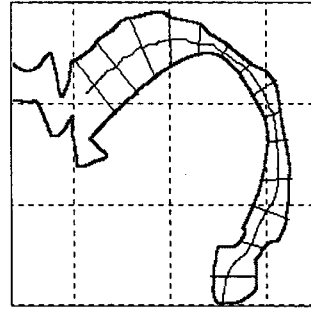


그림 3. 방법3



그리고 조음합성시 모음의 음가에 영향을 미치는 가장 중요한 파라미터는 혀의 위치와 입술 모양이다. 혀의 위치와 입술모양의 변화를 통해서만 모음의 음가가 변한다.

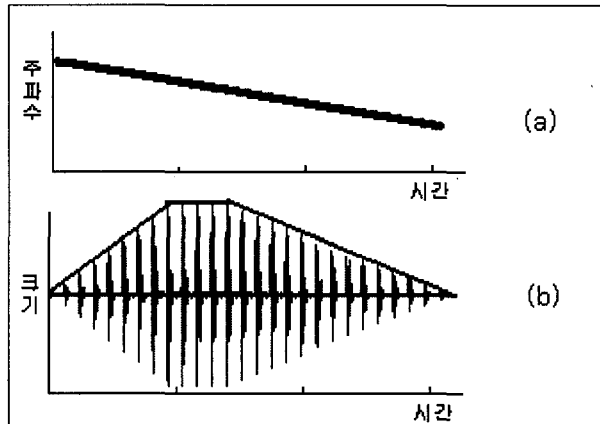
입술모양의 변화는 상하로 움직일 수 있게 하는 한 점을 이용해 입술을 변화시켰다. 입술부분의 단면적을 얻기 위해서는 Cutting 방법에 상관없이 수직으로 잘라서 만나는 점을 찾아 구하였다.

혀의 위치의 변화를 유동적으로 하기 위해 6개의 점들을 이용해 모두를 Interpolation하여 혀의 형상을 만들었다. 6개의 혀의 위치는 이뿌리 부근의 혀의 뿌리, 혓날, 혀의 최고점, 목젖근처의 혀, 후두개, 그리고 후두의 윗부분으로 나누어서 입력을 하였다.

2.3 실험용 단모음 합성조건

조음합성을 위해 성도의 단면적을 이용해 디지털 필터모형을 거쳐 나온 전달함수의 $V(z)$ 를 다시 원하는 주파수축 범위를 Resample하여 12차 All-Pole 형태로 근사화하여 파라미터를 구하였다. 이 파라미터를 이용해 실험용 단모음을 만들기 위해 그림 4와 같은 피치의 변화와 크기의 변화를 주어 합성음성을 만들었다..

그림 4. (a) 피치의 변화 (b) 크기의 변화



2.4 실험 및 검토

본 실험에서는 본 실험실에서 만든 'ArtSim'이란 시뮬레이터를 이용해 Cutting 방법에 따른 6가지 모음의 변화를 실험하였다. 실험방법에 대하여 시뮬레이터의 실행 순서에 따라 설명하겠다.

여기서는 시뮬레이터에 대한 상세한 언급은 하지 않을 것이며, 상세한 내용은 별도로 발표될 논문에서 논할 것이다.

먼저 이미지에서 성도의 형태를 얻은 후, 6가지 모음에 대한 혀의 위치를 입력한다. 그리고 Cutting 방법을 선택하면 방법에 따른 단면적을 구한다. 구한 단면적을 이용해 성도와 등가적인 디지털 필터를 구성하고 응답을 구하면 합성된 모음이 나온다.

이렇게 하여 나온 모음의 특성을 알아보기 위해 이번 실험에서는 성도함수인 단면적과 스펙트럼을 이용해 3가지 Cutting 방법에 따른 단면적과 스펙트럼의 형태를 그래프를 통해 비교하고, 또한 포먼트도 비교한다.

모음 /ㅏ/의 Cutting 방법으로 방법1을 나타낸 그림 5. (a)와 성도 면적을 나타낸 그림 6. (a)에서 보면 Glottis 부근과 이 부근에서의 단면적이 다른 두 방법보다 크고, 모음 /ㅓ/의 단면적을 나타낸 그림 6. (b)에서 보는 바와 같이 방법 2가 후두개 부근에서 다른 두 방법보다 단면적이 크게 나타난다. 그리고 두 모음의 방법 3을 나타낸 그림 5.(c)와 그림 6.(a), (b)에서는 대체적으로 방법 2와 같은 단면적을 나타내고 있다.

그림 6. (c), (d)의 포먼트 주파수를 비교해 보면, 제 1포먼트 주파수는 대체적으로 두 모음(/ㅏ/,

/기/)의 방법이 일치한다, 그 이유는 모음의 음가에 영향을 주는 혀의 위치(전설, 중설, 후설)에 따른 단면적의 크기가 모든 방법에서 거의 일치하기 때문이다.

방법1이 제 2포먼트 주파수에서 보면, 모음 /ㅏ/는 다른 두 방법보다 작지만, 모음 /기/는 다른 두 방법보다 크다. 표 2에서도 보면 모음 /기/를 제외한 거의 모든 모음에서의 제 2포먼트 주파수는 다른 두 방법보다 크다는 것을 알 수 있다. 모음 /기/는 성도의 총면적을 나타낸 표 1에서보면 방법 2의 면적이 방법 1보다 크다(진하게 표시된 부분). 모음 /ㅣ/를 제외한 다른 모음들은 방법1의 총면적이 크다. 모음 /ㅣ/는 다른 방법들보다 제 1포먼트 주파수가 크다. 총단면적의 크기로 포먼트 주파수들과 연관시키는 것보다는 각 조음점의 위치에 따른 단면적의 크기로 포먼트 주파수를 분석하는 것이 좋다. 하지만 여기서 총단면적의 크기로 포먼트 주파수를 비교해 보면 표 2에서 모음 /기/와 모음 /ㅣ/의 포먼트 주파수의 변화가 다른 모음들과 차이를 보인다는 것을 알 수 있다. 총단면적의 크기가 작아지면 제 1포먼트와 제 2포먼트 주파수 사이의 간격이 길어진다는 것을 알 수 있다.

그림 5. 모음 /a/의 Cutting 방법

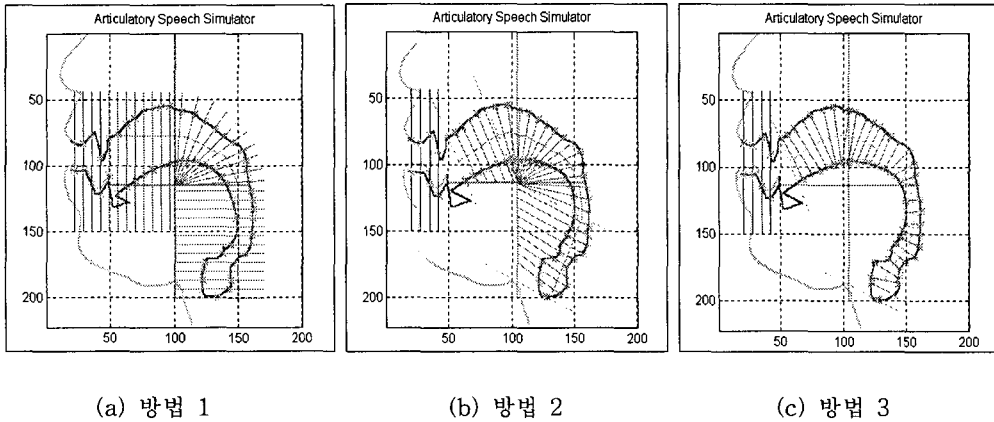
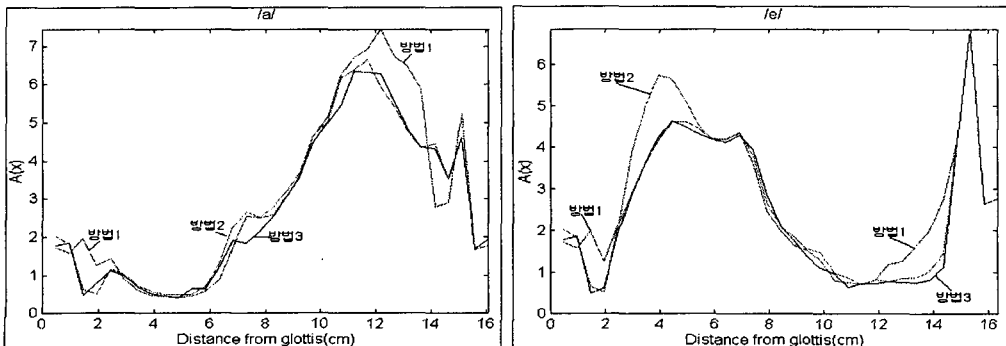
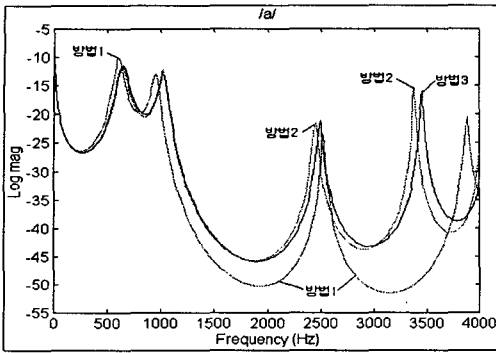


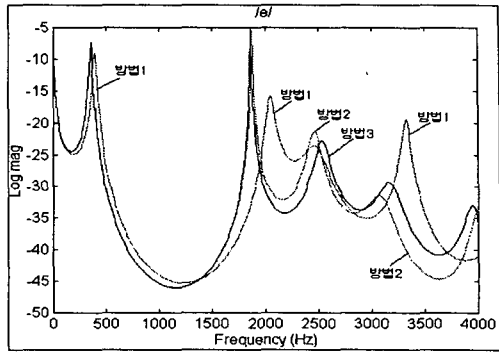
그림 6. /ㅏ/,/기/에 대한 성도 면적과 스펙트럼





(a) 모음 /ㅏ/의 성도면적

(c) 모음 /ㅏ/의 스펙트럼



(b) 모음 /ㅓ/의 성도면적

(d) 모음 /ㅓ/의 스펙트럼

전체적으로 볼 때, 모음의 음가에 영향을 많이 미치는 제 1포먼트 주파수와 제 2포먼트 주파수는 대체적으로 일치하는 것을 알 수 있으며, 제 3포먼트 주파수와 제 4포먼트 주파수가 변하는 것을 알 수 있다. 이것은 입술과 Glottis 부근에서 Cutting 방법들 간에 단면적의 크기의 변화가 커지기 때문이다.

표 1. 성도의 총면적

모음의 종류	성도의 총면적(cm ²)		
	방법 1	방법 2	방법 3
/ㅏ/	99.37	94.83	91.50
/ㅓ/	69.21	64.19	62.06
/ㅗ/	60.47	55.73	53.40
/ㅜ/	86.81	88.21	80.48
/ㅡ/	62.67	56.96	54.04
/ㅣ/	69.27	72.92	65.37

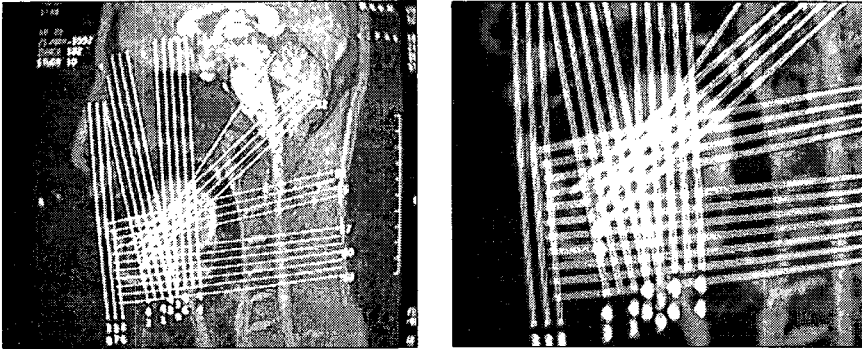
표 2. 포먼트의 값

모음 cutting		/ㅏ/	/ㅑ/	/ㅓ/	/ㅕ/	/ㅣ/	/ㅜ/
		F ₁	601	591	333	388	231
방법 1	F ₂	952	1367	831	2051	1478	1266
	F ₃	2513	2310	2550	2467	2226	2199
	F ₄	3880	3547	3760	3316	3298	3455
	F ₁	637	572	333	360	203	490
방법 2	F ₂	998	1386	915	1875	1811	1312
	F ₃	2448	2171	2134	2467	1986	2125
	F ₄	3372	3298	3363	3067	2808	3206
	F ₁	647	591	351	360	203	517
방법 3	F ₂	1025	1404	942	1866	1912	1321
	F ₃	2485	2208	2291	2540	2079	2143
	F ₄	3455	3381	3390	3150	2993	3298

3. 3차원 성도 모델의 구현

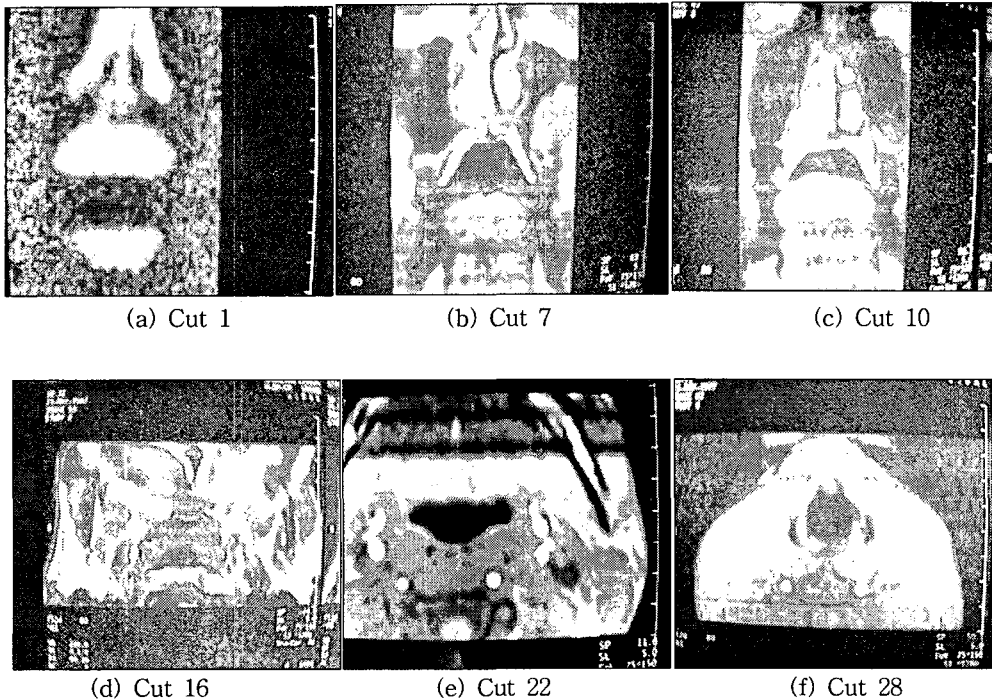
3차원으로 성도의 형태를 구성하기 위해 MR 이미지를 이용하였다. MRI 이미지는 평면으로 얻어지기 때문에 다시 재구성을 하여야 한다. MRI 이미지를 얻을 때 사람에게 해로운 방사능이 없기 때문에 선호되고 있으나, 오랜 스캐닝 시간이 걸리므로 정지된 이미지만을 얻을 수 있는 단점이 있다. 이러한 이유로 이미지를 얻을 때 발음을 일정시간 유지한 채 이미지를 얻어야 했다. 이미지를 얻기 위해 단면을 어떻게 나누어야 하는지도 주요한 문제이다. 단면을 나누는 방법은 성도의 형태에 따라 일정한 간격으로 얻는다면 정확한 이미지를 얻을 수 있겠지만, MRI 이미지를 얻는 장치의 한계로 인하여 일정한 간격이 아닌 성도의 형태에 따라 이미지를 재구성할 때 중요한 부분에 따라 즉, 성도의 중심을 기준으로 수직인 면으로 단면을 얻었다. 장치의 한계는 단면을 얻는 부분이 겹칠 경우 이미지가 흐리게 나와 분석하기가 어려워진다는 것이다. 그래서 그림 7에서 보는 바와 같이 자르는 단면을 정하여 MRI 이미지를 얻었다. 자르는 단면은 원하는 대로 설정할 수 있지만 수작업에 의해서 설정되기 때문에 차후 재구성하여야 할 필요가 있다.[3][4]

그림 7. MRI영상을 얻기 위한 단면 자르기



이렇게 하여 구한 단면영상을 그림 8에 나타내었다.

그림 8. 구해진 MRI영상의 일례



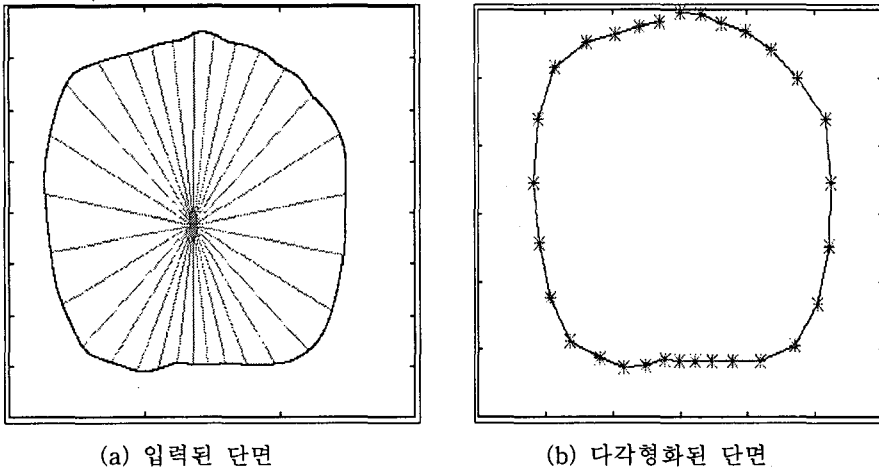
3차원 성도모델을 시각적으로 표현하기 위해서는 우선 MRI 영상을 입력할 때 설정한 기준선을 바탕으로 각 단면의 형태를 구하고 이 형상들을 연결하는 좌표를 구하여 3차원 입체형상으로 구성하게 된다. 이러한 과정에서 가장 첫 번째 발생하는 문제는 구해진 영상으로부터 어떻게 단면의 영상을 추출하여 좌표화 하는가 하는 것이다. 기존에는 수작업으로 입력하는 경우

도 있었으나 이 방법은 시간이 많이 걸리고 다시 컴퓨터에 입력할 경우 좌표평면으로의 변환이 필요하므로 본 연구에서는 읽어들이는 영상을 직접 마우스로 추적하여 면적을 구하고 제시된 중심점을 기준으로 자동으로 좌표변환을 하는 방법을 사용하였다.

그림 9는 입력도구에 의해 한 개의 단면을 도형으로 입력하는 화면이다. 그림에서와 같이 마우스 포인터를 이용하여 입력된 도형을 주어진 중심점을 따라 일정한 각도로 잘라서 동일한 개수로 이루어진 다각형으로 근사화하여 단면을 구하였다.

입력된 단면은 좌표변환에 의해서 3차원 평면으로 표현되게 된다. 기존의 2차원 도형으로 입력된 단면을 초기에 설정된 단면 기준선을 바탕으로 3차원평면으로 좌표변환하게 된다.

그림 9. 단면의 입력



좌표변환에는 다음과 같은 식을 이용한다.

$$\begin{aligned}
 (x_1, y_1, z_1) &\xrightarrow{y\text{축}(-90\text{도})} (z_1, y_1, x_1) \\
 (z_1, y_1, x_1) &\begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \rightarrow (x_2, y_2, z_2) \\
 (x_2, y_2, z_2) &+ \begin{pmatrix} x\text{trans} \\ y\text{trans} \\ 0 \end{pmatrix} \rightarrow (X, Y, Z)
 \end{aligned}$$

그림11은 구현된 3차원 입력 및 영상화 프로그램의 화면이다. 실행과정은 다음과 같다. 3차원 영상구현 프로그램에서 먼저 어떤 모음에 대해 구현할 것인지 결정한 후 Cut 이미지의 옆에 있는 축척을 입력한다. 입력하고 나면 첫 번째 Cutting 선에 따른 성도의 단면이 오른쪽에 나타난다. 그리고 성도 단면의 하단에 있는 버튼을 누르면 그 내용에 따라 Cutting 선에 따른 성도

단면이 변하거나, 성도의 단면의 재입력을 위한 성도 단면의 MRI 이미지가 나타난다. Cutting 단면을 나타내는 이미지의 하단의 버튼은 3D를 나타내기 위한 좌표의 변환 및 3D의 영상화를 할수 있다. 또한 하단에 성도의 단면에 따라 각각의 성도단면의 면적 및 단면의 높이를 나타내고 있다.

그림 10에 3D로 성도를 나타내고 있다.

그림 10. 구현된 성도의 3D 영상 (모음/e/)

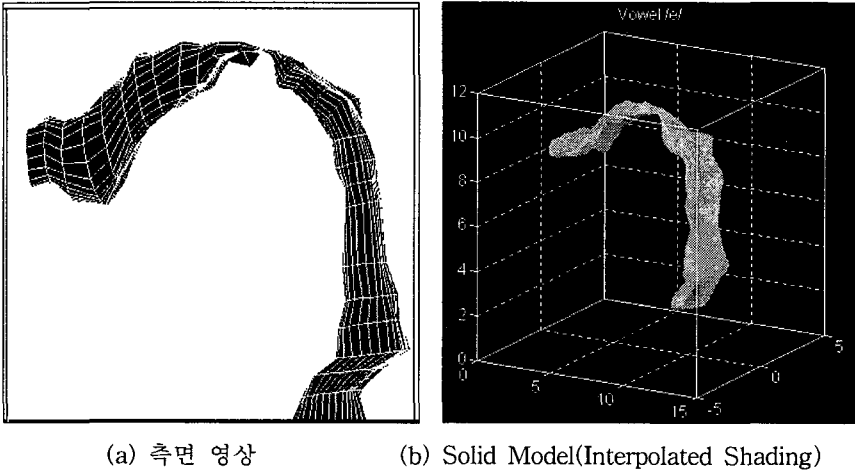
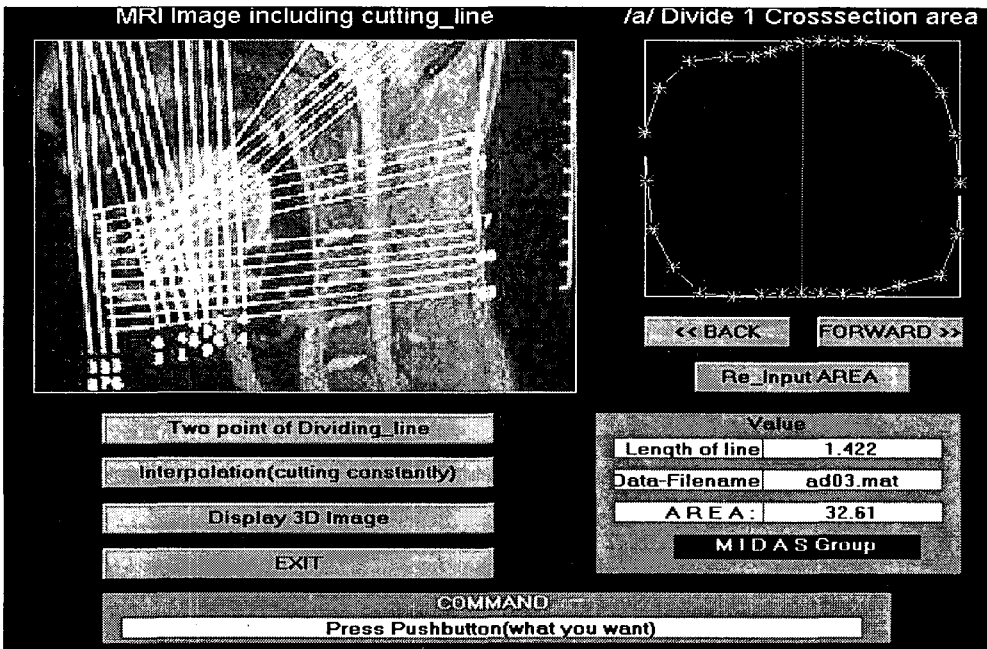


그림 11. 3차원 영상구현 MATLAB 프로그램의 화면



4. 결 론

본 논문에서는 MATLAB을 도구로 MRI 영상을 이용하여 2차원 성도모델을 구성하고 합성음을 생성하는 과정 및 3차원 성도모델로 영상화하는 과정에 대하여 기술하였다. 구현된 2차원 성도모델은 종단면 영상만으로 성도를 구성하므로 실제 성도형태와 오차가 발생하게 되나 음성학적인 조음점의 위치에 따른 합성음의 변화를 관찰할 수 있게 해 준다.

3차원 성도모델은 과정이 매우 복잡하나 MATLAB의 GUI기능을 이용하여 쉽게 입력 및 구현이 가능하였다. 구해진 영상은 전체적으로 단면의 수가 작은 관계로 매끄러운 영상을 보여주는 못하나 성도의 형상을 잘 나타내 주고 있다. 차후의 연구에서는 이 3차원 모델로부터 등가적인 음향모델을 구하고 이로부터 음성을 생성하는 과정을 구현할 예정이다.

감사의 글

본 논문에 사용된 MRI영상을 제공해 주시고 많은 도움을 주신 부산대학교 이비인후과 왕수진 교수님과 관련된 여러분께 감사의 말씀을 드립니다.

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

참 고 문 헌

- [1] 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, pp. 623-626 (in Japanese).
- [2] M. Matsumura, T. Niikawa, K. Shimizu, Y. Hashimoto, T. Morita. 1994. "Measurement of 3D shapes of Vocal Tract, Dental Crown and Nasal Cavity Using MRI, Vowels and Fricatives." ICSLP '94, YOKOHAMA, pp. 619-622.
- [3] T. Baer, J.C. Gore, L.C.Gracco and P.W.Nye. 1991. "Analysis of Vocal Tract Shape and Dimension Using MRI." Journal Acoust. Soc. Am, vol. 90(1), pp. 798-828.
- [4] A. R. Greenwood, C. C. Good and P. A. Martin. 1992. "Measurements of the Vocal Tract Shapes Using Magnetic Resonance Imaging." in IEEE Proceedings-I, Dec vol. 139(6), pp. 553-560.

접수일자 : '97. 10. 4.

게재결정 : '97. 11. 10.

▲ 경남 창원시 사림동 9번지
창원대학교 공과대학 제어계측공학과 (우: 641-773)
Tel : (0551) 79-7550 (O), 62-6174 (H)
Fax : (0551) 62-5064
e-mail: cwjo@sarim.changwon.ac.kr