A Simulation Study of the Vocal Tract in Tracheoesophageal Speaker

Cheol-Soo Kim, Soo-Geun Wang, Hwan-Jung Roh,
Eui-Kyung Goh, Kyong-Myong Chon, Byung-Joo Lee,
Soon-Bok Kwon[*], Suck-Hong Lee, Hak-Jin Kim[**], Byung-Gon Yang[***]

ABSTRACT

The vocal tract shapes were measured from tracheoesophageal speakers during the sustained phonation of five Korean vowels /u/, /o/, /a/, /e/, /i/ using magnetic resonance image(MRI). The subject's original vowel utterances with speech intelligibility and the synthesized vowels from MR images were analyzed. The results were as follows: (1) The vowels /a/, /e/, /i/ were perceived as the same sounds of actual subject's speech, but the vowels /o/ and /u/ were perceived as / ∂ / and strained /u/, respectively. (2) The synthesized vowels /a/ and /e/ from the MR images were perceived as the same sounds, but the vowels /u/, /o/, /i/ were perceived as different sounds. (3) The synthesized vowel by the expanded pharyngeal segment of 3 times in vowel /o/ was perceived as more natural than that of 2 times. The pharyngeal areas with varied sizes should be experimented to secure better speech production because the correct shapes of the vocal tract lead to distinct vowel production.

Keywords : Vocal tract, MRI, Vowel synthesis, Total laryngectomy, Tracheoesophageal speaker

Introduction

The accurate measurement of vocal tract shape during phonation is important in the study of speech production. In a majority of the early studies, measurements were made from lateral radiographic images (1), and transverse areas of the airway were obtained by applying transformations to widths measured from these lateral projections. Some data has been derived from casts of the vocal tract (2) and from measurements of cadavers (3). There have been a few tomographic studies of vocal tract shape using

---

[*] Depart. of Otolaryngology, College of Medicine
[**] Depart. of Diagnostic Radiology, College of Medicine, Pusan National University
[***] Depart. of English, College of Humanities, Dongeui University

radiography. Fant (4) used a conventional radio-tomograph to visualize cross sections of the pharynx. Johansson et al. (5) used computed tomography (CT). Although CT has only a transaxial imaging capability and some damaging effects, CT is best suited to imaging the pharyngeal airway of a supine subject because there is only sufficient adjustment to obtain images in planes that are orthogonal to the pharyngeal axis. Point tracking methods, such as x-ray microbeam by Fujimura et al. (6) and magnetometers by Schonle et al. (7), can supply dynamic information about the movement of structures in the oral cavity, but they are designed to obtain measurements only in the midsaggital plane and cannot supply detailed information about shape. Ultrasound scanning by Stone (8) has been used to generate dynamic images of the tongue surface, either in midsagittal or transverse planes, but it has been useful in only a limited part of tongue.

Magnetic resonance imaging (MRI) is free from many of the disadvantages associated with the methods we have already mentioned. Its advantages are that MRI produces tomographic images that appear similar to those produced by CT without using ionizing radiation; it also produces in three orthogonal planes without tilting the table or repositioning the subject. These techniques have no known harmful side effects. One of the major disadvantages of the MR technique is the time required to perform the imaging process. Therefore, if speech productions are to be studied, they must be sustainable images during the course of the acquisition process and may be subject to fatigue effects. Additional difficulties stem from the fact that calcified structures such as bone and teeth contain little mobile hydrogen and are indistinguishable from the airway in images. However, MR technique appears to offer the best opportunity to obtain data on the vocal tract shapes associated with sustainable speech sounds such as vowels and fricatives.

The articulatory synthesizers have gained importance as research tools and have grown in computational complexity. These make it possible to synthesize sounds by changing the original.

The authors analyzed the vocal tract shapes and the acoustic characteristics of their voices in the tracheoesophageal speakers, and then divided them into two groups with speech intelligibility, and compared vocal tract shapes and their acoustic output of the synthesized vowels. Furthermore, vowel sound was synthesized by changing the profile of the vocal tract. In this study, we suggest basic data for correct simulation of the vocal tract and voice rehabilitation in tracheoesophageal speakers.

Methods

Subjects

4 subjects were selected among 24 patients who received a total laryngectomy, with

or without neck dissection, and tracheoesophageal shunt operation for voice rehabilitation between January, 1993, and December, 1995. They included 3 males and 1 female ranging in age from 51 to 68 (Table 1). The 4 subjects had no interruption of artificial teeth on MR images.

Speech Intelligibility Test

Subjects slowly and accurately phonated the Korean vowels /u/, /o/, /a/, /e/, /i/ and 50 words from Kim's (9) bisyllabic word list. The voices were recorded in a soundproof chamber with a sampling rate of 11 kHz. They pronounced the word list 15 cm away from the microphone in the supine position. 10 adults with normal hearing ability listened subject's voice recording at the most comfortable level, three times each, and wrote them down. The results were analyzed. The speech intelligibility levels were graded by percentiles from the number of the bisyllabic words that the listeners could recognize correctly. Average speech intelligibility was computed for the four subjects. The cases with higher or lower speech intelligibility than the average were divided into two groups.

MRI Acquisition of the Vocal Tract

MRI acquisition of the vocal tract was obtained from 1.5T MR, VISION (Siemens Co.) while the subjects produced the five Korean vowels /a/, /e/, /i/, /o/, /u/, in the supine position, consecutively and continuously. All the segments were imaged at the right angle to the midsagittal line. In order to collect the full length of the vocal tract, the images were covered from the supraclavicular level to the lips (Fig. 1).

The pulse sequence of MRI acquisition was FLASH (fast low angle shot) 2D, TR (repetition time) 170 msec, TE (echo time) 5 msec, flip angle 30o, matrix 78×128 pixels, FOV 130 mm, NEX 1, and slice thickness 5mm. It took 19 seconds for an image acquisition.

Tooth Correction

The coronal CT was checked in the cases with teeth, cases II and III, and the space between upper and lower teeth was measured. The space was compensated by the measurement of the profile on MR image. Cases I and IV were totally edentulous (Fig. 2). Moreover, a new sequence of MRI acquisition (Turbo Spin Echo T1 pulse sequence) was employed to show the teeth of the subject clearly (Fig. 3). However, the border between the skin and air appeared to be unclear in the image. A composite view of the cross-sectional area taken by the previous sequence and Spin Echo sequence would be desirable to obtain the area accurately.

Calculation of Cross-Sectional Area

The boundary of vocal tract area on the MR images was identified with a scanner in the computer system. The cross-sectional area was selected and computed automatically by a graphic software, Area Properties (V3.2) on each segment.

Formant Calculation

Formant values were calculated from the cross-sectional area on the MR images using AreatoFormant (AF) and Formfrek (FF). AF is the formant program with compensation of the loss of vocal tract by Sondhi (10), which used the formants of Japanese basic vowels adopted by Yang and Kasuya (11), and transforms the value of the cross section of the vocal tract and calculates the formant. FF is a program for formant derivation from using Russian basic vowels by Fant (12).

Vowel Synthesis and Analysis

The vowel synthesis was done from formants by varying such parameters as vocal tract area, amplitude, duration and so on. The formant synthesizer, SenSyn 1.0 (Sensimetrics Co.) was used. The reference parameters for formant synthesis were adopted from the normal values of the amplitude and fundamental frequency of natural sounds, which was used in the synthesis of Yang's (13) Korean monophthongs. The vowels were synthesized by inputting the formant values of AF and FF into SenSyn 1.0. Also we enlarged the pharyngeal segments from the model of the vocal tract (the part 4-8 cm away from the pseudoglottis) in the back vowel /o/, 2 and 3 times, in case I with the poor speech intelligibility and case III with good speech intelligibility. Table 2 shows the input file of a given parameter for synthesizing the vowel /o/. Each synthesized vowel was repeated 3 times with a comfortable level for analysis. 10 listeners with normal hearing ability judged the quality of the synthesized vowel.

Results

Speech Intelligibility

Table 3 shows that the auditory impressions of the five Korean vowels are different in each case. The speech intelligibility of the list of the Korean bisyllabic words for cases I, II, III, IV were 72.0%, 61.0%, 83.1%, 82.6%, respectively. Average speech intelligibility in all cases was 74.7%. The average of cases III and IV was 82.9% as a good group. The cases I and II with 66.5% formed a poor group.

Formant Values from MR Images

The formant values were estimated from AF and FF. The formants show more change in F2 than in F3. In case II, the frequency of F1 for vowel /i/ was relatively lower than in the other cases. The formant values of all vowels using AF were higher than those calculated from FF. The reference values of normal formants were adopted from Yang's (13) report (Tables 4 and 5).

In the synthesized file of the cross-sectional area of the expansion with 2 and 3 times in the pharyngeal segment in cases I and III, the formants were obtained using AF and FF. The frequency of F1, F2, and F3 was lower in FF than in AF; in particular, the frequency of F2 was relatively lower than that of F1 and F3 (Table 6).

Analysis of Synthesized Vowels

The auditory impression of the synthesized vowels was different between AF and FF (Table 7). It is significant that the synthesized vowels /a/ and /e/ were perceived as the original sounds, while the synthesized vowels /u/, /o/, and /i/ were perceived as quite different sounds from the original sounds.

The synthesized vowel by enlarged pharyngeal segment of 3 times in vowel /o/ was perceived as more natural than that of 2 times in FF (Table 8).

Comparisons of the Cross-Sectional Area

The synthesized vowels were perceived as various sounds in each case. To find out the cause of the differences, we measured the cross-sectional area in each case from FF in the same vowel, and compared it with the distance from the peudoglottis on the same chart.

1. Vowel /a/

Cases III and IV phonated as clear /a/ because they had proper a cross-sectional area in the oral segment, near 14 cm from the pseudoglottis. Case I phonated as /∂/ with a large cross-sectional area in the oral segment, near 12 cm from the pseudoglottis. Case II phonated as /e/ with a similar width of the cross-sectional area as case I, especially in the pharyngeal segment (Fig. 4).

2. Vowel /e/

Case I phonated as /∂/ by having a large cross-sectional area of the whole oral segment, while case III phonated as a sound close to /e/ by having a very narrow pharyngeal segment. Cases II and IV phonated as clear /e/ by having an equal cross-sectional area in the oral and pharyngeal segment, symmetrically (Fig. 5).

3. Vowel /i/

Case I phonated as a sound close to /∂/ by having a wide oral segment, near 7 cm from the pseudoglottis, despite enough cross-sectional area in the lip. Case IV phonated as /e/. It was supposed that the maintenance of the given width in the midportion of the segment brought it out. Case II phonated as /i/ by having the pharyngeal segment with a relatively wide space, near 6 cm from the pseudoglottis. Case IV phonated as a sound close to /e/ (Fig. 6).

4. Vowel /o/

Case I phonated as /a/, while case III phonated as strained /a/. Cases II and IV phonated as /e/ because of a wide cross-sectional area in the oral segment. Each case made different sounds in vowel /o/ because the subjects were not able to make enough cross-sectional area in the pharyngeal segment due to the loss of the supraglottic region and pharynx during the operation (Fig. 7).

5. Vowel /u/

Cases I and III phonated as strained /a/ by making relatively sufficient space in the oral segment as opposed to the pharyngeal segment. The shape of the oral cavity in both cases was similar. Cases II and IV phonated as clear /e/ by making a proper cross-sectional area in the pharyngeal and oral segment. The shape of the vocal tract in both cases was similar (Fig. 8).

Discussion

An early laryngeal cancer could be cured completely with radiotherapy or partial laryngectomy, but more advanced laryngeal cancer requires total laryngectomy. After laryngectomy the peristaltic movement of the digestive tract remains, but phonation as the essential function of the larynx is lost. The loss of phonation can cause patients to be depressed and to disable them socially, physically and mentally. Thus, voice rehabilitation is necessary after laryngectomy. Numerous researches have been done for voice rehabilitation and various methods of voice rehabilitation have been suggested. The artificial larynx, esophageal speech and tracheoesophageal shunt speech are widely used.

The artificial larynx or electrolarynx is an electrically driven buzzer or a sound transducer. It has the advantages of low cost, availability, short learning time, and loudness. Its disadvantages are the dependence on batteries, mechanical sound, conspicuous appearance, and hygiene of the intraoral tubes. Nevertheless, many laryngectomized patients use the artificial devices effectively and depend on them. The esophageal speech is a popular and commonly recommended method for alaryngeal speech rehabilitation. Successful esophageal voice is preferable to the artificial larynx because it is less conspicuous and does not require the hands, it is more natural sounding; and the patients do not need devices that fail over time. The critical problem with esophageal speech is its low success rate and prolonged period required to learn it. The characteristics of esophageal speech include low fundamental frequency, short duration, and hard production (14).

The tracheoesophageal shunt speech is easier to use and has more natural sound than the esophageal speech, but it also has disadvantages. For example, the tracheostoma should be closed when making sound and can cause saliva or food to enter the airway. However, the operation technique is simple and it takes less time to learn for proper speech. Furthermore, it allows for continuous speech production, at present, it is more widely used than other methods of voice rehabilitation after total laryngectomy. This operation method was popular by Amatsu (15). The lung air is sent to the esophagus or hypopharynx via a tracheoesophageal shunt and vibrates the pseudoglottis formed by the mucosal fold with about 4 cm thickness in the pharyngoesophageal segment. The vibratory mechanism of the pseudoglottis was revealed using a combination of videolaryngoscopy, laryngeal stroboscopy and laryngeal electromyography (16). The aspiration problem after this operation was considerably improved with Singer's (17) voice prosthesis and the inferior-based esophageal muscle flap by Amatsu (15). For more successful rehabilitation, the probability of a cricopharyngeal muscle spasm can be predicted by an esophageal insufflation test before the operation. In the esophageal insufflation test or Taub Test (18), a tube is inserted into the hypopharynx at the tracheostoma level through the nasal cavity, and air is blown strongly into it to make sounds. Unless a sound is not made in the tracheostoma during the test, the tracheoesophageal shunt operation is not recommended. However, sometimes the test results are not conclusive to determine a successful operation. If the spasm of the cricopharyngeal muscle is suspected clearly, cricopharyngeal myotomy (19) or pharyngeal plexus neurectomy (20) can be performed for the successful rehabilitation.

The acoustic analysis of subject's actual speech is important. For it, Yang and Kasuya (11) directly measured the subject's speech in the supine position at the MRI room. The authors recorded the subject's speech in a soundproof room in the same posture as for the MR imaging process. Vowels /a/, /e/, /i/, /u/ were perceived as similar sounds. However, vowel /o/ sounded quite different, which might be caused by the loss of the supraglottic region and the pharyngeal segment during the operation.

In addition, speech intelligibility is an index that is important in the analysis of the result of voice rehabilitation and is used in evaluating the capability of speech. Baek et al. (21) reported that tracheoesophageal and esophageal speech showed the same result in speech intelligibility. Wang (22) reported that speech intelligibility is 76.3% in tracheoesophageal speech, which is lower than normal, but higher than in esophageal speech. In this study, speech intelligibility is 82.9% in the good group, and 66.5% in the poor group (average 74.7%). This shows little difference from Wang's (22) report.

The accurate measurement of the vocal tract is very important in the evaluation of voice rehabilitation. Therefore, various methods of measurement have been suggested. MR imaging capable of noninvasively obtaining transverse, coronal and sagittal sections of the body is widely used for examinations of the vocal tract. Baer et al. (23) measured the vocal tract using a fixed grid plane system. However, this method cannot measure exactly due to the simple configuration of the tongue and the palate. The length of the vocal tract by this method was 1.5 cm shorter than by others. Yang and Kasuya (11) calculated the cross sectional area of the oral segment from the lips to the atlas in the coronal image, and that from the glottis to the palate in the axial section of the midsagittal image. In our study, all images of the cross section were taken in the orthogonal plane to the midsagittal line of the entire vocal tract.

It is difficult for subjects to maintain a sustained phonation during the imaging process considering the tracheoesophageal speaker's maximum phonation time. However, Yang and Kasuya (11) reported almost the same results. The familiar vowels showed little difference in pitch and vocal tract shape when subjects phonated them repetitively, maintaining the same form of the lips after a practice session.

The time for obtaining images from MR images varies according to the studies; 204 seconds for Baer et al. (23), 123 seconds for Matsumura et al. (24), and 39 seconds for Yang and Kasuya (11). It took 19 seconds in this study. This may be enough time for a subject to maintain the same form of the vocal tract during the imaging process.

The boundary of the vocal tract on MR images should be defined to calculate its real cross-sectional area. It has been defined in various ways. Lakshminarayanan et al. (25) drew it by hand, Martelli (26) found that the image of air is darker than that of the limbic muscle or other tissue and he developed a computer program that can define the boundary automatically from the 50% shade level. Baer (23) used Martelli's (26) program, while he drew by hand only the part that was interrupted by the teeth. We inputted the images into the computer by scanner, and automatically identified the boundary of the vocal tract using a computer program, Area Properties (V 3.2). The teeth contained little water and could not be seen on the image. Matsumura (24) defined the water space between the dental crown and the teeth as the boundary on the image. Yang and Kasuya (11) made a gypsum dental impression in advance and corrected the original image of the oral cavity. Baer (23) measured the boundary of the teeth using vinyl polysiloxane (3M) impression molds. We measured the area of the teeth from its CT image and corrected it in the MR images.

Generally, the peak of energy filtered by the vocal tract resonance in each vowel was referred to as formant which is classified into three basic formants, F1, F2, and F3 counting from the lowest frequency. Yang and Kasuya (11) found the length of the vocal tract indirectly using physical characteristics of sound and considering that formant is affected by the length of the pharynx and oral cavity. The formant is considered vocal tract resonance and the change of the vocal tract shape causes the change of formant. The more the length of the vocal tract is shortened, the more the value of the formant increases. Formant values from MR images were calculated using the cross section and distance from the pseudoglottis on each image. Rubin (27) measured the cross-sectional area of the vocal tract by an empirical mathematical formula. Yang and Kasuya (11) used Sondhi and Schroeter's (28) algorithm to implement a computer program. Two formant programs, AreatoFormant and Formfrek, were used in this experiment. From the area file, the authors changed the value of the cross-sectional area in vowel /o/.

The acoustic synthesizer is helpful in the analysis of the rehabilitated alaryngeal speech. The acoustic synthesizer produces sound as a device modeled on the human vocal organ; the glottis or the pseudoglottis works as a source and the vocal tract filters the source to produce a speech sound. The early stage of the synthesizer was an articulatory synthesizer modeled on the human articulator. In the 1950s, a pattern playback was devised and it transformed the spectrographical images drawn by a brush into sounds. In the 1960s, a digital computer software simulated a formant style synthesizer that produced sounds very similar to the original sound by controlling 60 parameters (13). In this study, a formant synthesizer (SenSyn 1.0) was used in synthesizing vowels. The synthesized vowels /a/ and /e/ from MR images were perceived as close to the subject's original utterances, but the vowels /u/, /o/, and /i/ were different from them.

The differences of sounds between the subject's original utterances and the synthesized vowels were large in several cases. Also the differences between the good group and the poor group in speech intelligibility were noticeable. The differences should not have been surprising because many simplifying assumptions were made. Among these assumptions is the notion that vocal tract shape is well kept in the beginning of the imaging process, but the shape is slightly changed due to the interruption of magnetostrictive noise during the process. Another assumption is the different phonation method from normal adults because tracheoesophageal speakers should press the tracheostoma with fingers during phonation. Also the modification of lips and tongue with indirect compensation by the cervical muscles might be assumed. These assumptions should be considered in the synthesis of vowels from MR images in future research. Also, new operative conceptions for better rehabilitated speech should be considered with enlarged pharyngeal segments.

REFERENCES

[1] Abramson AS, Cooper FS : Slow motion x-ray pictures with stretched speech as a research tool. *J Acoust Soc Am* 1963; 35:1888-1889; also *ACLS Newsletter* 1963; 14:7.

[2] Ladefoged P, Anthony JFK, Riley C: Direct measurement of the vocal tract. *UCLA Working papers in phonetics* 1971; 19:4-13; also in *J Acoust Soc Am* 1971; 49:104.

[3] Heinz JM, Stevens KN: On the derivation of area functions and acoustic spectra cineradiographic films of speech. *J Acoust Soc Am 1964*; 36:1037.

[4] Fant G: Formants and cavities, in *Processings of the fifth international congress of phonetic sciences,* edited by E. Zwirner and W. Bethge(Karger, Basel), pp. 120-140.

[5] Johansson C, Sundberg J, Wilbrand H, Ytterbergh C: From sagittal distance to area: A study of transverse, cross-sectional area in the pharynx by means of computer tomography. *R Inst Technol STL-QPSR* 1983; 4:39-49.

[6] Fujimura O, Kiritani S, Ishida H: Computer controlled radiography for observation of movements of articulatory and other human organs. *Comput Biol Med* 1973; 3:371-84.

[7] Schonle PW, Grabe K, Wenig P, Hohne J, Schrader J, Conrad B: Electromagnetic articulography: use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract. *Brain Lang* 1987; 31:26-35.

[8] Stone M: A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data. *J Acoust Soc Am* 1990; 87:2207-17.

[9] Kim CJ: Relationship between pure tone and speech hearing level utillizing meaningful monosyllabic words list. *J Pusan Med College* 1976; 16(2):1-16.

[10] Sondhi MM: Model for wave propergation in lossy vocal tract. *J Acoust Soci Am* 1974; 55:1070-5.

[11] Yang CS, Kasuya H : Accurate measurement of vocal tract shapes from magnetic resonance images of child, female and male subjects. *ICSLP*: Yokohama press 1994: 623-6.

[12] Fant G: The relations between area functions and the acoustic signal. *Phonetica* 1980; 37:55-86.

[13] Yang BG: A perceptual study of Synthesized Korean Monophtongs. *Korean J Linguistics* 1995; 20-3:127-46.

[14] Gate GA, Hearne EM III: Predicting esophageal speech. *Ann Otol Rhinol Laryngol* 1982; 91:454-57.

[15] Amatsu M: A new one-stage surgical technique for postlaryngectomy speech. *Arch Oto Rhino Laryngol* 1978; 220:149-52.

[16] Omori K, Kojima H, Nonomura MI: Mechanism of T-E shunt phonation. *Arch Oto laryngol Head Neck Surg* 1994; 120:648-52.

[17] Singer MI, Hamaker RC, Blom ED: Application of the voice prosthesis during laryngectomy. *Ann Otol Rhinol Laryngol* 1989; 98:921-5.

[18] Taub S, Spiro RH: Vocal rehabilitation of laryngectomees: Preliminary report of a new technique. *Am J Surg* 1972; 124:87-90.

[19] Chodosh PL, Giancarlo HR, Goldstein J: Pharyngeal myotomy for vocal rehabilitation

postlaryngectomy. *Laryngoscope* 1984; 94:52-7.

[20] Singer MI, Blom ED, Hamaker RC: Pharyngeal plexus neurectomy for alaryngeal speech rehabilitation. *Laryngoscope* 1986; 96:50-4.

[21] Baek MJ, Oh IJ, Wang SG, Chon KM: Comparison of the Amatsu tracheoesophageal shunt operation for speech and esophageal speech after total laryngectomy. *Korean J Otolaryngol* 1993; 36:102-9.

[22] Wang SG: An experience of Amatsu tracheoesophageal shunt operation for speech rehabilitation after total laryngectomy. *Korean J Otolaryngol* 1993; 36:795-803.

[23] Baer T, Gore JC, Gracco LC, Nye PW: Analysis of vocal tract shape and dimensions using magnetic resonance imaging : *Vowels. J Acoustic Soc Am* 1991; 90(2): 799-827

[24] Matsumura M, Nikawa T, Shimizu K: Measurement of 3D shape of vocal tract, dental crown and nasal cavity: Vowels and Fricatives. *ICSLP*: Yokohama press 1994: 619-22.

[25] Lakshminarayanan AV, Lee SB, McCutcheon MJ : MR imaging of the vocal tract during vowel production. *J Magn Reson Image* 1991; 1:71-6.

[26] Martelli A: An application of heuristic search methods to edge and contour detection. *Comm ACM* 1976; 19:73-83.

[27] Rubin P, Baer T, Mermelstein P: An articulatory synthesizer for perceptual research. *J Acoust Soc Am* 1981; 70:321-8.

[28] Sondhi MM, Schroeter J: A hybrid time-frequency domain articulatory speech synthesizer. *IEEE Trans Acoust Speech Signal Process* 1987; 35:955-67.

TABLE 1. *The subjects of study*

case(sex/age)	Operation	Postop. period(months)
I(M/54)	TL. Lt. MND with T-E shunt	17
II(M68)	TL. Rt. MND with T-E shunt	36
III(F/51)	TL. Lt. MND with T-E shunt	29
IV(M67)	TL. Lt. MND with T-E shunt	24

* TL: Total laryngectomy
MND: Modified neck dissection, T-E: Tracheoesophageal

TABLE 2. *Synthetic parameters for vowel /o/ (SenSyn 1.0)*

(Total number of waveform samples = 8000)

SYM	V/C	MIN	VAL	MAX	SYM	V/C	MIN	VAL	MAX
DU	C	30	400	5000	UI	C	1	5	20
SR	C	5000	20000	20000	NF	C	1	6	6
SS	C	1	2	3	RS	C	1	8	8191
SB	C	0	1	1	CP	C	0	0	1
OS	C	0	0	20	GV	C	0	60	80
GH	v	0	60	80	GF	C	0	60	80
F0	v	0	1000	5000	AV	v	0	60	80
OQ	v	10	50	99	SQ	v	100	200	500
TL	v	0	0	41	FL	v	0	0	100
DI	v	0	0	100	AH	v	0	0	80
AF	v	0	0	80	F1	v	180	540	1300
B1	v	30	90	1000	DF1	v	0	0	100
DB1	v	0	0	400	F2	v	550	900	3000
B2	v	40	110	1000	F3	v	1200	2600	4800
B3	v	60	150	1000	F4	v	2400	3500	4990
B4	v	100	100	1000	F5	v	3000	3700	4990
B5	v	100	100	1500	F6	v	3000	4500	4990
B6	v	100	100	4000	FNP	v	180	280	500
BNP	v	40	900	1000	FNZ	v	180	280	800
BNZ	v	40	900	1000	FTP	v	300	2150	3000
BTP	v	40	900	1000	FTZ	v	300	2150	3000
BTZ	v	40	900	2000	A2F	v	0	0	80
A3F	v	0	0	80	A4F	v	0	0	80
A5F	v	0	0	80	A6F	v	0	0	80
AB	v	0	0	80	B2F	v	40	250	1000
B3F	v	60	300	1000	B4F	v	100	320	1000
B5F	v	100	360	1500	B6F	v	100	1500	4000
ANV	v	0	0	80	A1V	v	0	60	80
A2V	v	0	60	80	A3V	v	0	60	80
A4V	v	0	60	80	ATV	v	0	0	80

Varied Parameters:

Time	F0	AV	Time	F0	AV
0	1766	30	50	1750	48
5	1767	31	:	:	:
10	1769	33	360	1390	35
15	1771	34	365	1390	34
20	1773	36	370	1390	32
25	1770	37	375	1390	29
30	1776	41	380	1350	26
35	1762	42	385	1350	23
40	1758	45	390	1350	20
45	1754	47	395	1350	17

TABLE 3. *Results of sound from actual subject's speech*

Case	Vowel				
	/u/	/o/	/a/	/e/	/i/
I	/u/	/ə/	/a/	/e/	/i/
II	/u/	/ə/	/a/	/e/	/i/
III	/u/	/ə/	/a/	/e/	/i/
IV	/u/	/ə/	/a/	/e/	/i/

TABLE 4. *Results of formant from Formfrek (Hz)*

Case	Vowel	F1	F2	F3
Normal	/u/	370	730	2600
	/o/	540	900	2600
	/a/	800	1345	2710
	/e/	550	2100	2900
	/i/	330	2520	3230
I	/u/	880	1539	2589
	/o/	952	1525	3002
	/a/	539	1214	2595
	/e/	644	1494	2595
	/i/	604	1345	2391
II	/u/	464	1920	2530
	/o/	593	1899	2415
	/a/	724	1632	2515
	/e/	576	1688	2903
	/i/	300	2224	3357
III	/u/	827	1456	2426
	/o/	918	1336	2463
	/a/	883	1414	2563
	/e/	878	1747	2679
	/i/	619	1769	2565
IV	/u/	590	1867	2551
	/o/	660	1539	2531
	/a/	800	1483	2816
	/e/	611	1635	2450
	/i/	430	1736	2894

* F: Formant

TABLE 5. *Results of formant from AreatoFormant (Hz)*

Case	Vowel	F1	F2	F3	B1	B2	B3
Normal	/u/	370	730	2600	80	90	60
	/o/	540	900	2600	90	110	150
	/a/	800	1345	2710	110	120	120
	/e/	550	2100	2900	60	90	150
	/i/	330	2520	3230	150	70	80
I	/u/	929	1794	2476	110	552	314
	/o/	983	1844	3104	97	2146	404
	/a/	820	1231	2403	91	83	98
	/e/	929	1685	2733	80	118	169
	/i/	690	1635	2312	69	216	153
II	/u/	497	2371	3035	61	169	1808
	/o/	612	2062	2915	64	194	2270
	/a/	756	1932	2662	70	340	1122
	/e/	598	1653	2924	94	143	240
	/i/	353	2136	3357	70	94	134
III	/u/	981	1686	2440	69	102	120
	/o/	1036	1591	2386	74	154	126
	/a/	1003	1472	2524	90	145	148
	/e/	1007	1730	2783	86	104	172
	/i/	683	1690	2764	61	85	128
IV	/u/	654	1942	2731	61	96	118
	/o/	757	1971	2789	63	111	139
	/a/	843	1749	2940	72	247	179
	/e/	706	1763	2426	61	92	107
	/i/	527	1629	2600	60	81	118

* F: Formant B: Formant bandwidth

TABLE 6. *Results of formant after expansion of pharyngeal segment in vowel /o/(Hz)*

Expansion of area		2x			3x		
case	Method	F1	F2	F3	F1	F2	F3
I	AF	901	1820	3117	837	1807	3106
	FF	614	1037	2845	600	987	2805
III	AF	921	1614	3049	826	1600	3152
	FF	771	1147	2874	708	1083	2858

* AF: AreatoFormant FF: Formfrek

TABLE 7. *Results of synthesized vowels*

Case	Vowel	F1	F2
I	/u/	/e/	/a/
	/o/	/e/	/a/
	/a/	/a/	/∂/
	/e/	/e/	/∂/
	/i/	/u/	/∂/
II	/u/	/e/	/e/
	/o/	/e/	/e/
	/a/	/a/	/e/
	/e/	/e/	/e/
	/i/	/i/	/i/
III	/u/	/a/	/a/
	/o/	/a/	/a/
	/a/	/a/	/a/
	/e/	/e/	/e/
	/i/	/e/	/e/
IV	/u/	/e/	/e/
	/o/	/e/	/e/
	/a/	/e/	/a/
	/e/	/e/	/e/
	/i/	/e/	/e/

* AF: AreatoFormant FF: Formfrek

TABLE 8. *Results of synthesized sound after expansion of pharyngeal segment in vowel /o/*

Expansion of area	2x		3x
case	Method	Sound	Sound
I	AF FF	/e/ /∂/or/o/	/e/ /o/
III	AF FF	strained /a/ /a/	/a/ /∂/or/o/

* AF: AreatoFormant FF: Formfrek

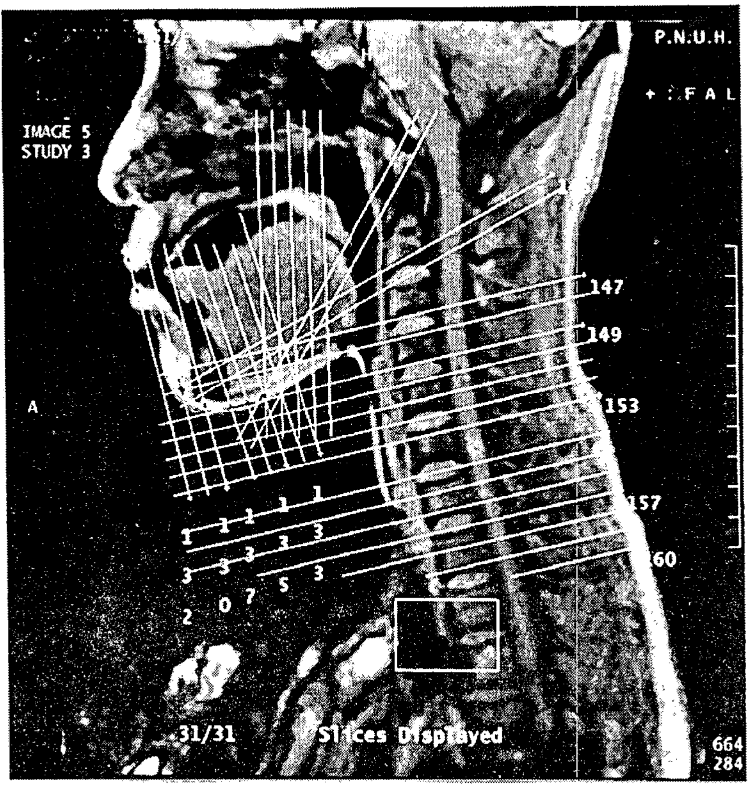
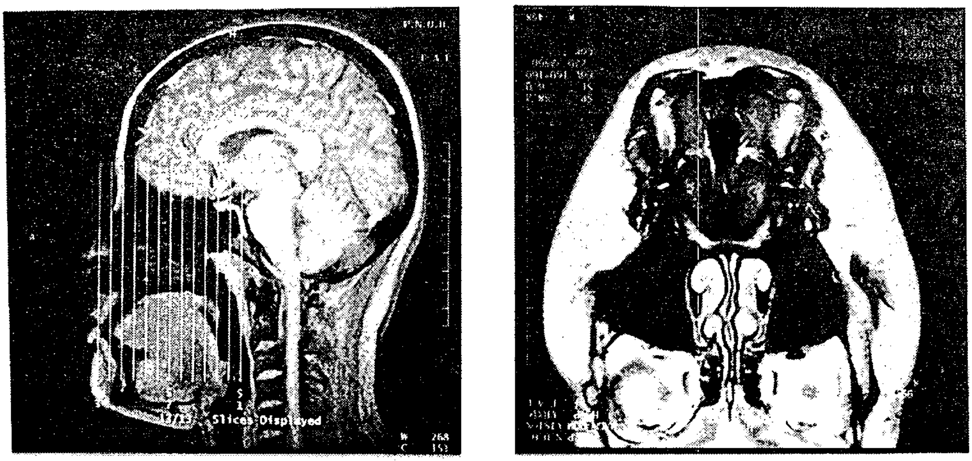
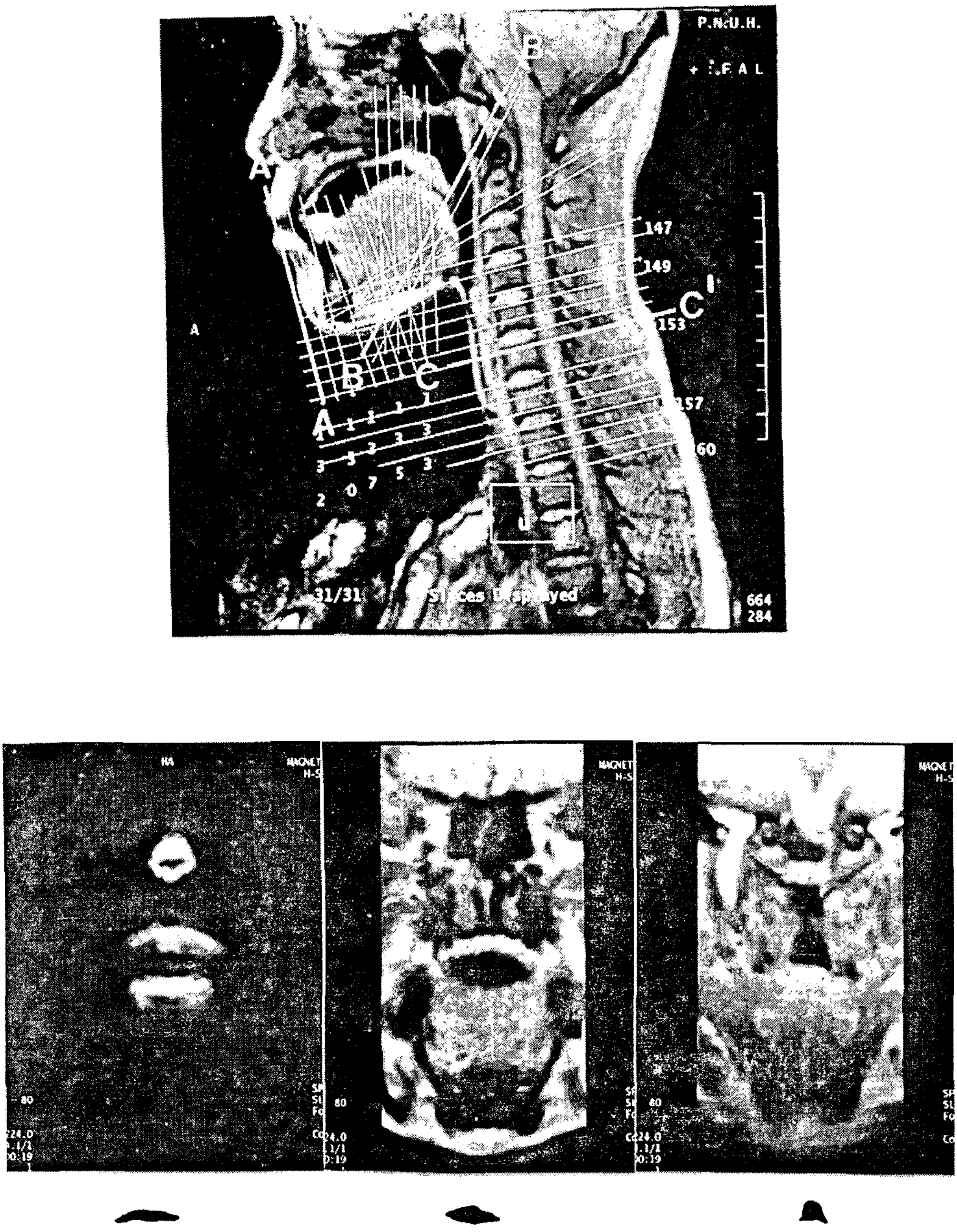


Fig. 1. Midsagittal section of vocal tract on MRI in case III



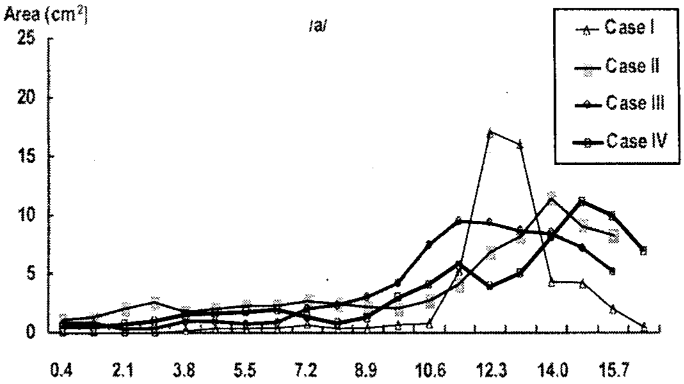
(a) (b)

Fig. 2. Cross section of mid-oral cavity by turbo spin echo MRI



(a) (b) (c)

Fig. 3. Cross sectional area in case III. a: A-A', b: B-B', c: C-C'



Distance from pseudoglottisto lip (cm)

Fig. 4. Cross-sectional area in /a/ phonation.

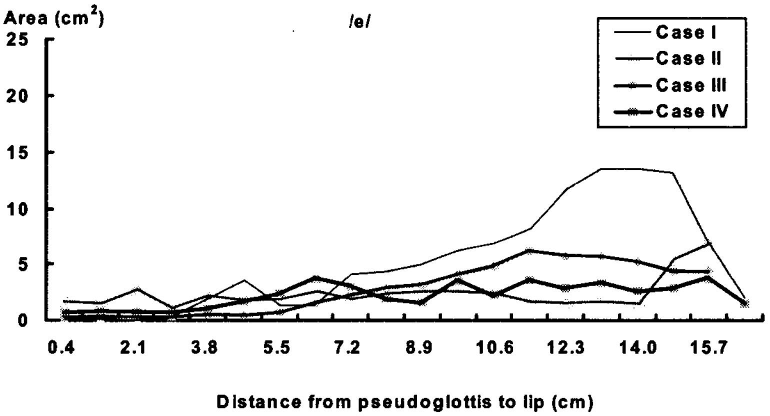


Fig. 5. Cross-sectional area in /e/ phonation.

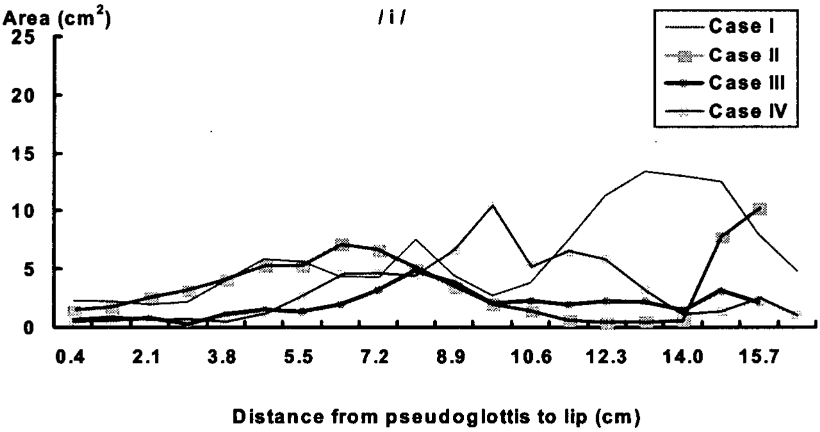


Fig. 6. Cross-sectional area in /i/ phonation.

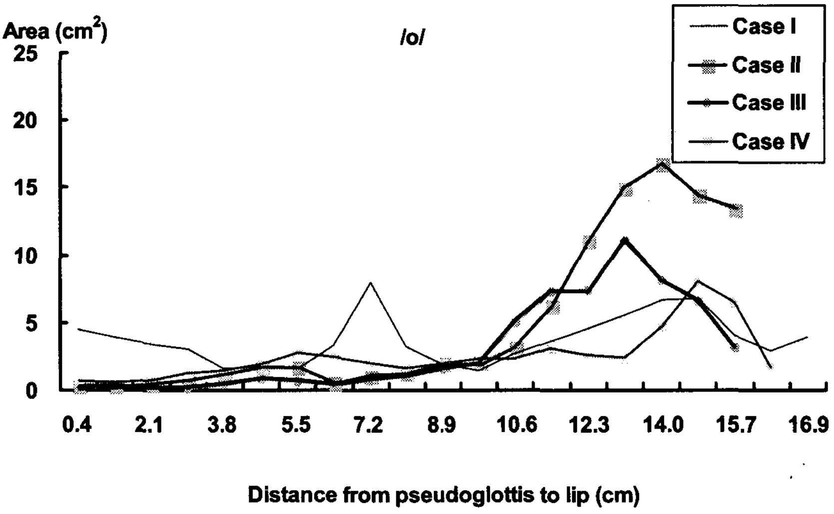


Fig. 7. Cross-sectional area in /o/ phonation.

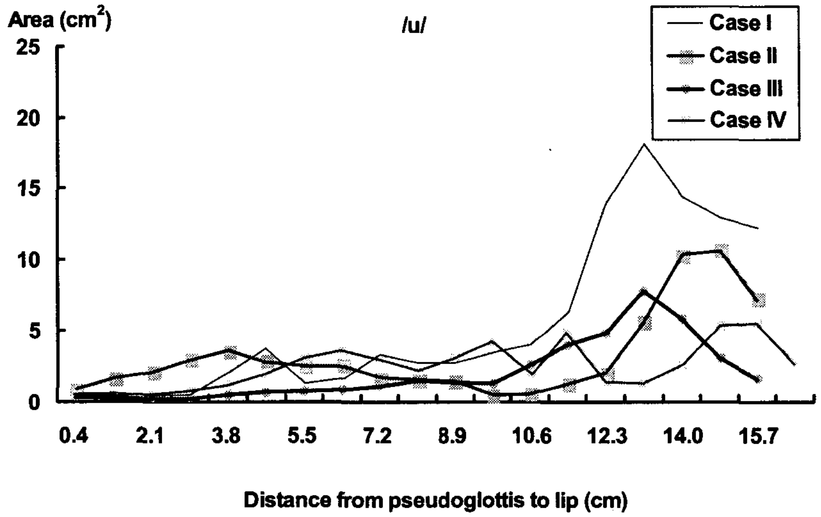


Fig. 8. Cross-sectional area in /u/ phonation.

Received : July 25, 2000.
Accepted : September 2, 2000.

▲ Cheol-Soo Kim,
M.D., Department of Otolaryngology, College of Medicine,
Pusan National University, Ami-Dong 1-10, Seo-Gu, Pusan 602-739, Korea
Tel: +82-51-240-7328, Fax: +82-51-248-1248

▲ Soo-Geun Wang,
M.D., Department of Otolaryngology, College of Medicine,
Pusan National University, Ami-Dong 1-10, Seo-Gu, Pusan 602-739, Korea
Tel: +82-51-240-7331, Fax: +82-51-248-1248
e-mail : wangsg@pusan.ac.kr

▲ Hwan-Jung Roh,
M.D., Department of Otolaryngology, College of Medicine,
Pusan National University, Ami-Dong 1-10, Seo-Gu, Pusan 602-739, Korea
Tel: +82-51-240-7333, Fax: +82-51-248-1248

▲ Eui-Kyung Goh,
M.D., Department of Otolaryngology, College of Medicine,
Pusan National University, Ami-Dong 1-10, Seo-Gu, Pusan 602-739, Korea
Tel: +82-51-240-7332, Fax: +82-51-248-1248

▲ Kyong-Myong Chon,
M.D., Department of Otolaryngology, College of Medicine,
Pusan National University, Ami-Dong 1-10, Seo-Gu, Pusan 602-739, Korea
Tel: +82-51-240-7330, Fax: +82-51-248-1248

▲ Byung-Joo Lee,
M.D., Department of Otolaryngology, College of Medicine,
Pusan National University, Ami-Dong 1-10, Seo-Gu, Pusan 602-739, Korea
Tel: +82-51-240-7335, Fax: +82-51-248-1248

▲ Soon-Bok Kwon,
M.D., Department of Otolaryngology, College of Medicine,
Pusan National University, Ami-Dong 1-10, Seo-Gu, Pusan 602-739, Korea
Tel: +82-51-240-7543, Fax: +82-51-248-1248
e-mail : sbkwon@intizen.com

▲ Suck-Hong Lee,
M.D., Department of Diagnostic Radiology, College of Medicine,
Pusan National University, Ami-Dong 1-10, Seo-Gu, Pusan 602-739, Korea
Tel: +82-51-240-7372. Fax: +82-51-248-1248

▲ Hak-Jin Kim,
M.D., Department of Diagnostic Radiology, College of Medicine,
Pusan National University, Ami-Dong 1-10, Seo-Gu, Pusan 602-739, Korea
Tel: +82-51-240-7371, Fax: +82-51-248-1248

▲ Byung-Gon Yang,
M.D., Department of English, College of Humanities, Dongeui University,
24 Kaya-dong, Pusanjin-gu, Pusan 614-714, Korea
Tel: +82-51-890-1227, Fax: +82-51-890-1209
e-mail : bgyang@hyomin.dongeui.ac.kr