# The Voiceless Stop Distinction in the Alaryngeal Speech

**Ki Hwan Hong*** · **Hyun Ki Kim****


## ABSTRACT

Theoretically, alaryngeal speakers have difficulty in accomplishing the production of voiceless consonants. However, the perceptual studies often reveal a clear production of voiceless consonants giving good articulation scores in skilled alaryngeal speakers. The purpose of the present study was to clarify the production of voiceless stops in mode of articulation to normal speakers and skilled alaryngeal speakers. The acoustic characteristics of alaryngeal speech compared to the normal speech were investigated with special reference to the voiceless stop consonants. The surface electromyography from neck is used to monitor pharyngeal activity during speech. The general result is, that esophageal, shunt and neoglottal speakers realize the distinctions between the three types of [p] in a manner parallel to normals, whereas those using an electric voice generator do not.


**Keywords : Voiceless stops, Alaryngeal speakers, Acoustic, Electromyographic.**

## 1.0 INTRODUCTION

For the production of voiceless stop consonants, normal laryngeal speakers systematically vary the voice onset time (VOT) values to characterize the prevocalic stops [1]. The appropriate control of abduction and adduction of the vocal folds produces various VOT values for stops. In Korean stop consonants of normal laryngeal speech there is a three-way distinction in both manner and place of articulation classified as glottalized, lenis and aspirated consonants. The glottalized stop is characterized by the completely adducted state of the vocal folds, the stiffened vocal folds and the abrupt decreasing of the stiffness near the voice onset at the explosion, and the aspirated type is characterized by the extensively abducted state of the vocal the vocal folds are moderately abducted at the explosion. However, alaryngeal speech (esophageal, neoglottal, shunt, and electric speech) differ from normal laryngeal speech primarily with voicing source.

* Dept. of Otolaryngology-Head and Neck Surgery, Chonbuk Univ.
* Dept. of French Language and Literature, Chonbuk Univ.

The acoustic studies for alaryngeal speech have widely been carried out with respect to voice quality, suprasegmental features, and segmental features [5]. It has been shown that the voice quality of esophageal speech is characterized by very low fundamental frequency, highly restricted range of the fundamental frequency, and a large amount of pertubations of the fundamental frequency, such as jitter and shimmer [6]. As for the supra-segmental features, some Korean fluent esophageal speakers can produce a clear production of stops resulting in good articulation scores [7], but the voice onset time and vowel duration of esophageal speech were different from those of normal speech. The tracheoesophageal shunt speech is achieved when pulmonary air is directed through a prosthesis into the upper esophagus to vibrate the pharyngoesophageal segment and produce voice. These speakers sustain phonation longer, maintain faster speaking rates with less pause time, and speak with greater intensity than esophageal speakers [8, 9].

The purpose of this study is to evaluate the three-way distinction of Korean voiceless stops in the alaryngeal speech using acoustic and EMG studies, and to investigate the characteristics of alaryngeal speech compared to the normal speech with special reference to the voiceless stop consonants.

## 2.0 MATERIALS AND METHODS

### 2.1 Subjects and speech materials

Four groups of superior alaryngeal speakers participated in this study. Each group of speakers consisted of five men who had undergone total or near total laryngectomy two to seven years earlier aged from 54 to 65 years. The esophageal speaker group began using esophageal speech within 1 year postlaryngectomy, and the tracheoesophageal shunt prosthesis (PROVOX) was inserted within about three month after laryngectomy. The neoglottal speech after near total laryngectomy and electric speech could start immediately after surgery. They were all judged as superior alaryngeal speaker groups by the speech pathologists perceptually evaluated in terms of general effectiveness of speech [10]. The normal group consisted of five men aged from fifty to sixty years.

Each speaker was tape-recorded at the same recording level, decibel, in a quiet room on a digital audio-tape recorder. A head-mounted Sony ECM50 microphone was held at a constant distance of 5 cm in front of the speaker's mouth. Test words were prepared so as to place the [p] consonants in different phonological environments (/CVCV/ syllables, C; consonant, V; vowel), and were all nonsense words. The following are test words : 1. the labial lenis stop followed by vowel /e/; /pepe/, 2. the labial aspirated stop followed by vowel /e/; /phephe/, 3. the labial glottalized stop followed by vowel /e/; /p'ep'e/. For the ideal signals, samples of phonation at comfortable pitch and loudness were recorded with repetitive phonation.

### 2.2 Acoustic Analysis

All acoustic measures were obtained using a speech analysis system made by Tokyo university [11]. This method of analysis and computer hardware/software package were selected because it is commercially available and capable of performing the selected acoustic measures with little training. All speech samples were played from a digital audio-tape recorder at a constant playback level. For phonetic analysis, the tested words were sampled at 8 kHz with 10-bit accuracy. Each word was displayed as a wide or narrow band spectrogram and acoustic waveform and, using a cursor, the duration of voice onset time (VOT) in the word initial position (/CVCV/), vowel duration during vocal fold vibration (/CVCV/), oral closure time (/CVCV/) were measured accurately. An amplitude display of the wave-form was used to measure the explosive sound intensity at the consonant's burst in the word initial position (/CVCV/). VOT in the /CVCV/ syllables is defined as the measurement between burst onset and the identifiable periodic vibration for a following vowel within the acoustic wave. Duration of oral closure time can be measured as the time from offset of vowel periodicity in a speech wave to consonant burst onset (/CVCV/), not at the word-initial position.

### 2.3 Electromyographic analysis

The Viking II EMG system (Nicolet) was used. The surface electrode was applied to the midline neck area (just above the tracheostoma) with a ground electrode to the lateral neck area. The acoustic signals using the test words were recorded simultaneously with EMG signals. On the EMG signals, the measurement of electrical activity of appropriate points for the explosion of consonants and vocal fold vibration was possible when comparing the corresponding points of the acoustic signals. The electrical activities were derived from the midline strap muscles in the normal subjects and from the remnant pharyngeal muscle in the alaryngeal speakers.

## 3.0 RESULTS

### 3.1 Acoustic findings

The voice onset time, vowel duration, closure time and explosive intensity during reading of test words were analyzed. As in the (Table 1), the mean VOT indicate that alaryngeal speech group except electrolaryngeal speech had a shorter VOT in the lenis (43.4, 39.4, 37.8 msec) and aspirated consonants (53.1, 50.2, 45.9 msec) than normal lenis (56.8 msec) and aspirated (68.9 msec), but longer in the glottalized consonants (27.6, 28.4, 32.5 msec) than normal glottalized (20.7 msec). There was no difference among alaryngeal speech except electric speech. The mean values of vowel duration (Table 2) indicate that alaryngeal speech group had a longer duration in the aspirated (176.0, 174.5, 182.9 msec) and glottalized (160.4, 174.6, 183.2 msec) than normal aspirated (145.9 msec) and glottalized (144.5 msec). In the normal speech, the vowel duration of lenis consonant were not measured because of a voiced /b/ phenomenon of voiceless /p/ in the word medial position (/CVCV/). But in the alaryngeal speech the lenis consonant appeared as a voiced or voiceless consonant and revealed longer vowel duration than the aspirated and glottalized when it was voiceless. The vowel duration was not different among alaryngeal speech except electric speech. The mean explosive sound intensity of consonants indicate that alaryngeal speech group had a larger intensity in the lenis (35.9, 32.3, 31.0 dB) than normal (24.7 dB) and a smaller intensity in the aspirated than normal, but similar to the normal group in the glottalized (Table 3). In the normal speech, the explosive intensity is higher in the aspirated (41.8 dB) and glottalized (34.2 dB) than lenis (24.7 dB), but in the alaryngeal speech the intensity is not different among consonants. The intensity is not different among alaryngeal speech except electric speech. The mean closure duration of consonants indicate that alaryngeal speech group except electric speech had a shorter duration in the aspirated (107.9, 102.3, 121.6 msec) and glottalized (107.1, 122.5, 121.5 msec) than normal aspirated (159.0 msec) and glottalized (161.0 msec), and were not different between alaryngeal speech. The closure duration of the lenis in the normal were not measured also because of a voiced /b/ phenomenon, but voiced or voiceless sometimes in the alaryngeal speech (Table 4). When the lenis consonant was voiceless, the closure duration revealed shorter than the aspirated and glottalized.

Table 1. Overall Values(msec) and Statistical Comparisons with Normal and Alaryngeal Speech for Voice Onset Time

| Consonants | Normal | Alaryngeal | | | |
|---|---|---|---|---|---|
| | | ES | SS | NS | ELS |
| Lenis | 56.8 | 43.4 | 39.4 | 37.8 | >20.0 |
| F-values with normal | | 88.1* | 199.2* | 176.0* | N/C |
| Aspirated | 68.9 | 53.1 | 50.2 | 45.9 | >20.0 |
| F-values with normal | | 98.2* | 223.5* | 330.7* | N/C |
| Glottalized | 20.7 | 27.6 | 28.4 | 32.5 | >20.0 |
| F-values with normal | | 86.1* | 59.1* | 128.6* | N/C |
| Among consonants | 829.3* | 321.5* | 161.6* | 64.8* | 0.0 |

ES:esophageal speech, SS:shunt speech, NS:neoglottal speech, ELS: electrolaryngeal speech, *: statistical significant(p<.01)
Degree of freedom(df) is 2,36 in the among consonants, and 1,24 in the between consonants with normal, NC; not checked.

Table 2. Overall Values(msec) and Statistical Comparisons with Normal and Alaryngeal Speech for Vowel Duration

| Consonants | Normal | Alaryngeal | | | |
|---|---|---|---|---|---|
| | | ES | SS | NS | ELS |
| Lenis | N/C | NC | N/C | N/C | N/C |
| F-values with normal | | N/C | N/C | N/C | N/C |
| Aspirated | 145.9 | 176.0 | 174.5 | 182.9 | N/C |
| F-values with normal | | 223.0* | 97.1* | 112.1* | 194.1* |
| Glottalized | 144.5 | 160.4 | 174.6 | 183.2 | N/C |
| F-values with normal | | 31.6* | 118.8* | 159.3* | 208.2* |
| Among consonants | N/C | N/C | N/C | N/C | N/C |

ES:esophageal speech, SS:shunt speech, NS:neoglottal speech, ELS: electrolaryngeal speech, *: statistical significant(p<.01)
Degree of freedom(df) is 1,24 in the between consonants with normal, NC; not checked.

Table 3. Overall Values(dB) and Statistical Comparisons with Normal and Alaryngeal Speech for Explosive Power

| Consonants | Normal | Alaryngeal | | | |
|---|---|---|---|---|---|
| | | ES | SS | NS | ELS |
| Lenis | 24.7 | 35.9 | 32.3 | 31.0 | >10.0 |
| F-values with normal78.3* | | 22.0* | 38.3* | | N/C |
| Aspirated41.8 | 44.1 | 38.3 | 38.0 | | >10.0 |
| F-values with normal | | 7.2* | 29.0* | 15.7* | N/C |
| Glottalized | 34.2 | 33.4 | 33.3 | 34.7 | >10.0 |
| F-values with normal | | 0.4 | 0.3 | 0.5 | N/C |
| Among consonants | 77.4* | 32.7* | 1.5 | 4.8* | 0.0 |

ES:esophageal speech, SS:shunt speech, NS:neoglottal speech,
ELS: electrolaryngeal speech, *: statistical significant(p<.01)
Degree of freedom(df) is 2,36 in the among consonants, and 1,24
in the between consonants, NC; not checked.

Table 4. Overall Values(msec) and Statistical Comparisons with Normal and Alaryngeal Speech for Closure Duration

| Consonants | Normal | Alaryngeal | | | |
|---|---|---|---|---|---|
| | | ES | SS | NS | ELS |
| Lenis | N/C | 69.5 | N/C | N/C | N/C |
| F-values with normal | N/C | N/C | N/C | N/C | |
| Aspirated | 159.0 | 107.9 | 102.3 | 121.6 | N/C |
| F-values with norma | 1547.6* | 347.4* | 111.1* | 218.3* | |
| Glottalized | 161.0 | 107.1 | 22.5 | 121.5 | N/C |
| F-values with norma | 1561.9* | 281.1* | 190.3* | 236.4* | |
| Among consonants | N/C | N/C | N/C | N/C | N/C |

ES:esophageal speech, SS:shunt speech, NS:neoglottal speech,
ELS: electrolaryngeal speech, *: statistical significant(p<.01)
Degree of freedom(df) is 2,36 in the among consonants, and 1,24
in the between consonants, NC; not checked.

3.2 Electromyographic findings

In the normal laryngeal speakers, the electrical activities were derived from the midline strap muscle. However, in the alaryngeal speakers, although the strap muscles were removed already, the electrical activities of the muscle were also derived from the remnant pharyngeal muscle. Instead of the quantitative measurement of electrical activities of the midline strap muscle and remnant pharyngeal muscle in the normal and alaryngeal speakers, the patterns of electrical activity were observed at the explosion of stop consonants. In the normal speakers, the electrical activities of the midline neck muscle during the production of aspirated and glottalized consonants were higher than for the lenis consonant (Fig. 1). In the esophageal, tracheoesophageal and neoglottal speakers, the electrical activities during production of the aspirated and glottalized consonants were higher than for the lenis consonant just like as in the normal (Fig. 2, 3). But in the electric speech the electrical activities for the three types of consonant were not different as in Fig. 2, 3. These phenomenon were shown in all normal subjects and all fluent alaryngeal speakers.

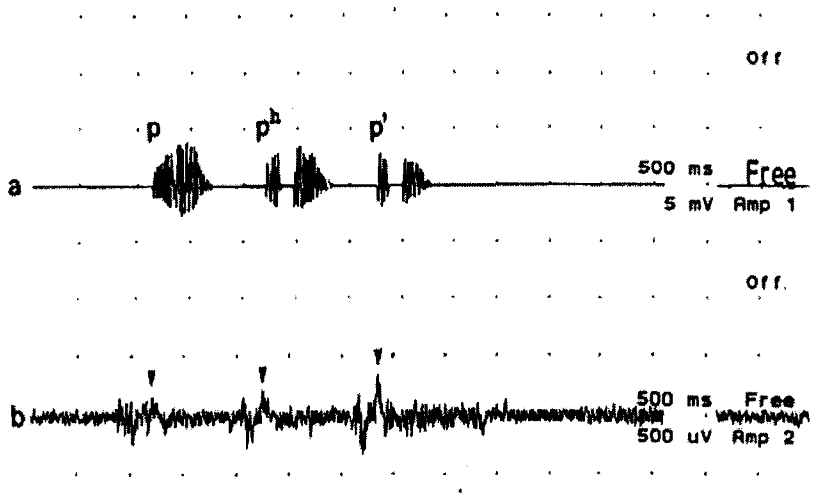


Fig. 1. Acoustic (a) and electromyographic (b) signals during production of Korean bilabial stops followed by vowel /i/ in normal speech. Electromyographic activities are highest in the glottalized followed by the aspirated and lenis stops.

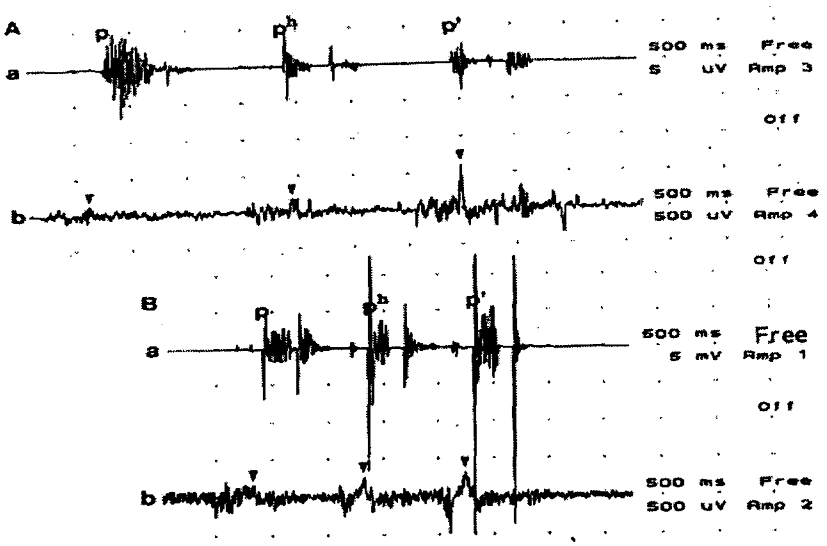


Fig. 2. Acoustic(a) and electromyographic(b) signals during production of Korean bilabial stops followed by vowel /i/ in esophageal (A) and shunt (B) speech. Electromyographic activities are highest in the glottalized followed by the aspirated and lenis stops.

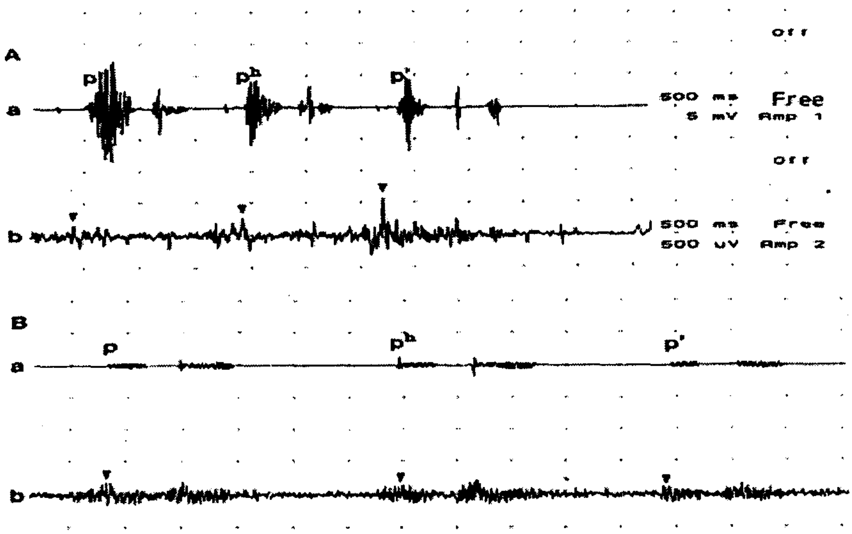


Fig. 3. Acoustic (a) and electromyographic (b) signals during production of Korean bilabial stops followed by vowel /i/ in neoglottal (A) and electrolaryngeal (B) speech. Electromyographic activities are highest in the glottalized followed by aspirated and lenis stops, but not in the electric speech.

## 4.0 DISCUSSION

Total laryngectomy results in a removal of tissue necessary to normal vocal function and in change of the anatomy and physiology of the vocal tract for speech production. As a result of removal of the larynx, the processes of speech production in the alaryngeal speech are changed by using alternative methods of voicing. In the esophageal and tracheoesophageal speech, the pharyngoesophageal segment is apparently not supported by an abductor-adductor system, and there is no evidence to support the view that laryngectomized individuals are capable of altering the level of muscular activity within the pharyngoesopheal segment.

In this study, VOT values of alaryngeal speech except electric speech are shorter than normal speech in the aspirated and lenis stops, but longer in the glottalized stop than normal. But the mean VOT values for the lenis and aspirated stops in the esophageal, tracheoesophageal and neoglottal speech are sufficient for achieving voiceless stops and quite important cues for voiceless stop distinction. Also the esophageal and tracheoesophageal speakers did effect systematic variation in the timing of voice onset during the production of stops, and the VOT values associated with the production of prevocalic voiceless stops exhibited lag intervals which were significantly shorter than those used by normal speakers [12]. Although the alaryngeal speaker does not possess the abductor-adductor properties of phonatory apparatus, it would be reasonable to suppose that alaryngeal speakers might exhibit differences in VOT. It is well known that vowel duration are conditioned by the voicing features of their consonant contexts. Christensen (12) measured vowel duration from symmetric CVC syllables produced by ten esophageal speakers. They revealed that duration of vowels spoken by esophageal speakers within voiceless consonant contexts were consistently longer than vowels spoken in similar contexts by normal speakers. In addition, vowel duration spoken by esophageal speakers in voiced consonant contexts were always longer that those uttered in voiceless consonant environments. On the normal Korean stops, Hong et al. [13] reported that vowel duration after explosion of the aspirated and glottalized stops were not different, but could not be measured in the lenis stop due to voiced /b/ phenomenon. In this study, vowel duration for stops in the alaryngeal speech except electric speech show consistently longer than normal speakers. The lenis stop shows a voiced or voiceless phenomenon in the alaryngeal speech, and shows longer duration than other stops when it is a voiceless. These data of vowel duration suggest that the alaryngeal speakers presumably makes compensatory adjustments in the timing control system in order to realize these variations in vowel length before voiceless consonants

The oral closure duration is called " phonation-off time", and this measure includes both consonant closure and release. As in normal Korean speech, the phonation-off times are about 160 msec in the aspirated and glottalized stops, but not are shown in the lenis stop due to the voiced /b/ phenomenon in the word medial position (13). This study shows shorter phonation-off times of the aspirated and glottalized in the alaryngeal speech than normal. At lower intelligibility levels, phonation-off times were decreased as in the alaryngeal speech. The explosive sound intensities differ in the normal Korean stops, significantly higher intensities in the aspirated and glottalized than the lenis. This study shows higher intensities in the aspirated followed by the glottalized and lenis, but is not significant between the glottalized and lenis. Murry et al. [14] reported that the speakers showing high intelligibility had lower explosive power than the normal. In this study, the explosive power in the lenis is higher than normal and results in low intelligibility level, but similar levels in the aspirated and glottalized. These values also are quite important cues for voiceless stop distinction.

Hong et al. [15] reported the normal activities of posterior cricoarytenoid (PCA) muscle and thyroarytenoid (TA) muscle during production of Korean stops using hook-wire electrodes. In the word initial position, the aspirated stop showed the most marked and earliest activation of the PCA muscle associated with a steep reactivation of the TA muscle before the voice onset. The glottalized stop showed little or no activation of the PCA muscle associated with the earliest and most marked reactivation of the TA muscle. On the other hand, the lenis stop showed more moderate activation of the PCA muscle than for the glottalized stop, and the least reactivation of the TA muscle. However, Total laryngectomy needs sacrifice of all internal laryngeal structures. The source of vibration in the alaryngeal speech is the pharyngoesophageal segment. The pharyngoesophageal segment is apparently not supported by an abductor-adductor system, and there is no evidence that the pharyngoesophageal segment controls or influences the vibratory rate of this sphincter on a systematic basis. In the electromyographic study of muscular activity of the inferior constrictor and cricopharyngeus in laryngectomized speakers using esophageal speech, Shipp [16] failed to observe any consistent typical or modal patterns of muscle activity, and noted that there was substantial inter-subject variation in the patterning of muscular activity of these muscles during voicing and no single type of activity of muscle behavior during the phonatory portion of alaryngeal voice. Schwart [17] mentioned that the esophageal speaker may produce voiceless sounds in isolation through the use of tongue and lip and intraoral air pressure in syllable initial position, which can be followed by phonation without undue delay. However, this study shows, in the esophageal, tracheoesophageal and neoglottal speakers, that the electrical activities of the remnant pharyngeal muscle during production of the aspirated and glottalized consonants were higher than for the lenis consonant like as in the normal, even though the tested muscles were different.

But in the electric speakers the electrical activities for the three types of consonants were not different. These result supports the suggestion that alaryngeal speakers may use the remnant pharyngeal muscle for producing voiceless stop.

By way of summary for this study, there are two major findings of the acoustic data. First, alaryngeal speakers systematically varied VOT for the production of phonetically representative stops, and the general pattern of these variations paralleled that observed for normal speakers. Second, common characteristics of fluent alaryngeal speakers compared to the normal are longer vowel duration, shorter closure duration, and higher explosive power in the lenis. The differences of acoustic data between alaryngeal speech except electric speech are not shown. On the EMG data, this study also indicates that the different activities of the remnant pharyngeal muscle between stops may be important for the production of voiceless distinction.

## REFERENCES

[1] Lisker L, Abramson A. S. 1967. "Some effects of context on voice onset time in English stops." *Language Speech 10*, 1-28.

[2] Han M. S., Weitzman R. S. 1970. "Acoustic features of Korean /P,T,K/, /p,t,k/ and /p$^h$, t$^h$, k$^h$/." *Phonetica* 22, 112-8.

[3] Kagaya R. A. 1974. "Fiberscopic and Acoustic Study of the Korean Stops, Affricates and Fricates." *J. Phonetics* 2, 161-80.

[4] Hirose H., Lee, C. Y., Ushijima, T. 1974. "Laryngeal control in Korean Stop Production." *J. Phonetics 2,* 145-52.

[5] Clark J. G., Stemple J. C. 1982. "Assessment of three modes of alaryngeal speech with a synthetic sentence identification (SSI) task in varying message-to-competition ratios." *J. Speech Hear Res 25,* 333-8.

[6] Robbins L., Fisher H. B., Blom E. C., Singer M. I. 1984. "A comparative acoustic study of normal, esophageal, and tracheoesophageal speech production." *J. Speech Hear Dis 49,* 202-10.

[7] Hong, K. H., Jung, W. C., Youn, H. W., Kim, H. K. 1997. "Acoustic and electromyographic characteristics of fluent alaryngeal speech." Korean *J. Otolaryngol 40,* 879-87.

[8] Baggs T. W., Pine S. J. 1983. "Acosutic charateristics : tracheoesophageal speech." *J. Communication Dis 16,* 299-307.

[9] Wetmore S. J, Krueger K., Wesson K. 1981. "The Singer-Blom speech rehabilitation procedure." Laryngoscope 91, 109-17.

[10] Snidecor J. C., Curry E. T. 1959. "Temporal and pitch aspects of superior esophageal speech." *Ann Otolaryngol 68,* 623-6.

[11] Imagawa H., Kiritani S. 1989. "High speed speech analysis system using a personal computer with DSP and its applications to pronounciation training." *Ann Bull RILP(Tokyo university) 23,* 173-85.

[12] Christensen J. M., Weinberg B., Alfonso P. J. 1978. "Productive voice onset time characteristics of esophageal speech." *J. Speech Hear Res 21,* 56-62.

[13] Hong, K. H., Seiji N., Hajime H. 1991. "Laryngeal adjustments for the Korean stops, Affricates and Fricatives(EMG study)." *Ann Bull RILP (Tokyo univ.) 25,* 17-31.

[14] Murry T., Brown W. S. 1975. "Intraoral air pressure variability in esophageal speakers." *Folia Phoniat 27,* 237-49.

[15] Hong, K. H., Chon, D. S., Kim, Y. J., Jung, K. Y. 1992. "Laryngeal adjustments for Korean stop." "acoustic, electromyographic and fiberoptic study." *Korean J. Otolaryngol 35,* 770-82.

[16] Ship T. 1967. "Frequency, duration, and perceptual measures in relation to judgements of alaryngeal speech acceptability. *J. Speech Hear Res 10,* 417-427.

[17] Schwart R. J. 1982. "Articulatory adjustments in esophageal speech." Paper presented at *the III World Congress of Laryngectomees,* Tokyo, August.

Received : Jan. 25, 2000.
Accepted : Feb. 29, 2000.

▲ Ki Hwan Hong
Dept. of Otolaryngology-Head and Neck Surgery,
Chonbuk National University, Medical School,
Chonju, Chonbuk, KOREA 560-182,
Tel : +82-652-250-1990
Fax : +82-652-250-1986
e-mail : khhong@moak.chonbuk.ac.kr

▲ Hyun Ki Kim
Dept. of French Language and Literature,
Chonbuk National University,
Chonju, Chonbuk, KOREA 560-182,
Tel: +82-652-270-3671
Fax: +82-652-250-1986
e-mail : hyungk@moak.chonbuk.ac.kr