Voicing Distinction in Tracheoesophageal Speech -an Aerodynamic Study-

N. Nishizawa*, N. Sakai*, Y. Inuyama*
K. Tanaka**
M. Kato***, M. Takahashi*** and T. Ifukube***

*Department of Otolaryngology, Hokkaido University, School of Medicine, Kita 15 Nishi 7, Kita-ku, Sapporo 060, Japan
**ENT Clinic, Sapporo National Hospital, 4-2, Kikusui, Shiroishi-ku, Sapporo 003, Japan
***Research Institute for Electronic Science, Hokkaido University, Kita 14 Nishi 7, Kita-ku, Sapporo 060, Japan

Introduction

Esophageal and tracheoesophageal (TE) speakers have difficulty in accomplishing the voiced/voiceless distinction for consonants (Hirose et al, 1983; Nishizawa et al,1992). This difficulty must be attributed to the pathological structure of their articulatory system.

The laryngectomized patients maintain the normal vocal tract but lack the laryngeal mechanism. The vibrating source of esophageal and TE speech is called "pseudoglottis". It is the mucosal fold arising from the posterior wall of the reconstructed pharyngoesophagus. There is a general agreement that formation of the pseudoglottis is associated with the activity of the remnant of the inferior pharyngeal constrictors. It has been suggested that activities of the pharyngeal constrictors have some influence on physical characteristics of the pseudoglottis(Shipp, 1970; Komiyama et al, 1976; Nishizawa et al; 1993). However, from a viewpoint of speech production, adjustments of the pseudoglottis must not be so fine as of the normal larynx.

Moreover, the airflow source driving the pseudoglottal vibration is served differently from normal laryngeal speech. In esophageal speech, the pseudoglottis is carried into vibration by the regurgitating air injected into the esophagus. Although TE speakers utilize pulmonary airflow for phonation, it is introduced through the TE fistula into the subpseudoglottal cavity.

Thus, these alaryngeal speakers have different mechanisms for voicing control from normal speakers in two ways; lack of the normal laryngeal adjustment and modified airflow. Nevertheless, literatures have revealed that skilled esophageal and TE speakers produce voiced and voiceless cognates with sharp distinction (Hirose et al, 1983; Nishizawa et al, 1992). In these alaryngeal speech, the timing of the pseudoglottal vibration during consonant articulation has been suggested to be the major acoustic cue for voicing distinction as in the case of normal laryngeal speech (Hirose et al, 1983). However the physiological mechanism for on-off control of the pseudoglottal vibration is not yet clear.

In the present study, mechanisms of voicing distinction for consonants were studied using three excellent TE speakers. It was supposed that transpseudoglottal pressure gradient is built up differently from normal laryngeal speech, partially because of the modified airflow and partially because of the lack of the normal laryngeal mechanism. Special interest was taken in the effect of transpseudoglottal pressure gradient on the vibration of the pseudoglottis.

Subjects and methods

1. Subjects

The subjects were three Japanese laryngectomees who were in good health, KM (68 y-old male), KT (52y-old female) and YM(58 y-old male). TE shunt was constructed at the same time as the total laryngectomy according to the method described before (modified Amatsu-tracheoesophageal shunt operation (Tanaka et al, 1987)). All of them had good intelligibility of speech without voice prostheses.

2. Speech materials

The speech materials for the experiment were eight nonsense disyllables consisting of vowel-consonant-vowel sequences with /p/,/b/, /k/ and /g/ for consonants and /a/ and /i/ for vowels(Table 1). The subjects were required to pronounce each test word in a frame sentence "hai .... desu (Yes, it is ....)". Each test word was repeated 40 times by subject KT and 20 times by subjects KM and YM .

Table 1.speech materials

/apa/	/aba/	/aka/	/aga/
/ipi/	/ibi/	/iki/	/igi/

3. Data acquisition and processing (Figure 1)

Sub- and suprapseudoglottal air pressure recordings were obtained simultaneously during the production of the speech samples. Suprapseudoglottal air pressure was measured with a miniature pressure transducer (Camino 420 XP) inserted through a nostril into the oropharynx. In order to measure subpseudoglottal air pressure, another pressure transducer was inserted through the tracheostoma into the subpseudoglottal cavity via TE fistula. The outputs were amplified and recorded on a multichannel data recorder (TEAC 210-A) together with acoustic signals.The air pressure recordings were later digitized at a sampling rate of 2kHz for a computer processing.

4. Classification of the speech samples

The subjects in the present study did not always produce voiced and voiceless cognates with definite distinction. Judgment was made upon the voiced or voiceless feature of the consonant in each utterance sample. A phonetically trained listener judged auditorily the test syllables. Separately from the auditory judgment, the audio waveforms were examined. The segment concerned was categorized as voiced when the audio waveform revealed periodic pulse at the time of consonant release and as unvoiced when the periodic pulse was not detectable at release. Only the recordings on which these separate judgments agreed were chosen for processing. Thus, the speech samples were divided into four categories; voiceless consonant with voicelessness, voiced consonant with voicing, voiceless consonant replaced by voiced cognate and voiced consonant replaced by voiceless cognate.

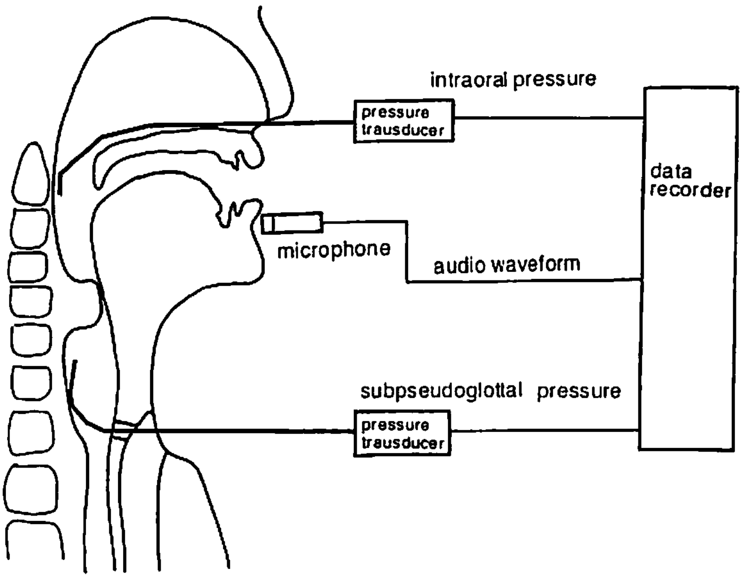

Fig.1 Schematic representation of the experiment

Results

1. Voiced/ voiceless confusions

Table 2 shows the rate of voiced/ voiceless confusion for the consonant part of the speech materials in each subject. Subject KM produced voiced and voiceless cognates definitely without any confusion. Subject KT and YM showed some confusions in voicing distinction. In subject KT, replacement of voiceless sound by voiced cognate was much more frequent than that of voiced sound by voiceless cognate (18.8:8.8%). In subject YM, all of the confusions were replacement of voiceless sound by voiced cognate (47.9%). Replacement of voiceless sound by voiced cognate appeared more frequently for /p/ than /k/ in both subjects who showed this type of confusion (Pr=0.01). In subject KT, replacement of voiceless sound by voiced cognate appeared more frequently when the vowel part of the speech material was /a/ than /i/ (Pr=0.025). Subject

Table 2 Comparison of the rate of voiced/ voiceless confusion (confused samples/ total samples) with respect to difference in preceding and succeeding vowel and place of articulation

consonant	p		k		b		g	
vowel	a	i	a	i	a	i	a	i
subject								
KM	-	-	-	-	-	-	-	-
KT	14/34	8/35	4/34	-	-	-	3/35	3/33
YM	9/12	11/11	2/12	2/13	-	-	-	-

Table 3 Results of the Ps and Po measurement with respect to voiced/voiceless feature of consonant part of the speech material

speech material	voiced consonant				voiceless consonant			
	Ps(cmH2O)		Po(cmH2O)		Ps(cmH2O)		Po(cmH2O)	
	mean	SD	mean	SD	mean	SD	mean	SD
subject								
KM	67.1	8.8	8.8	8.6	69.4	10.4	48.0	10.6
KT	18.4	4.8	4.9	3.3	20.1	5.0	13.7	6.1
YM	56.3	11.0	7.1	3.8	57.4	14.4	17.6	9.3

KT replaced voiced sound by voiceless cognate for a few cases of /g/ (8.8%). No devoicing occurred for /b/.

2. Sub- and suprapseudoglottal pressure for voicing distinction

Tables 3 and 4 summarize the variations of intraoral air pressure (Po) and subpseudoglottal air pressure (Ps) at the time of consonant release. When the pressures are compared irrespective of accuracy for voicing distinction (Table 3), the mean values for Ps measurements did not differ statistically between voiceless and voiced consonants whereas mean values for Po measurements were significantly higher for voiceless consonants than for voiced consonants (Pr=0.01). Mean Po values for voiceless consonants with voicelessness were significantly higher than that for voiceless consonants replaced by voiced cognates (Table 4) ; the differences were statistically significant in utterances /apa/ and /ipi/ for KT and in /apa/ and /aka/ for YM (Pr=0.01), whereas they were not significant in utterances /aka/ for KT and /iki/ for YM probably because of the small number of samples with voicing. When the voiceless consonants were produced with voicelessness (Table 4), mean Po values were significantly higher for utterance /aka/ than for /apa/ in all the subjects (Pr=0.025). In subjects KM and KT , mean Po values were significantly higher for utterance /ipi/

Table 4 Results of the Ps and Po measurement with respect to the speech material and to the voiced (V+)/ voiceless (V-) feature of actually produced sound

K.M.

consonant		p		b		k		g	
vowel		a	i	a	i	a	i	a	i
Ps (cmH2O)									
V+	mean	-	-	61.5	65.1	-	-	70.3	71.1
	SD	-	-	5.8	10.3	-	-	8.0	7.3
V-	mean	60.3	74.7	-	-	64.4	78.1	-	-
	SD	7.3	8.1	-	-	6.3	8.4	-	-
Po (cmH2O)									
V+	mean	-	-	14.1	19.6	-	-	0.3	2.2
	SD	-	-	2..9	4.9	-	-	0.9	1.3
V-	mean	37.5	51.0	-	-	45.5	58.1	-	-
	SD	7.7	6.4	-	-	6.2	9.6	-	-

K.T.

consonant		p		b		k		g	
vowel		a	i	a	i	a	i	a	i
Ps (cmH2O)									
V+	mean	25.0	22.3	19.7	20.7	25.5.	-	18.9	15.0
	SD	4.3	8.3	5.1	5.3	2.1	-	3.2	3.2
V-	mean	16.0	21.2	-	-	20.5	19.4	14.3	10.5
	SD	3.4	6.2	-	-	4.8	2.8	3.4	0.5
Po (cmH2O)									
V+	mean	4.4	4.9	3.8	5.7	13.9	-	3.8	7.2
	SD	1.9	0.9	2.4	2.0	8.8	-	3.4	3.8
V-	mean	11.1	17.0	-	-	18.3	17.3	9.7	6.3
	SD	3.6	5.1	-	-	4.5	3.1	1.1	1.1

Y.M.

consonant		p		b		k		g	
vowel		a	i	a	i	a	i	a	i
Ps (cmH2O)									
V+	mean	58.8	46.4	63.2	51.2	51.5	67.1	58.6	53.8
	SD	23.5	4.4	14.1	10.5	1.9	9.6	8.5	8.4
V-	mean	50.0	-	-	-	65.2	64.5	-	-
	SD	0.67	-	-	-	13.7	7.5	-	-
Po (cmH2O)									
V+	mean	7.7	9.3	11.9	9.8	16.9	24.6	4.5	3.7
	SD	2.7	2.9	2.6	2.3	2.2	6.5	1.3	0.6
V-	mean	14.9	-	-	-	27.9	26.5	-	-
	SD	1.4	-	-	-	2.3	4.2	-	-

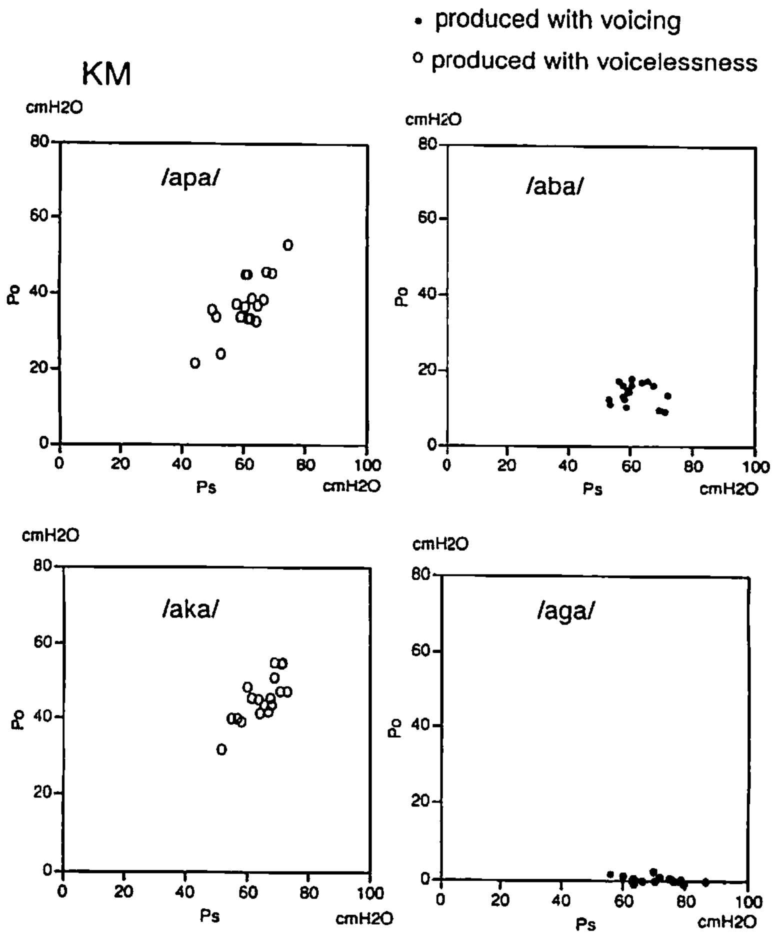

Fig.2-1 Relationships of Ps and Po values for individual utterances with respect to the speech material. Open circles represent the value for sounds actually produced with voicelessness and closed circles represent the value for sounds actually produced with voicing. (Subject KM)

than /apa/when they were produced with voicelessness (Pr =0.01) ,whereas the comparison could not be made in subject YM because all the samples of /ipi/ were produced with voicing.

The relationships between Ps and Po values for individual utterances are presented in figures 2-1,2,3. In subjects KM and KT, a positive relationship was seen for voiceless consonants with voicelessness. In subject Y.M., such positive relationship was not found. In each utterance, Ps value was constantly higher than Po not only for voiced sound but for voiceless sound.

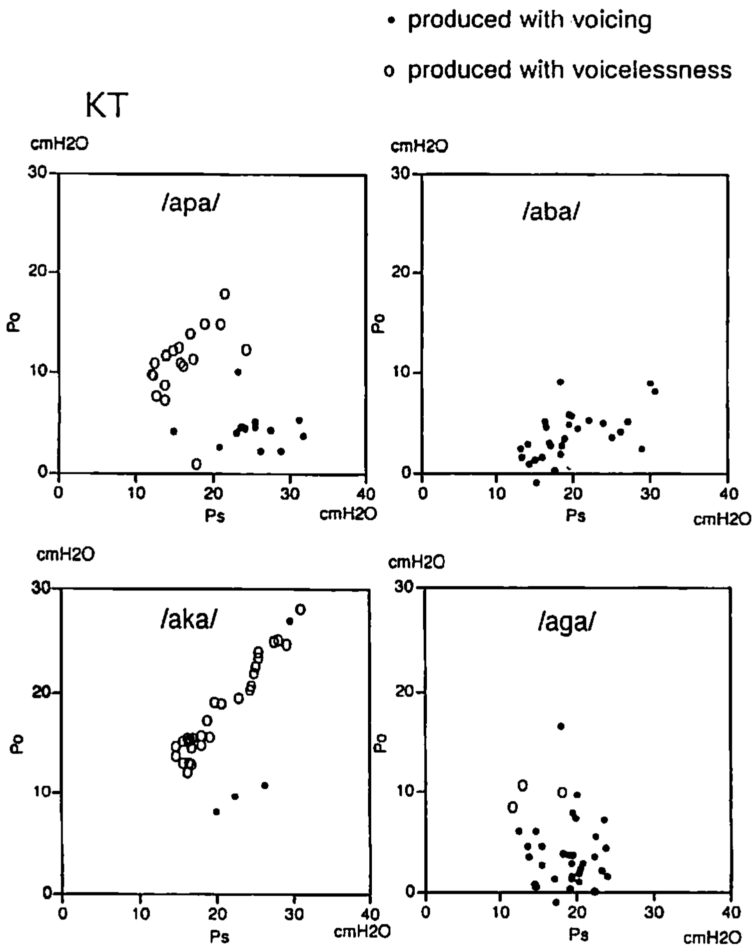

Fig.2-2 Relationships of Ps and Po values for individual utterances with respect to the speech material. Open circles represent the value for sounds actually produced with voicelessness and closed circles represent the value for sounds actually produced with voicing. (Subject KT)

Discussion

1. Aerodynamics of voicing distinction in laryngeal and alaryngeal speech

There is a general agreement that, in the normal laryngeal speech, voiceless consonants have greater peak intraoral pressure values than their voiced cognates (Subtelny et al,1966; Arkebauer et al, 1967; Brown & McGlone, 1969; Lisker, 1970; Lubker & Parris, 1970) .The intraoral pressure difference between the voiced and voiceless consonants has been attributed

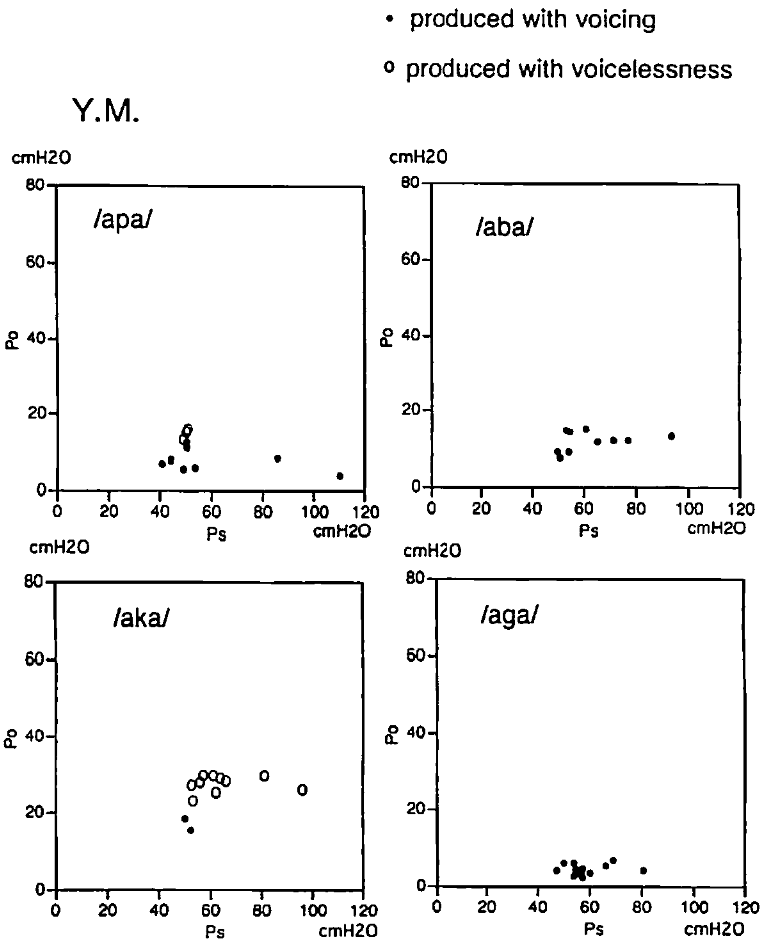

Fig.2-3 Relationships of Ps and Po values for individual utterances with respect to the speech material. Open circles represent the value for sounds actually produced with voicelessness and closed circles represent the value for sounds actually produced with voicing. (Subject YM)

mainly to the glottal impedance (Arkebauer et al, 1967; Malecot, 1966; Netsell, 1969; Klich, 1982). Glottal impedance is higher when the vocal folds are adducted and vibrating for the production of voiced consonants than it is when the vocal folds are abducted and not vibrating for the production of voiceless consonants. Thus, there is less transglottal airflow in voiced consonants and lower intraoral pressure results. On the other hand, it has been suggested that there are some supralaryngeal adjustments contributing to the intraoral pressure difference for the voicing distinction (Kent & Moll,1969; Bell-Berti, 1975; Westbury, 1983). Active control for the volumetric expansion of the supralaryngeal cavity is suggested to contribute to the buildup of the intraoral air pressure. These supralaryngeal adjustments of the intraoral pressure are supposed, in turn, to influence the vibration of the vocal fold by altering the transglottal pressure gradient and by passive abduction-adduction of the glottis (Bell-Berti, 1975; Stevens,1991). We can point out that these aerodynamics of the voicing distinction for the normal laryngeal speech are applicable for some results of the present study on TE speech.

In the present study, the speech materials of voiceless consonants were associated with higher Po measurements than their voiced cognates. Ps measurements showed no significant difference between voiced and voiceless consonants. It is suggested from these results that, as in the normal laryngeal speech (Netsell, 1969), the active adjustment for voicing distinction in TE speakers resides not in the subpseudoglottal structure but in the pseudoglottal or suprapseudoglottal structure.

Po measurements for voiceless consonants were significantly higher when the speech materials were produced with voicelessness than when they were replaced by voiced cognates. The pseudoglottal resistance against upcoming airflow is thought to be higher when it is in vibration as in the normal laryngeal speech. Consequently, Po seems to be lowered when the pseudoglottis is in vibration regardless of the subject's aim of voicing distinction.

In the normal laryngeal speech, the intraoral pressure for stop consonant is thought to be affected by volume expansion of the vocal tract. If the respiratory output and laryngeal complexity is the same, higher intraoral pressure is expected for the smaller oral cavity size (Brown & McGlone, 1969; Isshiki & Tanabe, 1972). The present study agrees with the above consideration in that , in voiceless consonants produced with voicelessness, Po values tend to be higher when the consonant part of the speech material was /k/ than /p/ and when the vowel part was /i/ than /a/. As formerly mentioned, in the normal laryngeal speech, high intraoral pressure is thought to suppress the vibration of the vocal fold by lowering the transglottal pressure gradient and by passive abduction of the glottis (Bell-Berti, 1975; Stevens, 1991). It follows that, if the respiratory and laryngeal adjustments are the same, the suppressive effect of intraoral pressure on the vocal fold vibration is larger when the occlusion occurs more backward and when the preceding and succeeding vowel is narrower. The voiced/voiceless confusions seen in subjects KT and YM agree well with the considerations above. These subjects showed that replacement of voiceless consonant by voiced cognate occurred most frequently when the speech material was /apa/ and that replacement of voiced consonant by voiceless cognate never occurred when the speech material was /aba/. For these two subjects, we can assume that the pseudoglottal vibration is directly affected by intraoral pressure.

Difference of the absolute values exists between the mean Ps measurements of the present study and that of the normal laryngeal speech. The mean Ps values are higher in the TE speakers examined than the comparable values for normal speakers(9-11cmH2O (Netsell,1969)).The higher Ps values in the present study are thought to be the result of higher pseudoglottal impedance and higher respiratory effort (Kinishi & Amatsu, 1986).

The glottal impedance is minimum in the normal laryngeal speech for voiceless stop consonants. Consequently, peak intraoral pressure is almost the same as subglottal pressure for voiceless stops(Netsell,1969; Loefqvist, 1975). Whereas in the TE speakers examined, Po measurements were always lower than Ps measurements not only for the voiced consonants but for the voiceless consonants with voicelessness. This means that ,in TE speech, the pseudoglottal impedance is not negligible even when the pseudoglottis is not vibrating for production of voiceless consonants.

2. Mechanism of voicing distinction in TE speech

Perceptual studies have often revealed that skilled esophageal and TE speakers produce consonants with clear voicing distinction (Hirose et al, 1983; Nishizawa et al, 1992; Nichols, 1976). Hirose et al.(1983) suggested that, for syllable medial position, the presence or absence of voicing during the closure period is the most significant acoustic cue for voicing distinction in these substitute voices as in normal laryngeal speech. In the present study, the subjects showed to be able to produce intervocalic stops with effective voicing distinction as has been reported. The utterance samples in the present study were chosen following the criteria that perceptual judgment of voicing agrees with the acoustic judgment of presence or absence of voicing at the time of consonant release. It follows that , in the subjects examined, cessation of vibration is effectively achieved in accurately devoiced voiceless consonants.

The physiological mechanism of achieving "voicelessness" is not yet clear for esophageal and TE speakers. Hirose et al.(1983) and Nishizawa et al.(1992) reported through fiberoptic observations that cessation of the vibration is substantiated by the transient opening of the pseudoglottis in some esophageal and TE speakers. Some tension control mechanism has been supposed to operate around the region of the pseudoglottis even though the opening can not be elicited by the normal mechanism of "abduction" . The results of the present study are compatible with the supposition that there might be some active control of the pseudoglottis. According to this supposition, the higher intraoral pressure of voiceless consonants with voicelessness is interpreted to be the "result" of decreased pseudoglottal impedance through transient opening and cessation of vibration. Positive relationships between Po and Ps seen in consonant with voicelessness produced by subjects KM and KT are also explainable by the supposition of pseudoglottal opening.

On the other hand, Isshiki et al. (1972) and Kobayashi (1987) showed that, in laryngectomized speakers using electrolarynges, higher intraoral pressure for voiceless consonant is achieved. In these experiments using electrolarynges, there was no upcoming airflow. Therefore the intraoral pressure was thought to be controlled by suprapseudoglottal articulatory mechanism without pseudoglottal resistance to the airflow. It seems reasonable to assume that suprapseudoglottal adjustment can be active for intraoral pressure control also in TE speakers. In the literatures mentioned, authors have suggested that the variation of intraoral pressure is associated to some acoustic cues for voicing distinction other than on-off control of the voice source, for example, burst intensity. In the present study, however, it was showed that intraoral pressure exerts influences on the subject's pseudoglottal vibration directly. Thus, if suprapseudoglottal adjustment is active for intraoral pressure control in TE speakers, this adjustment might be active for achieving "voicelessness" as well as for higher burst intensity.

The aerodynamic data in the present study are insufficient to elucidate the mechanism of achieving voicelessness in TE speech. The results are compatible at the same time with the suppositions that there might be active control of the pseudoglottis, of the vocal tract,or of both. However , if the TE speakers can achieve the control of intraoral pressure by active adjustments of the vocal tract through rehabilitation procedures, these adjustments are expected to be effective for on-off control of pseudoglottal vibration in voicing distinction.

Acknowledgment

This work was supported in part by the Grtant-in-Aid for Cancer Research (4-17) from the Ministry of Health and Welfare. We thank Professor Minoru Ohshima in Otaru University of Commerce for his critical review of this paper.

References

Arkebauer, H., Hixon, T. & Hardy, J. (1967). Peak intraoral air pressure during speech. *J. Speech Hear. Res.*, 10,196-208.

Bell-Berti, F. (1975). Control of pharyngeal cavity size for voiced and voiceless stops. *J. Acoust. Soc. Am.*, 57,456-461.

Brown, W. S. & McGlone, R. (1969). Relation of intraoral air pressure to oral cavity size. *Folia Phoniatr.*, 21,321-331.

Hirose, H., Sawashima, M. & Yoshioka, H. (1983). Voicing distinction in esophageal speech -Perceptual, fiberoptic and acoustic studies-. *Ann. Bull. Res. Inst. Logoped. Phoniatr. Univ. Tokyo,* 17,187-199.

Isshiki,N. & Tanabe, M. (1972). Acoustic and aerodynamic study of a superior electrolarynx speaker. *Folia Phoniatr.*, 24,65-76.

Kent, R. & Moll, K. (1969). Vocal tract characteristics of the stop cognates. *J. Acoust. Soc. Am.*, 46,1549-1555.

Kinishi, M. & Amatsu, M. (1986). Aerodynamic studies of laryngectomee after the Amatsu tracheoesophageal shunt operation. *Ann. Otol. Rhinol. Laryngol.*, 181-184.

Klich, R. J.(1982). Effect of speech level and vowel context on intraoral air pressure in vocal and whispered speech. *Pholia Phoniatr.*, 34,33-40.

Kobayashi, N.(1987). Intraoral air pressure variations for consonantal voicing distinction in speech without larynges. *Ann. Bull. Res. Inst. Logoped. Phoniatr. Univ. Tokyo,* 21,193-201.

Komiyama, S., Watanabe, H., Ryu, S. & Hiroto, I. (1976). Study of a pseudoglottis in laryngectomized patients. *Otologia (Fukuoka),* 22,435-441.

Lisker, L. (1970). Supraglottal air pressure in the production of English stops. *Lang. Speech,* 13,215-230.

Loefqvist, A. (1975). A study of subglottal air pressure during the production of Swedish stops. *J.Phonet,* 3,175-189.

Lubker, J. F. & Parris, P. J. (1970). Simultaneous measurements of intraoral pressure, force of labial contact and labial electromyographic activity during production of the stop consonants cognates /p/ and /b/. *J. Acoust. Soc. Am.,* 47,625-633.

Malecot, A. (1966). The effectiveness of intraoral air-pressure-pulse parameters in distinguishing between stop cognates. *Phonetica,* 14,65-81.

Netsell, R. (1969). Subglottal and intraoral air pressure during the intervocalic contrast of /t/ and /d/. *Phonetica,* 20,68-73.

Nichols, A. C. (1976). Confusions in recognizing phonemes spoken by esophageal speakers : I. Initial consonants and clusters. *J. Communicat. Dis..*,9,27-41.

Nishizawa, N. , Tanaka, K., Sakai, N., Maguchi, S., Mikuni, H., Ogasawara, M., Kobashi, M. & Mesuda, Y (1992). Adjustments in tracheoesophageal speech production. *Japan J. Logoped. phoniatr,* 33,248-255.

Nishizawa, N., Tanaka, K., Sakai, N., Inuyama, Y., Takahashi, M. & Kato, M. (1993). Esophageal speech and TE shunt speech -A comparison using the same subjects. *J Otolaryng. Japan,* 7 (in press).

Shipp, T. (1970). EMG of pharyngoesophageal musculature during alaryngeal voice production. *J. Speech Hear. Res.*, 13,184-192.

Stevens, K.N. (1991). Vocal fold vibration for obstruent consonants. In J. Gauffin & B. Hammarberg (Eds), *Phonatory mechanisms: Physiology, acoustics and assessment* (pp. 29-36). San Diego: Singular Pub. Group, Inc..

Subtelny, J., Worth, J. & Sakuda, M. (1966). Intraoral pressure and rate of flow during speech. *J. Speech Hear. Res.*,9,498-518.

Tanaka,K., Suenaga, T., Ohashi, M., Yoshimura, S.& Satoh, N. (1987). The evaluation of TE shunt operation in patients with high-dose irradiation. *J. japan. Bronchoesophagol. Soc.* 38,361-365.

Westbury, J. R.(1983). Enlargement of the subglottal cavity and its relation to stop consonant voicing. *J. Acoust. Soc. Am.*, 73,1332-1336.