

## **INFORMATION TO USERS**

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the original text directly from the copy submitted. Thus, some dissertation copies are in typewriter face, while others may be from a computer printer.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyrighted material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is available as one exposure on a standard 35 mm slide or as a 17" x 23" black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. 35 mm slides or 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.



300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA



**Order Number 8820874**

**Aerodynamic characteristics of the voicing distinction in electric  
larynx speech**

**Kobayashi, Noriko, Ph.D.**

**City University of New York, 1988**

**Copyright ©1988 by Kobayashi, Noriko. All rights reserved.**

**U·M·I**  
300 N. Zeeb Rd.  
Ann Arbor, MI 48106



**PLEASE NOTE:**

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark .

1. Glossy photographs or pages \_\_\_\_\_
2. Colored illustrations, paper or print \_\_\_\_\_
3. Photographs with dark background \_\_\_\_\_
4. Illustrations are poor copy \_\_\_\_\_
5. Pages with black marks, not original copy
6. Print shows through as there is text on both sides of page \_\_\_\_\_
7. Indistinct, broken or small print on several pages
8. Print exceeds margin requirements \_\_\_\_\_
9. Tightly bound copy with print lost in spine \_\_\_\_\_
10. Computer printout pages with indistinct print \_\_\_\_\_
11. Page(s) \_\_\_\_\_ lacking when material received, and not available from school or author.
12. Page(s) \_\_\_\_\_ seem to be missing in numbering only as text follows.
13. Two pages numbered \_\_\_\_\_. Text follows.
14. Curling and wrinkled pages \_\_\_\_\_
15. Dissertation contains pages with print at a slant, filmed as received
16. Other \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

**U·M·I**



**AERODYNAMIC CHARACTERISTICS OF THE VOICING DISTINCTION**

**IN ELECTRIC LARYNX SPEECH**

by

**NORIKO KOBAYASHI**

A dissertation submitted to the Graduate Faculty in  
Speech and Hearing Sciences in partial fulfillment of  
the requirements for the degree of Doctor of Philosophy,  
The City University of New York.

1988

**COPYRIGHT BY**  
**NORIKO KOBAYASHI**  
**1988**

This manuscript has been read and accepted for the Graduate Faculty in Speech and Hearing Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Jan. 25, 1988  
Date

Katharina S. Harris  
Chair of Examining Committee

Jan. 25, 1988  
Date

Jimmy Horbhep  
Executive Officer

Thomas Baer, Ph.D.  
\_\_\_\_\_

Lawrence J. Raphael, Ph.D.  
\_\_\_\_\_

\_\_\_\_\_  
Supervisory Committee

The City University of New York

## ABSTRACT

AERODYNAMIC CHARACTERISTICS OF THE VOICING DISTINCTION  
IN ELECTRIC LARYNX SPEECH

by

Noriko Kobayashi

Advisor: Professor Katherine S. Harris

Two major hypotheses were tested in this study. One was to test whether intraoral air pressure is controlled by supralaryngeal maneuvers in electric larynx talkers, for whom respiratory and laryngeal controls are not available. The other was to test whether electric larynx talkers are able to produce perceptually differentiable voiced and voiceless stops in intervocalic position.

Three alaryngeal talkers who use an electric larynx of the transcervical type served as subjects. These talkers have good speech intelligibility with the electric larynges. The speech material consisted of /əCVC/ structure with /p/, /b/, /t/, /d/, /k/, and /g/ for consonants, and /ɔ/ and /I/ for vowels. Simultaneous recordings were obtained for intraoral air pressure, electromyography, i.e., signals from the genioglossus muscles and the middle constrictor muscle (for one subject only), movement of the jaw and the lower

lip, and acoustic signals, for twenty repetitions of the twelve utterance types.

For all three subjects, voiceless stops were associated with significantly higher intraoral air pressure than their voiced cognates. Data on EMG and articulatory movements suggested that oropharyngeal cavity was enlarged for voiced stops compared to voiceless stops. These differences in oropharyngeal cavity size seemed to be associated with higher intraoral air pressure for voiceless stops than for voiced stops. Perceptual results showed that subjects could produce grossly differentiable intervocalic voiced and voiceless stops, although only one third of the voiceless stops were perceived as voiceless. Acoustic data showed longer consonantal closure duration for voiceless stops than for voiced stops, but the difference is not as large as in normals. Relative differences in peak intraoral air pressure was strongly associated with perception of the voicing distinction.

Thus, alaryngeal talkers who speak with an electric larynx controlled intraoral air pressure by active manipulation of the supraglottal system. These variations in intraoral air pressure apparently provide useful perceptual cues for the consonant voicing distinction.

## ACKNOWLEDGEMENTS\*

I wish to express my deep thanks to those who enabled me to continue and complete this dissertation.

I am genuinely grateful to my advisor, Katherine S. Harris, without whose understanding, support, and encouragement, this project could not have been completed. I feel myself very fortunate not only because I learned so many things from her, but also because we have developed a wonderful friendship in the process of completing the work.

I would like to thank the members of my advisory committee. I thank Thomas Baer for his very valuable knowledge and understanding of the functioning of the larynx. I also thank Lawrence J. Raphael for his expertise in acoustic phonetics. He was always practical and humorous, making the task a pleasure.

There are many other people to whom I am grateful:

Hiroshi Muta, for his preparation of the subjects for the experiment and for his help with figures.

Judith A. Rubin-Spitz and Sara Basson, for helping to use the Biomedical Data Processing System.

Jan Edwards, for typing and editing the early drafts.

Jo Estill, for her invaluable last minute help in typing the final edition of the dissertation and for her

friendship and insight into cultural differences that were often hard to comprehend.

Many other friends provided emotional and spiritual strength when things were not so easy.

Finally, I would like to thank my parents and my late grandmothers, whose love and belief in me have supported me over many years, especially when I was experiencing difficulties in a foreign country.

\*This research was supported in part by  
NINCDS Grant NS-13870 to Haskins Laboratories.

## TABLE OF CONTENTS

ABSTRACT. . . . .	iv
ACKNOWLEDGEMENTS. . . . .	vi
TABLE OF CONTENTS . . . . .	viii
LIST OF TABLES. . . . .	.xii
LIST OF FIGURES . . . . .	.xiv
Chapter 1 INTRODUCTION . . . . .	1
1.1 Purpose. . . . .	1
1.2 Acoustics of the consonantal distinction for stop consonants in normal speech . . . . .	2
1.2.1 Voice onset time . . . . .	3
1.2.2 Vowel duration . . . . .	4
1.2.3 Consonant closure duration . . . . .	4
1.2.4 Burst intensity. . . . .	5
1.3 Speech mechanisms of the consonantal voicing distinction in normal speech . . . . .	6
1.3.1 Behavior of the vocal folds for the voicing distinction. . . . .	6
1.3.2 Intraoral air pressure variations for the voicing distinction. . . . .	8
1.3.2.1 Glottal resistance . . . . .	9
1.3.2.2 Supraglottal cavity size . . . . .	10

1.3.2.3	Supraglottal articulatory activity . . . . .	12
1.3.2.4	Incomplete velopharyngeal closure. . . . .	13
1.4	Alaryngeal speech. . . . .	14
1.4.1	Laryngectomy . . . . .	14
1.4.2	Alaryngeal speech. . . . .	16
1.4.2.1	Esophageal speech. . . . .	18
1.4.2.2	Artificial larynx speech . . . . .	20
1.4.3	Perception of alaryngeal speech. . . . .	22
1.4.4	Acoustic characteristics of alaryngeal speech.	23
1.4.4.1	Voice onset time . . . . .	23
1.4.4.2	Consonantal closure. . . . .	25
1.4.4.3	Vowel duration . . . . .	26
1.4.4.4	Burst intensity. . . . .	27
1.4.5	Speech mechanism underlying the production of the voicing distinction . . . . .	27
1.4.5.1	Speech mechanisms for the voice onset time production . . . . .	28
1.4.5.2	Speech mechanisms underlying the production of durational difference . . . . .	29
1.4.5.3	Intraoral air pressure . . . . .	30
Chapter 2	METHODS. . . . .	33
2.1	Subjects . . . . .	33
2.2	Speech materials . . . . .	35
2.3	Electric larynx. . . . .	38
2.4	Data acquisition . . . . .	38
2.4.1	Intraoral air pressure . . . . .	39

2.4.2	Electromyography (EMG) . . . . .	40
2.4.3	Articulatory movement. . . . .	42
2.4.4	Acoustic recordings. . . . .	42
2.4.5	Perceptual experiment. . . . .	43
2.5	Data processing. . . . .	44
2.5.1	Data segmentation. . . . .	45
2.5.2	Smoothing and averaging. . . . .	45
2.6	Data analysis. . . . .	47
2.6.1	Measurements of intraoral air pressure . . . . .	47
2.6.2	Measurements of EMG. . . . .	47
2.6.3	Measurements of articulatory movements . . . . .	49
2.6.4	Acoustic measurements. . . . .	52
2.6.5	Perceptual analysis. . . . .	52
2.6.6	Statistical analyses . . . . .	54
Chapter 3	RESULTS. . . . .	55
3.1	Intraoral air pressure . . . . .	55
3.2	Electromyographic analysis . . . . .	71
3.3	Movement . . . . .	79
3.4	Acoustic results . . . . .	97
3.4.1	Consonant closure duration . . . . .	97
3.4.2	Vowel duration . . . . .	99
3.5	Perception . . . . .	101
3.5.1	Intelligibility. . . . .	101
3.5.2	Relationships between speech perception and production . . . . .	105

**Chapter 4 DISCUSSION . . . . .110**  
**4.1 Intraoral air pressure . . . . .110**  
**4.2 Electromyography . . . . .119**  
**4.3 Articulator movement . . . . .123**  
**4.4 Acoustic results . . . . .127**

**REFERENCES. . . . .134**

## LIST OF TABLES

TABLE	
2.1	Utterance types . . . . . 36
2.2	The word forms of the utterance types presented to the subjects to read . . . . . 37
3.1	Analysis of Variance: effect of intraoral air pressure (cm-H <sub>2</sub> O) . . . . . 64
3.2	Analysis of Variance: effect of anterior genioglossus muscle amplitude (microV). . . . . 77
3.3	Analysis of Variance: Effect of posterior genioglossus muscle amplitude (microV). . . . . 78
3.4	Analysis of variance: effect of middle pharyngeal constrictor muscle amplitude (microV) for DB. . . 80
3.5	Analysis of variance: effect of jaw vertical displacement (cm) . . . . . 86
3.6	Analysis of variance: effect of lower lip vertical displacement (cm). . . . . 91
3.7	Analysis of variance: effect of lower lip horizontal displacement (cm). . . . . 96
3.8	Analysis of variance: effect of consonantal closure duration (ms) . . . . . 98
3.9	Analysis of variance: effect of vowel duration (ms) . . . . . 100
3.10	Intelligibility scores for three alaryngeal talkers in percent correct. . . . . 102
3.11	Overall probability of a "voiced" response, in percent, for each speaker. . . . . 103
3.12	Confusion matrix for FB . . . . . 104
3.13	Confusion matrix for AS . . . . . 106
3.14	Confusion matrix for DB . . . . . 107
3.15	F-values for the significant differences between the physiological and acoustic values of the voicing contrasts, based on tokens sorted into three groups. . . . . 109

## LIST OF FIGURES

FIGURE	
1.1	Anatomical structure of the speech organs before and after laryngectomy . . . . . 17
2.1	Lineup point . . . . . 46
2.2	Sample waveforms of intraoral air pressure and the corresponding acoustic signals . . . . . 48
2.3	Examples of smoothed EMG signals for the anterior genioglossus muscle . . . . . 50
2.4	Waveforms of vertical jaw movement . . . . . 51
2.5	An example of durational measurement in the acoustic waveform. . . . . 53
3.1	Averaged waveforms of intraoral air pressure for FB . . . . . 56
3.2	Averaged waveforms of intraoral air pressure for AS . . . . . 58
3.3	Averaged waveforms of intraoral air pressure for DB . . . . . 59
3.4	Distribution of peak intraoral air pressure values for FB. . . . . 61
3.5	Distribution of peak intraoral air pressure values for AS. . . . . 62
3.6	Distribution of peak intraoral air pressure values for DB. . . . . 63
3.7	Mean intraoral air pressure values for voiced and voiceless stops as a function of vowel context for FB . . . . . 66
3.8	Mean intraoral air pressure values for voiced and voiceless stops as a function of place of articulation for FB . . . . . 67

3.9	Mean intraoral air pressure values for voiced and voiceless stops as a function of vowel context for AS . . . . .	69
3.10	Mean intraoral air pressure values for voiced and voiceless stops as a function of place of articulation for AS . . . . .	70
3.11	Averaged waveforms of the activity of the anterior genioglossus muscle for FB. . . . .	72
3.12	Averaged waveform of the activity of the anterior genioglossus muscle for AS. . . . .	73
3.13	Averaged waveforms of the activity of the posterior genioglossus muscle for FB . . . . .	74
3.14	Averaged waveforms of the activity of the posterior genioglossus muscle for AS . . . . .	75
3.15	Averaged waveforms of vertical jaw displacement for FB. . . . .	82
3.16	Averaged waveforms of vertical jaw displacement for AS. . . . .	83
3.17	Averaged waveforms of vertical jaw displacement for DB. . . . .	84
3.18	Averaged waveforms of vertical displacement of the lower lip for FB . . . . .	87
3.19	Averaged waveforms of vertical displacement of the lower lip for AS . . . . .	88
3.20	Averaged waveforms of vertical displacement of the lower lip for DB . . . . .	90
3.21	Averaged waveforms of horizontal displacement of the lower lip for FB . . . . .	92
3.22	Averaged waveforms of horizontal displacement of the lower lip for AS . . . . .	93
3.23	Averaged waveforms of horizontal displacement of the lower lip for DB . . . . .	95
4.1	Individual waveforms of intraoral air pressure for AS and FB. . . . .	115

CHAPTER 1  
INTRODUCTION

1.1 Purpose

The purposes of this study were twofold: First this study was designed to test the hypothesis that intraoral air pressure is controlled by supralaryngeal maneuvers. Second, it was designed to test the hypothesis that alaryngeal talkers, who are not able to manipulate transglottal air pressure or flow, nonetheless produce perceptually differentiable intervocalic voiced and voiceless stop consonants, using the mechanisms available to them.

These two purposes are linked by the investigation of the production of voicing contrasts in alaryngeal talkers who use electronic larynges. If such talkers do, in fact, consistently produce supraglottal air pressure contrasts, these maneuvers can be presumed to have some value in producing a perceptible voicing contrast. By a detailed examination of the relationship between intraoral air

pressure and the perceptual adequacy of the resulting stop consonants, we may be able to gain some insights into the importance of the acoustic consequences of the pressure differences.

Turning to the alaryngeal talkers themselves, it is important to investigate the mechanisms available to them in producing a contrast between "voiced" and "voiceless" stop consonants, although, of course, the normal means for producing such contrasts are not available.

## 1.2 Acoustics of the Consonantal Voicing Distinction for Stop Consonants in Normal Speech

In this section, acoustic features associated with the perception and production of the voicing distinction in normal speech will be discussed. It has been known that the simple difference in phonation (i.e., absence and presence of voicing) is an acoustic characteristic and a perceptual cue to stop voicing, especially in intervocalic and other medial positions and contexts. Much research has focussed on the acoustic and perceptual cues for the consonantal voicing distinction.

### 1.2.1 Voice onset time

In initial prestressed position in English, the timing of vocal fold vibration relative to the stop release is longer for voiceless stops than for voiced stops (Lisker and Abramson, 1967). Using synthetic speech, Liberman, Delattre, and Cooper (1958) and Lisker and Abramson (1970) reported that a stop consonant was identified as voiceless or voiced by the time from the release of the closure to the onset of voicing (VOT). If VOT is longer than about 25-40 msec (depending upon the place of articulation of the consonant), the consonant is identified as voiced. However, researchers have disagreed about the adequacy of VOT as providing a sufficient cue for the consonantal voicing distinction.

Klatt (1975) suggested that the voicing distinction might not be perceived based solely on VOT and he suggested some other acoustic cues for the distinction. Winitz, LaRiviere and Herriman (1975) found that stops were identified as voiceless when only the burst and the first 10, 20, or 30 msec of aspiration were present. They postulated that aspiration was the primary cue and VOT was a relatively secondary cue for the voicing distinction. When the aspiration noise was present, the consonant was perceived as voiceless, while it was perceived as voiced when the aspiration noise was absent. Stevens and Klatt (1974) postulated that the starting frequency of F1 also provides an important perceptual cue for voicing, in that

there is a rapid formant transition after the onset of voicing for the production of voiced consonants. The formant transition is completed prior to the voicing for the production of voiceless consonants. However, Lisker (1978) claimed that the presence or absence of voicing is neither necessary nor sufficient for the voicing distinction. He pointed out that this hypothesis (a F1 transition detector) is not more reliable than the absolute duration of the time interval between burst and onset of voicing and F1 for the voicing distinction, although it is probably that no single acoustic parameter would account for the identification of the voicing distinction for initial stop consonants.

#### 1.2.2 Vowel duration

It has been known for some time that vowel duration is longer before voiced consonants than before voiceless consonants (Peterson and Lehiste (1961). Denes (1955), House (1961), and Raphael (1972). have reported that durational differences are perceptually important. For postconsonantal positions Peterson and Lehiste (1960) did not find significant differences for the voicing contrast.

#### 1.2.3 Consonant closure duration

Consonant closure duration has been measured between the offset of the periodic acoustic pulses of the preceding

vowel and the onset of the burst release of the consonant. Lisker (1978) found that a 70-100 msec closure duration was required by the listeners to identify "rapid" rather than "rabid." In a study by Zue and Lafferiere (1979), 129 msec consonant closure duration for /t/ and 75 msec for /d/ were reported in pre-stress contexts. Lisker (1978) pointed out that the absolute duration was important only in the absence of closure voicing. Using French pre-stressed VCV contexts, Bijjter and Serniclaes (1974), Serniclaes (1974), and Wajskop and Sweerts (1973) showed that consonant closure duration as well as VOT was an important cue for the voicing distinction.

#### 1.2.4 Burst intensity

Zue and Lafferiere (1979) reported, in a study with English speaking male talkers, that there is a tendency of voiced stops to have a weaker burst intensity than voiceless stops. Repp's experiment (1979) with synthetic speech showed similar results. He suggested that, "It is entirely possible that burst intensity constitutes an important cue when aspiration is absent ...." The number of studies on burst intensity is still small and further investigation would be fruitful for discovering if there are additional as yet unrecognized acoustic cues and for understanding the relative importance of the cues already known.

### 1.3 Speech Mechanisms of the Consonantal Voicing Distinction in Normal Speech

The physiology of two major mechanisms of the voicing distinction, vocal fold behavior and intraoral air pressure variation in normal speech, will be discussed in this section.

#### 1.3.1 Behavior of the vocal folds for the voicing distinction

The usual observation about vocal fold behaviors for the production of voiced stops, at least in intervocalic positions, is that vocal fold vibration continues for some time during the articulatory closure. For voiceless stops, on the other hand, the vocal folds are not vibrating during closure. Pickett (1980) notes that the vocal folds are open during the closed phase of the voiceless stops, while the vocal folds remain closed to produce voicing during the closed phase of the voiced stops. The open glottis for voiceless stops presents little or no resistance to pulmonic air flow into the mouth, allowing intraoral air pressure to increase until it is equal to subglottal air pressure. At the end of the closure for the voiceless stops, when the articulatory closure is released, there is a release of intraoral air pressure. An intense transient sound is produced followed by turbulent flow (i.e., friction and

aspiration). The transient and turbulent noises are called "the burst" of the stop consonants. The vocal folds then move into voicing position. The period of closure is shorter for voiced stops than for voiceless stops. At the release of the closure, voiced stops are accompanied by a weak transient.

The muscle activity underlying the different vocal fold behaviors for the voicing contrast has been studied by Hirose and Gay (1972). They found that the posterior cricoarytenoid muscle (PCA), an abductor muscle for the vocal folds was active during closure for voiceless stops. The intra-arytenoid muscle (IA), an adductor muscle, was active during voiced stops. The activities of the PCA and the IA were found to be reciprocal, in that when the PCA was active for the production of voiceless stops, the IA was inactive. Flanagan, Rabiner, Christopher, Boch, and Shipp (1976) observed the activities of the PCA and the IA with glottal openings simultaneously recorded using photoglottography (PGG). They found the PCA started to be active 20 to 30 msec before glottal opening and that its activity continued until IA started to be active. These studies indicated that the laryngeal muscles precisely control the vocal fold positioning (i.e., abduction and adduction) for the production of the voicing contrast.

### 1.3.2 Intraoral air pressure variations for the voicing distinction

In the production of consonants, variable amounts of air pressure in the mouth are generated when there is some articulatory resistance to the pulmonary air flow. After the earliest study by Rousselot (1897), much research has been conducted in an attempt to associate variation in intraoral pressure with different theories of phonology and physiology in consonant production (Black, 1950; Stetson, 1951; Subtelny, Worth, and Sakuda, 1966; Malecot, 1966a, 1966b, 1970; Arkebauer, Hixon, and Hardy, 1967; Soda, 1967; Brown and McGlone, 1969; Lisker, 1970; Lubker and Paris, 1970; Prosek and House, 1975; Murry and Brown, 1976; Muller and Brown, 1980). The results of these studies show general agreement in that it has been found that voiceless consonants have greater peak intraoral pressure values than their voiced cognates, although there are no differences in subglottal pressure (Netsell, 1969). However, Lisker (1970) pointed out that the values are highly variable depending on the factors of stress and consonant position.

Some mechanisms that could cause intraoral pressure differences between voiced and voiceless stop consonant pairs have been postulated based on physiological findings and theoretical or mechanical models of speech production. Differences in laryngeal resistance, oropharyngeal cavity size, degree of velopharyngeal closure, and respiratory effort are some of these mechanisms.

### 1.3.2.1 Glottal resistance

In voiced consonant production, the vocal folds are adducted, glottal resistance is increased and the transglottal air flow from the respiratory system into the oral cavity is reduced. The slower airflow rate, together with shorter consonant closure duration (Lisker, 1978; Zue and Lafferiere, 1979), results in less time for pressure to increase in the oral cavity. In voiceless consonants, there is more transglottal airflow with less glottal resistance due to the open glottis, and higher intraoral pressure is generated. Therefore, glottal resistance and intraoral pressure difference are related.

Support for this concept came from Rothenberg (1968) and Muller (1974) who reported that the dominant factor for the pressure difference is glottal resistance, based on the calculations of variables using mechanical models. Arkebauer, et al (1967) and Malecot (1966a, b) also support the notion with their physiological data. In those conditions where there is less glottal resistance contrast, as in whispering, Soda, et al, (1967), Lindqvist (1969), Malecot (1970), Schwartz (1971), Murry and Brown (1976), and Klitch (1982) reported less intraoral pressure difference between voiced and voiceless consonants. Moreover, the studies by Netsell (1969), McGlone and Shipp (1972), and Lofqvist (1975), who measured subglottal air pressure for voiced-voiceless consonants, showed similar subglottal and supraglottal pressure values in voiceless consonants, but in

voiced consonants, there was less pressure supraglottally than subglottally. These data support the theory that glottal resistance during the production of voiced consonants is one of the major factors causing the intraoral pressure difference between voiced-voiceless cognates.

#### 1.3.2.2 Supraglottal cavity size

A major reason why researchers expect a relationship between supraglottal cavity size and consonantal voicing is that the expansion of the supraglottal cavity during the articulatory closure period can explain the maintenance of vocal fold vibration in voiced stops. In stop consonant production, complete closure is formed at some point above the glottis and prevents air flow from passing through the mouth. As the pressure in the oral cavity is rapidly increased, the transglottal pressure gradient, which is necessary for the vocal folds to maintain vibration, will be decreased. Increased volume of supraglottal cavity in voiced stops is one of the possible mechanisms for sustaining vocal fold vibration by maintaining a sufficient transglottal pressure gradient.

Measurements have been made of the position of the larynx during voiced and voiceless stops using cinefluorography, cineradiography, and a thyroumbrometer by Perkell (1969), Kent and Moll (1969), Houde (1967), Lindqvist, et al, (1973), Ewan and Krones (1974), and

Westbury (1983). They reported that the larynx is higher for the voiceless stops than for their voiced counterparts. Perkell (1969) noted the higher velum height for voiced stops. Kent and Moll (1969) reported that the hyoid bone is drawn downward during voiced stop production, possibly expanding the pharyngeal cavity anteroposteriorly and vertically. Houde (1967) reported, by cineradiographic observation, that the tongue is lowered in the production of /b/, and Westbury (1983) reported the frontward movement of the tongue for voiced stops in cinefluorography. X-ray microbeam data by Kiritani, Ito, and Fujimura (1976) showed that tongue pellets were higher in position for /t/ and /k/ than for /d/ and /g/. Fujimura (1985) observed that the tongue pellets moved slightly downward immediately before the lip closure was released in /b/. Data from the dynamic palatograph have shown that the lingua-palatal contact area was greater for /t/ than for /d/ (Fujii, 1970; Fujimura, Tatsumi and Kagaya (1972). Bell-Berti (1975) reported two mechanisms relating to the expansion of pharyngeal cavity in voiced stops. Her EMG data suggest that there is an "active" mechanism in which levator palatini and sternohyoid muscles work to expand the cavity, and that there is also a "passive mechanism" in which the superior and middle constrictors and the faucal pillars suppress the constriction of the pharyngeal wall.

### 1.3.2.3 Supraglottal articulatory Activity

It has been posited that the feature underlying the difference between voiced and voiceless stops may not reside entirely in differences at the laryngeal level, though research by Lisker and Abramson (1964, 1971) and Halle and Stevens (1971) suggested that the primary difference is at the laryngeal level. In searching for an articulatory basis for the phonetic feature "tenseness" (thought to be associated with the voicing distinction) many physiological studies were conducted to examine the degree of activity of the relevant upper articulators in voiced and voiceless stops. These measurements do not directly show the cavity size differences, but some of them imply change of cavity size due to various articulatory maneuvers.

The activity of the orbicularis oris muscle during the production of labial stops was measured using an electromyographic technique (Harris, Lysaught and Schvey 1965; Fromkin, 1966; Tatham and Morton 1969; Lubker and Parris 1970). They did not find consistent differences in EMG peak amplitude between /p/ and /b/, although Tatham, et al, (1969) found small but significant differences in muscle activity at the point of release of /p/ and /b/. Mechanical pressure of lips was measured by Malecot (1966b) and Lubker, et al, (1970) and the results indicated that /p/ and /b/ had equivocal pressure values, although there was a slight tendency for /p/ to be associated with more pressure than /b/.

Measurements of mechanical pressure of the tongue were reported by Stetson (1928), McGlone, et al, (1967) and Brown (1973). The results generally agree that there is a tendency for /t/ to be associated with more lingual pressure than /d/, although statistically significant differences were not found. Thus, overall, there does not seem to be a substantial difference in supraglottal articulatory activity associated with "tenseness" per se, but there is a tendency in that direction.

#### 1.3.2.4 Incomplete velopharyngeal closure.

Malecot (1955), Isshiki and Ringel (1964), and Arkebauer, et al, (1967) have suggested that an open velopharyngeal port could explain the continuation of vocal fold vibration during voiced stops, because a transglottal pressure gradient could be maintained with an open passage to the nasal cavity. Yanagisawa and Hyde (1966), and Kent and Moll (1969) observed velopharyngeal opening during voiced stops, using cinefluorography. Rothenberg (1968) supported this view with his acoustic studies. However, the nasal airflow data by Lubker (1973) and Westbury (1979) showed insignificant nasal airflow for non-nasal stops and they suggested there might be palatal elevation during the voiced stops not resulting in the opening of the port.

Thus, the relationship between the consonantal voicing distinction and velopharyngeal closure is controversial.

Moreover, as mentioned before, studies of cavity size suggest that the velum is more elevated for voiced stops than for their voiceless counterparts. Individual variability (Bell-Berti, 1975) might explain the discrepancy among the studies, since manner of articulation and voice quality could differ among subjects while maintaining the voiced-voiceless contrast. It is, however, unlikely that the velopharyngeal port has a critical role in the production of consonantal voicing.

#### 1.4 Alaryngeal Speech

##### 1.4.1 Laryngectomy

Laryngeal cancer is the growth of malignant cells in the larynx. Different kinds of medical treatments are given to a patient with laryngeal cancer, depending of the stage, location, and physiology of the cancer. These treatments are radiation therapy, chemotherapy, and laryngectomy (partial or total). If the cancer is extensive, a total laryngectomy is performed, rather than a partial laryngectomy, which involves the removal of the part of the larynx where the cancer is located. A total laryngectomy involves the surgical removal of the entire body of the larynx, including the removal of the hyoid bone, the epiglottis and its inferior and superior attachments, the extrinsic strap muscles and the upper tracheal rings. If

cancer cells have metastasized into the cervical lymph nodes, a radical neck dissection is also performed. A radical neck dissection involves the removal of the cervical lymph nodes and the associated veins, the accessory nerve, and some of the muscles in the neck.

The surgical procedure of a total laryngectomy is preceded by a tracheostomy, which allows the patient to breathe and to inhale anesthesia through a tracheostoma during the surgery. A tracheostoma is constructed by creating a small opening on the neck at the level of the upper tracheal rings. After the tracheostomy, an incision is made vertically in the midline or in a T-shape on the skin of the neck. The incision extends from above the hyoid bone to the sternum. After elevating the platysma muscle, the sternohyoid, thyrohyoid, and omohyoid muscles are divided. The hyoid bone is separated from the suprahyoid muscles. The inferior constrictor muscles are dissected near their attachments to the thyroid cartilage. The trachea is then sectioned above the tracheostoma. The upper rings of the trachea and the cricoid cartilage are separated from the esophagus. Above the hyoid bone, the pharynx is opened and further incisions are continued until the larynx can be freed and removed.

The suturing procedure follows the removal of the larynx. First, the tracheal stump is sutured to the base of the neck. The pharyngeal wall is approximated, with the hypopharynx connected to the upper esophagus. The superior

portion of the pharynx is sutured to the base of the tongue. The suture is made in layers; sutures of the submucous tissue and sutures of the muscle tissue. In closing the pharyngeal defect, muscles of the inferior constrictor and cricopharyngeus are sutured.

A total laryngectomy creates some peculiar patterns of respiration, phonation, and, sometimes, of swallowing for the patient. Respiration is performed through the tracheostoma, rather than through the nasal, oral, and pharyngeal cavities, and the larynx (Fig. 1.1). Phonation is carried on either by vibrating the pharyngo-esophageal (P-E) segment, or by using an artificial larynx, which generates a sound source for speech. According to Van den Berg and Moolenaar-Bijl (1959) and Damste (1958), the cricopharyngeus or cricopharyngeal sphincter constitutes the P-E segment, which is also called the neoglottis. Sometimes, the patient exhibits difficulties in swallowing due to tight closure of the esophago-pharyngeal segment. This may also prevent the patient from acquiring esophageal speech. Singer and Blom (1981) proposed the dissection of the pharyngeal constrictor, which they claimed relieved swallowing and esophageal speech problems.

#### 1.4.2 Alaryngeal speech

A laryngectomized talker needs a new voice because normal voice production is impossible after the surgical

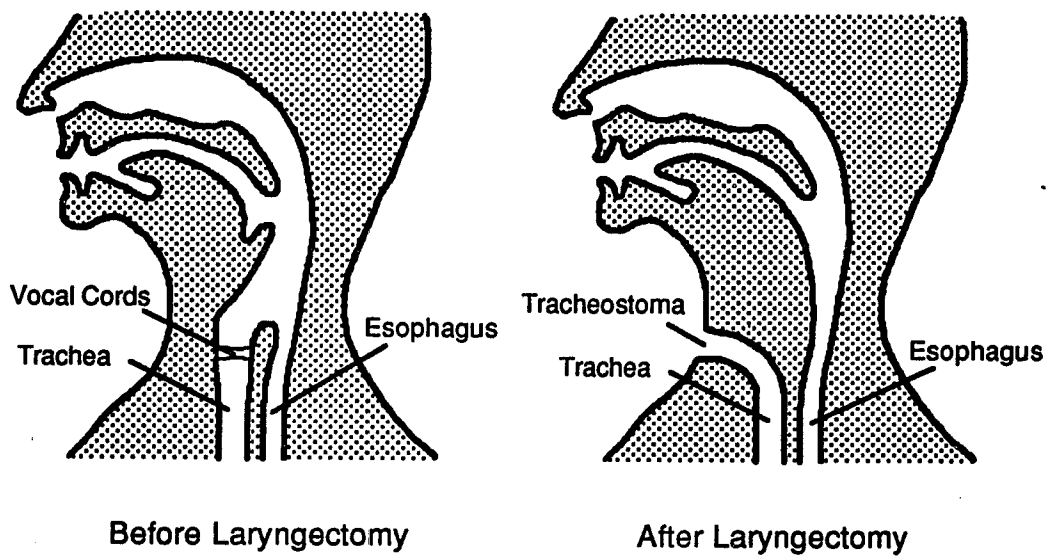


Figure 1.1 Anatomical structure of the speech organs before and after laryngectomy.

removal of the larynx, although articulatory gestures remain relatively intact. There are three means for voice production after a total laryngectomy: esophageal speech, artificial larynx speech, and surgical reconstruction of the voice. The selection of the means of voice production to be used is usually done by the laryngectomized patient, the speech pathologist, and the surgeon, based on the structural and functional characteristics of the peripheral speech mechanism after the surgery, the physical condition of the patient, his or her age, motivation, communicative needs, and other factors. In the following sections, two types of alaryngeal speech, esophageal speech and artificial larynx speech, will be described.

#### 1.4.2.1 Esophageal speech

In esophageal speech, air is taken into the esophagus and sound is generated by the vibration of the P-E segment on the release of the air. It should be noted that esophageal speech differs from normal belching in that the talker has acquired voluntary control of the initiation and duration of the esophageal sound. Three methods for air intake are used by esophageal talkers: inhalation, injection, and consonant injection. Esophageal talkers may choose one method or combine two methods, depending on what is easiest for them. In inhalation, the talker inhales sharply and quickly through the nose or mouth and negative

intrathoracic pressure is created around the outer surface of the esophagus. The difference in air pressure between the inside and the outside of the esophagus allows air to be taken down into the esophagus. In injection, the air within the oropharyngeal cavity is forced down into the esophagus by creating high oropharyngeal air pressure by the compression of the lips or the tongue against the alveolar ridge or the palate. In consonant injection, air is taken into the esophagus by creating high air pressure of the oropharyngeal cavity while producing consonants with articulatory closures or narrow constriction (i.e., obstruents). The air behind the articulatory constriction is then forced down into the esophagus, overcoming the muscle tension of the neoglottis and the air pressure in the esophagus. The major advantage of esophageal speech is that the talker uses his own organ for phonation without the requirement of any mechanical or prosthetic device. However, good anatomical structure and function of the oropharyngeal organs and esophagus is needed in order to acquire esophageal speech. If the neoglottis is too tight or too loose, esophageal speech is very difficult to produce. With a tight neoglottis, air is not easily taken into the esophagus. By contrast, with a loose neoglottis, it is difficult to excite the neoglottis into vibration when the air is released. Another disadvantage of esophageal speech is that it generally takes a relatively long time to learn: although there is much individual variation, it

usually requires over six months of intensive speech therapy. Financial, motivational, emotional, and other factors may make it difficult for a patient to maintain speech therapy for the needed amount of time. The success rate for the acquisition of esophageal speech has been reported to be as low as 30% (Gates, et al, 1982) and as high as 40 to 60% (Snidecor, 1975), depending on the criterion for success and the method of subject selection.

#### 1.4.2.2 Artificial larynx speech

In artificial larynx speech, sound generated by an external source is combined with the patient's supraglottal articulation to produce speech. The two types of artificial larynges are the pneumatic type and the electronic type. In the pneumatic type, vibration is produced by a reed or membrane in the device by pulmonary air from the stoma. The stoma is connected to the vibrator with a cuff and a tube. The sound generated at the vibrator is introduced to the oral cavity and this source is filtered by the shape of the upper vocal tract in the usual manner. In the electronic type of artificial larynx, the sound is generated by a battery-driven vibrator as a buzzing noise. There are two types of electronic larynges: an intraoral type and a transcervical type. In the intraoral electronic larynx, such as the Cooper-Rand artificial larynx, sound is generated in a small transducer and is transmitted through a

plastic tube to the talker's mouth. In the transcervical artificial larynx, a battery-powered vibrator is activated by a button switch, generating a continuous buzzing noise. The device is pressed against the neck and the sound which is generated by the talker's manual operation of the button is transmitted through the neck tissue into the oropharynx. Again, this sound source is filtered by the shape of the upper vocal tract in the usual manner. One of the advantages of artificial larynx speech, as compared to esophageal speech, is that it can be learned in a relatively short period of time. Almost anyone can communicate effectively with an artificial larynx almost immediately after its operation has been explained. Another advantage is that artificial larynx speech is reported to be more easily discriminated than esophageal speech in noise and over the telephone (Clark and Stample, 1982). The artificial larynx can be used by patients during the period when they are learning esophageal speech and can be used by those individuals who are unable to learn esophageal speech. The major disadvantage of artificial larynx speech is a negative attitude toward it on the part of many patients for various reasons. It is considered by some to be a sign of failure to learn esophageal speech (Lauder, 1968; Duguay, 1978; Goldstein, 1978); others dislike the mechanical sound, the need to use one hand to operate it, or the possibility of mechanical breakdown or battery failure.

### 1.4.3 Perception of alaryngeal speech

Reports from perceptual studies of esophageal speech have shown much individual variability with respect to overall intelligibility scores, presumably depending on the population of subjects (Anderson, 1950; DiCarlo, Amster, and Herer, 1955; McCroskey and Mulligan, 1963; Christensen, et al, 1978; Hirose, et al, 1983; Connor, et al, 1985). Scores vary between 40% and 70%. The largest number of errors were found in voicing features during stop consonant production in these studies, and general agreement that voicing errors for voiceless consonants were higher than those for voiced consonants.

In electronic larynx speech, researchers have reported overall intelligibility scores of 50%-70% (Barney, Haworth, and Dunn, 1959; McCroskey and Mulligan, 1963; Weiss, Yeni-Komshian, and Heinz (1979); and Weiss and Basili (1985). Although there are widespread individual differences, an excellent electronic larynx talker in the study of Isshiki and Tanabe (1972) exhibited better than 90% intelligibility. The predominance of voicing errors were reported for this talker, as well as for esophageal talkers. In electronic larynx talkers, voiceless consonants were frequently perceived as voiced consonants. Some voiced consonants were perceived as voiceless, although the substitutions occurred less frequently than those of voiced consonants for voiceless.

Differences in intelligibility between esophageal speech and electronic larynx speech were reported to be negligible (McCroskey and Mulligan, 1963; Kalb and Carpenter, 1981).

#### 1.4.4 Acoustic characteristics of alaryngeal speech

A small number of studies have been conducted on the acoustic features of alaryngeal speech. The investigations were carried out mainly for the purpose of explaining the perceptual characteristics of alaryngeal speech from the acoustic data, and also for the purpose of finding the relationships between the production characteristics of the alaryngeal speech and the acoustic consequences of the characteristics. Four major acoustic characteristics for the production of the voicing contrasts have been reported. They are voice onset time (VOT), consonantal closure duration, vowel duration, and burst intensity of consonantal closure release.

##### 1.4.4.1 Voice onset time

Studies for the voice onset time in esophageal speech during the production of initial stop consonants have been reported (Sacco, Mann, and Schultz 1968; Christensen, et al, 1978; Hirose, et al, 1983). The results show that esophageal talkers systematically differentiate the time

between the onset of the voice production by the neoglottis and the articulatory closure release for the voicing distinction. The values of VOT were significantly longer for the voiceless consonants than for their voiced cognates. Christensen, et al, (1978) also reported that voicing lag of the voiceless consonants was shorter in esophageal speech than in normal speech. Hirose, et al, (1983) reported that greater VOT values for voiceless stops than for voiced stops were important for the perception of voicing contrasts in word-initial positions. In word-medial positions, they reported that absence of voicing for voiceless stops, and presence of voicing for voiced stops had a significant effect on the perception of voicing contrasts.

For electronic larynx speech, Weiss, et al, (1979) found significant differences between the VOT values for the voicing contrast. Some of the values for the voiceless consonants, however, were in the range of the voiced consonants of normals. The authors suggested that the short VOT values of voiceless stops could cause the perceptual confusion of the voicing contrast in electronic larynx speech. There are reports which suggested some acoustic evidence for factors which would provide some perceptual cues for the voicing distinction, although these studies do not present substantial data. For example, Rothman (1978) reported that the electronic larynx talker with good intelligibility regulated acoustic intensity during the consonant production by attenuating the device's vibration

for the voiceless consonant. The maneuver can be considered as providing relevant or pseudo-VOT for the production of the voiceless consonants. Isshiki and Tanabe (1972) reported that their electronic larynx talker differentiated initial /p/ and /b/ by the durational differences between the initiation of the electronic larynx and the release of the mouth closure for the stops.

#### 1.4.4.2 Consonantal closure

Studies about differences in consonantal closure duration for the production of the voicing distinction in alaryngeal speech are scarce, as are such studies of normal speakers. For esophageal speech, Connor, et al, (1985) reported that longer phonation off-times (i.e., consonantal closure) were found for /t/ than for /d/ in talkers with high intelligibility, although talkers with low intelligibility showed no differences. The durational values for /t/ and /d/ in esophageal talkers were within the range of the values for /t/ found by Lisker (1978) for normal talkers. For electronic larynx speech, similarly, Rothman (1978) found that consonant closure duration was significantly different between /t/ and /d/ in talkers with high intelligibility.

#### 1.4.4.3 Vowel duration

For esophageal speech, studies by Christensen and Weinberg (1976) and Gandour, Weinberg, and Rutkowski (1980) showed that vowel durations before voiced consonants were longer than those before voiceless consonants in high intelligibility speech. Connor, et al, (1985) reported no significant durational difference between vowels before voiced consonants and vowels before voiceless consonants. They suggested that their results were different from the other studies because their talkers had low intelligibility. In a study of post-consonantal vowel duration, Swisher (1980) reported longer duration for vowels following voiceless consonants than for those following voiced consonants in esophageal speech, although no comparable differences were found for normal talkers. She suggested that the increased phonation length as a therapy technique caused the longer vowel duration, reducing the perceptual salience for the production of voiceless consonants.

For electronic larynx speech, Weiss, et al, (1979) reported the importance of vowel duration differences for the perception of the voicing distinction. They found that the durations were longer for vowels preceding voiced stops than for those preceding voiceless stops in word-final context.

#### 1.4.4.4 Burst intensity

The important role of burst intensity on the release of consonantal closure for the production of the voicing distinction has been studied for electronic larynx speech. Isshiki and Tanabe (1972) found a greater burst intensity for voiceless consonants than for voiced consonants in an electronic larynx talker who was able to produce the consonantal voicing distinction. They suggested that the increased intensity or sharpness of the voiceless consonants overlaid on the duty-cycle of the electronic larynx distinguished them from the voiced consonants. Weiss, et al, (1979) also reported greater burst intensity for /k/ than for /g/.

#### 1.4.5 Speech mechanisms underlying the production of the voicing distinction

The speech production mechanisms associated with the voicing distinction in alaryngeal speech have been studied. Among the acoustic characteristics for the voicing contrasts, differentiation of VOT values and durational characteristics of the vowel were related to the physiological findings. Studies of intraoral air pressure also have provided physiological evidence underlying the production of the voicing distinction, although the acoustic consequences are still unclear.

#### 1.4.5.1 Speech mechanisms for voice onset time production

Sacco, et al, (1968) reported that esophageal talkers have problems in initiating and terminating voicing and as an acoustic consequence of the problem, they have peculiar patterns of voice onset time differences for voicing contrasts. Hirose, et al, (1983) observed the opening gestures of the neoglottis in esophageal talkers, using a fiberoptic technique. They observed a transient opening of the neoglottis for the production of voiceless consonants. They also found that the time course of the opening gesture varied considerably even for the same utterance type. These findings seem to provide the physiological basis underlying the short or absent voicing for the production of voiceless consonants. Simultaneous recordings of the EMG signals from surface electrodes on the skin of the anterior neck slightly above the level of the neoglottis, in their study, revealed a transient increase in activity immediately before the stop closure for /t/ but none for /d/. These results could indicate that the opening gesture of the neoglottis was due to muscle activity in the neoglottis region, although the precise muscular mechanism remains unknown. Observations using x-ray cinematography by Sawada (1959) were cited in the study of Hirose, et al, (1983). He found that the base of the tongue was elevated postero-superiorly during the production of voiceless consonants in esophageal speech. These movements of the tongue and the hyoid bone were considered to help stop the

vibration of the neoglottis causing the hypopharyngeal lumen to close.

For electronic larynx speech, cessation of voicing is controlled by manually deactivating the vibration of the device. Rothman and Goldstein (1976) reported a case of an electrolarynx talker with excellent intelligibility. They observed that the good talker manually stopped the vibration of the device for the aspiration period before the following vowel for the production of the voiceless consonants. They also reported that the talker manually supplied a slight pressure of the vibrating disc of the electronic larynx for about 20 msec for the intervocalic /t/, attenuating the vibratory duty-cycle of the device. This technique seemed to have produced a slightly audible aspiration noise following consonant release.

#### 1.4.5.2 Speech mechanisms underlying the production of durational difference

As an explanation of the longer vowel duration in esophageal speech, Christensen and Weinberg (1976) claimed that the phonatory mechanisms of esophageal talkers are not capable of controlling adduction-abduction of the neoglottis. They pointed out that the neoglottis was continuously adducted and, therefore, the esophageal talkers demonstrated differences in vowel duration "that reflect alterations in the time-dependent relationships necessary to

effect appropriate trans-pseudoglottal pressure associated with voicing onset and termination." (PP.687) No major study has reported on the physiological basis for vowel duration differences in the production of the voicing distinction in electronic larynx speech.

#### 1.4.5.3 Intraoral air pressure

In early research of DiCarlo, et al, (1955) and Damste (1958), intraoral air pressure in esophageal talkers was found to be greater for /p/ than for /b/. Murry and Brown (1975) reported that peak intraoral air pressure was greater for voiceless consonants than for their voiced cognates in esophageal speech. Their data, however, revealed considerable variability among the talkers and between contexts (i.e., word-initial vs word-medial). In fact, two out of the seven talkers exhibited greater pressure values for /b/ than for /p/. It was suggested that different maneuvers for air taking have effects on individual variability. They also investigated the relationship between intelligibility scores and the articulatory stability reflected in intraoral air pressure variation. The results revealed that there was no clear trend showing that talkers with small standard deviations in intraoral air pressure and with preserved voice-voiceless differentiation in pressure value had high intelligibility. They stated that the reasons for intelligibility in stop cognate

production in esophageal speech is not clear, especially as it relates to variation of peak intraoral air pressure. Hirose, et al, (1983) reported that voiceless consonants had higher intraoral air pressure than their voiced cognates and that the pressure values for intervocalic voiceless consonants were much greater in the esophageal talkers than in the normal talkers. The authors postulated that the mechanism which causes high intraoral air pressure for the voiceless consonants in esophageal talkers could be ascribed to the result of increasing intraesophageal pressure, since the intraesophageal pressure of laryngectomees has been reported to be markedly higher than the subglottal pressure of normal talkers in sustained phonation (Dey and Kirchner, 1961; van den Berg and Moalenaar-Bijl, 1959). Hirose, et al, (1983), through fiberoptic observation reported that the neoglottis started to open almost simultaneously with the increase in intraoral air pressure. This finding may provide additional evidence that the transesophageal air pressure is responsible for the increase in intraoral air pressure in this group.

For electronic larynx speech, Rothman and Goldstein (1976) observed that good electronic larynx talkers used buccal air (i.e., air trapped in the oral cavity with the articulatory closure) to increase pressure behind the tongue. Isshiki and Tanabe (1972) reporting aerodynamic data from their excellent electronic larynx talker found that intraoral air pressure was very high

(40-60 cm-H<sub>2</sub>O) for /p/ and was variable and low (or even negligible, frequently) for /b/. They also reported that those samples of /b/ produced with relatively high intraoral air pressure tended to be heard as /p/. As to the physiological basis for the production of different intraoral air pressure values for the voicing contrasts, the authors claimed that the upper articulatory mechanism must be responsible because the electronic larynx talker did not have a laryngeal control mechanism. Concerning the acoustic consequence of high intraoral air pressure, they postulated that higher intraoral air pressure for electronic larynx speech than for normal speech was required so that plosive bursts would be produced loudly enough not to be masked by the continuous electronic larynx sound during the production of the voiceless stop. Finally, the study suggested that high intraoral pressure must be generated by a quick and tense articulatory movement for the production of voiceless consonants, whereas relaxation of articulatory muscles should be required for a talker who substitutes voiceless consonants for the production of voiced consonants.

## CHAPTER 2

### METHODS

#### 2.1 Subjects

The subjects for this study were three male laryngectomees, FB, AS, and DB, who speak Standard American English with electronic larynges of the transcervical type. They were selected because they were clinically assessed as having good speech intelligibility with electronic larynges and because they do not use esophageal speech for daily communication. It was believed that they did use air supplied from the esophagus for speech based upon the following observations: (1) without the electronic larynges, they did not produce any sounds during multiple repetitions of stop consonants and affricates (e.g. /papapapapa/ and /tatatata/ ), which sometimes lead even poor esophageal talkers to the production of esophageal speech (Moolenaar-Bijl, 1953). (2) When they swallowed liquid, "clanking" noises were heard, indicating that their pharyngo-esophageal junctions were so tight that they had

difficulties in swallowing. The tightness of the pharyngoesophageal junction frequently prevents laryngectomees from acquiring esophageal speech because they cannot take air into the esophagi" (Singer and Blom, 1981).

Subject FB was 63 years old and had had a total laryngectomy with radical neck dissection two years and six months earlier due to cancer of the larynx. He received six months of radiation therapy after the surgery. Speech therapy for esophageal speech was also tried during that period. He reported that he could not inject or inhale air into the esophagus and the therapy was terminated. An electronic larynx has been used for two years since the surgery and FB reported that he feels comfortable with it for daily communication.

Subject AS was 66 years old and had had a total laryngectomy five years earlier due to cancer of the larynx. Speech therapy was initiated one or two months after the surgery but it was terminated because air injection was very difficult and esophageal sounds were produced only sporadically and with great difficulty. AS has been using an Aurex electric larynx for 5 years and he and his wife reported that people understand AS almost perfectly in daily conversation.

Subject DB was 65 years old and had had a total laryngectomy with radical neck dissection 2 years and seven months earlier due to cancer of the larynx. He reported that he had a surgical removal of some tissue in the base of

the tongue and the pharynx from under the chin 21 months after the laryngectomy followed by radiation therapy for a month. He started speech therapy for esophageal speech one month after the laryngectomy, but it was terminated after one year because he had difficulties taking air into the esophagus. Speech intelligibility with the electronic larynx (the Servox) was excellent in daily communication.

## 2.2 Speech Materials

The speech materials were 12 utterance types consisting of CVC sequences with /p/, /b/, /t/, /d/, /k/, and /g/ for consonants and /ɔ/ and /I/ for vowels (Table 2.1). These CVC sequences were preceded by schwa, /ə/ in order to ensure that the electronic larynx sound had been produced prior to the initial consonants and remove possible onset delay of operation (VOT) as a cue. The final consonants were always voiceless with the same place of articulation as the initial ones for all the utterance types in order to eliminate the influences of the final consonant on the initial one for the voicing distinction acoustically and perceptually. All utterance types were presented in English orthography in forms which looked like English words (Table 2.2). Twenty repetitions of each utterance type were elicited; therefore, each subject produced 240 tokens during the experiment. The order of presentation of the 240 tokens was randomized.

Table 2.1  
Utterance Types

/əpɔp/	/əpɪp/
/əbɔp/	/əbɪp/
/ətɔt/	/ətɪt/
/ədɔt/	/ədɪt/
/əkɔk/	/əkɪk/
/əgɔk/	/əgɪk/

Table 2.2  
The word forms of the utterance types presented to the subjects to read.

APOP (= /əpɔp/)	APIP (= /əpɪp/)
ABOP (= /əbɔp/)	ABIP (= /əbɪp/)
ATOT (= /ətɔt/)	ATIT (= /ətɪt/)
ADOT (= /ədɔt/)	ADIT (= /ədɪt/)
ACOCK (= /əkɔk/)	AKICK (= /əkɪk/)
AGOCK (= /əgɔk/)	AGICK (= /əgɪk/)

Prior to data recording, there was a practice session to familiarize the subject with the nonsense syllables and the experimental setting in general. The subjects read the randomized lists of tokens after the practice session. The electronic larynx was hand-held by the subject in his normal manner, and was placed on his neck before he started producing the tokens.

### 2.3 Electronic Larynx

Subject FB and DB used the SERVOX Electronic Speech Aid (usually called "the Servox"), placing the devices with their hands against the side of the neck. Subject AS used the NEOVOX Electro-Larynx by Aurex (Model 520T), also holding the device with his hand against the side of the neck.

### 2.4 Data Acquisition

Simultaneous aerodynamic, EMG, kinematic, and acoustic recordings were obtained from the three subjects during their productions of the 240 utterance tokens. All three experiments were conducted at Haskins Laboratories.

#### 2.4.1 Intraoral Air Pressure

For subjects FB and AS, a Catheter Tip Pressure Transducer (MMI-Gaeltec) was used to measure intraoral air pressure. The transducer has a 80 cm length flexible catheter whose tip was mounted with a miniature pressure transducer of 1.33 mm diameter. The transducer was inserted through the nostrils into the nasopharynx. Topical anesthesia was applied to the nasal mucosa in order to minimize subject discomfort. Calibration signals were recorded by inserting the tip of the catheter into a glass cylinder which contained water at a temperature slightly below 37 degrees Celsius. Calibration recordings were made at three levels of water depth: 0 cm, just below the surface of the water, 3 cm, and 10 cm. For subject DB, intraoral air pressure was measured by inserting a polyethylene tube with an inner diameter of .068 inch and an outer diameter of 0.100 inch through the nostril into the nasopharynx. The outer end of the tube was connected to a pressure transducer (Setra Systems Model, 236L). This different pressure transducer was used for subject DB due to damage to the original miniature pressure transducer after the second experiment. A calibration recording at 3 cm of water was recorded during the data recording for DB.

#### 2.4.2 Electromyography (EMG)

Muscle action potentials were recorded to obtain data on activity of tongue muscles for speech gestures. Electromyographic (EMG) signals were picked up by hooked-wire electrodes inserted to the anterior genioglossus muscle for the three subjects, the posterior genioglossus muscle for FB and AS, and the middle pharyngeal constrictor muscle for DB. EMG techniques used in the experiments are described by Basmajian and Stecko (1962), Hirose et al, (1971), and Harris (1981).

For the preparation of the electrodes, the tips of small insulated wires 0.05 mm in diameter were bent back about 2 mm and small sections of insulation from the hooks were burnt off in order to provide electrical contact with muscle tissue. The wires were passed through a fine hypodermic needle of 26 or 27 gauge, and only the hooked ends were left outside of the needle.

After topical anesthesia was applied at the site of insertion, the sterilized needle was inserted percutaneously into the muscles. The insertions to the anterior genioglossus muscle was made perpendicularly to the surface of the skin at the midpoint between the site where the hyoid bone existed before the laryngectomy and the mandibular ridge in the paramedial line. The needle was inserted to the depth at which the physician could palpate it with his finger on the floor of the subject's mouth. The needle was then withdrawn with the wire electrode left in place. For

the posterior genioglossus muscle, the insertion was made by aiming the needle slightly more posteriorly than the direction of insertion for the anterior genioglossus muscle.

In order to verify the correct placement of an electrode in a specific muscle, muscle activity was monitored during the performance of verification gestures (Hirose, 1971; Smith and Hirano, 1968). For both anterior and posterior genioglossus muscles, prominent muscle activity was observed when the subjects' tongues were protruded and when they produced /i/, /u/, /ta/, /ti/, /ka/, and /ki/. No activity was observed with low vowels /a/ and /ɔ/. More vigorous activity was observed with the anterior genioglossus during the production of /t/ than during /k/, and vice versa with the posterior genioglossus. In the case of subject DB, however, no activity was monitored during the production of /i/ or /u/ when the deep insertion was made aiming at the posterior genioglossus muscle. Muscle activity of great amplitude was monitored during /t/ and /k/ from the electrode. Considering the direction and the depth of the needle and also the unique oropharyngeal structure of the subject, who had had the surgical removal of some part of the tongue and pharyngeal tissue, it is assumed that the electrode was inserted into the upper part of the middle pharyngeal constrictor rather than the posterior genioglossus muscle. However, we must regard this muscle assignment as somewhat tentative.

At several times during each experiment, a calibration signal of 300 microvolts was recorded on EMG channels and the amplitudes of muscle activities were later calibrated in reference to this signal.

#### 2.4.3 Articulatory Movement

Jaw, lower lip and head movements were recorded using an opto-electronic tracking system (Kay, et al, 1985). A Selspot camera equipped with a Schottky planar diode located in its focal plane monitored the modulated light from infra-red light emitting diodes (LEDs) attached to the subject's nose, lower lip, and jaw. Electronics associated with the camera decoded the output signals of the electrodes and provided position data in X and Y coordinates. A calibration signal was recorded by moving the LEDs attached to a measurement caliper through 2 cm in the field of view. The articulatory movements, which are shown as the amount of displacement of the LEDs were later calibrated by using this calibration signal as a reference.

#### 2.4.4 Acoustic Recordings

The audio signals of the subjects during the experiments were recorded using a directional microphone (Sennheiser Model MKH816T), which was placed approximately 10 cm away from the subject's mouth. Although the

experiment room was designed for physiological recording rather than for audio recording, maximum efforts were made to reduce the ambient noise during the recordings.

#### 2.4.5 Perceptual Experiment

The tokens produced by the alaryngeal talkers were used as stimuli for a perceptual experiment. Three listening tapes were made, each of which consisted of 239 to 240 tokens (12 utterance types x 19-20 repetitions) for each talker. Tokens were randomized and an interval of five seconds of silence was inserted between the tokens. An interval of 10 seconds of silence was given after every 20 tokens. Seven listeners heard all three tapes. The criterion for the selection of the listeners was that they have had clinical experience in speech pathology. Practice sessions were provided for the listeners prior to the experiments in such a way that the listeners heard 30 randomized tokens of each alaryngeal talker while the listeners were looking at the "correct" utterance type of each token, in order to identify or become familiar with what the talker intended to produce. The listeners were then asked to identify the initial consonant and to transcribe it on a response sheet. The experiments were conducted in a sound-proofed room and the tapes were heard monaurally using headphones.

It should be noted that the listeners were advised to judge a token as "voiceless" when they thought that the token could be voiceless or when it sounded different from the voiced cognate. This advice was given because the alaryngeal talkers were not always successful in producing very clear voiceless stops and sometimes the perceptual difference between voicing contrasts was small.

## 2.5 Data Processing

The physiological and acoustic signals were DC offset and amplified through adjustments of the amplifier in order to achieve an optimal recording level for each channel. The amplified signals were recorded on FM magnetic tape, using a 14-channel instrumentation data recorder. These recorded physiological and acoustic data were digitized on a VAX 11/780 computer for further processing and analysis. The acoustic signal was sampled at the rate of 10,000 samples per second and the EMG and movement signals were sampled at 200 samples per second. All signals were quantized using an A/D converter with 12 bit precision for full range. The simultaneously-recorded EMG, intraoral pressure, kinematic, and acoustic signals were time-locked via a timing code generator/reader interfaced to the computer.

### 2.5.1 Data Segmentation

The Haskins Laboratories Physiological Signal Processing (PSP) program (Gulisano, 1982) was used to identify and extract the 19 to 20 tokens of each utterance type, using the acoustic waveforms. The release of the initial consonant was used as a temporal reference point (or "lineup point") for each token for subsequent processing. The accurate measurement of lineup points was important because the tokens were aligned at these points for further processing (ensemble averaging) and also because physiological measurements were made with reference to these points. The onset of the consonant release was defined as the onset of the periodic acoustic pulses of the vowel or as the onset of the release burst of the initial consonantal closure. Figure 2.1 gives an example of the lineup point which was located at the release burst of the initial /t/ in the production of /ətɔt/ by subject DB.

After measuring the lineup points, all the tokens were labeled by utterance type. For further processing, signals were restricted to a range of 500 ms preceding and 500 ms following the lineup point for each token.

### 2.5.2 Smoothing and Averaging

The physiological signals were smoothed using the Haskins Laboratories PSP program. The smoothing procedure was considered necessary because the raw waveforms of the

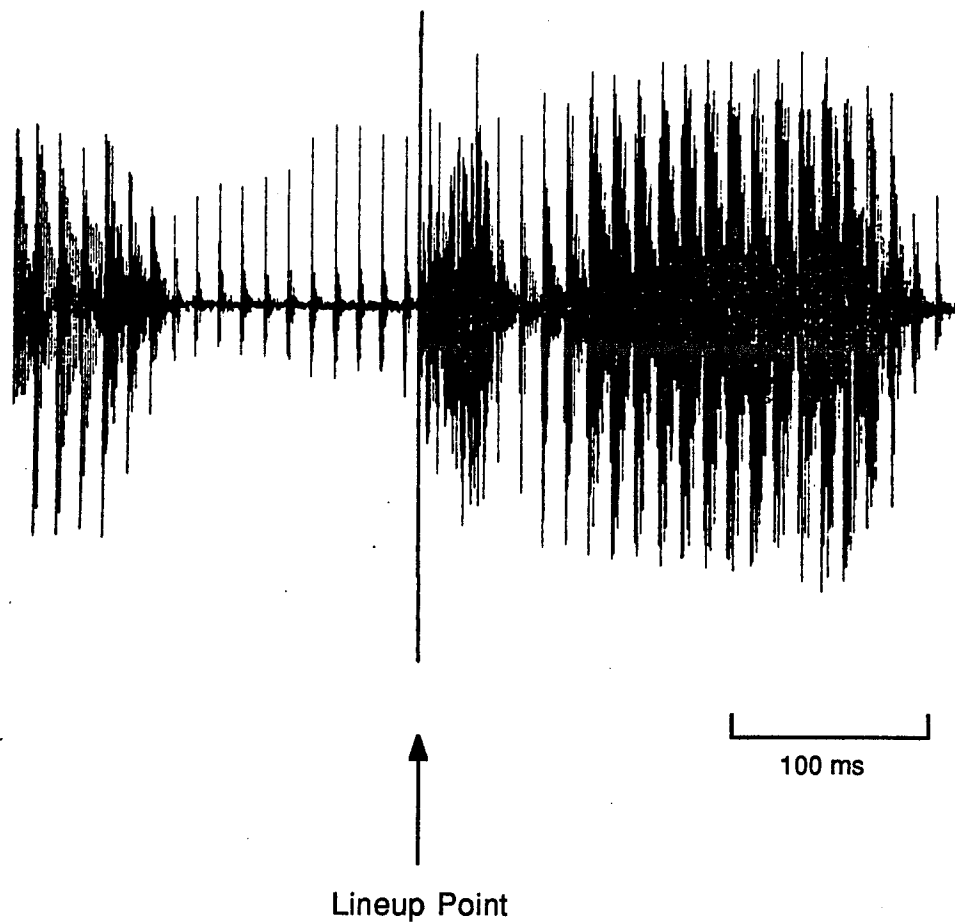


Figure 2.1 Linup point.

The vertical line labelled U3T3 is the lineup point located at the consonant release for the initial /t/ in the production of /ətət/ by subject DB.

physiological signals often exhibited minute fluctuations, possibly due to the mechanical artifact of the recordings and also due to the nature of the signals. This noise made it difficult to measure amplitude peaks and valleys in a signal. The integration time for smoothing was 35 ms for the EMG signals and 15 ms for the pressure and movement signals.

## 2.6 Data Analysis

### 2.6.1 Measurements of Intraoral Air Pressure

Peak values of intraoral air pressure for the initial consonants were obtained by subtracting the pressure value of the baseline from the greatest value before the lineup point. The baseline was identified as the stable value during the acoustic interval for /ə/. For the few tokens which did not have flat baselines, the lowest pressure values during the /ə/ were used as the baseline value. Figure 2.2 gives examples of the pressure waveforms during the production of /əpIp/ and /ət)t/ by subject DB.

### 2.6.2 Measurements of EMG

The peak values for the muscle activity associated with the production of the initial consonant in each token were obtained simply by finding the greatest EMG value before the

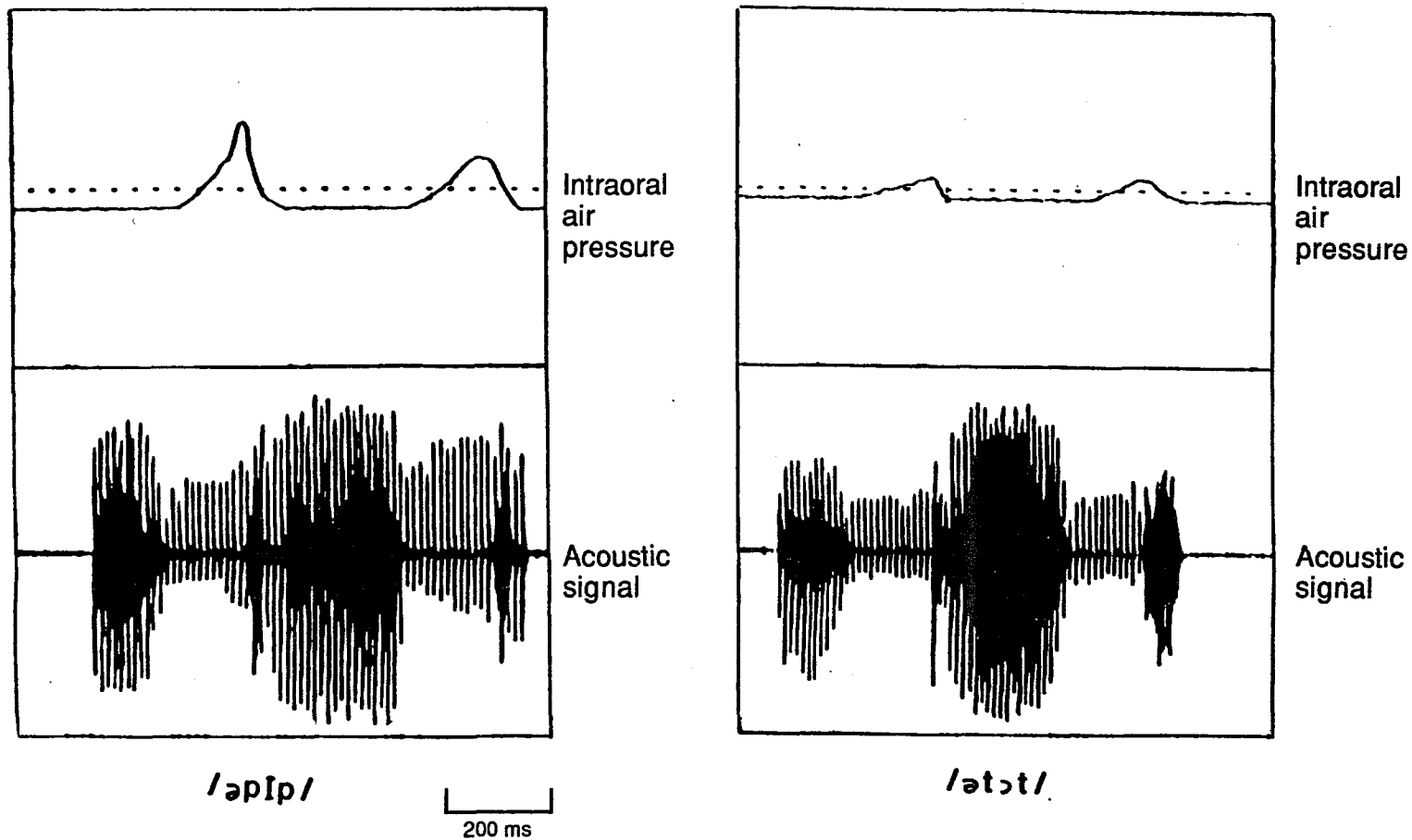


Figure 2.2 Sample waveforms of intraoral air pressure and the corresponding acoustic signals during the production of /əpɪp/ (on the left) and /ətɔt/ (on the right) by Subject D.B. Intraoral air pressure is displayed on the top and acoustic signals on the bottom. The initial intraoral air pressure peaks represent the pressure increase for initial consonants.

lineup point. Since EMG signals fluctuate, a single peak value of the EMG signal for a token has little significance in terms of discussing representative muscle behavior. However, averaging peak values of the tokens provides a more comprehensive value for the muscle activities associated with certain articulatory gestures. Figure 2.3 shows examples of smoothed EMG signals for the genioglossus anterior for /ətɔt/ for FB.

### 2.6.3 Measurements of Articulatory Movements

At the onset of production of the utterance types in this study, the speaker's mouth was open for the /ə/, subsequently, the jaw and lower lip began moving upward for the production of the stop consonant. Therefore, the amount of vertical displacement of the articulator was measured by calculating the distance between the lowest value during /ə/ and the peak value during the consonant closure. The amount of horizontal displacement of the lower lip (for /p/ and /b/ only) was measured by calculating the distance between the peak value, which corresponded to the maximum advancement of the articulator for consonant lip protrusion and the lowest value, which corresponded to the maximum retraction of the articulator. Figure 2.4 gives an example of the jaw movement for the production of /əpɔp/ by subject FB.

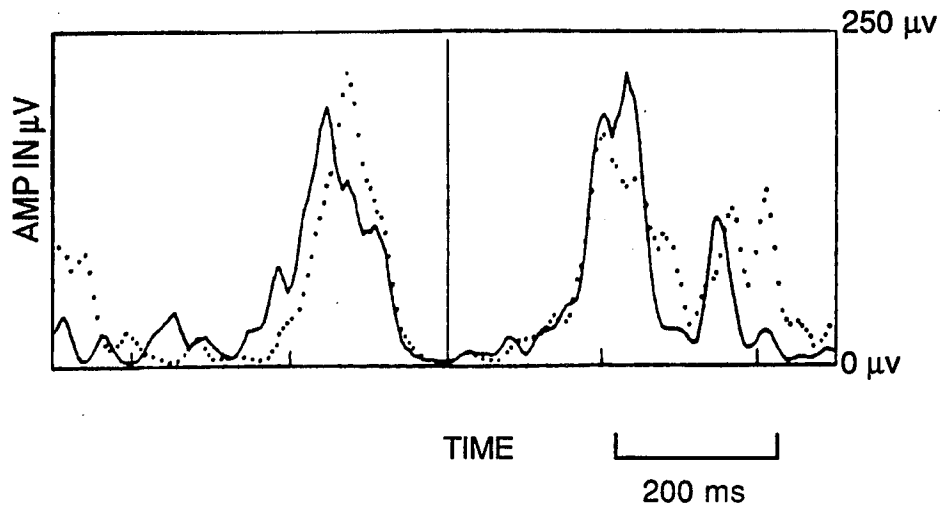


Figure 2.3 Examples of smoothed EMG signals for the anterior genioglossus muscle for the production of /ətʃt/ for FB. The solid line is the signal for the eleventh repetition of the utterance, the dotted line is the signal for the twentieth repetition of the same utterance. The vertical line in the center of the figure is the lineup point, which corresponds to the consonant release for the initial /t/.

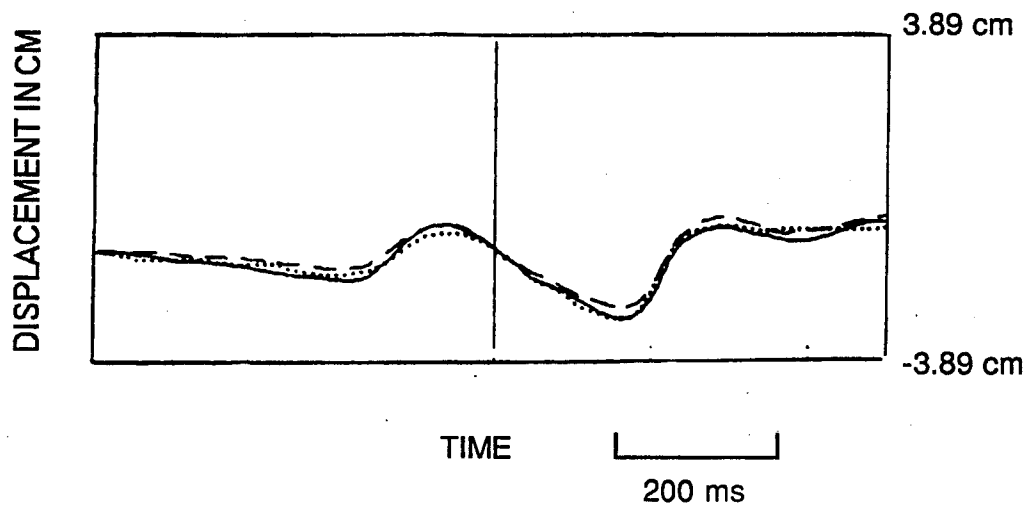


Figure 2.4 Waveforms of vertical jaw movement for the production of /əpɔp/ by Subject FB. The three lines represent the waveforms of the fourth, fifth and sixth repetitions of the utterance. The y axis indicates jaw y displacement.

#### 2.6.4 Acoustic Measurements

The duration of the closure for the initial consonant and the duration of the first vowel were measured. The onset of the consonant was defined as the offset of the periodic acoustic pulses for /ə/, followed by the periodic pulses of the electronic larynx. The offset of the consonant closure was defined either as the onset of the release burst of the consonant, if a release burst was visible in the acoustic waveform, or else as the onset of periodic acoustic pulses of the following vowel. Vowel duration was measured from the onset of periodic acoustic pulses for the vowel to the termination of these periodic pulses and the onset of the periodic pulses of the electronic larynx. Figure 2.5 gives an example of how the durational measurements were made.

#### 2.6.5 Perceptual Analysis

Perceptual analyses were performed for two purposes. One purpose was to obtain formal speech intelligibility scores for the speakers; the second purpose was to classify the tokens into three categories: "correct tokens," "incorrect tokens," and "ambiguous tokens." Seven listeners were asked to identify the consonant produced in each utterance for all tokens. Tokens were labeled as "correct" when five out of the seven listeners perceived the token as having the same voicing status as the target utterance. For example, if a speaker intended to produce the voiceless

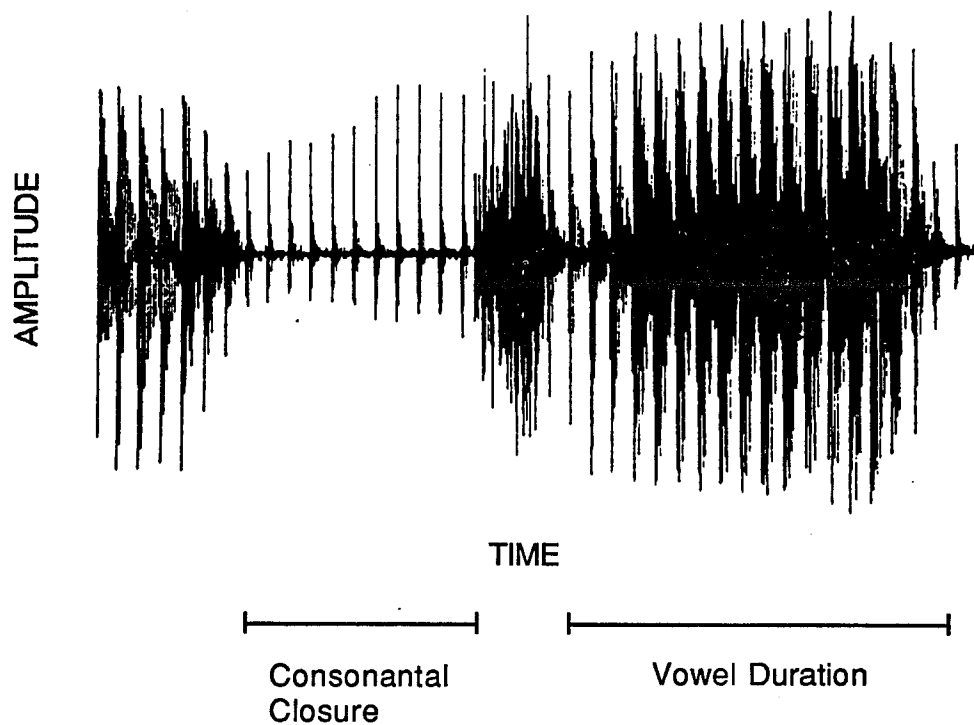


Figure 2.5 An example of durational measurement in the acoustic waveform for the production of /stɔt/ by subject DB. The period of consonantal closure corresponds to the initial /t/ and that of vowel duration, /ɔ/.

consonant /p/ and more than five listeners perceived it as /p/, the token was labeled as "perceptually correct." If a speaker intended to produce the voiceless consonant /p/ and more than five listeners perceived it as /b/, the token was labeled as "perceptually incorrect." When a token was perceived by three listeners as voiceless and by four other listeners as voiced, or vice versa, the token was labeled as "perceptually ambiguous." A confusion matrix was constructed for each speaker, based on the perceptual judgements.

#### 2.6.6 Statistical Analyses

Physiological and acoustic values were examined by analyses of variance in order to determine which differences between the voicing pairs were significant. Discriminant analysis was also used in order to investigate the underlying physiological and acoustic events which had contributed to the perceptual judgements. That is, the discriminant analysis was performed in order to determine which physiological and acoustic factors contributed to the listeners' perceptions of the consonants as either voiced or voiceless. The statistical analyses were performed on a VAX 11/780 computer using the Biomedical Data Processing System (BMDP).

## CHAPTER 3

## RESULTS

3.1 Intraoral Air Pressure

Mean values of intraoral air pressure for each consonant type for each subject were obtained by averaging the peak values of the intraoral air pressure for the 19 or 20 tokens of each consonant type produced by individual speakers. Figure 3.1 shows the averaged waveforms of intraoral air pressure for subject FB. The wave forms are displayed as sets of each voicing pair (i.e. /p/ vs /b/), /t/ vs /d/), /k/ vs /g/), /pI/ vs /bI/), /tI/ vs /dI/), and /kI/ vs /gI/) since the major purpose of this study is to compare the differences in speech mechanisms between these voicing pairs. In these waveforms, the first peak corresponds to the increase in intraoral air pressure for the first consonant of /əCVC/. The second peak corresponds to the second consonant. The vertical lines in the middle of the displays are lineup points, which correspond to the consonant release. It was consistently observed that

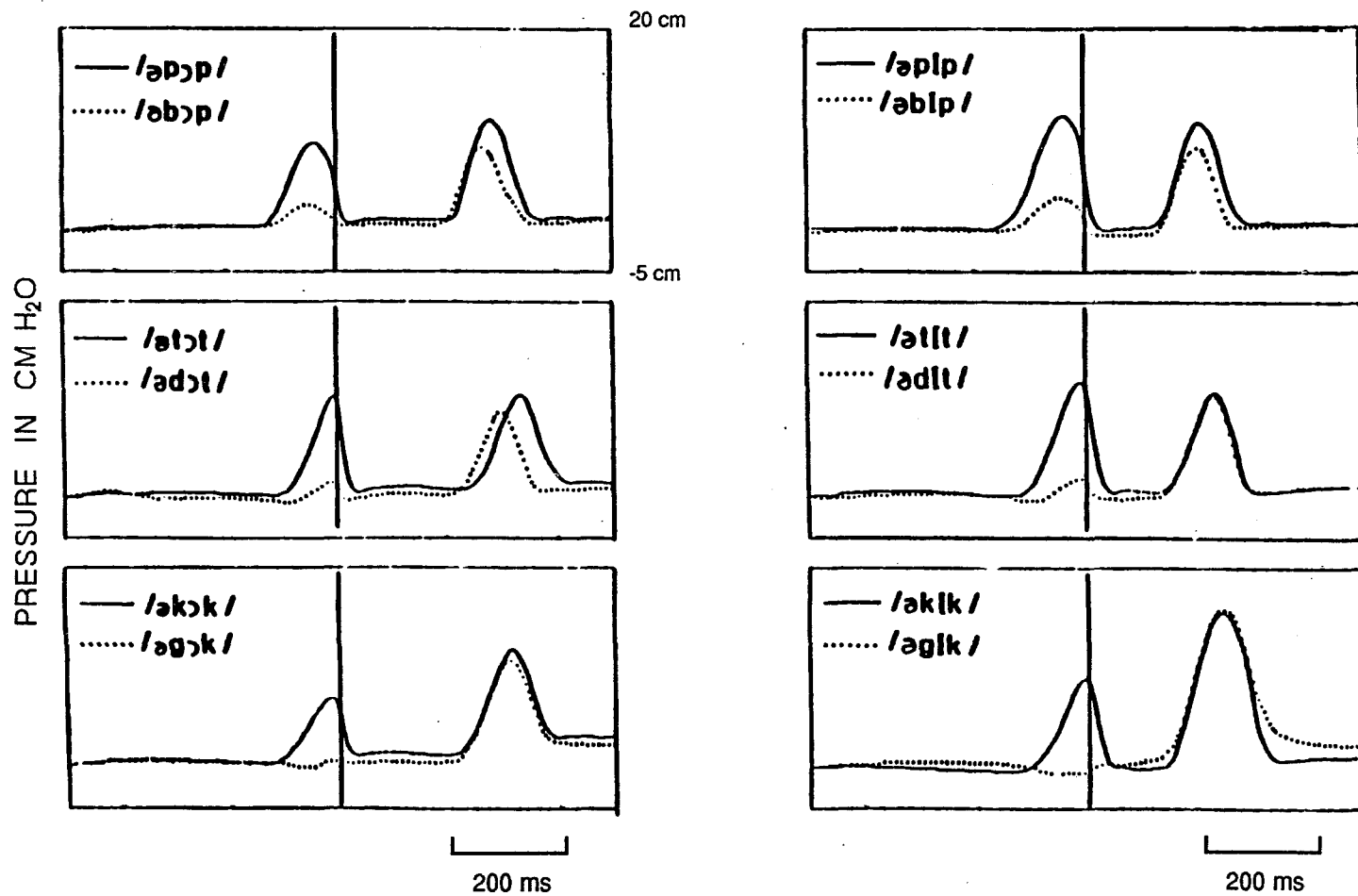


Figure 3.1 Averaged waveforms of intraoral air pressure for Subject FB. Solid lines represent the voiceless stops and dotted lines their voiced counterparts. The solid line represents the lineup points (see text).

voiceless consonants are associated with higher pressure peaks, as compared to their voiced cognates for the initial stops. It was also noted that mean pressure amplitudes for the initial voiceless stops were very high, ranging from 6.39 cm H<sub>2</sub>O (for /əKʔk/) to 11.89 cm H<sub>2</sub>O (for /ətIt/). Pressure amplitudes for voiceless stops produced by normal speakers are reportedly around 5 to 6 cm H<sub>2</sub>O.

Figure 3.2 shows the averaged waveforms of intraoral pressure for subject AS. For the initial stops, voiceless consonants are always associated with higher pressure amplitudes than their voiced cognates. Mean pressure amplitudes ranges from 17.32 cm H<sub>2</sub>O (for /əpɔp/) to 28.19 cm H<sub>2</sub>O (for /əkIk/ for voiceless stops and from 1.33 cm H<sub>2</sub>O (for /əgIk/) to 8.06 cm H<sub>2</sub>O (for /əbɔP/) for voiced stops. In these waveforms, it is noted that the second peak which corresponds to the pressure increase for the second consonant is not observed for labial and alveolar consonants for this subject.

Figure 3.3 shows the averaged waveforms of intraoral air pressure for subject DB. This subject produced a pressure increase for both voiceless and voiced stops. However, voiceless consonants are consistently associated with higher pressure amplitudes than their voiced cognates. Mean pressure amplitudes range from 37.45 cm H<sub>2</sub>O (for /əpIp/) to 49.86 cm H<sub>2</sub>O (for /əkIk/) for voiceless stops and from 13.53 cm H<sub>2</sub>O (for /əbIP/) to 30.15 cm H<sub>2</sub>O (for /əgɔk/).

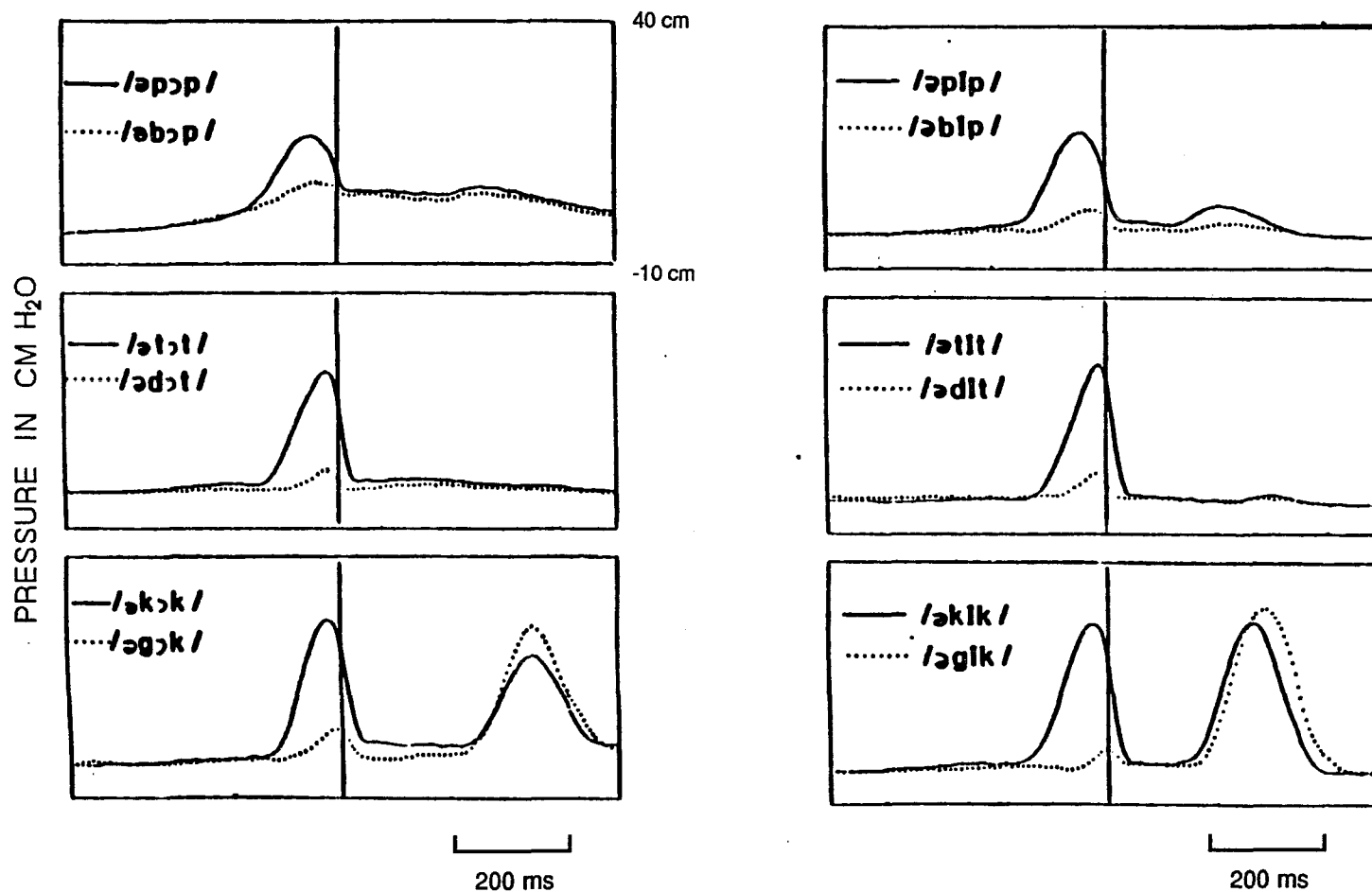


Figure 3.2 Averaged waveforms of intraoral air pressure for Subject AS. Solid lines represent the voiceless stops and dotted lines their voiced counterparts. The solid line represents the lineup points (see text).

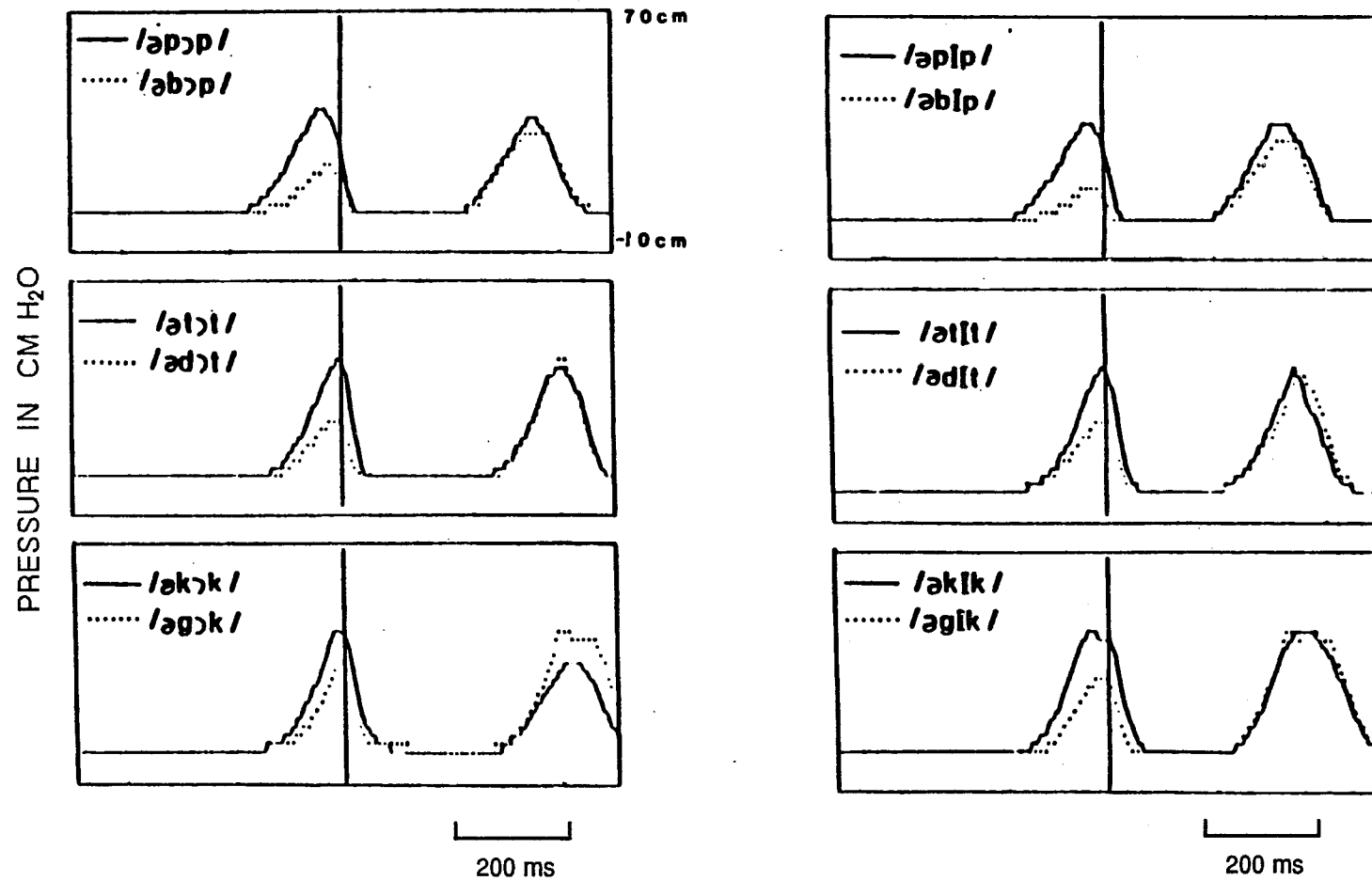


Figure 3.3 Averaged waveforms of intraoral air pressure for Subject DB. Solid lines represent the voiceless stops and dotted lines their voiced counterparts. The solid line represents the lineup points (see text).

Similarly to FB and AS, these pressure amplitudes are extremely high, as compared to normal speakers.

Figure 3.4 gives the frequency distribution of peak intraoral pressure values for subject FB. The frequency distribution was made for each consonant type with the two vowels grouped together, based on 40 peak pressure values (20 for /ɔ/ and 20 for /I/) for each consonant type. The shaded columns correspond to voiceless stops and the white columns to their voiced counterparts. It is observed that the peak pressure values for the voiceless stops and their voiced cognates are distributed differently. For /k/ and /g/, however, the values overlap somewhat at around 0 cm H<sub>2</sub>O, with some of the values for /k/ being lower than others.

Figure 3.5 shows the comparable frequency distributions for subject AS. It can be observed that these are rather clear normal distributions with very little overlap for the three voicing pairs. Figure 3.6 shows the same frequency distributions for subject DB. These distributions are generally well-separated for the voicing pairs, although the values overlap slightly for the velar pair /k,g/.

A three-way analysis of variance (ANOVA) with repeated measures with place of articulation, vowel and voicing type as the independent variables and peak intraoral air pressure as the dependent variable was performed separately for each subject. Table 3.1 shows the results of these analyses of variance for subjects FB, AS, and DB. A significant main

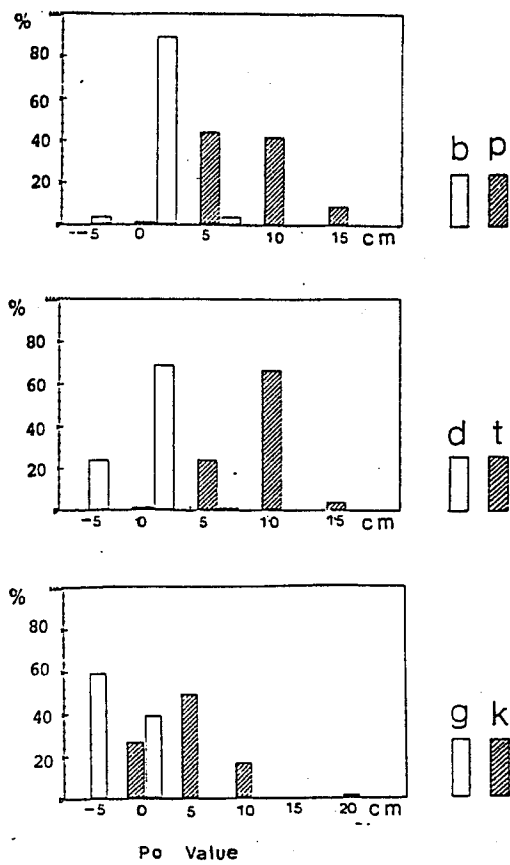


Figure 3.4 Distribution of peak intraoral air pressure values for Subject FB.

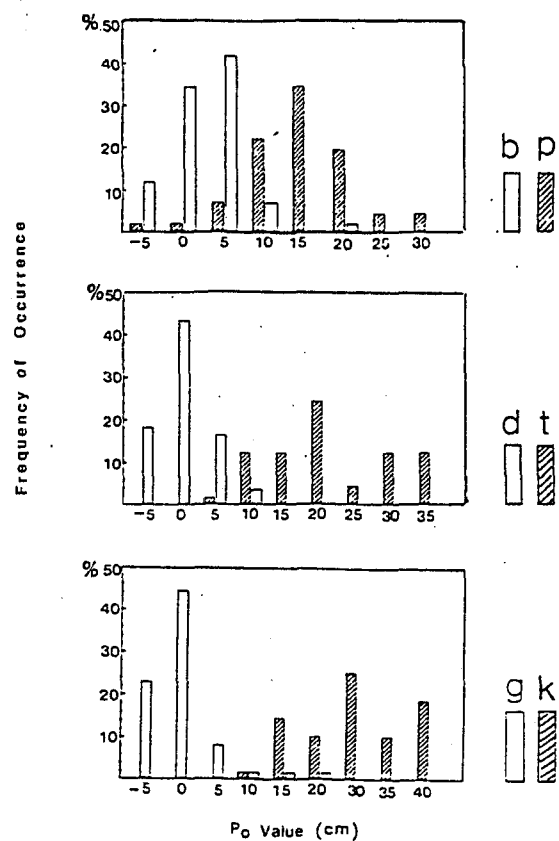


Figure 3.5 Distribution of peak intraoral air pressure values for Subject AS.

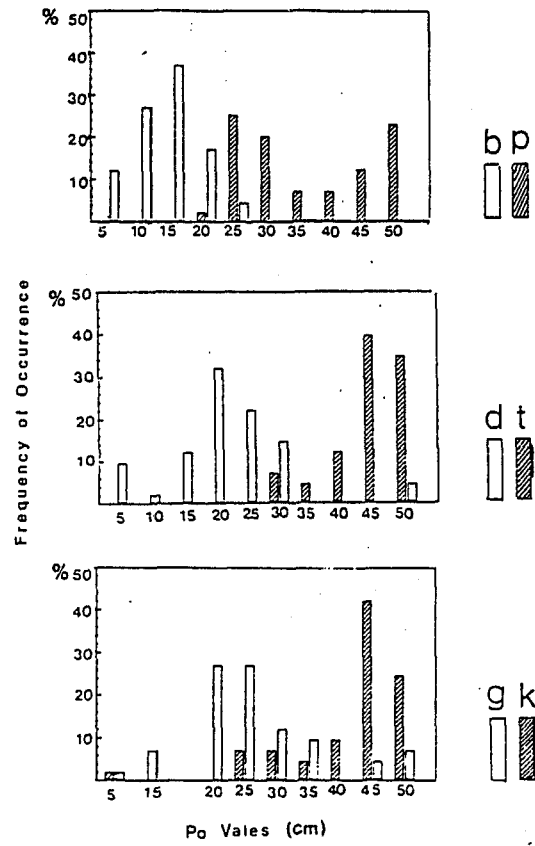


Figure 3.6 Distribution of peak intraoral air pressure values for Subject DB.

Table 3.1  
Analysis of Variance: Effect of Intraoral Air Pressure (cm-H<sub>2</sub>O)

Subj	Source	Sum of Squares	Degrees of Freedom	Mean Square	F	Tail Prob.
FB	Voicing	4127.698	1	4127.698	571.47	0.0000**
	Place	378.749	2	189.374	26.22	0.0000**
	Vowel	66.886	1	66.886	9.26	0.0026**
	VowxVoi	64.436	1	64.436	8.92	0.0031**
	PlaxVoi	45.734	2	22.867	3.17	0.0441*
	PlaxVow	16.253	2	8.127	1.13	0.3264
	PlaxVowxVoi	15.799	2	7.899	1.09	0.3367
	Error	1632.374	226	7.223		
AS	Voicing	23140.269	1	23150.269	485.85	0.0000**
	Place	591.510	2	295.755	6.21	0.0024**
	Vowel	20.001	1	20.001	0.42	0.5177
	VowxVoi	260.636	1	260.636	5.47	0.0202*
	PlaxVoi	1783.546	2	891.773	18.72	0.0000**
	PlaxVow	184.044	2	92.022	1.93	0.1474
	PlaxVowxVoi	23.039	2	11.520	0.24	0.7854
	Error	10720.982	225	47.64		
DB	Voicing	26279.645	1	26279.645	261.29	0.0000**
	Place	4434.487	2	2217.243	22.05	0.0000**
	Vowel	0.255	1	0.255	0.00	0.9599
	VowxVoi	99.382	1	99.382	0.99	0.3213
	PlaxVoi	474.074	2	237.037	2.36	0.0970
	PlaxVow	496.160	2	248.081	2.47	0.0871
	Error	22830.772	227	100.576		

\*\* p < 0.01  
\* p < 0.05

effect was found for voicing status for FB ( $F(1,226) = 571.47, p < 0.01$ ), for AS ( $F(1,225) = 485.85, p < 0.01$ ), and for DB ( $F(1,227) = 261.29, p < 0.01$ ). A significant main effect was also found for place of articulation for FB ( $F(2,226) = 26.22, p < 0.01$ ), for AS ( $F(2,225) = 6.21, p < 0.01$ ), and for DB ( $F(2,227) = 22.05, p < 0.01$ ). A significant main effect for vowel was found for FB ( $F(1,226) = 9.26, p < 0.01$ ), but not for AS or DB.

The interaction between vowel and voicing was significant for FB ( $F(1,226) = 8.92, p < 0.01$ ). Figure 3.7 illustrates the change in mean intraoral pressure for the voicing status as a function of vowel context. Although intraoral air pressure is higher for voiceless stops than for voiced stops in both vowel contexts, it is higher when followed by /I/, than when followed by /ɔ/ for voiceless stops, whereas there is no difference in pressure values across vowels contexts for voiced stops.

The interaction between voicing and place of articulation was also found significant for FB ( $F(2,226) = 3.17, p < 0.05$ ). Figure 3.8 shows the change in mean intraoral air pressure for the two voicing conditions as a function of place of articulation. It is observed that the pattern for all places of articulation is similar in that voiceless stops have higher intraoral air pressure values than their voiced cognates. However, the two voicing conditions differ in that for voiceless stops, the alveolar /t/ has the highest mean air pressure value, followed by the

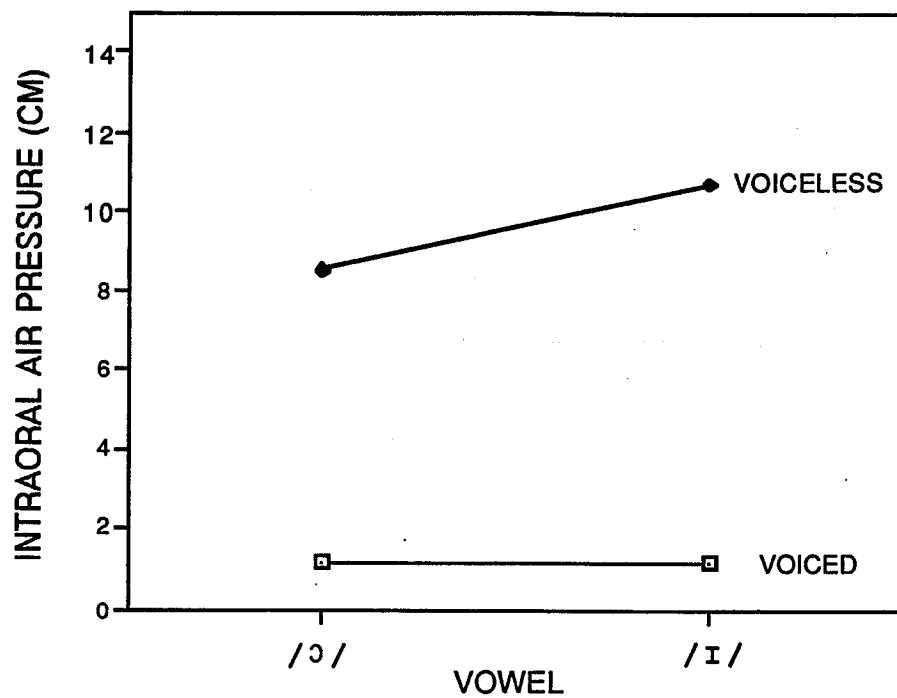


Figure 3.7 Mean intraoral air pressure values for voiced and voiceless stops as a function of vowel context for Subject FB.

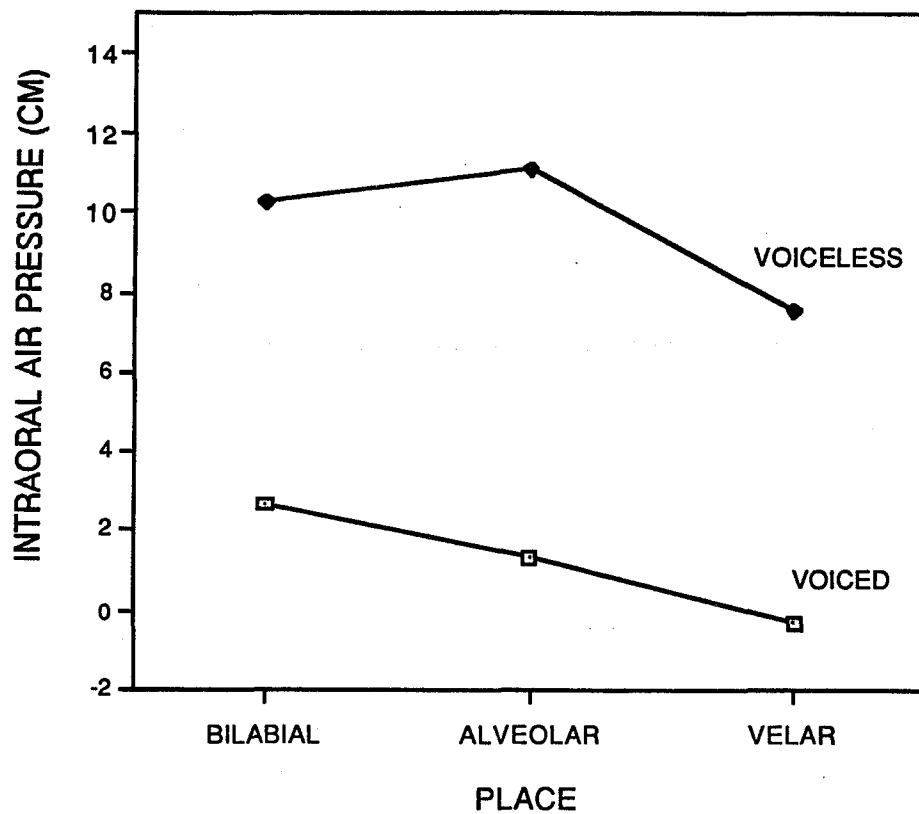


Figure 3.8 Mean intraoral air pressure values for voiced and voiceless stops as a function of place of articulation for Subject FB.

bilabial stop /p/, whereas for voiced stops, the bilabial stop /b/ has the highest mean air pressure value, followed by the alveolar stop /d/ and the velar stop /g/.

Subject AS also exhibited a significant interaction between vowel and voicing ( $F(1,225) = 5.47, p < 0.05$ ). Figure 3.9 shows the change in mean intraoral air pressure for the voicing status as a function of vowel contexts for AS. It can be observed that although voiceless stops are always associated with higher pressure values than voiced stops, for voiceless stops, the values are higher when followed by /I/, than when followed by /O/. The opposite pattern is observed for voiced stops.

Figure 3.10 illustrates mean intraoral pressure values for the voicing status as a function of place of articulation for AS. The interaction between voicing and place of articulation is again significant ( $F(2,225) = 18.72, p < 0.01$ ). The pattern for both voicing conditions is similar, in that voiceless stops have higher mean intraoral air pressure values than voiced stops. The two voicing conditions differ in that for voiceless stops the velar stop has higher pressure than the alveolar stop which has higher pressure than the bilabial stop, whereas for voiced stops, the bilabial has the highest pressure value, followed by the velar and alveolar stops.

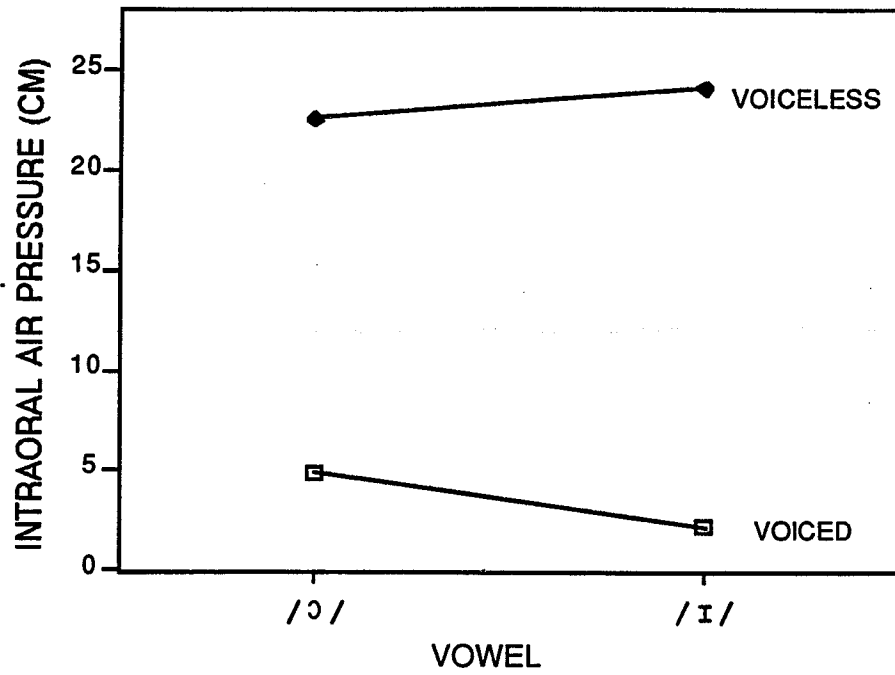


Figure 3.9 Mean intraoral air pressure values for voiced and voiceless stops as a function of vowel context for Subject AS.

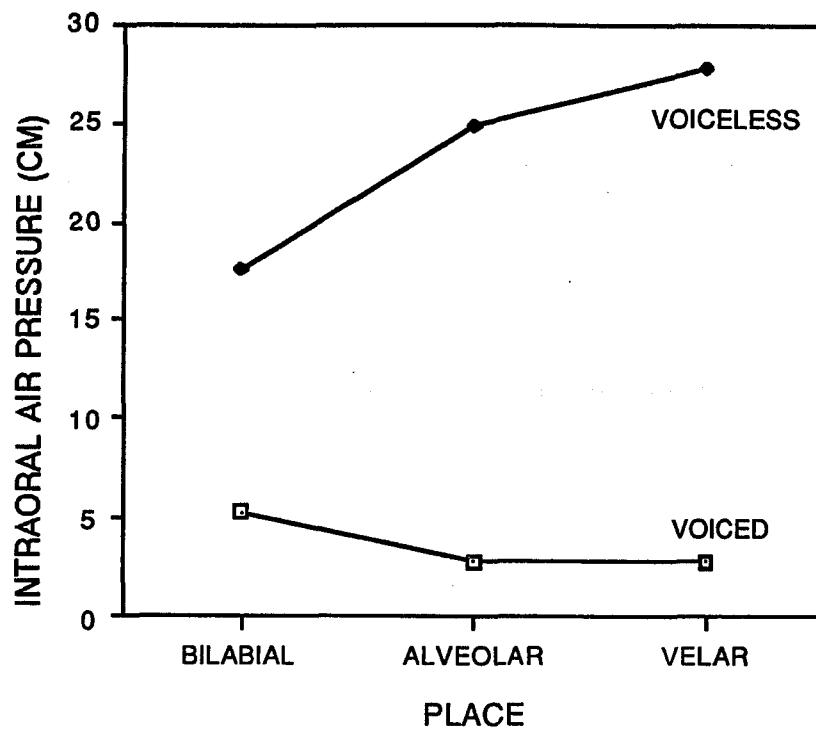


Figure 3.10 Mean intraoral air pressure values for voiced and voiceless stops as a function of place of articulation for sSubject AS.

### 3.2 Electromyographic Analysis

EMG potentials were measured in order to assess the pattern of tongue activity associated with consonantal voicing distinctions. Graphic displays of EMG patterns are given in Figs. 3.11, 3.12, 3.13, and 3.14. Figure 3.11 shows averaged waveforms of the activity of the anterior genioglossus muscle for subject FB. Solid lines represent muscle activity for the voiceless consonants and dotted lines represent muscle activity for the voiced consonants. The vertical solid lines show the acoustic lineup points. The amplitude of muscle activity is given in microvolts. Significantly greater muscle activity is observed for /g/ than for /k/, for /gI/ than for /kI/, and for /pI/ than for /bI/.

Figure 3.12 shows averaged waveforms of the activity of the anterior genioglossus muscle for subject AS. Significantly greater muscle activity is observed for /d/ than for /t/ and for /g/ than for /k/ in both vowel contexts. Also, muscle activity is significantly greater for /pI/ than for /bI/.

Figure 3.13 shows averaged waveforms of the activity of the posterior genioglossus muscle for subject FB. There is no significant difference in the level of muscle activity for any of the voicing pairs.

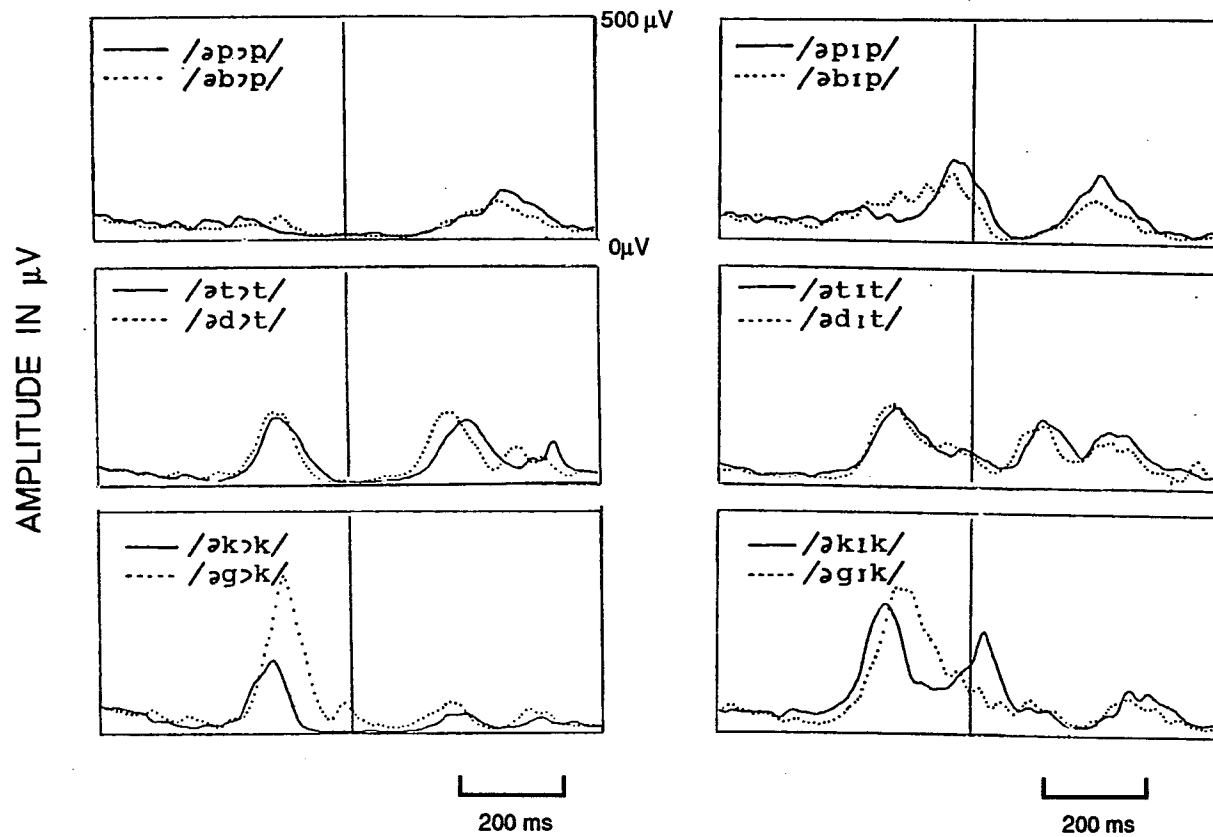


Figure 3.11 Averaged waveforms of the activity of the anterior genioglossus muscle for subject FB. Solid lines represent muscle activity for the voiceless stops and dotted lines activity for their voiced counterparts. See solid line corresponds to the acoustic lineup points.

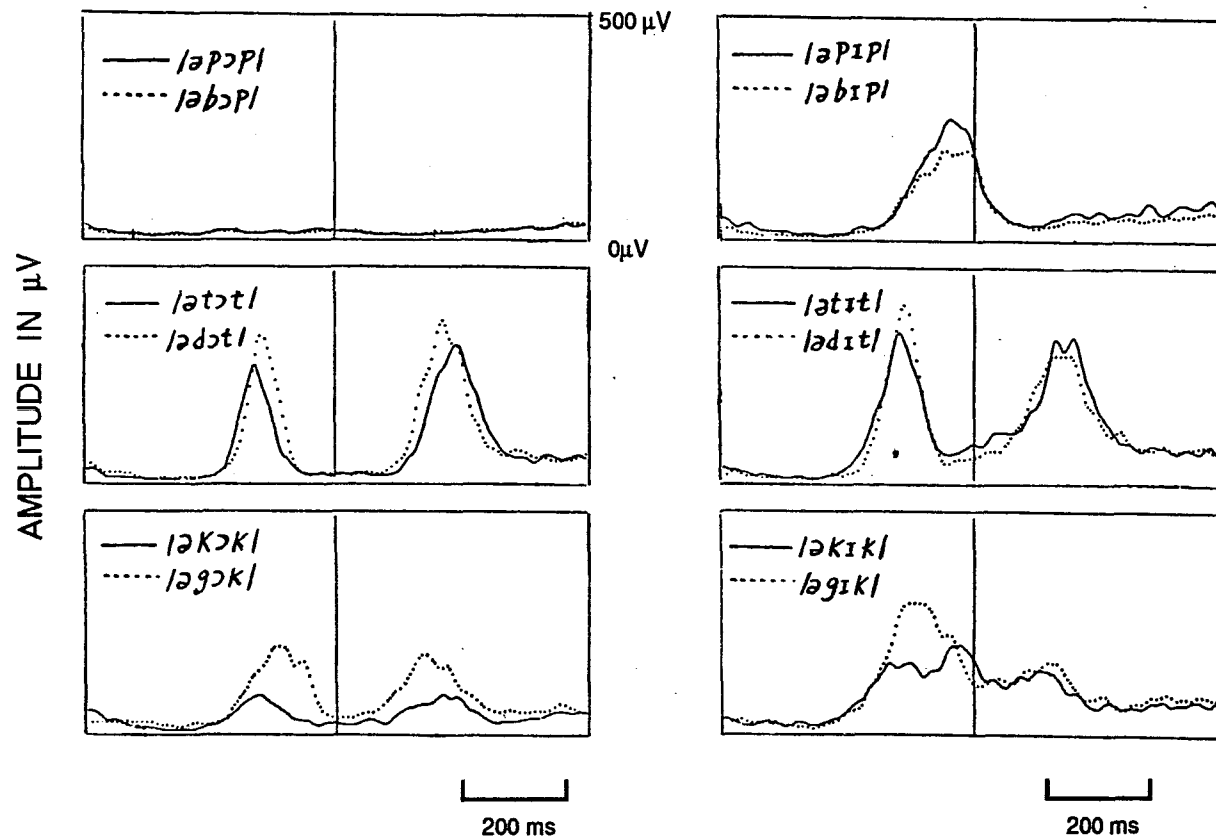


Figure 3.12 Averaged waveforms of the activity of the anterior genioglossus muscle for subject AS. Solid lines represent muscle activity for the voiceless stops and dotted lines activity for their voiced counterparts. See solid line corresponds to the acoustic lineup points.

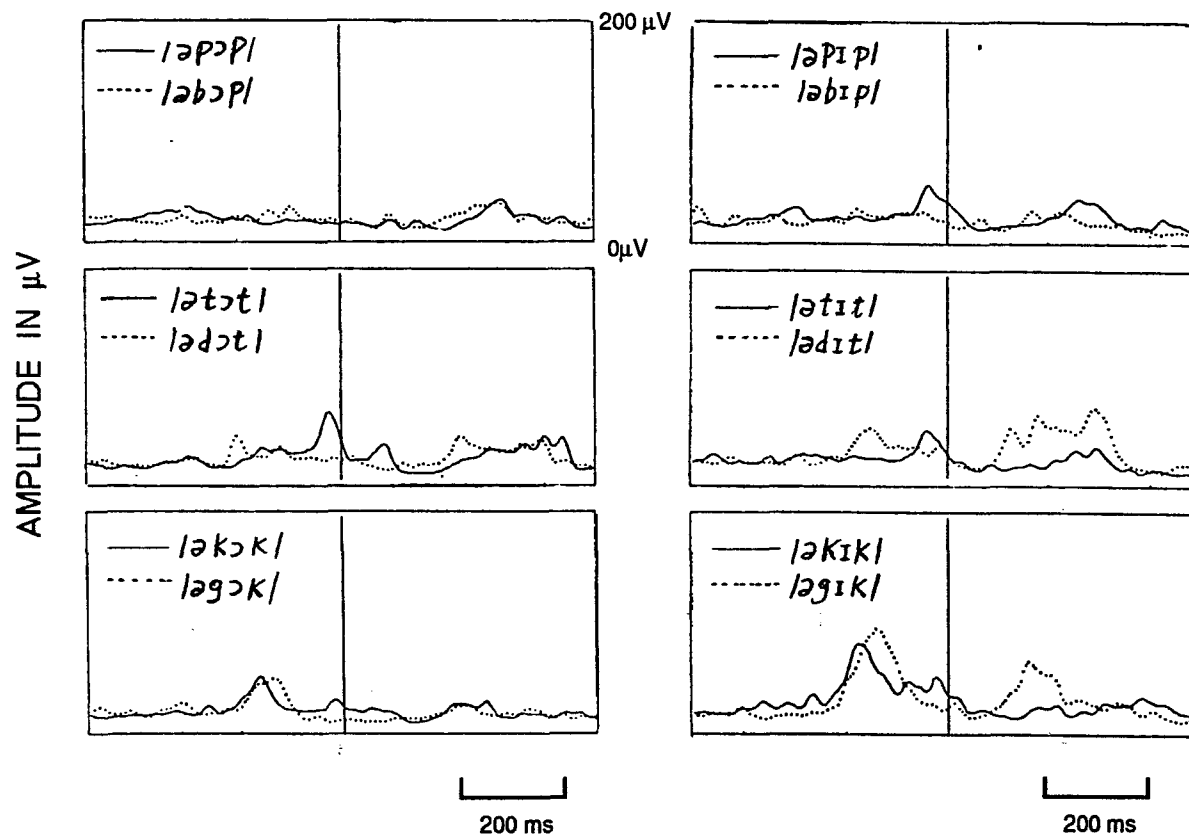


Figure 3.13 Averaged waveforms of the activity of the posterior genioglossus muscle for subject FB. Solid lines represent muscle activity for the voiceless stops and dotted lines activity for their voiced counterparts. See solid line corresponds to the acoustic lineup points.

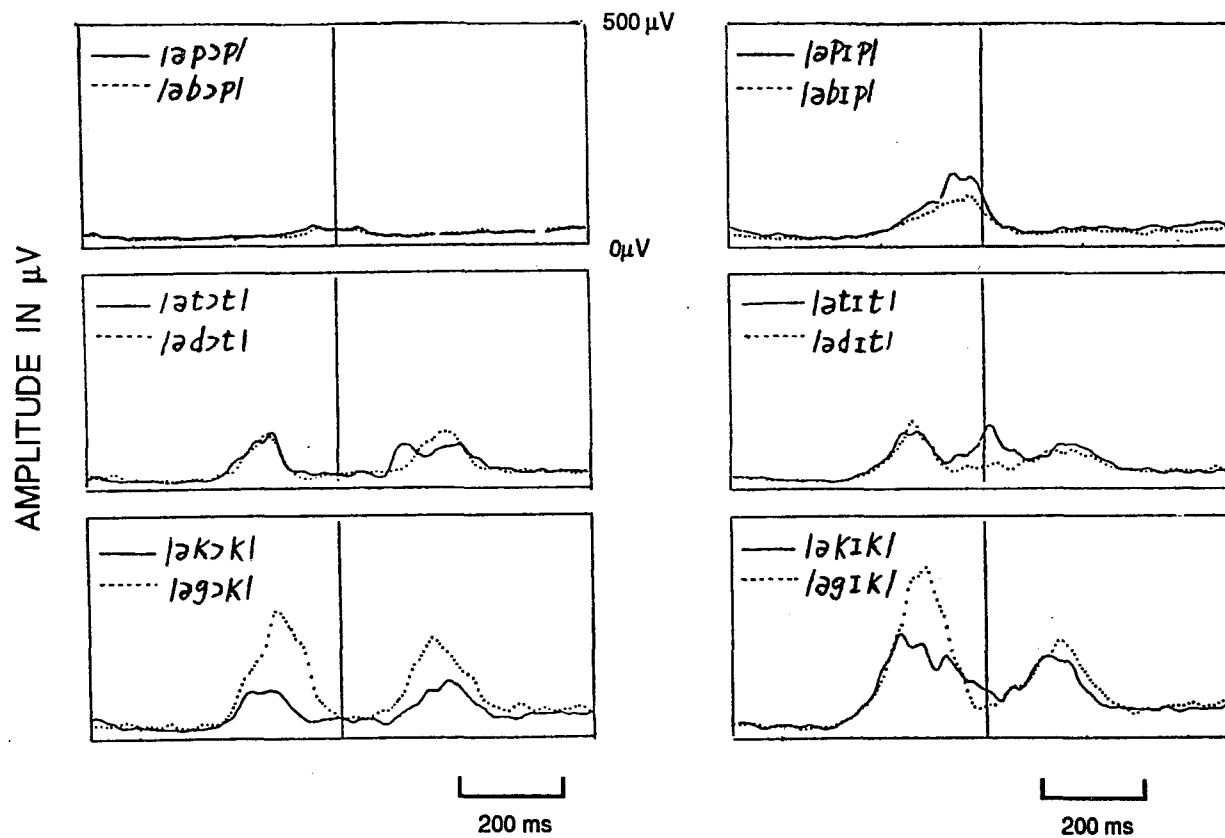


Figure 3.14 Averaged waveforms of the activity of the posterior genioglossus muscle for subject AS. Solid lines represent muscle activity for the voiceless stops and dotted lines activity for their voiced counterparts. See solid line corresponds to the acoustic lineup points.

Figure 3.14 shows averaged waveforms of the activity of the posterior genioglossus muscle for subject AS.

Significantly greater muscle activity is observed for /g/ than for /k/ in both vowels contexts, and also for /pI/ than for /bI/.

Table 3.2 gives the results of the analyses of variance for the EMG potentials of the anterior genioglossus muscle for all three subjects. A significant main effect was found for voicing status for FB ( $F(1,226) = 13.97, p < 0.01$ ), for AS ( $F(1,225) = 20.38, p < 0.01$ ), and for DB ( $F(1,227) = 5.10, p < 0.05$ ). For alveolar and velar stops, voiced stops always have greater EMG potentials than their voiceless cognates. For bilabial stops, voiceless stops have greater EMG potentials than their voiced cognates.

Table 3.3 gives the results of the analyses of variance for EMG potentials of the posterior genioglossus muscle for subjects FB and AS. (EMG recordings of this muscle were not usable for subject DB because of experimental error. There was a significant main effect for voicing status for AS ( $F(1,225) = 32.08, p < 0.01$ ). Similarly to the muscle activity pattern of the anterior genioglossus muscle, greater EMG potentials were observed for /d/ than for /t/ and for /g/ than for /k/. By contrast, for bilabial stops, greater EMG potentials were observed for /p/ than for /b/. For FB, no significant main effect of voicing status was observed.

Table 3.2  
Analysis of Variance:  
Effect of Anterior Genioglossus Muscle Amplitude (microV)

Subj	Source	Sum of Squares	Degrees of Freedom	Mean Square	F	Tail Prob.
FB	Voicing	52324.208	1	52324.208	13.97	0.0002**
	Place	1342858.572	2	671429.286	179.32	0.0000**
	Vowel	383284.835	1	383284.835	102.37	0.0000**
	VowxVoi	52453.316	1	52453.316	14.01	0.0002**
	PlaxVoi	151105.484	2	75552.742	20.18	0.0000**
	PlaxVow	136506.073	2	68253.036	18.23	0.0000**
	PlaxVowxVoi	38077.034	2	19038.517	5.08	0.0069**
	Error	846199.461	226	3744.245		
AS	Voicing	231868.497	1	231868.497	20.38	0.0000**
	Place	2506587.228	2	1253293.614	110.16	0.0000**
	Vowel	1748656.350	1	1748656.350	153.71	0.0000**
	VowxVoi	20306.196	1	20306.196	1.78	0.1829
	PlaxVoi	450224.778	2	225112.389	19.79	0.0000**
	PlaxVow	271662.818	2	135831.409	11.94	0.0000**
	PlaxVowxVoi	46948.198	2	23474.099	2.06	0.1294
	Error	2559717.749	225	11376.523		
DB	Voicing	87494.642	1	87494.642	5.10	0.0249*
	Place	541377.837	2	270688.919	15.78	0.0000**
	Vowel	91582.283	1	91582.283	5.34	0.0218*
	VowxVoi	15104.360	1	15104.360	0.88	0.3491
	PlaxVoi	163691.221	2	81845.610	4.77	0.0093**
	PlaxVow	4487.988	2	2243.994	0.13	0.8775
	PlaxVowxVoi	966.992	2	483.496	0.03	0.9722
	Error	3894884.063	227	17158.080		

\*\* p < 0.01

\* p < 0.05

Table 3.3  
Analysis of Variance:  
Effect of Posterior Genioglossus Muscle Amplitude (microV)

Subj	Source	Sum of Squares	Degree of Freedom	Mean Square	F	Tail Prob.
FB	Voicing	692.961	1	692.961	0.18	0.6731
	Place	62763.411	2	31381.706	8.08	0.0004**
	Vowel	3934.809	1	3934.809	1.01	0.3152
	VowxVoi	131.958	1	131.958	0.03	0.8539
	PlaxVoi	13207.928	2	6603.964	1.70	0.1849
	PlaxVow	31570.872	2	15785.436	4.06	0.0184*
	PlaxVowxVoi	2377.842	2	1188.921	0.31	0.7366
	Error		226			
AS	Voicing	231099.093	1	231099.093	32.08	0.0000**
	Place	1861204.365	2	930602.182	129.16	0.0000**
	Vowel	541641.867	1	541641.867	75.18	0.0000**
	VowxVoi	35715.261	1	25715.261	3.57	0.0601
	PlaxVoi	762529.779	2	381264.890	52.92	0.0000**
	PlaxVow	132833.944	2	66416.972	9.22	0.0001**
	PlaxVowxVoi	17853.415	2	8926.708	1.24	0.2916
	Error	1621090.841	225			

\*\* p < 0.01

\* p < 0.05

Table 3.4 gives the results of the analyses of variance for EMG potentials of the middle constrictor muscle for DB. A significant main effect was found for voicing status ( $F(1,227) = 46.75, p < 0.01$ ).

To summarize the results of the analysis of the EMG data, the anterior genioglossus muscle shows greater EMG potentials for voiced stops than for voiceless stops when the place of articulation of the stops is alveolar or velar for all three subjects. When the place of articulation is bilabial, however, voiceless stops have greater EMG potentials than voiced stops. For the posterior genioglossus muscle for subjects FB and AS, a pattern similar to that observed for the anterior genioglossus muscle was found. The middle constrictor muscle of subject DB shows greater EMG potentials for voiceless stops than for voiced stops, regardless of place of articulation.

### 3.3 Movement

The amount of vertical jaw displacement was calculated for each token by subtracting the value of vertical displacement of the electrode attached to the nose from the vertical displacement of the electrode attached to the jaw. This value will be referred to hereafter as "vertical jaw displacement."

Table 3.4  
Analysis of Variance:  
Effect of Middle Pharyngeal Constrictor Muscle Amplitude (microV) for DB

Source	Sum of Squares	Degree of Freedom	Mean Square	F	Tail Prob.
Voicing	131027.469	1	131027.469	46.75	0.0000**
Place	298185.079	2	149192.540	53.20	0.0000**
Vowel	22937.038	1	22937.038	8.18	0.0046**
VowxVoi	3502.192	1	3502.192	1.25	0.2648
PlaxVoi	50136.671	2	25068.336	8.94	0.0002**
PlaxVow	16069.753	2	8034.877	2.87	0.0589
PlaxVowxVoi	1669.892	2	834.946	0.30	0.7427
Error	636213.271	227	2802.702		

\*\* p < 0.01

Figure 3.15 shows averaged waveforms of vertical jaw displacement for subject FB. Solid lines represent jaw movement for voiceless stops and dotted lines represent jaw movement for voiced stops. Solid vertical lines correspond to the acoustic lineup points. Jaw displacement was measured between peak jaw displacement for the first consonant and the relatively flat line which corresponds to the jaw position for the initial /ə/ before the jaw starts closing for the first consonant. It is observed in Fig. 3.15 that voiceless stops tend to be associated with greater jaw displacements than their voiced cognates. This difference, however, is statistically significant only for the /kI/ - /gI/ pair ( $p < 0.01$ ).

Figure 3.16 gives averaged waveforms of vertical jaw displacement for subject AS. Subject AS, like subject FB, tends to exhibit greater jaw displacements for voiceless stops, as compared to their voiced counterparts. This difference is statistically significant for three voicing pairs: /t/ - /d/ ( $p < 0.01$ ), /tI/ - /dI/ ( $p < 0.01$ ), and /kI/ - /gI/ ( $p < 0.05$ ).

Figure 3.17 shows averaged waveforms of vertical jaw displacement for subject DB. Similarly to the other two subjects, most voiceless stops tend to be associated with greater jaw displacement than their voiced cognates except for the voicing pairs of /p/ - /b/ and /kI/ - /gI/. This difference was statistically significant only for the pair /t/ - /d/.

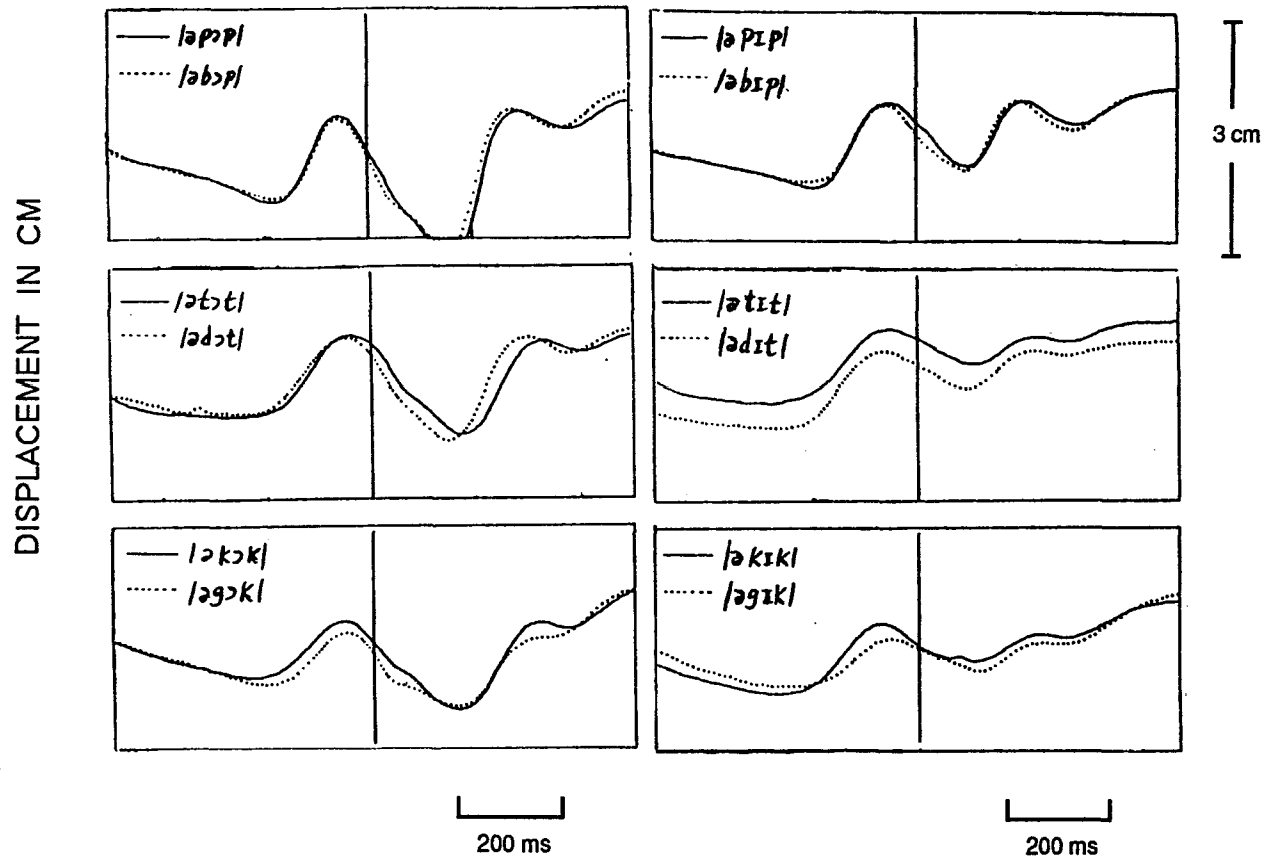


Figure 3.15 Averaged waveforms of vertical jaw displacement for Subject FB. Upward movement of the jaw is displayed as upward displacement of the waveform. Solid lines represent the movements for the voiceless stops and the dotted lines those for their voiced counterparts.

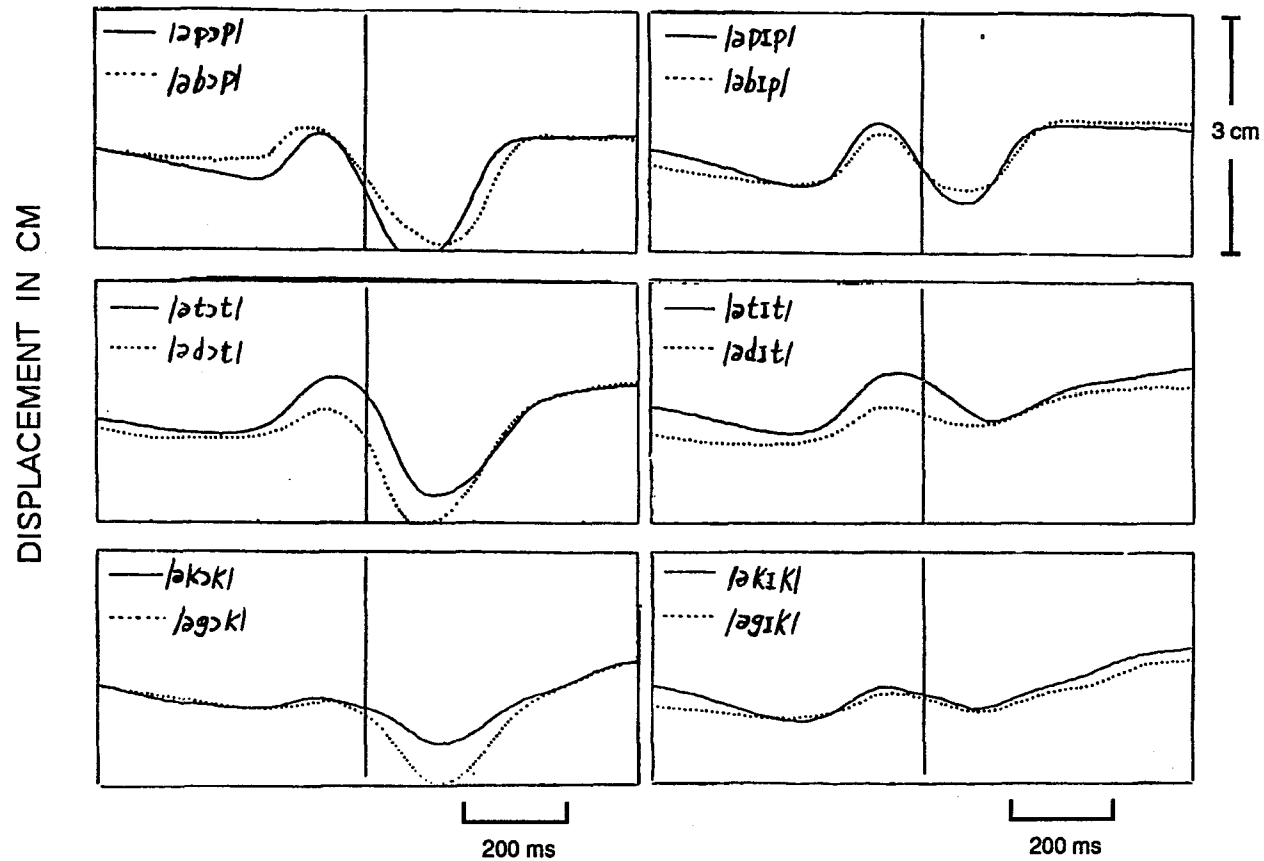


Figure 3.16 Averaged waveforms of vertical jaw displacement for Subject AS. Upward movement of the jaw is displayed as upward displacement of the waveform. Solid lines represent the movements for the voiceless stops and the dotted lines those for their voiced counterparts.

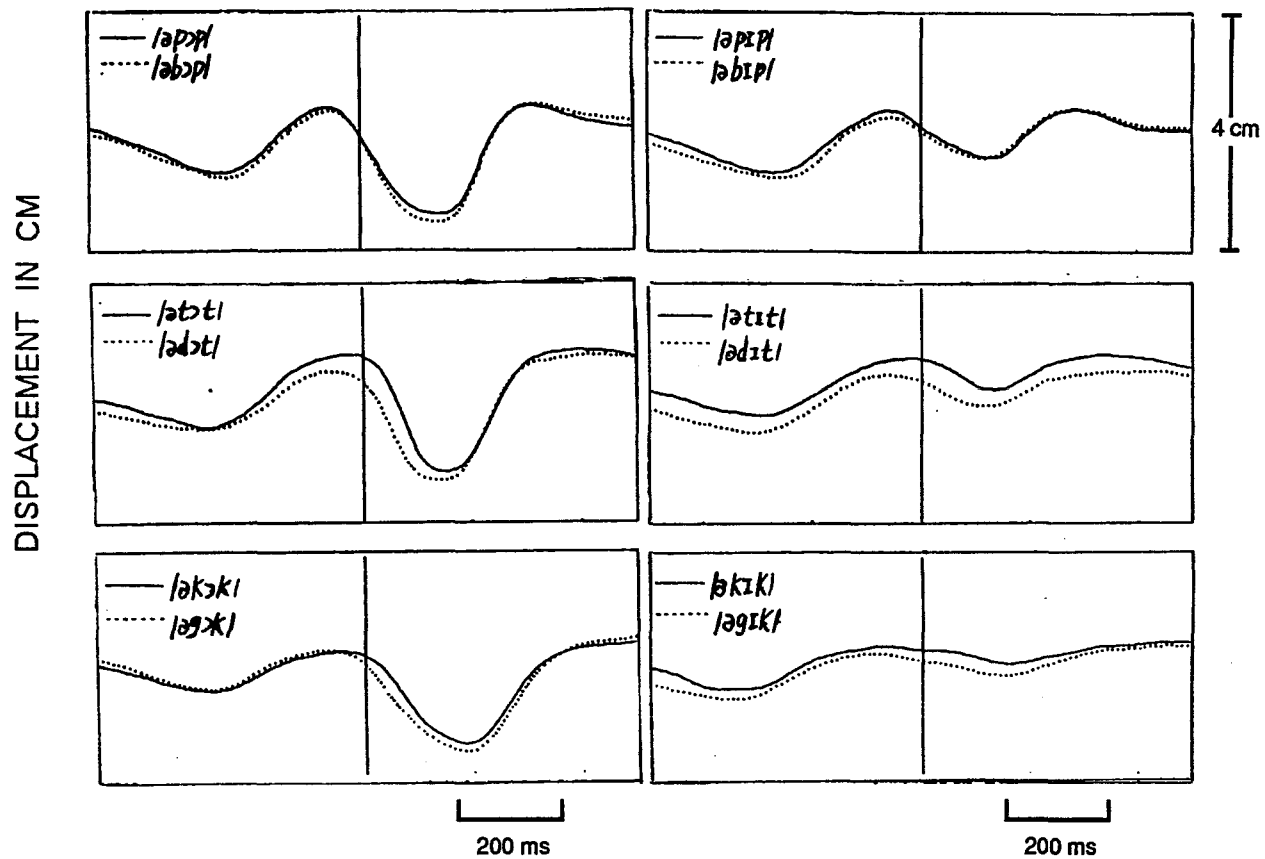


Figure 3.17 Averaged waveforms of vertical jaw displacement for Subject DB. Upward movement of the jaw is displayed as upward displacement of the waveform. Solid lines represent the movements for the voiceless stops and the dotted lines those for their voiced counterparts.

The results of the analyses of variance with jaw displacement as the dependent variable are given in Table 3.5 for all three subjects. A significant main effect was found for voicing status for FB ( $F(1,226) = 17.03, P < 0.01$ ) and AS ( $F(1,225) = 20.83, p < 0.01$ ). No significant main effect for voicing status was found for DB.

Values for the vertical displacement of the lower lip independent of jaw movement were also analyzed for bilabial stops. Lower lip displacement was calculated by subtracting vertical jaw displacement, as defined above, from the values of vertical lower lip displacement recorded from the electrode attached to the midline of the lower lip. The values considered here indicate the amount of upward displacement of the lower lip starting from a relatively flat line, which corresponds to the lower lip position for /ə/ before the lineup points.

Figure 3.18 shows averaged waveforms of vertical displacement of the lower lip for subject FB. The solid line represents lip movement for the voiceless stops and the dotted line represents lip movement for the voiced stops. For the pair of /p/ - /b/, /p/ has significantly greater lip displacement than /b/ ( $p < 0.05$ ). No significant difference between /pI/ and /bI/ was found.

Averaged waveforms of vertical lower lip displacement for AS is shown in Fig. 3.19. In contrast to FB, the voiced stop /b/ has significantly greater lip displacement than the voiceless stop /p/ in both vowel contexts. ( $p < 0.01$ ).

Table 3.5  
Analysis of Variance:  
Effect of Jaw Vertical Displacement (cm)

Subj	Source	Sum of Squares	Degree of Freedom	Mean Square	F	Tail Prob.
FB	Voicing	112.650	1	112.650	17.03	0.0001**
	Place	533.946	2	266.973	40.35	0.0000**
	Vowel	19.748	1	19.748	2.98	0.0854
	VowxVoi	0.060	1	0.060	0.01	0.9241
	PlaxVoi	37.599	2	18.800	2.84	0.0604
	PlaxVow	20.883	2	10.442	1.58	0.2086
	PlaxVowxVoi	29.785	2	14.893	2.25	0.1076
	Error	1495.148	226	6.616		
AS	Voicing	70.848	1	70.848	20.83	0.0000**
	Place	307.713	2	153.856	45.24	0.0000**
	Vowel	75.302	1	75.302	22.14	0.0000**
	VowxVoi	2.949	1	2.949	0.87	0.3527
	PlaxVoi	31.645	2	15.822	4.65	0.0105*
	PlaxVow	18.422	2	9.211	2.71	0.0688
	PlaxVowxVoi	4.351	2	2.176	0.64	0.5284
	Error	765.202	225	3.401		
DB	Voicing	14.065	1	14.065	3.17	0.0765
	Place	480.302	2	240.151	54.08	0.0000**
	Vowel	1.589	1	1.589	0.36	0.5503
	VowxVoi	11.641	1	11.641	2.62	0.1068
	PlaxVoi	31.198	2	15.599	3.51	0.0314*
	PlaxVow	57.586	2	28.793	6.48	0.0018**
	PlaxVowxVoi	25.740	2	12.870	2.90	0.0572
	Error	1008.016	227	4.441		

\*\* p < 0.01

\* p < 0.05

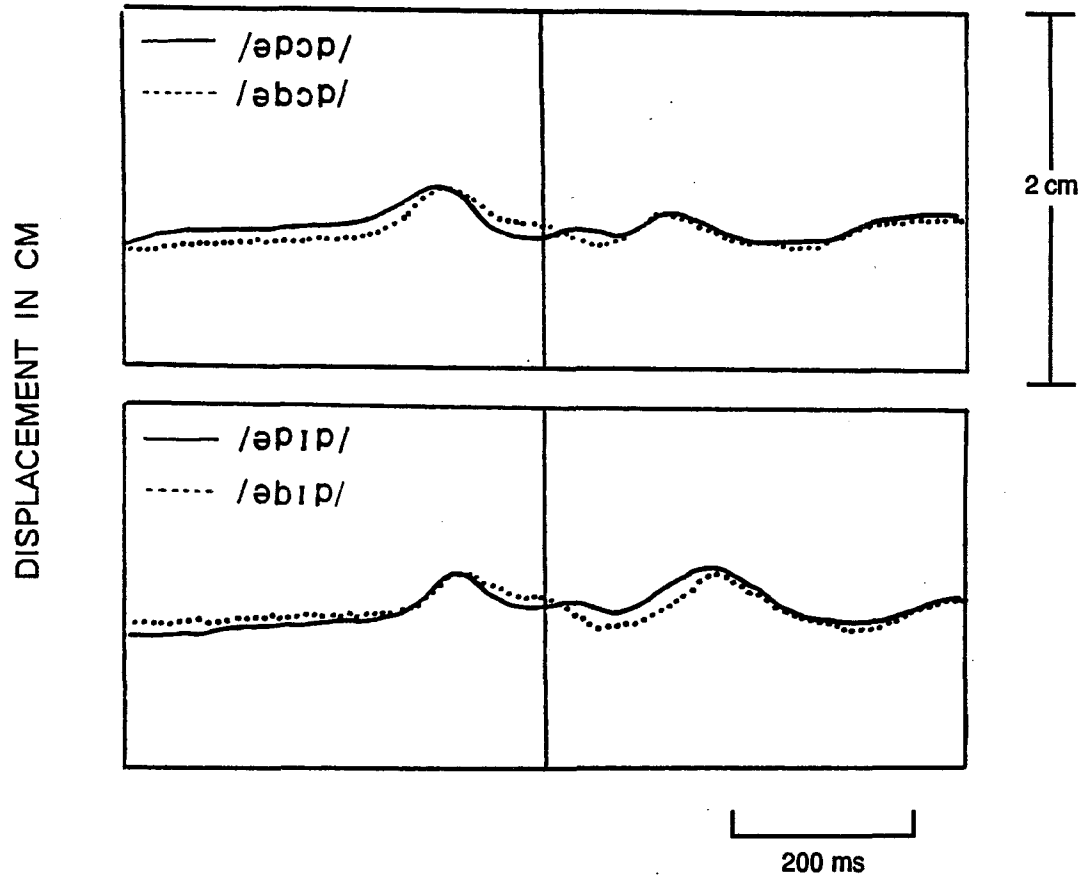


Figure 3.18 Averaged waveforms of vertical displacement of the lower lip for Subject FB. Upward movement of the lip is displayed as upward displacement of the waveform. Solid lines represent the movements for the voiceless stops and dotted lines those for their voiced counterparts.

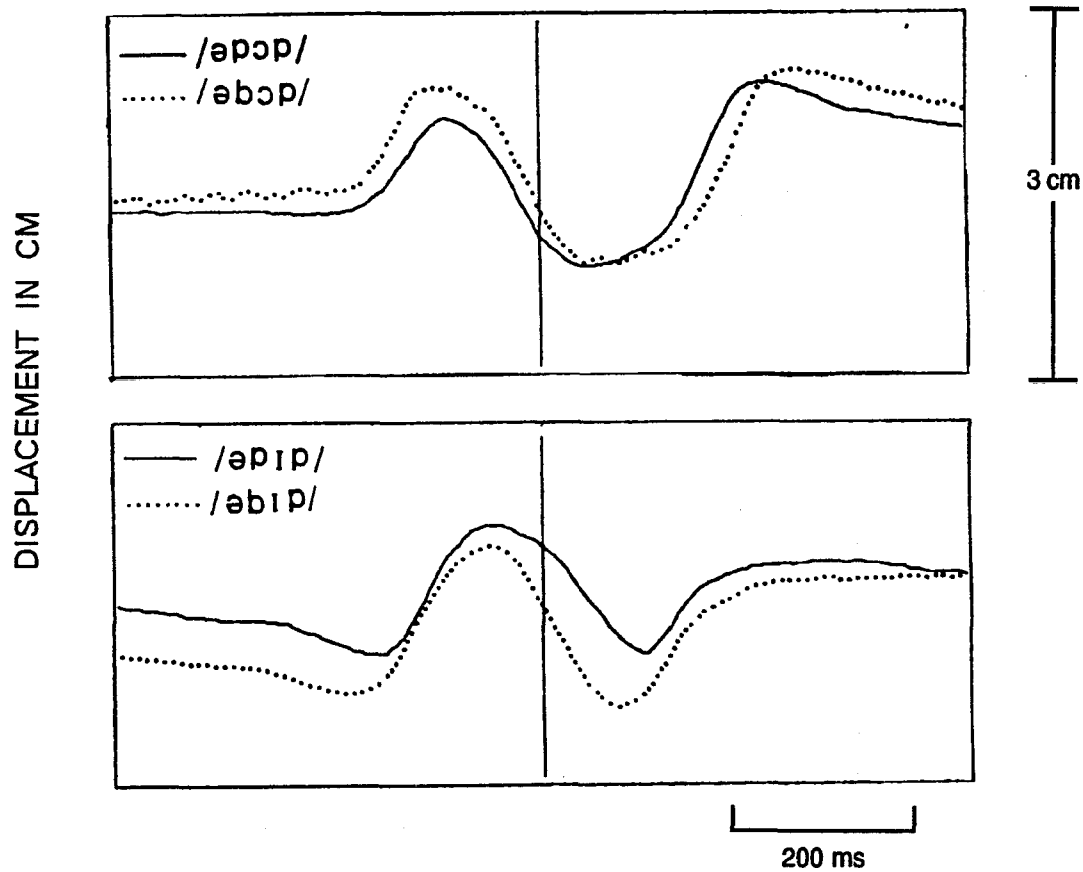


Figure 3.19 Averaged waveforms of vertical displacement of the lower lip for Subject AS. Upward movement of the lip is displayed as upward displacement of the waveform. Solid lines represent the movements for the voiceless stops and dotted lines those for their voiced counterparts.

Figure 3.20 shows averaged waveforms of vertical lower lip displacement for subject DB. Similarly to FB, the voiceless stop /p/ is associated with significantly greater lip displacement than the voiced stop /b/ ( $p < 0.01$ ).

The results of the analyses of variance with lower lip displacement as the dependent variable are given in Table 3.6 for all three subjects. A significant main effect of voicing status was found for AS ( $F(1,76) = 38.89, p < 0.01$ ) and DB ( $F(1,75) = 31.54, p < 0.01$ ), but not for FB.

Values for the horizontal or frontal displacement of the lower lip independent of the jaw were also calculated. These values were obtained by subtracting the real (i.e. independent of head movement) horizontal jaw movement from the horizontal lower lip displacement. The values indicate the amount of frontal displacement of the lower lip for the production of bilabial stops. The measured distance is between the peak of the waveform and the relatively flat line before the lineup point, which corresponds to the horizontal lower lip position for the initial /ə/.

Figure 3.21 shows averaged waveforms of horizontal displacement of the lower lip for subject FB. The lower lip has significantly greater frontal displacement for /p/ than for /b/ in both vowel contexts ( $p < 0.01$ ).

Figure 3.22 shows averaged waveforms of horizontal lower lip displacement for subject AS. The lower lip tends to show greater frontal displacement for /p/ than for /b/, but this difference is not statistically significant.

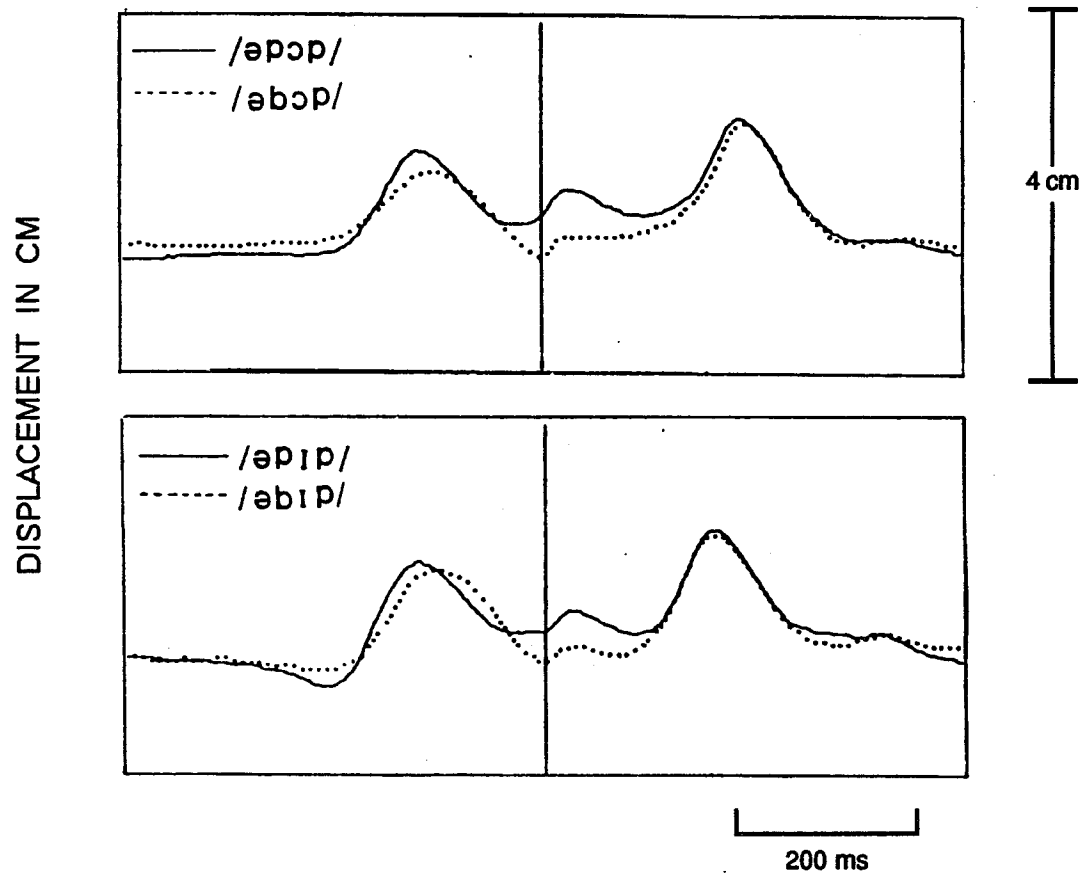


Figure 3.20 Averaged waveforms of vertical displacement of the lower lip for Subject DB. Upward movement of the lip is displayed as upward displacement of the waveform. Solid lines represent the movements for the voiceless stops and dotted lines those for their voiced counterparts.

Table 3.6  
Analysis of Variance:  
Effect of Lower Lip Vertical Displacement (cm)

Subj	Source	Sum of Square	Degree of Freedom	Mean Square	F	Tail Prob.
FB	Voicing	1.458	1	1.458	2.60	0.1111
	Vowel	0.265	1	0.265	0.47	0.4945
	VoixVow	4.050	1	4.050	7.22	0.0089**
	Error	42.647	76	0.561		
AS	Voicing	10.011	1	10.011	38.89	0.0000**
	Vowel	0.406	1	0.406	1.58	0.2130
	VoixVow	0.091	1	0.091	0.35	0.5536
	Error	19.566	76	0.257		
DB	Voicing	10.848	1	10.848	31.54	0.0000**
	Vowel	3.672	1	3.672	10.68	0.0016**
	VoixVow	0.014	1	0.014	0.04	0.8425
	Error	25.794	75	0.344		

\*\* p < 0.01

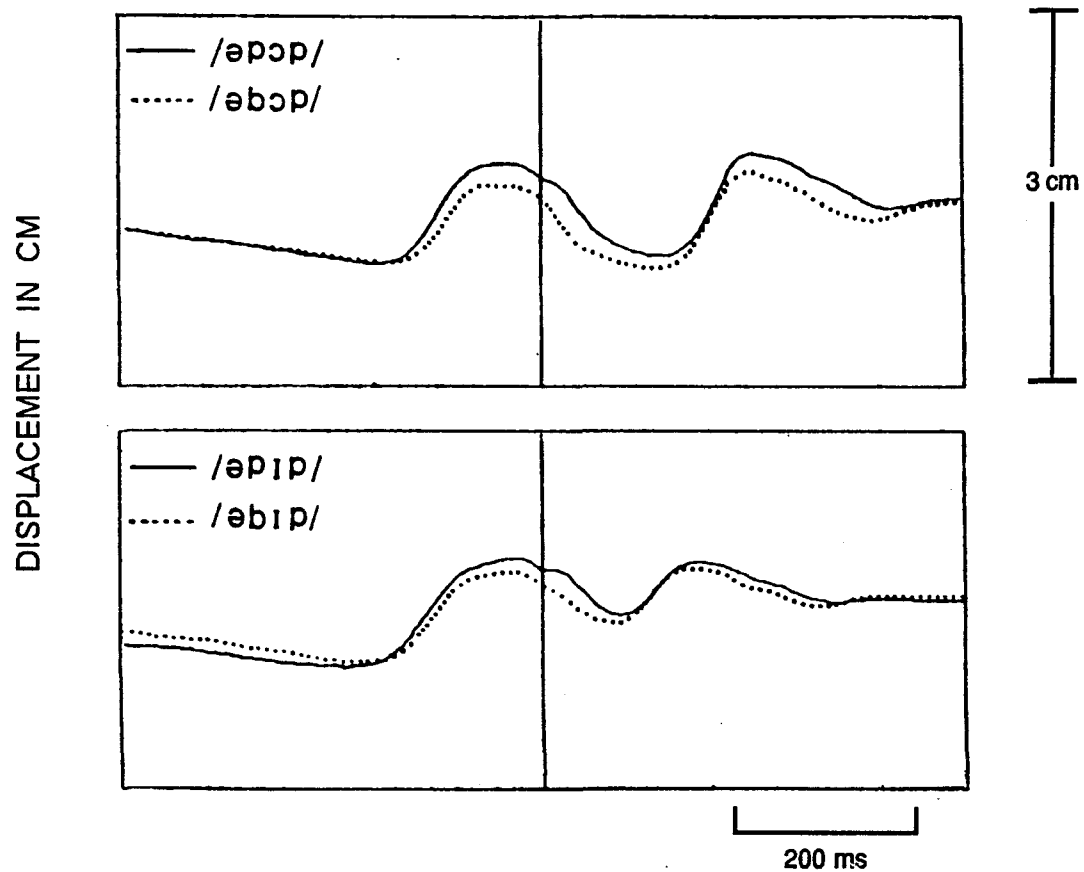


Figure 3.21 Averaged waveforms of horizontal displacement of the lower lip for Subject FB. Forward movement of the lip is displayed as upward displacement of the waveform. Solid lines represent the movements for the voiceless stops and dotted lines those for their voiced counterparts.

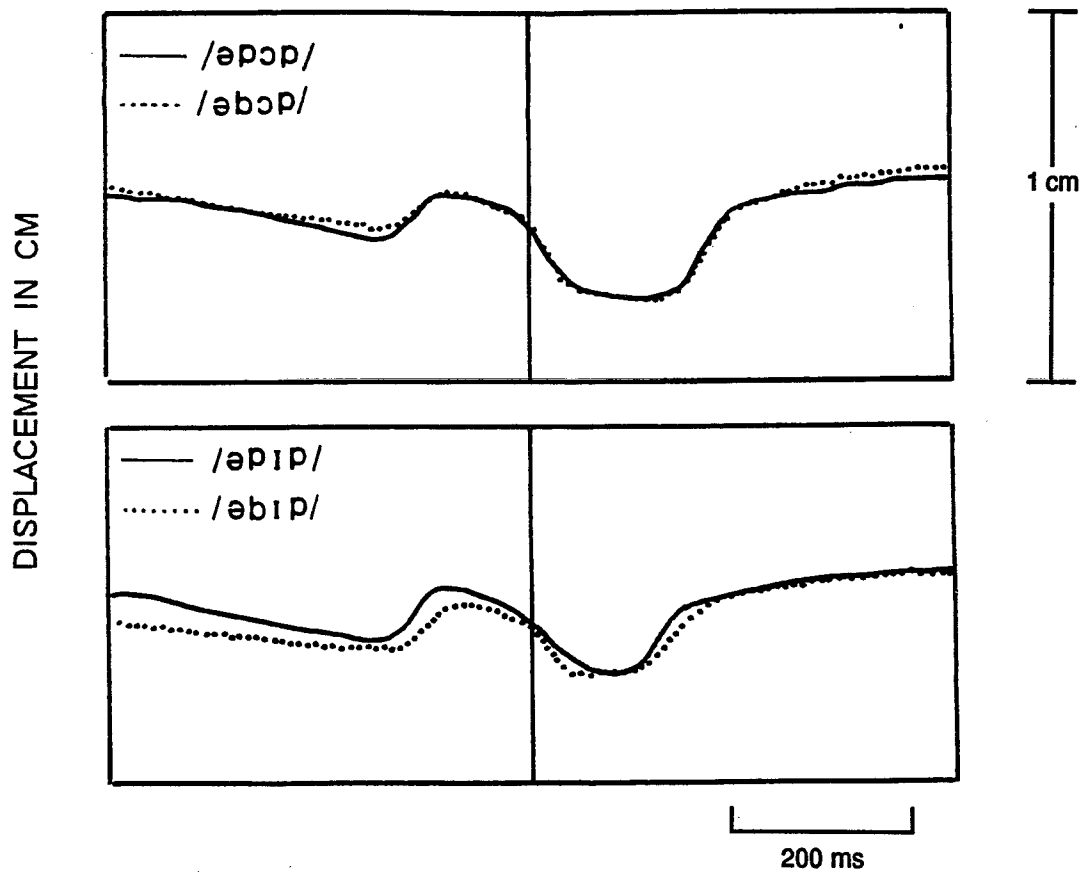


Figure 3.22 Averaged waveforms of horizontal displacement of the lower lip for Subject AS. Forward movement of the lip is displayed as upward displacement of the waveform. Solid lines represent the movements for the voiceless stops and dotted lines those for their voiced counterparts.

Figure 3.23 shows averaged waveforms of horizontal displacement of the lower lip for subject DB. Similarly to FB, the lower lip has significantly greater frontal displacement for /p/ than for /b/ in both vowel contexts ( $p < 0.01$ ).

Table 3.7 gives the results of the analyses of variance with horizontal lower lip displacement as the dependent variable for all three subjects. A significant main effect of voicing status was found for FB ( $F(1,76) = 89.62$ , ( $p < 0.01$ ) and for DB ( $F(1,75) = 149.54$ ,  $p < 0.01$ ), but not for AS. To summarize the results for the movement data, voiceless stops tend to have greater vertical jaw displacement than their voiced counterparts, although not all possible comparisons are statistically significant. Exceptions to this pattern are the /p/ - /b/ and /k/ - /g/ pairs for subject DB. For vertical lower lip displacement, the results are not consistent. Voiceless stops have greater displacement than their voiced counterparts for FB and DB, but the opposite pattern is observed for AS. For horizontal lower lip displacement, all three subjects exhibit greater displacements for voiceless stops than their voiced cognates, with these differences reaching statistical significance for subjects FB and DB.

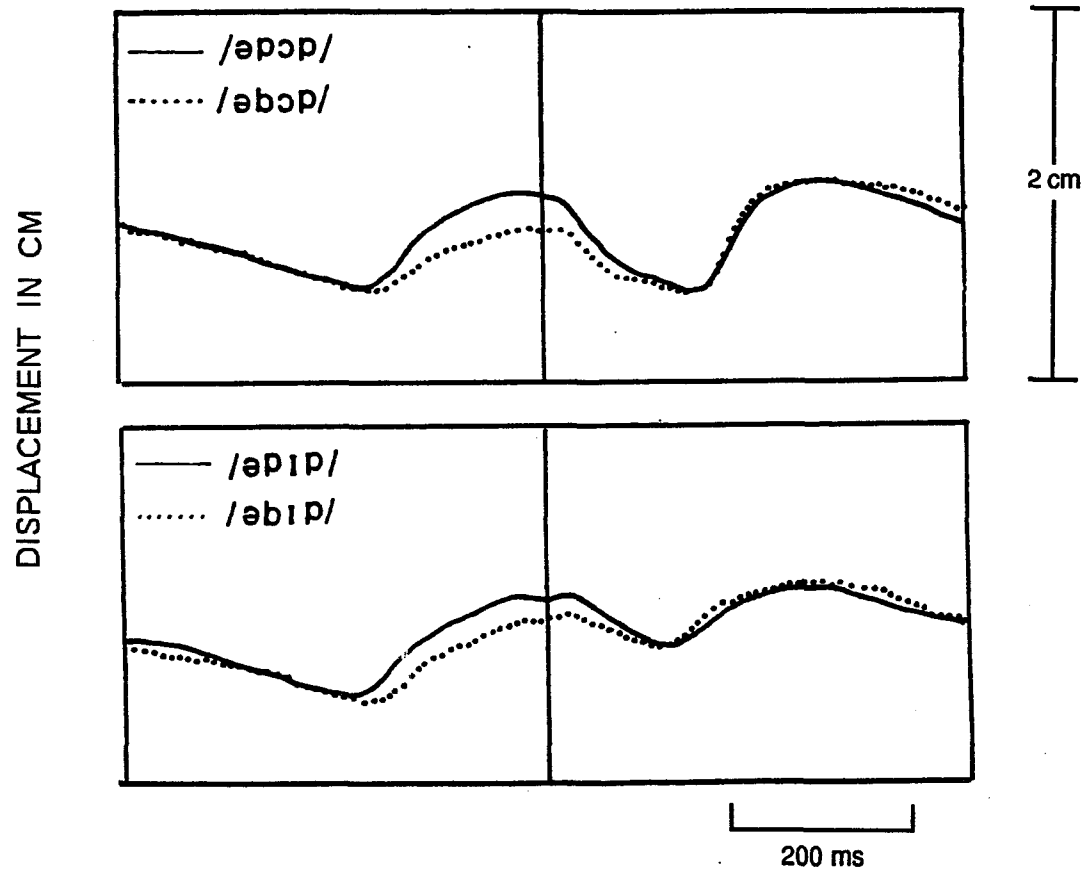


Figure 3.23 Averaged waveforms of horizontal displacement of the lower lip for Subject DB. Forward movement of the lip is displayed as upward displacement of the waveform. Solid lines represent the movements for the voiceless stops and dotted lines those for their voiced counterparts.

Table 3.7  
Analysis of Variance:  
Effect of Lower Lip Horizontal Displacement (cm)

Subj	Source	Sum of Squares	Degree of Freedom	Mean Square	F	Tail Prob.
FB	Voicing	86.528	1	86.528	89.62	0.0000**
	Vowel	28.085	1	28.085	29.09	0.0000**
	VoixVow	0.145	1	0.145	0.15	0.6999
	Error	73.375	76	0.965		
AS	Voicing	1.301	1	1.301	2.81	0.0980
	Vowel	2.888	1	2.888	6.23	0.0147*
	VoixVow	0.145	1	0.145	0.31	0.5782
	Error	35.255	76	0.463		
DB	Voicing	63.613	1	63.613	149.54	0.0000**
	Vowel	4.938	1	4.938	11.61	0.0011**
	VoixVow	3.316	1	3.316	7.80	0.0066**
	Error	31.905	75	0.425		

\*\* p < 0.01

\* p < 0.05

### 3.4 Acoustic Results

#### 3.4.1 Consonant Closure Duration

Consonant closure duration of the target consonants (i.e., the first consonant of the /əCVC/ sequence) was measured from the acoustic waveform. The results of the analysis of closure durations were not straightforward in that the durational differences were observed only for some voicing pairs in each subject and also in that each subject exhibited durational differences for voicing pairs different from the other subjects. Table 3.8 gives the results of the analyses of variance with the closure duration of the first consonant as the dependent variable for all three subjects. A significant main effect of voicing status was found for FB ( $F(1,226) = 14.48, p < 0.01$ ), for AS ( $F(1,225) = 60.74, p < 0.01$ ), and for DB ( $F(1,227) = 7.34, p < 0.01$ ). For FB, the mean closure duration is longer for voiceless stops than for voiced stops in the following pairs: /p/ - /b/ (/p/ = 110.15 ms, /b/ = 99.31 ms,  $p < 0.05$ ); /k/ - /g/ (/k/ = 113.17 ms, /g/ = 97.94 ms,  $p < 0.01$ ); and /kI/ - /gI/ (/kI/ = 125.52 ms, /gI/ = 107.50 ms,  $p < 0.01$ ). For AS, similar results were found for three voicing pairs: /t/ - /d/ (/t/ = 122.29 ms, /d/ = 102.33 ms,  $p < 0.01$ ), /k/ - /g/ (/k/ = 111.74 ms, /g/ = 85.30 ms,  $p < 0.01$ ) and /kI/ - /gI/ (/kI/ = 130.43 ms, /gI/ = 83.20 ms,  $p < 0.01$ ). For DB, there is a tendency for voiceless stops to have longer closure durations than their voiced counterparts, with this

Table 3.8  
Analysis of Variance:  
Effect of Consonantal Closure Duration (ms)

Subj	Source	Sum of Squares	Degree of Freedom	Mean Square	F	Tail Prob.
FB	Voicing	3790.705	1	3790.705	14.48	0.0002**
	Place	3816.140	2	1908.070	7.29	0.0009**
	Vowel	16361.987	1	16361.987	62.51	0.0000**
	VowxVoi	0.004	1	0.004	0.00	0.9969
	PlaxVoi	3126.309	2	1563.155	5.97	0.0030**
	PlaxVow	944.579	2	472.289	1.80	0.1670
	PlaxVowxVoi	164.921	2	82.461	0.32	0.7301
	Error	59156.850	226	261.756		
AS	Voicing	22480.932	1	22480.932	60.74	0.0000**
	Place	26663.285	2	13331.643	36.02	0.0000**
	Vowel	934.012	1	934.012	2.52	0.1136
	VowxVoi	349.604	1	349.604	0.94	0.3321
	PlaxVoi	9124.242	2	4562.121	12.33	0.0000**
	PlaxVow	574.373	2	287.187	0.78	0.4615
	PlaxVowxVoi	2871.940	2	1435.970	3.88	0.0220*
	Error	83274.279	225	370.108		
DB	Voicing	2684.834	1	2684.834	7.34	0.0073**
	Place	4799.003	2	2399.501	6.56	0.0017**
	Vowel	3344.442	1	3344.442	9.15	0.0028**
	VowxVoi	100.080	1	100.080	0.27	0.6014
	PlaxVoi	2020.156	2	1010.078	2.76	0.0653
	PlaxVow	1874.048	2	937.024	2.56	0.0794
	PlaxVowxVoi	1074.975	2	537.488	1.47	0.2322
	Error	83013.849	227	365.700		

\*\* p < 0.01

\* p < 0.05

difference reaching statistical significance only for /pI/ - /bI/ (/pI/ = 174.68 ms, /bI/ = 157.60 ms,  $p < 0.01$ ).

### 3.4.2 Vowel Duration

The durations of the vowels following the initial consonants were measured from the acoustic waveforms. Table 3.9 shows the results of the analyses of variance with vowel duration as the dependent variable for all three subjects. A significant main effect of voicing status was found for FB ( $F(1,226) = 18.99$ ,  $p < 0.01$ ), but not for AS and DB. For AS, however, the tail probability has a marginal status ( $F(1,225) = 3.52$ ,  $p < 0.062$ ). For FB, the mean vowel duration is significantly longer for voiceless stops than for voiced stops for the following voicing pairs: /p>/ - /b>/ (/p>/ = 231.29 ms, /b>/ = 220.75 ms,  $p < 0.01$ ); /t>/ - /d>/ (/t>/ = 257.76 ms, /d>/ = 234.37 ms,  $p < 0.01$ ); and /tI/ - /dI/ (/tI/ = 164.52 ms, /dI/ = 153.97 ms,  $p < 0.01$ ). For AS, the opposite pattern to FB was observed, with differences reaching statistical significance for the following voicing pairs: /pI/ - /bI/ (/pI/ = 158.47 ms, /bI/ = 179.37 ms,  $p < 0.01$ ) and /kI/ - /gI/ (/kI/ = 176.61 ms, /gI/ = 197.07 ms,  $p < 0.01$ ). For DB, a significant difference in mean vowel duration was found only for the /k>/ - /g>/ pair (/k>/ = 277.68 ms, /g>/ = 291.68 ms,  $p < 0.05$ ).

**Table 3.9**  
**Analysis of Variance:**  
**Effect of Vowel Duration (ms)**

Subj	Source	Sum of Squares	Degree of Freedom	Mean Square	F	Tail Prob.
FB	Voicing	3518.227	1	3518.227	18.99	0.0000**
	Place	14891.674	2	7445.837	40.20	0.0000**
	Vowel	372193.655	1	372193.655	2009.43	0.0000**
	VowxVoi	1023.718	1	1023.718	5.53	0.0196*
	PlaxVoi	2958.359	2	1479.180	7.99	0.0004**
	PlaxVow	1862.369	2	931.184	5.03	0.0073**
	PlaxVowxVoi	268.623	2	134.312	0.73	0.4854
	Error	41860.554	226	185.224		
AS	Voicing	1602.222	1	1602.222	3.52	0.0620
	Place	56092.045	2	28046.023	61.57	0.0000**
	Vowel	214236.179	1	214236.179	470.29	0.0000**
	VowxVoi	1562.107	1	1562.107	3.43	0.0654
	PlaxVoi	6904.296	2	3452.148	7.58	0.0007**
	PlaxVow	1785.539	2	892.770	1.96	0.1433
	PlaxVowxVoi	881.939	2	440.969	0.97	0.3814
	Error	102496.672	225	455.541		
DB	Voicing	4.332	1	4.332	0.02	0.8886
	Place	80629.031	2	40314.515	184.85	0.0000**
	Vowel	143933.074	1	143933.074	659.97	0.0000**
	VowxVoi	645.679	1	645.679	2.96	0.0867
	PlaxVoi	759.521	2	379.760	1.74	0.1776
	PlaxVow	1719.862	2	859.931	3.94	0.0207*
	PlaxVowxVoi	1180.824	2	590.412	2.71	0.06
	Error	49506.554	227	218.091		

\*\* p < 0.01

\* p < 0.05

### 3.5 Perception

#### 3.5.1 Intelligibility

Intelligibility scores and confusion matrices were obtained by counting the responses of the listeners which were in agreement with what the alaryngeal talkers intended to produce. There were 19-20 tokens for one utterance type and seven listeners heard all tokens. Tokens with the same consonant and different vowels (e.g., /pɔ/ and /pI/) were grouped together. There were, therefore, either 273 responses [(19 tokens + 20 tokens) x 7 listeners] or 280 responses [(20 tokens + 20 tokens) x 7 listeners]. Table 3.10 shows the intelligibility scores for the alaryngeal talkers. The voiced consonants were perceived as voiced slightly above chance and only a quarter of the voiceless consonants were perceived as voiced. In AS, the voiced consonants had high intelligibility compared to the voiceless ones. DB showed better intelligibility for voiced consonants than for voiceless ones, although both were close to the chance level. In all three talkers, it was found that voiceless consonants had very low intelligibility. Table 3.11 gives the percentage of the responses that were labeled either as voiced or as voiceless in FB, AS, and DB. There was a general tendency for many consonants to be heard as voiced and fewer consonants to be heard as voiceless.

The confusion matrix of FB (Table 3.12) revealed that voiced consonants were perceived with high percentages of

Table 3.10

Intelligibility scores for three alaryngeal talkers in percent correct. Each talker intended to produce 120 tokens for both voiced and voiceless consonants.

	<u>Subjects</u>		
	<u>FB</u>	<u>AS</u>	<u>DB</u>
Voiced	58%	76%	63%
Voiceless	23%	18%	54%

Table 3.11

Overall probability of a "voiced" response, in percent, for each speaker.

	<u>Subjects</u>		
	<u>FB</u>	<u>AS</u>	<u>DB</u>
Heard as voiced	62%	68%	55%
Heard as voiceless	38%	32%	45%

Table 3.12

Confusion matrix for FB. Correct responses of the seven listeners are represented as percent.

Produced as	Perceived as					
	p	b	t	d	k	g
p	36	64				
b	31	69				
t			60	40		
d			41	59		
k					36	64
g					25	75

correct responses; 69% for /b/, 59% for /d/, and 75% for /g/, whereas for voiceless consonants, listeners perceived them as voiceless for 36% of /p/ and /k/ and 60% of /t/. For subject AS, a similar pattern was found in the confusion matrix (Table 3.13). The voiced consonants were perceived correctly more often than the voiceless consonants. Among the voiceless consonants, the listeners perceived /k/ rather correctly (56%) compared to the scores for the other consonants, /p/ (34%) and /t/ (31%). The confusion matrix of DB (Table 3.14) had a slightly different pattern from FB and AS. The voiceless consonants were perceived more correctly than those of FB and AS; /t/ and /k/ had the percentages of 66% and 71%, respectively. The voiced consonants had better scores than voiceless consonants except for /g/, 40% of which were perceived as /k/. The confusion matrices showed a general tendency for the voiced consonants of the alaryngeal talkers' speech to be perceived correctly more than 75% of the time. The voiceless consonants, on the other hand, were correctly perceived less frequently than the voiced consonants, and their intelligibility was variable among the consonant types and among the talkers.

### 3.5.2 Relationships between speech perception and production

In discriminant analyses, each token was classified into one of three groups; perceptually correct, perceptually

Table 3.13

Confusion matrix for AS. Correct responses of the seven listeners are represented as percent.

Produced as	Perceived as					
	p	b	t	d	k	g
p	34	66				
b	24	76				
t	0.4	0.4	31	68.2		
d			12	88		
k					56	44
g					34	66

Table 3.14

Confusion matrix for DB. Correct responses of the seven listeners are represented as percent.

Produced as	Perceived as					
	p	b	t	d	k	g
p	42	58				
b	21	79				
t			66	34		
d			25	75		
k			5	1	71	23
g				2	38	60

ambiguous, and perceptually incorrect. The physiological and acoustic values among the groups were examined using three-way analyses for the voicing contrast. Table 3.15 gives the results of the analysis. The results indicated that the differences in intraoral air pressure were an important contribution to the perception of the voicing distinction. Nine of the 18 types of voicing contrast groups (6 types x 3 subjects) had the significant pressure values for the voicing distinction. The vertical jaw movement values were also useful for the correct perception of the voicing contrasts for 5 types of voicing contrast groups. Vowel duration, the anterior and posterior genioglossus muscles, and the middle constrictor muscles also contributed to the perception of some of the voicing contrasts.

Table 3.15

F-Values for the significant differences between the physiological and acoustic values of the voicing contrasts, based on tokens sorted into three groups; perceptually correct tokens, perceptually ambiguous tokens, and perceptually incorrect tokens. All the F-values in this table are significant at  $p < 0.05$ .

Subj	Contrasts	Po	Vowel	GGA	GGP	MC	Jaw
FB	t>/d>	7.33					4.05
	k>/g>		5.61				
	tI/dI	3.99					
	kI/gI	4.15	5.09				
AS	k>/g>	7.09					4.53
	pI/bI	4.34		4.28	6.35		
	tI/dI	17.36					7.41
	kI/gI	15.88		4.41			6.96
DB	p>/b>		4.23			4.09	
	t>/d>	27.75					5.40
	tI/dI	4.10					

Po: Peak intraoral air pressure  
 Vowel: Vowel duration  
 GGA: Anterior genioglossus muscle  
 GGP: Posterior genioglossus muscle  
 MC: Middle pharyngeal constrictor  
 Jaw: Vertical jaw movement

## CHAPTER 4

### DISOUSSION

#### 4.1 Intraoral Air Pressure

It had been thought that, in normal speech, the level of intraoral air pressure must depend on articulatory resistance to the outgoing pulmonary air stream, and therefore, that the pulmonary air source must be a prerequisite for for intraoral air pressure. In this study, however, a substantial increase in intraoral air pressure above atmospheric pressure was found in the production of the voiceless stops by all three subjects, and in the production of the voiced stops by Subject DB. It is obvious that pulmonary air source was not used for the generation of intraoral air pressure in these subjects. Some possible mechanisms for accomplishing the control of intraoral air pressure will be discussed in this section.

The values of intraoral air pressure produced by the subjects were extremely high compared with the values of normal talkers. In conversational speech, a typical value

of intraoral air pressure is 5-7 cm-H<sub>2</sub>O for the production of voiceless consonants and 3-5 cm-H<sub>2</sub>O for the production of voiced consonants (Subtelny, et al, 1966; Arkebauer, et al, 1967; Brown and McGlone, 1969; Lubker and Parris, 1970; Prosek and House, 1975; Muller and Brown, 1980; Brown and McGlone, 1979). For esophageal speech, another type of alaryngeal speech, the values of intraoral air pressure have been reported to be significantly higher than for those of normal speech (DiCarlo, et al, 1955; Damste, 1958; Murry and Brown, 1975; Hirose, 1984). Some researchers ascribed the greater values to the smaller oropharyngeal cavities of the alaryngeal talkers. In esophageal speech, however, the values varied among subjects. The authors explained the high variability by appealing to general articulatory instability and the interaction between sound generation and the maneuvers of air intake performed by the articulators. In the present study, the values of intraoral air pressure had little variability except for the production of /kIk/ in Subject FB. The stability of the values may be due to the difference between esophageal talkers, in that the electronic larynx does not have to use the articulators for air intake either by injection or by inhalation of air. Isshiki and Tanabe (1972) reported extremely high intraoral pressure values for the production of /p/ by an electronic larynx user with excellent intelligibility. The values were 20-30 cm-H<sub>2</sub>O for /p/ and 10-15 or almost negligible for /b/. They also reported that the values were more variable

for /b/ than for /p/, and values were sometimes higher and sometimes lower than those for normals. Their findings for the intraoral air pressure values and variability seem quite compatible with the findings of this study.

It should be noted that in the present study, peak pressure values were used instead of some other possible measure of intraoral air pressure, such as duration of pressure increase or decrease. It was thought that the measurement of peak intraoral air pressure was reasonable and valid because the values usually showed little overlap between voicing contrasts (as shown in Figs. 3.4, 3.5, and 3.6) and they seemed representative of the pressure characteristics of consonant production in the electronic larynx talkers.

The differences in the values of intraoral air pressure between voiceless and voiced stops were substantial in all subjects, and indeed were greater than the differences in value reported in studies with normal talkers. These findings suggest that the alaryngeal talkers made some effort to differentiate the intraoral air pressure values for the voicing contrasts. Considering the fact that they had no larynges to control the pressure differences, they must have used some other maneuver for the achievement of the difference.

Concerning the interaction between vowel context and voicing status for FB and AS, a consistent pattern was observed in that the values for voiceless stops are always

greater than those for voiced stops, although the pattern of change in mean intraoral air pressure for the voicing status as a function of vowel context differs slightly between the two talkers (Figs. 3.7 and 3.9). In normal speech, however, as Lisker (1970) pointed out, the value of intraoral air pressure tends to be variable, depending on stress and vowel context with values for voiced and voiceless consonants sometimes overlapping.

The pattern of the change in mean intraoral air pressure for the voicing status as a function of place of articulation is also different between FB and AS. Especially for voiceless stops, there is no particular place of articulation which consistently has the maximum or minimum pressure value for the two subjects. It is a consistent finding, however, that the mean values of intraoral pressure for voiceless stops are higher than for voiced stops.

Two questions may be asked about the production of intraoral air pressure in electronic larynx speech. The first question is how electronic larynx talkers were able to increase the pressure in the absence of a pulmonary air source. The usage of air behind the consonantal closure air is one way of increasing pressure in the oral cavity. The talker traps the standing air in the oral cavity by closing the cavity with the articulators during consonantal closure. Rothman and Goldstein (1976) reported that electronic larynx talkers with good intelligibility produced consonants (/t/,

/d/, and /s/) used such a maneuver. It is possible that the pressure of the trapped air can be increased by reducing the oropharyngeal cavity size. Some articulatory maneuvers for the reduction of the oropharyngeal size will be discussed in the following two sections.

The second question to be asked is whether intraoral air pressure is increased less for voiced consonants or whether there was an active gesture to decrease intraoral pressure for voiced cognates in electronic larynx speech. Brown and McGlone (1979), in their study of intraoral air pressure control during valsalva speech by normal talkers suggested the possibility that a pressure drop in the oropharyngeal cavity could be achieved by activities of articulatory organs, independent of the interaction with the closed larynx. Some strategies which would enlarge the oropharyngeal cavity size during closure for voiced stops have been suggested by Bell-Berti (1975). In this study, data from EMG and articulatory movement provided some evidence for enlargement of cavity size for voiced stops, and reduction, for voiceless stops. There will be further discussion below.

Another interesting finding concerning the control mechanism for intraoral air pressure is the timing of the pressure drop relative to consonantal closure release. Figure 4.1 shows the waveforms of intraoral air pressure for subject AS (top), and FB (bottom). These waveforms correspond to single productions of the utterance type

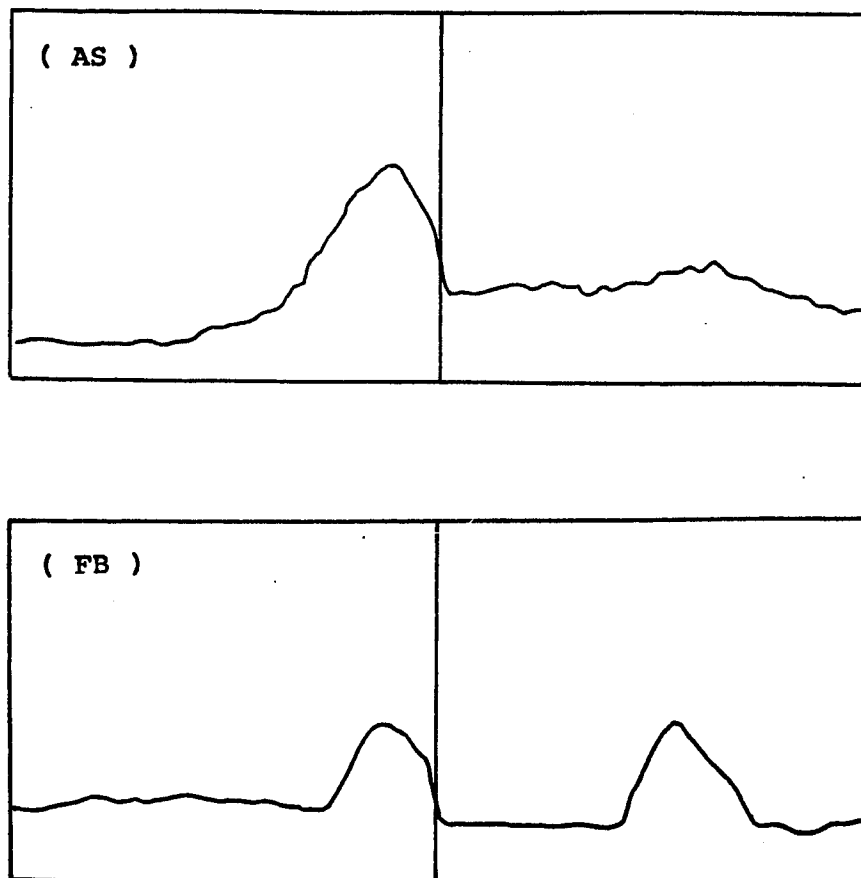


Figure 4.1 Individual waveforms of intraoral air pressure during production of /əpp/ by subject AS (top) and FB (bottom). The vertical in the center is the moment of labial closure release.

/əpɔp/, respectively. In these waveforms, intraoral air pressure is increased for the initial consonant /p/, and after reaching a peak, it gradually drops. In these waveforms, the pressure drop starts prior to the lineup point, the moment of release of lip closure, for the following vowel, /ɔ/. The pressure drop is almost completed, reaching to the baseline level, before the mouth is open. Similar patterns of pressure drop were observed in /əpɔp/, /əbɔp/, /əpɪp/, and /əbɪp/ for subject FB and in /əpɔp/ and /əpɪp/ for subject AS. In the other utterance types with alveolar and velar stops, this pattern was not observed. Since the pattern was not observed for subject DB, the pressure drop pattern seems to be peculiar to FB and AS. For normal talkers in previously reported studies, (e.g., Miller and Brown, 1980), intraoral air pressure starts increasing shortly before consonantal closure, reaching a peak towards the end of closure, after which the pressure drops slightly just before closure is released. After release, intraoral air pressure drops rapidly, for obvious reasons. In the case of alaryngeal talkers FB and AS, on the other hand, the drop of intraoral air pressure is not achieved by release of consonantal closure in some instances, since the pressure drop is almost completed during consonantal closure. Possibly the pressure dropped as the air in the oral pharyngeal cavity escaped into the nasal cavity due to release of velopharyngeal closure. However, careful inspection of audio waveforms revealed no

acoustic changes which might suggest nasalization in the output of the electric larynx during the consonantal closure. It is still unclear as to how intraoral air pressure drops in these instances. However, there must be some supraglottal articulatory mechanism responsible, because no other mechanism is available. No matter what mechanism they used, the important point here is that intraoral air pressure drop seemed to be deliberately controlled by the talkers rather than achieved naturally as a result of consonantal closure release.

In electronic larynx speech with these talkers, intraoral air pressure is generated, differentiated and sometimes dropped for stopped consonants, with active manipulation of supraglottal mechanisms. The fact that alaryngeal talkers managed to control intraoral air pressure using available mechanisms would suggest that intraoral pressure is an important speech attribute for consonant production. The alaryngeal talkers must have made some effort to preserve the speech attribute. The effort was realized by supraglottal manipulation. It is not known, however, whether the supraglottal manipulation they used is an exaggeration of a normal mechanism or whether it is a compensatory mechanism alaryngeal talkers have acquired after laryngectomy.

Since it was found that intraoral air pressure was generated and that the relationship of pressure values in voicing contrasts was maintained for consonant production in

electronic larynx speech as normal speech, the role of intraoral air pressure in consonant production becomes relevant. Whether or not the pressure generation and variations were achieved by the alaryngeal talkers as a residual or carry-over speech behavior from before the laryngectomy, it was a fact that they made successful efforts to preserve the pressure characteristics by manipulating the articulatory gestures for the production of consonantal voicing pairs. Brown and McGlone (1979) reported that normal talkers produced almost the same intraoral air pressure values for the consonants in both normal and valsalva speech. The behavior of our subjects and Brown and McGlone's normal talkers suggest that preservation of intraoral air pressure differences must be accomplished by upper articulatory maneuvers which all talkers use, although they may be exaggerated in the absence of transglottal flow. Similarly, the results of the present study suggest that the values of intraoral air pressure are specific to phonemes and, therefore, the talkers try to achieve the target using their speech organs as effectively as possible.

If we assume that the values of intraoral air pressure are phoneme specific, some possible acoustic events associated with the pressure variations become of special interest. In other words, how do these pressure variations produce perceptual cues for voicing contrasts? The present study has not been able to provide an answer to that

question. Details about perceptual cues will be discussed later in the section on perceptual studies.

#### 4.2 Electromyography

Data from electromyographic (EMG) signals have been used to assess the amount of tongue movement. It was assumed that the differences in tongue movement associated with the production of the voicing contrasts would reflect the volumetric differences of the oropharyngeal cavity for the contrasts. For example, if the tongue is positioned forward during the production of an alveolar stop, the size of the oropharyngeal cavity behind the tongue is greater than when the tongue is positioned less forward. The major interest in the electromyographic part of this study was to examine relative signal size to determine whether the difference in intraoral air pressure values associated with the voicing contrasts might have been accomplished by actively changing oropharyngeal cavity size.

The genioglossus muscles are reported to be able to draw the tongue roof forward and move the front of the tongue forward and upward (Smith and Hirano, 1968; Smith, 1971; Hirose, 1971; Miyawaki, 1974). Therefore, the muscle was active in the production of alveolar and velar consonants. These authors also reported that the genioglossus muscle exhibited different EMG activity

depending on the location of the electrodes inserted into the muscle. The anterior genioglossus muscle (GGA) was most active in producing /t/, while the posterior genioglossus muscle (GGP) was active during the articulatory gestures for /k/, /i/, and /u/. No research seems to have been conducted to compare the activity of the genioglossus muscle for voiced versus voiceless consonants, however.

There has been little research reporting EMG activity associated with the cavity enlargement for the intraoral air pressure drop in the production of voiced consonants. Bell-Berti (1975) reported active and passive modes for changing the pharyngeal cavity size by the articulatory muscles. The findings of the present study are in agreement with her assumption that there seem to be active articulatory maneuvers changing the oropharyngeal cavity size associated with the change in intraoral air pressure.

From the EMG data of this study, two major findings were obtained on a possible mechanism underlying the changing oropharyngeal cavity size. The first finding was a mechanism for reducing the cavity size, rendering intraoral air pressure large. The middle pharyngeal constrictor was more active during the production of voiceless stops than for their voiced cognates in Subject DB, the only subject for whom data was obtained for this muscle. As the contraction of the muscle causes the constriction of the pharynx, the results suggest that the oropharyngeal cavity was smaller for the voiceless stop than for the voiced

cognate, thus increasing the value of intraoral air pressure by compressing the air behind the consonantal closure in the cavity. This mechanism is plausible in these talkers, because they seemed to have a tight closure of the pharyngo-esophageal segment and so that air must have been trapped in the cavity unless there was a nasal leakage, which seems unlikely. Higher activities for /pI/ than for /bI/ were recorded from GGA of Subjects FB and AS, and from GGP of Subject AS. It is difficult to interpret the results in terms of the relationship between intraoral air pressure values and oral cavity size, estimated from tongue activity, since two contradictory interpretations are possible. One is that the oropharyngeal cavity must be smaller for /p/ than for /b/ with an elevated or fronted tongue, especially in the front part of the cavity, generating higher air pressure for /p/ than for /b/. The other interpretation of the EMG results is that the oropharyngeal cavity was larger for /p/ than for /b/ because the elevated or fronted tongue for /p/ enlarges the rear part of the cavity. However, if this were so, intraoral air pressure should be higher for /b/ than for /p/, which is contradictory with the pressure results of the present study. At this point, it seems reasonable that we await other evidence for the explanation of the greater genioglossus muscle activity for /p/ than for /b/. Of course, it is also possible that this muscle activity pattern difference is independent of cavity size.

The second major finding from the EMG data concerned a possible mechanism associated with the volumetric increase in the oropharyngeal cavity for the production of the voiced consonants. During the production of some of the alveolar and particularly the velar stops, the genioglossus muscle was more active for the voiced stops than for their voiceless cognates. This result suggests that the advancement and elevation of the genioglossus muscle enlarged the oropharyngeal cavity behind the articulatory closure for the tongue. The larger cavity for the voiced stops would reasonably explain the smaller values of intraoral air pressure, measured in the back of the cavity.

One important question remains to be asked: that is, whether tongue activity associated with the voicing contrast was specific to the alaryngeal talkers or whether it was typical articulatory behavior in normal speech, as well. The data from an x-ray microbeam study (Kiritani, et al, 1976) suggested higher tongue position for /t/ and /k/ than for their voiced cognates. This finding could be interpreted as providing contradictory results to the present study if a higher tongue position for voiceless stops than for voiced stops implies greater genioglossus muscle activity for the former. However, it is not appropriate to use X-ray microbeam data for estimation of tongue muscle activity, and moreover, their study does not provide statistical information. So, a conclusion for the

amount of tongue activity for the voicing contrast must be reserved at present.

Fujii (1970) and Fujimura, et al, (1972) reported in palatographic data that the tongue contact area was larger for /t/ than for /d/. This result may suggest that the oropharyngeal cavity is smaller for voiceless stops than for voiced stops if the tongue is more elevated in order to have a larger contact area with the palate for the former, as compared to the latter. Especially if there were data on vertical jaw movement showing greater jaw elevation for /t/ than for /d/, the relationship between tongue activity and oropharyngeal cavity size would be compatible with the findings of the present study. As for implications for future studies, it will be interesting to conduct experiments measuring the tongue activity with various techniques such as EMG, palatography, and ultrasound for voicing contrasts using normal talkers.

#### 4.3 Articulator movement

The purpose of obtaining the articulatory movement data was to investigate the difference in articulatory gesture associated with different values of intraoral air pressure in voicing contrasts. Some interesting evidence was obtained implying that articulators showed greater displacement for reducing oropharyngeal cavity size for the

voiceless consonants than for the voiced ones, possibly resulting in higher intraoral air pressure for the voiceless consonants than for their voiced cognates.

The results of the vertical movement of the jaw showing the greater upward displacement for /t/ and /k/ than for /d/ and /g/ in three subjects indicate that the distance between the floor of the oral cavity and the palate was reduced for voiceless stops than for voiced stops. The reduction of the distance would result in a smaller oral cavity size and greater intraoral air pressure for the voiceless stops than for the voiced cognates. As mentioned above, it is also interesting to assume that the larger tongue contact area for /t/ than for /d/ (Fujii, 1970; Fujimura, et al, 1972) and high tongue position for /t/ and /k/ than for /d/ and /g/ (Kiritani, et al, 1976) in normals resulted from the elevation of the oral cavity floor. If that were the case, their studies are compatible with the present study, suggesting smaller cavity size and higher intraoral air pressure for voiceless stops than for their voiced counterparts.

The data for vertical lip movement for the bilabial stops provided different results among the subjects. In FB and DB, the lower lip showed greater upward displacement for /p/ than for /b/. In AS, by contrast, the displacement was greater for /b/ than for /p/. As there does not seem to exist any substantial data for lip movement for the production of voicing contrasts in normal subjects, and as

this study had only three subjects, it is difficult to draw a decisive conclusion about the general pattern of vertical lip movement. The different patterns of the movement among the alaryngeal subjects might be due to the individual variability and the purposes for the distinctive lip movement for voicing contrasts remain unknown at this point.

The horizontal movement of the lower lip in FB and DB revealed that the lip was more protruded for /p/ than for /b/. It is assumed that the greater protrusion of the lip should not simply indicate the enlargement of the oropharyngeal cavity during the production of /p/. Activities of other organs which might be involved for the change of the cavity size behind the labial closure should be taken into consideration. Actually, the activity of the anterior genioglossus muscle of FB and that of the middle constrictor muscle of DB were higher for /p/ than for /b/, indicating the smaller oropharyngeal cavity for /p/ than for /b/. If the protrusion of the lip for voiceless stops is not achieved for the purpose of changing the cavity size, what would the physiological consequence of the labial protrusion be? One possibility is that it serves to increase the mechanical force of the lips in order to produce greater acoustic intensity of the burst sounds for /p/ than for /b/. Isshiki and Tanabe (1972) from their findings of higher orbicularis oris muscle activity for /p/ than for /b/, suggested that intense articulatory movement was important to increase high intraoral air pressure, which

consequently produces greater burst intensity for an acoustic cue for voiceless stops. For normals, Harris, et al, (1965) have shown that there is no significant difference in mechanical force of articulators between voiced and voiceless stops. For electronic larynx talkers, however, it is possible that they exaggerate mechanical force of articulators for the production of greater burst intensity in order to compensate for the lack of respiratory and laryngeal adjustment. Two future research studies are proposed. One is to obtain data for mechanical force of articulators for voicing contrasts in electronic larynx talkers. The other is to investigate the amplitude of the burst intensity for voicing contrasts both in normal and in alaryngeal subjects.

For all the articulatory movement data, measurements were made of the difference in articulatory displacement between the articulatory position for /ə/ and the peak of the displacement prior to the acoustic lineup point. It might have been preferable to measure the difference in displacement between the point of consonant closure and that of the peak displacement, since the major interest of the articulatory movement part of the present study was to assess the change in oropharyngeal cavity size to increase or decrease intraoral air pressure.

#### 4.4. Acoustic results

Acoustic measurements were obtained to investigate the following questions: (1) whether any systematic differences in acoustic characteristics were produced in the speech of alaryngeal talkers, when they are producing the voicing contrasts no matter whether the contrasts were perceptually distinguishable or not, and (2) what kind of acoustic measures might have been serving as the perceptual cue to the consonantal voicing distinction.

In this study, longer consonantal closure duration for the voiceless stops than the voiced stops was expected for the alaryngeal talkers to preserve the normal speech characteristics and to provide perceptual cues for the voicing distinction. The results indicated that difference in the closure duration for the contrasts were preserved by the alaryngeal talkers, although not seen in all the voicing contrast groups. The values of the closure duration for the voiceless stops were almost compatible with those reported for normal talkers (Lisker, 1978; Zue and Lafferiere, (1979), and for the talkers who use esophageal speech (Connor, et al, 1985). However, the values for voiced stops were sometimes greater than those in normal or esophageal talkers with high intelligibility. The smaller durational differences between the voicing contrasts could contribute to the low intelligibility scores of the talkers of this study.

Durational difference in vowels following the stops revealed inconsistent results among the subjects, in that vowel durations following the voiced stops were greater than vowel durations following the voiceless stops in 2 subjects, whereas the other subject showed the opposite durational pattern. The results suggest that the talkers systematically but idiosyncratically changed vowel duration following consonants with different voicing status, although normal talkers do not do so. Originally, post-consonantal vowel durations were measured because during the pilot studies some listeners in the perceptual experiments suggested that they might have been using this as a cue for the voicing distinction, even though previous research has shown that such durational differences do not correlate with the voicing distinction. It seems to be the case that, for the talkers with the electronic larynx, post-consonantal vowel duration was not an important acoustic feature for the production of the voicing distinction, despite the insufficient number of other acoustic features preserved for the distinction. As Swisher (1980) suggested, it is very likely that the individual variability for the pattern of vowel duration was the consequence of the talkers' effortful and exaggerated articulatory behaviors in their own attempt to increase the intelligibility. Data from a greater number of electronic larynx talkers may provide a clearer conclusion about the vowel duration characteristics. An attempt was made to measure burst intensity as a potential

acoustic feature of the voicing distinction. Moreover, it was felt that the measurement would be worth trying because there were some suggestions of the important role of the burst intensity as a cue in the literature (Isshiki and Tanabe, 1972; Rothman, et al, 1978). However, the measurement was not successful because, in many tokens, there was no identifiable release burst in the acoustic waveforms, and also because the amount of acoustic energy for the burst was obscured by the continuous acoustic energy of the vibratory duty-cycle of the electronic larynx. If audio recordings can be made using a technique to reduce the audio amplitude of the electronic larynx during and/or after the recording, some interesting data about burst intensity may be obtained.

The intelligibility scores of three electronic larynx talkers in this study were approximately 40-60% with individual variability. The scores are difficult to compare with those of other studies of electronic larynx speech because of the contextual differences among the studies. No detailed report about the intelligibility of stops in medial position has been found. However, the scores of the present study seem to be compatible with the scores of the initial consonants in the study by Weiss, et al, (1985). Considering the high intelligibility of the final consonants in their electronic larynx talkers, presumably using the cue of duration of the preceding vowels, the absence of such an

acoustic cue may explain the reason for the low intelligibility of the present talkers.

The perceptual confusion in this study was not an unexpected result, but rather, it clearly reflected the nature of electronic larynx speech. The result was in agreement with other studies (Barney, et al, 1959; McCroskey and Mulligan, 1963; Weiss, et al, 1985 ), in that, many voiceless consonants were perceived as voiced and that voiced consonants were correctly perceived more frequently than the voiceless consonants. This result indicated that voicing cues were not effectively used by the electronic larynx talkers. However, electronic larynx talkers succeeded in producing some voiceless consonants which were actually perceived as voiceless. The question is how they did it. The considerable individual variability for the voicing confusion may suggest that each talker used different maneuvers for the production of voicing contrasts. One may have used a maneuver which was more effective for the production of the voiceless stops (e.g., in the case of DB), and the another may have not (e.g., in the case of AS).

The issue about the response bias of the listeners should be discussed here. In this study, listeners were instructed to identify the consonant as voiceless if they thought that the consonant was possible to be voiceless even though it sounded more like voiced. Moreover, the listeners were given the information that 50% of the tokens were intended to be voiceless. Without the instruction and the

information, different results may have been obtained, in that less voiceless consonants might have been perceived. It should be pointed out, however, that some listeners reported that they perceived some, though not many, clear and obvious voiceless consonants.

Discriminant analysis was performed to find out what measures contributed to the perception of the voicing distinction. The relative differences of the peak intraoral air pressure seem to be most effective for the distinction in all three electronic larynx talkers. The result agrees with the claim of Connor, et al, (1985), who used esophageal talkers. It is suggested, from the results of the present study, that the control of the relative difference in intraoral air pressure for the voicing distinction could be advised as one of the successful technique of speech therapy. It would be interesting to conduct a future study looking for some relationship between intraoral air pressure and its perceptual affects in a condition that an electronic larynx talker is instructed to increase intraoral air pressure by some supralaryngeal maneuvers, during the production of voiceless consonants, while the pressure is monitored, and the produced consonants are later identified for the voicing status through perceptual experiments. This proposed research may provide a useful information as to whether high intraoral air pressure for voiceless consonants relative to voiced ones generates useful perceptual cues, or not. As for the very interesting question of the acoustic

and perceptual consequences of the pressure differences, future studies are required to provide the answer.

Post-consonantal durational differences also seemed to affect the perception of the voicing distinction, according to the results of discriminant analysis. The results, however, were not straightforward, in that, the results were not consistent among the talkers and among the utterance types. For example, longer vowel durations led the listeners to perceive the preceding consonants as voiced, and shorter vowel duration led them to perceive the sound as voiceless in FB. The opposite results were obtained in DB. It could be suggested that vowel duration differences themselves were not very useful for the perception of the voicing distinction.

The results of the discriminant analysis showed that jaw movements, activities of the posterior and anterior genioglossus muscles and, for one instance, the pharyngeal constrictor muscle contributed to the perception of the voicing distinction. The results, like those of vowel durations, were not consistent among the talkers and the contextual differences. It is assumed that these physiological measures did not directly provide strong perceptual cues for the voicing distinction.

In summary, electronic larynx speech, whose mechanism might be considered to be inadequate to provide effective acoustic cues for the production of the consonantal voicing distinction, was found to be capable of producing such

distinctions to some extent. Relative differences in peak intraoral air pressure were strongly associated with the perception of the voicing distinction. It is assumed that the pressure difference must have generated some effective acoustic cues, since other ordinarily used acoustic cues for normal speech were not available. Further research is required to find the acoustic consequences of intraoral air pressure differences.

The results of the research, together with Brown and McGlone's results for speech during valsalva maneuvers, suggest further that talkers actively control intraoral air pressure, per se. Detailed physiological comparisons of normals during valsalva maneuvers and electronic larynx talkers are required to see whether electronic larynx users preserve normal maneuvers, exaggerate normal maneuvers, or develop substitute maneuvers.

## REFERENCES

- Arkebauer, H. J., Hixon, T. J., and Hardy, J. C. (1967). Peak intraoral air pressure during speech. J. Speech Hear. Res., 10, 196-208.
- Barney, H. L., Haworth, F. E., and Dunn, H. K. (1959). An experimental transistorized artificial larynx. Bell Syst. Tech. J., 38, 1337-1356.
- Bejjter, P. and Serniclaes, W. (1974). Influence du contexte vocalique sur l'aperception du voicement des occlusives, (Repport d'Activites de l'Institut de Phonetique, ULB, Bruxelles, 8/1-2, pp. 101-108.)
- Bell-Berti, F. (1975). Control of pharyngeal cavity size for English voiced and voiceless stops. J. Acoust. Soc. Am. 57, 456-461.
- Black, J. W. (1950). The pressure component in the production of consonants. J. Speech Hearing Res., 15, 207-210.
- Brown, W. S., and McGlone, R. E. (1969). Relation of intraoral air pressure to oral cavity size. Folia Phoniatica, 21, 321-331.
- Brown, W. S., McGlone, R. E., and Proffit, W. (1973). Relationship of lingual and intraoral air pressures during syllable production. J. Speech Hear. Res. 16, 141-151.
- Brown, W. S., and McGlone, R. E. (1979). Supraglottal air pressure during a valsalva maneuver. (Speech communication Papers, Presented at the 97th Meeting of the Acoustical Society of America, ASA 50 Speech Communication Preprint Experiment. pp. 12-16) Cambridge, MA: MIT.
- Damste, P. H. (1958). Oesophageal speech (Thesis Print), Groningen: Gebr. Hoitsema.
- Danes, P. (1955). Effect of duration on the perception of voicing. J. Acoust. Soc. Am., 27, 761-764.
- Dey, F. L., and Kirchner, J. A. (1961). The upper esophageal sphincter after laryngectomy. Laryngoscope, 71, 99-115.

- DiCarlo, L. A., Amster, W. W., and Herer, G. R. (1955). Speech after laryngectomy, Syracuse University Press, Syracuse, NY.
- Diedrich, W. M., and Youngstrum, K. (1966). Alaryngeal speech. Springfield, Illinois: Thomas Publisher Co.
- Dugay, M. J. (1978). Why not both? (That is, why not esophageal speech and artificial larynx or speech aid.) In S.J. Salmon and L. P. Goldstein (Eds), The Artificial Larynx Handbook, Grune and Stratton Publishers.
- Ewan, W., and Kronen, R. (1974). Measuring larynx movement using the thyroumbrometer. J. Phonetics, 2, 327-335.
- Flanagan, J. L., Rabiner, L. R., Christopher, D., Boch, D. E., and Shipp, T. (1976). Digital analysis of laryngeal control in speech production. J. Acoust. Soc. Am., 60, 446-455.
- Fromkin, V. A. (1966). Neuromuscular specification of linguistic units. Language and Speech, 9, 170-199.
- Fujii, I. (1970). Phoneme identification with dynamic palatography. Ann. Bull. RILP, 4, 67-68. Tokyo: Research Institute of Logopedics and Phoniatics, University of Tokyo.
- Fujimura, O., Tatsumi, I. F., and Kagaya, R. (1972). Computational processing of palatographic patterns. J. Phonetics, 1, 47-54.
- Fujimura, O. (1985). [Personal communication].
- Gandour, J., Weinberg, B., and Rutkowski, D. (1980). Influence of post vocalic consonants on vowel duration in esophageal speech. Language and Speech, 23, 149-158.
- Gates, G., Ryan, W., Cantu, E., and Hearne, E. (1982). Current status of laryngectomy rehabilitation: Cases of failure. American Journal of Otolaryngology, 3 (2), 8-14.
- Goldstein, L. P. (1978). Listener judgment of artificial larynx speech. In S. J. Salmon and L. P. Goldstein (Eds.), The Artificial Larynx Handbook, Grune and Stratton Publishers.
- Halle, M. and Stevens, K. N. (1971). A note on laryngeal features. Quarterly Progress Report, No.101, Vol.18, 198-213, Research Laboratory of Electronics, M.I.T.
- Harris, K. S., Lysaught, G. F., and Schvey, M. M. (1965). Some aspects of the production of oral and nasal labial stops. Lang. Speech, 8, 135-147.

- Hirose, H. (1971). Electromyography of the articulatory muscles: current instrumentation and technique. Haskins Laboratories Statis Report on Speech Research, SR-25/26, 73-86, New Haven, Conn.: Haskins Laboratories.
- Hirose, H., and Gay, T. (1972). The activity of the intrinsic laryngeal muscles in voicing control. Phonetica, 25, 140-164.
- Hirose, H., Honda, K., and Sawashima, M. (1984). An intraoral pressure study of esophageal speech. Ann. Bull. RILP, 18, 153-159, Tokyo: Research Institute of Logopedics and Phoniatrics, University of Tokyo.
- Houde, R. A. (1967). A study of tongue body motion during selected speech sounds. Unpublished doctoral dissertation, Univ. of Michigan.
- House, A. S. (1961). On vowel duration in English. J. Acoust. Soc. Am., 33, 1174-1178.
- Hyman, M. (1955). An experimental study of artificial larynx and esophageal speech. J. Speech Hearing Dis, 20, 291-299.
- Isshiki, N., and Ringel, R. (1964). Air flow during the production of selected consonants. J. Speech Hearing Res., 7, 233-244.
- Isshiki, N., and Tanabe, M. (1972). Acoustic and aerodynamic study of a superior electrolarynx speaker. Folia Phoniatic., 24, 65-76.
- Kalb, M., and Carpenter, M. (1981). Individual speaker influence on relative intelligibility of esophageal speech and artificial larynx. J. Speech Hearing Dis., 46, 77-80.
- Karnell, M., and Willis, C. R. (1982). The effect of vowel context on consonantal intraoral air pressure. Folia Phoniatic., 34, 1-8.
- Kent, R. and Moll, K. (1969). Vocal tract characteristics of the stop cognates. J. Acoust. Soc. Am., 46, 1549-1555.
- Kiritani, S., Ito, K. and Fujimura, O. (1975). Tongue-pellet tracking by a computer-controlled X-ray microbeam system. J. Acoust. Soc. Am., 57, 1516-1520.
- Klatt, D. H. (1975). Voice onset time, friction and aspiration in word initial consonant clusters. J. Speech Hear. Res., 18, 686-706.
- Klich, R. J. (1982). Effects of speech level and vowel context on intraoral air pressure in vocal and whispered speech. Folia Phoniatic., 34, 33-40.

- Kobayashi, N., Horiguchi, S., and Baer, T. (1985). Aerodynamic and acoustic characteristics of the voicing distinction in electronic larynx speech. J. Acoust. Soc. Am., (Suppl, 1. 75), S86.
- Lauder, E. (1968). The laryngectomy and the artificial larynx. J. Speech Hearing Dis., 33, 147-157.
- Lindqvist, J., Sawashima, M., and Hirose, H. (1973). An investigation of the vertical movement of the larynx in a Swedish speaker. Ann. Bull. RILP, 7, 27-34, Tokyo: Research Institute of Logopedics and Phoniatrics, University of Tokyo.
- Lisker, L. (1970). Supraglottal air pressure in the production of English stops. Lang. Speech, 13, 215-230.
- Lisker, L. (1978). Rapid vs. ravid: A catalogue of acoustic features that may cue the distinction. Haskins Laboratories Status Report on Speech Research, SR-54, 127-132, New Haven, Conn.: Haskins Laboratories.
- Lisker, L. and Abramson, A. S. (1964). A cross-language study of voicing in initial stops; Acoustical measurements. Word, 20, 384-422.
- Lisker, L. and Abramson, A. S. (1971). Distinctive features and laryngeal control. Language, 47, 767-784.
- Lofqvist, A. (1975). A study of subglottal air pressure during the production of Swedish stops. J. Phonet., 3, 175-189.
- Lubker, J. F., and Parris, P. J. (1970). Simultaneous measurements of intraoral pressure, force of labial contact, and labial electromyographic activity during production of the stop consonant cognates /p/ and /b/. J. Acoust. Soc. Am., 47, 625-633.
- Lubker, J. F. (1973). Transglottal airflow during stop consonant production. a-oral air-pressure-pulse parameters in distinguishing between stop cognates. Phonetica, 14, 65-81.
- Malecot, A. (1966.b). Mechanical pressure as an index of force of articulation. Phonetica, 14, 169-180.
- Malecot, A. (1970). The lenis-fortis opposition: It's physiological parameters. J. Acoust. Soc. Am., 47, 1588-1592.
- McCroskey, R. L., and Mulligan, M. (1963). The relative intelligibility of esophageal speech and artificial larynx speech. J. Speech Hearing Dis., 38, 37-41.

- McGlone, R. E., and Shipp, T. (1972). Comparison of subglottal air pressures associated with /p/ and /b/. J. Acoust. Soc. Am., 51, 664-665.
- Menon, K. M. N. and Shearer, W. M. (1971). Hyoid positions during repeated syllables. J. Speech Hear. Res., 714, 858-864.
- Moolenaar-Bijl, A. (1953). Connection between consonant articulation and the intake of air in esophageal speech. Folia Phoniatica, 5, 212-216.
- Muller, E. (1974). Some new measures of supraglottal air pressure and their articulatory interpretation. Unpublished doctoral dissertation, University of Florida.
- Muller, E. C., and Brown, W. S. (1980). Variations in the supraglottal air pressure waveform and their articulatory interpretation. In N. J. Lass (Ed.) Speech and language: Vol.4, 317-389. New York: Academic Press.
- Murry, T., and Brown, W. S. (1976). Peak intraoral air pressures in whispered stop consonants. J. Phonetics, 4, 183-187.
- Netsell, R. (1969). Subglottal and intraoral air pressures during the intervocalic contrast of /t/ and /d/. Phonetica, 20, 68-73.
- Perkell, J. S. (1969). Physiology of Speech Production: Results and Implication of a Quantitative Cineradiographic Study. Cambridge, MA: MIT Press.
- Peterson, G. E., and Lehiste, I. (1960). Duration of syllable nuclei in English. J. Acoust. Soc. Am. 32, 693-703.
- Pickett, J. M. (1980). The Sound of Speech Communication: A Primer of Acoustic Phonetics and Speech Perception. Baltimore: University Park Press.
- Prosek, R. A., and House, A.S. (1975). Intraoral air pressure as a feedback cue in consonant production. J. Speech Hearing Res., 18, 133-147.
- Raphael, L. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. J. Acoust. Soc. Am., 51, 1296-1302.
- Repp, B. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable initial stop consonants. Language and Speech, 22, 173-189.
- Rothenberg, M. (1968). Breath-stream dynamics of simple-released-plosive production. Bibliographica, Phonetica, 6, 6-22.

- Rothman, H. B. and Goldstein, L. P. (1976). Analysis of speech produced with an artificial larynx. Paper presented at the American Speech and Hearing Association Convention, Houston, Texas.
- Rousselot, P. J. (1897). In H. Walter (Ed.) Principes de Phonetique Experimentale: Vol.1, Paris.
- Sacco, P. R., Mann, M. B., and Schultz, M. C. (1967). Perceptual confusions among selected phonemes in esophageal speech. J. Indiana Speech Hearing Ass., 26, 19-33.
- Sawada, K. (1959). an experimental study of alaryngeal speech. Oto-Rhino-Laryng. Clinic (Kyoto), 52, 42-72.
- Salmon, S. J. and Goldstein, L. P. (1978). The Artificial Larynx Handbook. Grunes Stratton, Inc.
- Schwartz, M. (1971). Air consumption values for oral and whispered plosive-vowel syllables. J. Acoust. Soc. Am., 49, 610.
- Serniclaes, W. (1974). Perceptual processing of acoustic correlates of the voicing feature. Speech Communication Seminar, Stockholm, August, 87-93.
- Singer, M. I., and Blom, E. D. (1981). Selective myotomy for voice restoration after total laryngectomy. Arch. Otolaryngol., 107, 670-673.
- Soda, T., Nishida, Y., and Suwoya, H. (1967). Intraoral pressure change in Japanese consonants. Otologica Fukuoka, 13, 34-43.
- Snidecor, J. C. (1975). Some scientific foundations for voice restoration. Laryngoscope, 85, 640-647.
- Stetson, R. H. (1951). Motor Phonetics (2nd ed.), Amsterdam, North-Holland Publishing Co.
- Stevens, K. N., and Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. J. Acoust. Soc. Am., 55, 653-659.
- Subtelny, J. D., Worth, J. H., and Sakuda, M. (1966). Intraoral pressure and rate of flow during speech. J. Speech Hear. Res. 9, 498-518.
- Swisher, W. E. (1980). Oral pressures, vowel durations and acceptability ratings of esophageal speakers. J. Communication Disorders, 13, 171-181.
- Tatham, M. A. A. and Morton, K. (1969). Some electromyography data towards a model of speech production. Language and Speech, 12, 39-53.

- Van den Berg, J., and Moolenaar-Bijl, A. J. (1959). Cricopharyngeal sphincter, pitch, intensity, and fluency of esophageal speech. Pract. Otorhinolaryng.(Basel), 21, 298-315.
- Wajskop, M. and Sweerts, J. (1973). Voicing cues in oral stop consonants. J. Phonetics, 1, 121-130.
- Warren, D. W., and Wood, M. T. (1969). Respiratory volumes in normal speech: A possible reason for intraoral pressure differences among voiced and voiceless consonants. J. Acoust. Soc. Am., 45, 466-469.
- Weiss, M. S., Yeni-Komshian, G. H., and Heinz, J. M. (1979). Acoustical and perceptual characteristics of speech produced with an electronic larynx. J. Acoust. Soc. Am., 65, 1298-1308.
- Weiss, M.S., and Basili, A.G. (1985). Electrolaryngeal speech produced by laryngectomized subjects: perceptual characteristics. J. Speech Hearing Res., 28, 294-300.
- Westbury, J. R. (1983). Enlargement of the subglottal cavity and its relation to stop consonant voicing. J. Acoust. Soc. Am., 73, 1322-1336.
- Winitz, Z., LaRiviere, C., and Herriman, E. (1975). Variations in VOT for English initial stops. J. Phonetics, 3, 41-52.
- Yanagisawa, N., and Hyde, C. (1966). An aerodynamic study of the articulatory mechanism in the production of bilabial stop consonants. Stud. Phonolog., 4, 20-27.
- Zlatin, M. A. (1974). Voicing contrast: Perceptual and productive voice onset time characteristics of adults. J. Acoust. Soc. Am., 56, 981-994.
- Zue, V. W. (1976). Acoustic characteristics of stop consonants: A controlled study (Tech. Rep. No. 523). Boston: M. I. T., Lincoln Laboratory.
- Zue, V. W., and Laferriere, M. (1979). Acoustic study of medial /t,d/ in American English. J. Acoust. Soc. Am., 66, 1039-1050.