Respiratory Constraints on Speech Production at Prosodic Boundaries

by

Janet Slifka

Submitted to the Harvard-MIT Division of Health Sciences and Technology

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Speech and Hearing Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2000

Author ....
Harvard-MIT Division of Health Sciences and Technology
May 17, 2000

Certified by ....
Kenneth N. Stevens, Sc.D.
Clarence J. LeBel Professor of Electrical Engineering
Thesis Supervisor

Accepted by ....
Martha L. Gray, Ph.D.
Edward Hood Taplin Professor of Medical and Electrical Engineering
Co-director, Harvard-M.I.T. Division of Health Sciences and Technology

Respiratory Constraints on Speech Production at Pr^sodic Boundaries

by

Janet Slifka

Submitted to the Harvard-MIT Division of Health Sciences and Technology on May 17, 2000, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Speech and Hearing Sciences

Abstract

This research characterizes the respiratory system dynamics at the initiation and termination of utterances and determines correlations of physiological measures with acoustic cues for these prosodic boundaries. The analysis includes boundaries within a breath as well as boundaries that are aligned with the initiation and termination of exhalation.

Simultaneous recordings of the acoustic signal, airflow, esophageal pressure and lung volume were collected during read isolated utterances and short paragraphs. These measures were used to derive estimates of recoil forces of the chest wall, net muscular forces, and the area of the airway constriction. Data are presented from four subjects (two men, two women), all native speakers of American English. Perceptual ratings for initial and final prominent syllables and the locations of pauses within the utterance were also collected.

For speech boundaries that are aligned with breath boundaries, utterance initiation occurs during a rapid transition in muscular effort. Sound begins as soon as conditions permit and these conditions consistently occur during net inspiratory muscular force. Alveolar pressure reaches an initial peak (P_{PI}) that is, in most cases, correlated to the relaxation characteristic of the chest wall. The timing of P_{PI} generally coincides with a prominent syllable if that syllable is the first or second syllable in the utterance and precedes later prominences. Pressure at phonation onset is, on average, near $0.3P_{PI}$ for utterances initiated with a voiced sonorant and is near $0. 8P_{PI}$ for utterances initiated with a voiceless fricative. Phonation termination results from an approximately 3-fold increase in glottal area and a 1-3 cm H_20 fall in pressure. Irregular fundamental frequency (F0) at the end of voicing, in many cases, does not fit the classical definition of glottalization. Instead, voicing terminates with increasing glottal area, and F0 becomes irregular during the increase. In some cases, regular F0 resumes as glottal area continues to increase. Distinct respiratory gestures are made at pauses within a breath. The pressure is reduced by 2-3 cm H_20, on average, during a period of relatively little volume change. The findings in this research show that the role of the respiratory system in speech production goes beyond a more traditional view of this role as one of simply providing a relatively constant driving pressure during speech.

Thesis Supervisor: Kenneth N. Stevens
Title: Clarence J. LeBel Professor of Electrical Engineering and Computer Science

Acknowledgements

This work was supported in part by NIH grants #5T32DC00038, #5R01-DC00266-14 and #DC00075, the Clarence J. LeBel Chair, HST Fellowship, and a Dissertation Enhancement grant from the National Science Foundation INT 98-21048.

I am very grateful to my advisor Ken Stevens. He has shown me the qualities that are truly valuable and remarkable in research as well as in the researcher. His standards of excellence and integrity are ones that I aspire to make my own. His kindness and generosity to everyone he meets have consistently brought out the best in them. I will treasure my years working with him.

I would also like to thank my committee members. I have so enjoyed and benefited from the insightful comments of Dr. Stephen Loring. Our discussions about the respiratory system have been an exciting learning process for me. Dr. Robert Hillman has been a wonderful source of perspective regarding the clinical arena. I am grateful for being welcomed into his laboratory during the data acquisition sessions. Dr. James Kobler was an essential part of those acquisition sessions and the calibration process that followed. Dr. Martha Gray has been inspiring me since I first met her shortly after my arrival at MIT. I am grateful that she brought her expertise to my committee.

This work would surely not have been possible without the generous participation of the many subjects in the various stages of this project. I am thankful for the time, concentration, and effort that everyone gave to this work.

I am grateful to the faculty and staff of the Harvard-MIT Division of Health Sciences and Technology, and especially those involved in the Speech and Hearing Sciences Program. I would like to thank Nelson Kiang for his motivating energy and Denny Freeman for the time that I spent as a member of his laboratory. I spent the majority of a summer learning and working in the laboratory of Gunnar Fant at KTH. His wisdom and that of Johan Liljencrants expanded my views on speech research and I am very thankful for those experiences. While at MIT, I have been fortunate to have some amazing teachers as mentors. I am particularly grateful for the time I have spent as a teaching assistant for Dr. Amar Bose. Those years have changed me in so many ways for the good.

There are many colleagues and friends who have contributed to this work in time, expertise, and support. While I cannot possibly name them all, I am particularly indebted to Stefanie Shattuck-Hufnagel for the world of prosody, A.J. Aranyosi for technical-wisdom and coffee-wisdom, Helen Hanson for never saying “no” when I came to her for help and Arlene Wint for general sanity. I would like to thank the many members of the Speech Group especially Jennell Vick, Kelly Poort, and Majid Zandipour, the members of the Voice and Speech Laboratory at MEEI including Geoff Meltzner and Harold Cheyne, and the many friends who have made my time in Cambridge joyful, especially Brenda Divelbliss, Joyce Rosenthal, Diane Ronan, Christopher Long, Carol Summers, Darla Villani, and Astrid Hagen. I am grateful to Ben Hammond for showing me excellence.

I am so lucky to have Steve Brown in my life. His truly unwavering faith in me, his belief in the worth of good work, and especially his laughter have made the last couple of years very special. My thanks to my sister Jane Lecian and my wonderful grandmother, Helen Slifka. My final thanks are reserved for my mother, Josephine Khoenle. She is the rock in my world.

Table of Contents

List of Figures

List of Tables

Chapter 1

Project Overview

Speech is created from the coordination of many different structures of the speaker's body; each of those with their own physical dynamics. The respiratory system, making use of many of the structures of the torso, provides a drive to create sound sources. Smaller, faster structures like the vocal folds and tongue are also involved in creating these sources as well as shaping the sounds of the message. In the process of generating the message, acoustic changes, many of them arising from changes in the sound source, also layer on cues as to which syllables should receive more attention (prominences) and where boundaries are located in the signal (phrasing). This organizational structure of speech is referred to as speech prosody. The major goal of this study is to characterize the respiratory system dynamics at speech boundaries and determine the correlations to prosodic acoustic cues for those boundaries.

The following sections will give an overview of speech prosody, the role of the respiratory system during speech, and state the specific goals of this research.

1.1 Prosody

Written language is generally discussed in terms of syntax, the sentence structure, or the arrangement of the words in a sentence. Spoken language is also discussed in terms of its organizational structure, the speech prosody. While syntax constrains the prosody that a speaker may select, it does not uniquely determine the prosody of a given sentence. (For a review, see Shattuck-Hufnagel and Turk,1996.) The prosody of spoken language consists of changes in the realization of the signal that provide a means for enhancing interpretation of the speech. Regional variations in the signal indicate interpretative landmarks for the listener. In this case, the term *regional* is used to refer to changes that may span more than one sound segment or phoneme. These regional cues may include such frequently studied and salient changes as lengthening in segment duration (e.g., Oller,1973 and Wightman et al.,1992), or manipulation of fundamental frequency (F0) (e.g. Nespor and Vogel,1986 and Beckman and Pierrehumbert,1986). The fundamental frequency is determined by the period of oscillation of the vocal folds during voicing. Other, less dominant cues include changes in vowel quality or spectral content (e.g. Stevens,1994) or a decrease in signal amplitude (e.g. Streeter,1978).

Planning and implementation of the prosody of an utterance is generally considered to result from purely cognitive choices. Methods have been proposed for determining and labeling the prosodic structure of an utterance as well as examining the role that the generation of prosody plays in utterance planning (for example, Silverman et al.,1992). Various factors may be weighed by each speaker in planning the prosody for a given utterance, and the weight given to each of these factors may change with each utterance. However, only some aspects of some of these factors are known.

Prosodic structure is proposed to be hierarchical. Two types of hierarchy are generally considered: (1) a hierarchy of constituents and (2) a hierarchy of prominences. In the hierarchy of constituents, the exact levels and their constituents are under much discussion. For examples of current theories see Nespor and Vogel (1986), Hayes (1989), Pierrehumbert and Beckman (1988), or Selkirk (1978, 1986). The constituents at each level may be defined in terms of such concepts as intonation (Beckman and Pierrehumbert, 1986), as in a “full intonational phrase”, or rhythmic prominence, such as a “foot”, or phonological terminology, such as an “utterance”. An utterance is generally considered to be the top level in the prosodic hierarchy. Its boundaries are formed by non-hesitation pauses (Hayes,1989).

Phrases within an utterance are also marked by prosodic boundaries. Intonational phrases are marked by changes in the intonation or tone contour. Tone contour is discussed in terms of the fundamental frequency (F0). For example, a declarative statement usually ends in a drop in F0 while a yes-or-no question usually ends with a rise in F0. Phrases also end with preboundary lengthening (e.g. Lehiste et al.,1976, Selkirk,1984), which is an increase in the relative duration of the sound segments prior to the end of the phrase. The degree of lengthening is considered to be the greatest at the highest level in the hierarchy, the utterance.

Some initial studies on the acoustic cues for prominence examined the set of possible cues for prominence or lexical stress and proposed that, of those cues, the strongest determinant of prominence is an F0 or pitch change (e.g. Bolinger,1958, Fry,1958). The other hierarchy, the hierarchy of prominences, is frequently discussed in terms of the concept of a pitch-accent. A pitch accent is considered to be a change in the intonation contour that brings to perceptual prominence a syllable in a phrase. Again, there are many theories as to the elements and structure of this hierarchy. In general, prominences are attached to vowels and all vowels are either reduced or full. Only full vowels can receive a pitch accent in a phrase. However, as stated, the perception of a prominence will include all of the acoustic cues, including not only changes in the fundamental frequency contour but also such changes as a rise in signal amplitude, an increase segment duration, and an increase in spectral contrast between a consonant and an adjacent vowel (Fant,1987).

Some recent research has expanded into determining the physical correlates of prosodic changes across an utterance. The basis for this vein of research is that there may be aspects to the prosody of an utterance that arise from the constraints of the physical production system rather than from purely cognitive choices. Some studies have focused on the physical correlates in the articulators. For boundaries within a breath group (the por-

Coupled to this general view of a constant mean subglottal pressure is the qualifier that there are variations in that pressure.

> "...momentary increases in mean subglottal pressure appear on syllables with prominences...[as well as] momentary variations which seem to be attributed to changes in glottal resistance and changes in vocal tract driving point impedance." (Atkinson,1973).

The general view is that the respiration system acts to generate a relatively constant subglottal pressure during speech and that pressure may have local rises and drops associated with prosodic prominences or sharp changes in the impedance to air flow from the lungs. In fact, current speech production models often use a constant source to represent subglottal pressure (e.g. Stevens and Bickley,1991). Yet, this pressure does vary. The movement of air from the lungs provides the energy source for speech and it is an energy source that must be repeatedly replenished. The speaker's lungs can only contain a finite amount of air and the speaker must also maintain respiration for gas exchange purposes.

Some studies have provided additional details about the specifics of this relatively constant subglottal pressure. (see, for example, Bouhys et al.,1966, Lieberman,1967, Atkinson,1973, Fant and Kruckenberg,1994, and Fant et al.,1997b). The results indicate that the pressure has initial rise and final decay segments of about 120-200 milliseconds in duration. Pressure has been found to remain relatively constant until the decay segment (e.g. Lieberman,1967) or to decline about 6 to 4 cm H_20 across the breath group. (Fant and Kruckenberg,1994, Fant et al.,1997b).

Respiratory system constraints on the realization of the acoustics may be present when the state of the respiration system is changing rapidly. These instances occur: immediately following an inhalation as the respiration system quickly generates sufficient pressure to initiate an utterance and at the end of an utterance as the respiration system finishes an tion of speech produced on a breath), the timing of the articulatory dynamics has been measured to slow in the boundary adjacent sound segments (Byrd and Saltzman,1998) and a change in the nature of the articulatory gesture, such as increasing contact area for the production of stop consonants has also been reported (e.g. Fougeron and Keating,1997). Other studies have linked prominences within an utterance to changes in subglottal pressure (Ladefoged,1967, Lieberman,1967, Stevens,1994, and others). Fant and colleagues have examined the mechanisms for intonational prominence through the relationship of fundamental frequency and glottal configuration to subglottal pressure changes (Fant et al.,1997a, Fant et al.,1997b).

The present study examines correlations between physiological measures and the acoustic cues for prosodic boundaries. Given that the actual planned prosody is not known, the alternative is to determine systematic regional changes in the acoustic signal, their relation to prosodic cues, and whether these changes may have resulted in response to respiratory constraints. The first step toward that goal is to determine which aspects of prosodic structure may interact most directly with respiratory system changes.

1.2 The role of the respiration system in speech production

The role of the respiratory system during speech production has been described in various contexts such as:

> "...The muscles regulating the air pressure during many utterances have to be operated in such a way as to maintain a constant mean background pressure in the lungs..." (Draper et al, 1959)
>
> "...the respiratory system provides an essentially constant air pressure and the laryngeal and upper airway structures move to modulate the escaping air stream..." (Netsell, 1973)
>
> "...the primary function of the pulmonic system during speech is simply to produce Ps [subglottal pressure] that is reasonably constant and above some minimal level." (Ohala,1990)

exhalation. In fact, it is common practice in the study of speech acoustics to embed the speech of interest into a carrier phrase to avoid the effects that occur at the beginnings and ends of utterances. Other instances may occur at boundaries within an utterance, and at other places where manipulation of the pressure may be required to produce perceptually significant changes in the signal strength, such as at prosodic prominences.

Whether cueing a prominence or a boundary, the speaker manipulates the sound source. Sounds in speech may be either voiced (involving oscillation of the vocal folds), unvoiced (involving noise generated at a constriction in the vocal tract), or a combination of both. In all cases, sound is generated from a pressure difference across a constriction in the vocal tract. (For a review see Stevens, 1998.) Basically, this means that there are two requirements for a sound source: a pressure difference and a constriction. It also means that in order to change the nature of the source, the speaker can control the pressure, the constriction, or both. In fact, the two are not independent. In order to generate a pressure in the lungs, there has to be a resistance to airflow from the lungs. This resistance comes from constricting the vocal tract.

There are some cases where the speaker does not have appropriate control of the respiratory system. Neurological disorders, such as the dysarthrias, disrupt control of the respiratory pattern and can degrade speech intelligibility. Recent studies on speech breathing in such cases have focused on documenting the types of variations in speech breathing (for example, Annoni et al,1993, Hammen and Yorkston,1996). Subjects with aphasia have been studied using specific speech tasks that target their ability to plan speech segments and incorporate appropriate planning for use of the respiratory system (e.g. Hoole and Ziegler,1997). In normal speech, there is a relatively precise coordination of the respiratory system and the articulators that occurs at the start of an exhalation. Studies have shown that this coordination is disrupted in the speech of some stutterers (Peters and Boves,1988).

1.3 Project Goals

Given a basic role of the respiratory system during speech as that of providing an essentially constant pressure, this research addressed the following issues: (1) what are the details of the rise and fall of this pressure as the speaker renews the energy source? (2) what is the correlation between the pressure changes and the resistance to airflow as determined by the area of the constriction in the vocal tract? (3) are these boundary regions in the pressure and the area of the constriction at the initiation and termination of exhalation related to acoustic cues for the prosodic boundaries? (4) if so, is the respiration system active during the generation of acoustic cues for boundaries within a breath group?

This research focused on normal speech production by native speakers of American English. A database was collected of the acoustic signal as well as several physiologically-related signals for four subjects for normal read speech. No special requests were made of the speakers regarding lung volume, syllable prominence, or the loudness of the speech. The subjects were simply asked to read the given text as naturally as possible.

The analysis segment evaluated changes in the acoustic signal and the respiratory system at the beginnings and ends of utterances that are aligned with the initiations and terminations of exhalation; those instances where the respiration system is changing state. The areas examined include: net muscular force and respiratory system dynamics, pressure and timing landmarks in relation to the sound segment or phoneme at the boundary, timing and coordination of area of the constriction controlling the sound source, and correlations to the perceived prominences nearest the boundaries. This methodology is extended, in some aspects, to speech boundaries that do not align with breath boundaries, such as pauses within a breath group.

Chapter 2

Respiratory Anatomy and Mechanics

The basics of the statics and dynamics of the respiratory system are reviewed in this section as a basis for the measurement and calibration methods of Chapter 3 as well as the analysis of pressure-volume relationships of Chapter 5. The respiratory system as a whole is much more complex than presented here but the pertinent relationships for this work are given.

2.1 Anatomy

The respiratory system consists of the chest wall, the lungs, and the airways connecting them to the outside atmosphere (Figure 2.1). The chest wall, which is composed of the ribcage or thorax and the diaphragm, surrounds the lungs and is coupled to the lungs via a fluid-filled space called the pleural space. The trachea provides a pathway for air into and out of the lungs. The vocal folds are located at the junction of the trachea and the vocal tract within a structure called the larynx. The space between the vocal folds is called the glottis. The vocal tract is the portion of the airway above the vocal folds. For speech, the articulators are defined as those structures that modify the transfer function of the vocal tract, and that control constrictions in the vocal tract that cause sound sources when there is a flow of air.

The lungs are composed of many sac-like structures called alveoli and, correspondingly, the pressure in the lungs is called alveolar pressure.

A wide range of muscles can be used to move the structures of the respiratory system, but the diaphragm and the external intercostals are the main muscles of inspiration. The internal intercostals and some of the abdominal muscles are considered the most commonly used muscles of expiration.

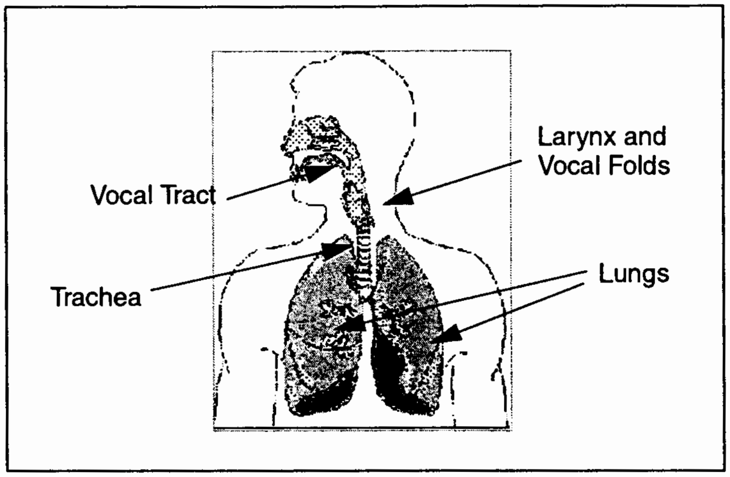


Figure 2.1: Air travels between the lungs and the surrounding air through the trachea, the glottis (the space between the vocal folds), and the vocal tract. The vocal folds are located in the larynx.

2.2 Mechanics

The function of the respiratory system requires the lungs to expand and contract. However, the body does not achieve both actions through muscular effort alone. The respiratory system is an elastic system and makes use of elastic recoil in its expansion/ contraction cycles (as in Rahn et al.,1946). There are two elastic elements: the lungs and the chest wall. When uncoupled from the chest wall, the lungs will collapse to a small resting volume. High surface tension within the alveoli of the lungs would cause the lung to collapse. In contrast, if the chest wall was uncoupled from the lungs, its volume would expand. For example, if the diaphragm was uncoupled from the lungs, the weight of the abdominal contents would pull it downward. If the ribcage was uncoupled from the lungs, it would expand outward. When coupled through the pleural space, the lungs/chest wall system reaches an equilibrium volume called relaxation volume (V_{rel}) where the force of the lungs to collapse is countered by an equal and opposite force of the chest wall to expand. At V_{rel}, the pressure in the pleural space, $P_{PL,}$ has a small negative value (near -5 cm H_20). Expanding above V_{rel} requires effort of the inhalation muscles and places the lungs in a position such that elastic recoil forces are available for exhaling back to V_{rel}. The muscles of exhalation can be used to compress the chest wall to volumes less than V_{rel}. As shown schematically in Figure 2.2, the lungs/chest wall system has a resting position at V_{rel} and can be maximally stretched to total lung capacity (TLC) and maximally compressed to residual volume (RV). Also shown in Figure 2.2 is the vital capacity (VC) which is the volume change between TLC and RV.

Typical values for TLC range between 4 and 7 liters. Generally, in normal resting breathing in an upright person, exhalation ends slightly above V_{rel}, (for example, Loring and Mead, 1982) at the functional residual capacity (FRC). FRC is generally in the range of 30-40% VC. Inspiratory muscles stretch the lungs above FRC by roughly one-half to one liter, which is generally about 8-12% VC. The airways are unobstructed and a maximal airflow rate of roughly 0.5 liters per second is reached. An average speaker takes 15-20 breaths per minute. During inhalation, the pressure in the lungs is less than atmospheric causing air to move into the lungs; during exhalation, the gradient is reversed. The pressure in the lungs may range between plus and minus 1 cm H_20. Detailed summaries of the actions involved in tidal or resting breathing can be found in (Agostoni and Mead,1964 and Mead and Agostoni,1964, among others). Read speech is generally initiated at lung volumes near 60% VC (e.g. Bouhys et al.,1966, Hixon et al.,1976) and conversational speech may be initiated at lower lung volume levels.

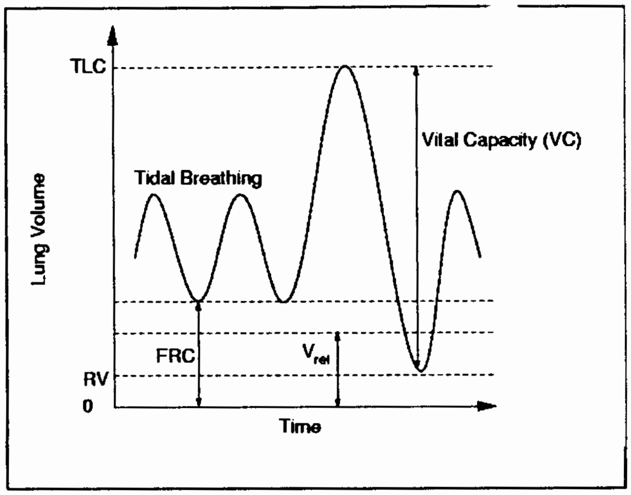


Figure 2.2: Lung volume levels. The sketched curve represents lung volume during two cycles of tidal breathing, an inhalation to total lung capacity (TLC), an exhalation to reserve volume (RV), and a return to tidal breathing. TLC is the volume of usable air in fully-inflated lungs. RV is the volume of air that cannot be expelled from the lungs and relaxation volume (Vrel) is the equilibrium point for the elastic elements: chest wall and lungs. Vital capacity (VC) is the volume difference between TLC and RV. Functional residual capacity (FRC) is the exhalation level for resting breathing.

The elastic elements, the chest wall and the lungs, are characterized in terms of their compliances. Compliance is a measure of the volume change created for a given change in pressure and is normally measured in liters per cm H_20:

$$Compliance = \frac{\Delta Volume}{\Delta Pressure} \tag{2.1}$$

These pressure-volume curves are not linear through the range of lung volumes and exhibit hysteresis depending on the volume history of the structures. Nonlinearity is introduced due to structural limits at high lung volumes (collagen in the lungs is stretched near its limit and increasing pressure does not increase volume). For the lungs, surface tension at low lung volumes introduces nonlinearity. Alveolar collapse at low lung volumes means that a larger pressure will be required to increase volume because pressure must be applied to overcome the surface tension of the surfaces in contact. Hysteresis is introduced due to factors such as stress-adaptation and plasticity (for example, Sharp et al.,1967). Also, since many of the muscles of respiration are part of the chest wall, when the muscles are active, they change the compliance of the chest wall, in part by altering the shape of the chest wall from the passive configuration. Due to these factors, among others, the compliances are characterized through a plot of lung volume versus pleural pressure for static conditions with muscles relaxed.

An example of measured static recoil curves for the chest wall and lung are shown in Figure 2.3 (adapted from Knowles et al.,1959). Each curve is based on over 100 measurements taken from four male subjects aged 28 to 45. Standard error is indicated by the horizontal bars at various lung volumes. The standard error in the range of values used during speech (40-60% VC) is ±0.3 cm H_20, indicating little variation in the recoil curves. The greatest error occurs at very low lung volumes.

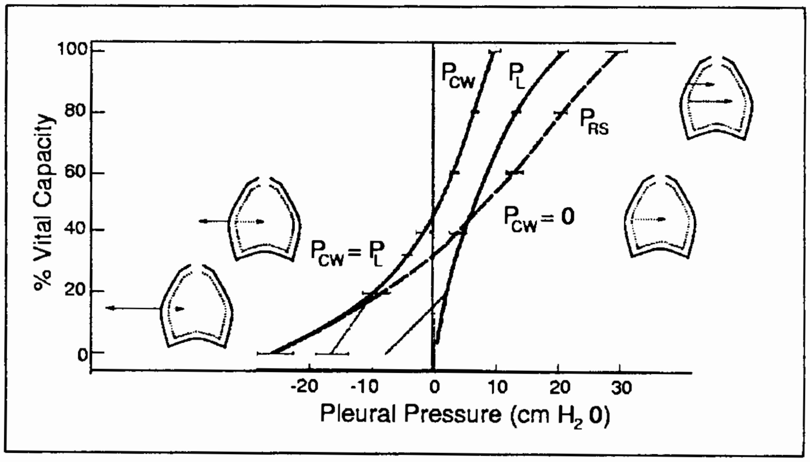


Figure 2.3: Recoil curves for the relaxed chest wall (P_{CW}), the lungs (P_L) and the respiratory system (P_{RS}). The horizontal brackets along the curves represent standard error. Below 20% VC, the recoil curves are extrapolated in shape since pressures were difficult to measure at such low lung volumes (Actual measures are given as the light lines below 20% VC.) Sketches of the chest wall show arrows that indicate the magnitude and direction of the recoil pressure for the chest wall (outer dark lines) and the lungs (inner gray lines). (Adapted from Knowles et al.,1959)

The pressure exerted by the respiratory system, P_{RS}, is the difference between pressure at the body surface (normally atmospheric pressure, P_{ATM}), and alveolar pressure:

$$P_{RS} = P_{ALV} - P_{ATM} \tag{2.2}$$

P_{RS} has two components (as pointed out in Kahn et al., 1946 and as shown in Figure 2.3), the pressure exerted by the chest wall and the pressure exerted by the lungs. For the case where all pressures given are referenced to P_{ATM}, $P_{RS}=P_{ALV}$, can be expressed as:

$$P_{ALV} = P_{MUS} + P_{CW} + P_{L} \tag{2.3}$$

where P_{MUS} is the applied muscular pressure. In order to attempt to estimate these quantities, pleural pressure is estimated using an esophageal balloon (see Chapter 3 for further discussion). This measurement is a reliable estimate of pleural pressure changes and can be expressed as:

$$P_{PL} = P_{MUS} + P_{CW} \tag{2.4}$$

From these relations, the recoil pressure of the lungs can be estimated from P_{PL}:

$$P_{L} = P_{ALV} - P_{PL} \tag{2.5}$$

The characterization of recoil forces and muscular forces will be examined in Chapter 5 for data collected during speech tasks. The next section describes the measurement and derivation of the data signals used in this study.

Chapter 3

Acoustic and Physiological Data Acquisition

3.1 Overview

Simultaneous recordings of the acoustic signal, electroglottography, airflow, and physiological signals used to estimate lung volume and lung pressure were made at the Voice and Speech Laboratory of the Massachusetts Eye and Ear Infirmary (MEEI) under the direction of Dr. Robert Hillman. During signal acquisition, subjects read isolated utterances and short paragraphs and gave spontaneous responses to questions. As a part of some calibration tasks, intraoral pressure was also recorded.

Data were collected from nine subjects (5 women, 4 men), but analysis was completed on four subjects (two women, two men). All of the data acquisition sessions were completed in a concentrated time period to attempt to create a high level of consistency in the protocol and equipment conditions. This pacing to the acquisition sessions did not allow all of the various data conversions and calibration procedures to be completed on one subject before the next session began. During the process, it was determined that some subjects would need to be excluded from the analysis stage. For this reason, a larger number of acquisition sessions were completed than was estimated to be needed for the analysis. In the end, two subjects were excluded because of very frequent evidence of peristaltic waves during the speech. Subjects were cautioned that if they swallowed prior to reading the utterances, it would invalidate the collection. If subjects indicated that they had swallowed, the speech tasks resumed after the peristaltic wave had passed through the data display. However, many of the utterances were still corrupted for two of the excluded subjects. Two other subjects were excluded due to inconsistencies across the duration of the acquisition session in lung volume estimation parameters. Four subjects were obtained with consistent calibrations and the remaining two subjects were not used.

The four subjects in the study range in age from 21 to 28, have normal hearing, have no history of respiratory ailments, are non-smokers, and speak American English as their first language. In this document, the speakers are referred to as Subject #4 through Subject #7. Subject #4 is a male speaker and is generally considered to use little intonation in his speech. Subject #5 is a female speaker and could be described as using a wider range of intonations and intensity levels in her speech. Subject #6 is a female speaker who spoke in a casual manner, and Subject #7 is a male speaker who spoke in a more careful and precise manner. Approval for the use of human subjects was granted through the Massachusetts Institute of Technology (MIT) under protocol number 2515, and through MEEI under protocol number 92-08-030. All subjects signed a consent form and were compensated for their participation.

3.2 Measurement Methods

Electroglottography provides information on the state of the glottal opening and produces a voltage that is roughly proportional to vocal fold contact area (for a review see Childers and Krishnamurthy,1985). The electroglottography sensors (Glottal Enterprises, Inc.) were held in contact with the front of the subject's throat at the level of the vocal folds using a velcro-sealed neckband. Airflow estimates were made using a circumferentially vented pneumotachograph mask (Rothenberg,1973) from Glottal Enterprises, Inc. Mounted to the exterior of this mask was a pressure-sensitive microphone (Sony Co.) located approximately 15 cm from the speaker's mouth and used to record the acoustic signal. Intraoral pressure was measured with a small open tube placed in the subject's mouth and connected at the other end to a pressure transducer.

Lung volume is controlled by the movements of two essentially independent structures: the ribcage and the abdomen (Konno and Mead,1967). For example, lung volume could be increased by either expanding the ribcage or lowering the diaphragm through expansion of the abdomen. Most people change lung volume through some combination of movement of the ribcage and movement of the abdomen. The cross-sectional areas of the ribcage and abdomen were estimated using respiratory inductive plethysmography (Respitrace, Ambulatory Monitoring, Alpsley, NY). These cross-sectional areas can be mapped to lung volume estimates (Konno and Mead,1967) (See "Calibration Procedures and Results" on page 35.) In order for this method to produce consistent results, any changes in ribcage and diaphragm cross-sectional areas must arise from lung volume changes and not from movement due to postural changes in the subject. For this reason, the subject was cautioned to maintain the same posture throughout the recording process. Each subject was seated in a chair with foot and head supports.

Alveolar pressure (P_{ALV}) was estimated from measurements of esophageal pressure (Van den Berg,1956). Esophageal pressure was measured using a thin latex balloon that was passed through the nasal cavity. Once the balloon entered the pharyngeal region, the subject repeatedly swallowed small sips of water to pull the balloon into the esophagus. The balloon was placed at a level below the trachea and above the diaphragm, and was inflated with 0.5 cm^3 of air. Placement was checked by looking for a strong response to a swallow. If the tip of the balloon was too far into the esophagus, that end of the balloon would cross into the abdomen when the subject's diaphragm was high in the ribcage. Pressure in the abdomen is positive. Placement with respect to whether the tip of the balloon had crossed the diaphragm was checked by having the subject make inspiratory and expiratory efforts against an occlusion, as in Baydur (1982). In this case, the occlusion was a mouthpiece that was connected to one port of a differential pressure transducer (Glottal

Enterprises, Inc.). The other port was connected to the esophageal balloon. During the maneuvers, the pressure measure was the difference between alveolar pressure and pleural pressure. This level remained essentially constant.

During the speech tasks, the distal end of the esophageal catheter was connected to a differential pressure transducer (Glottal Enterprises, Inc.) and the other port of the transducer was left open to atmospheric pressure. The balloon in the esophagus is separated from the lungs by the tissue of the esophagus and the pleural space. Pressure in the esophageal balloon will differ from that in the lungs because of these intervening structures. The esophagus itself does not normally have significant tone. For this reason, there is only a significant pressure drop across the esophageal tissue during swallowing (i.e. during a peristaltic wave). As stated earlier, the pleural space surrounds the lungs and contributes the coupling force between chest wall and lungs. The pleural space is a fluid-filled space and the pressure in that space is subject to the forces of gravity. A hydrostatic gradient in the pleural space can lead to variations in pleural pressure from 8-12 cm H_20 at low and high lung volumes respectively (Gold,1994). The esophageal balloon, then, takes a spatial sample of this pressure gradient. When muscular force is exerted, the lungs may deform non-uniformly; this nonuniformity will also contribute to the difference between alveolar pressure and the esophageal pressure measurement. In general, esophageal pressure will differ from pleural pressure and this difference is generally constant for the mid-range of lung volumes in an upright person (Wohl,1968).

Calibration maneuvers were performed before and after the speech task segment of the experiment. All data recordings were digitized on-line at a sample rate of 10 kHz using the Axoscope (Axon Instruments, Inc.) recording software and Digidata digitizers. Signals were filtered to prevent aliasing and amplified using a Cyberamp 380 Programmable Signal Conditioner (Axon Instruments) prior to digitizing.

3.3 Calibration Procedures and Results

3.3.1 Transducers

Voltage levels from all transducers were correlated to known pressure, flow, and amplitude values. Pressure transducers were calibrated using a water manometer for pressure across the range of the transducers. Airflow was calibrated using a compressed air source and an airflow meter. Sound pressure level (SPL) was calibrated using an electrolarynx (Cooper-Rand), which produces a wideband signal, held at two different locations from the microphone. An SPL meter (Bruel&Kujer) was held next to the microphone. The recorded voltage values from the microphone were converted to root-mean square (rms) values and correlated to the SPL values.

3.3.2 Lung Volume

Lung volume estimation involved the following calibration maneuvers performed by the subject: (1) resting breathing (2) isovolume maneuvers and (3) spirometer and spirobag maneuvers. First, the ribcage and abdomen voltage traces for resting breathing were examined to determine approximate voltage values for FRC under the assumption that during resting breathing, an exhalation ends at FRC. All abdomen and ribcage voltage traces were scaled such that zero volts corresponded to FRC. Isovolume maneuvers were used to determine the scale factors for ribcage and abdomen voltage traces. These maneuvers fixed the amount of air in the lungs by having the subject close the glottis and pass that volume back and forth between the ribcage and abdomen by moving the stomach in and out. The subject was asked to first take an easy exhalation or sigh before closing the glottis. The goal was to have the subjects perform the isovolume maneuvers near FRC. Example voltage traces for an isovolume maneuver are shown in Figure 3.1.

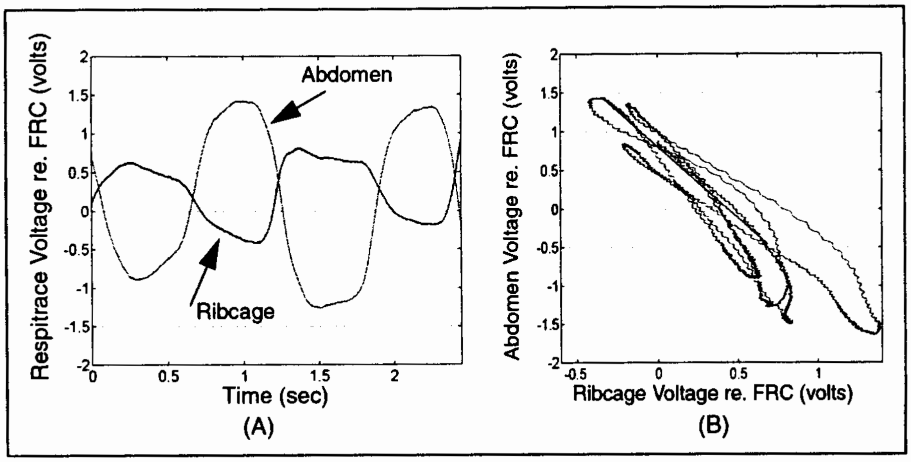


Figure 3.1: (a) Respitrace voltage traces for ribcage and abdomen signals during an isovolume maneuver. Voltages are scaled relative to FRC. (b) The same voltages traces as in (a) but with the ribcage signal plotted versus the abdomen signal in order to determine the scale factor such that the ribcage voltage will represent the same volume as an equal abdomen voltage.

As seen in Figure 3.1a, the abdomen voltage is larger at the peaks than the ribcage voltage, but each peak represents the same amount of air in the lungs. For this reason, the gain of ribcage signals was adjusted so that the values for the ribcage voltage and the abdomen voltage were the same for that fixed volume. This gain was determined by plotting the ribcage signal against the abdomen signal (Figure 3.1b), fitting a line to the data, and scaling the ribcage voltage by the inverse of the slope. For clarity, the examples given show one isovolume maneuver with several oscillations, but the actual calibrations fit the line to three such maneuvers. The two signals were then summed and converted to liters using values collected when the subject exhaled into a spirometer. Subjects were asked to inhale as much air as possible without disturbing their posture and then to exhale as much of the volume as possible into a mouthpiece attached to the spirometer. From these data, an estimate of each subject's vital capacity was made. The voltage output of the spirometer was calibrated to volume by putting a known 3 liters into the spirometer using a syringe. This liters/volt scale factor was the final step in the calibration. These final calibrated lung volume curves are displayed as volume in liters referenced to FRC, i.e., the difference between the volume and FRC.

For calibration verification purposes, the subject also breathed into a spirobag. The spirobag is a breathing tube attached to a flexible bag that will hold 0.8 liters of air when fully inflated. The subject slowly inflated and deflated the bag several times. An example of a calibrated spirobag maneuver for Subject #4 is shown in Figure 3.2. For additional verification of the calibration process, lung volume estimated from the respitrace signals was compared to the time integral of low-pass filtered airflow. For all isolated read utterances, the average difference between these two methods of estimating lung volume was less than 0.1 liters.

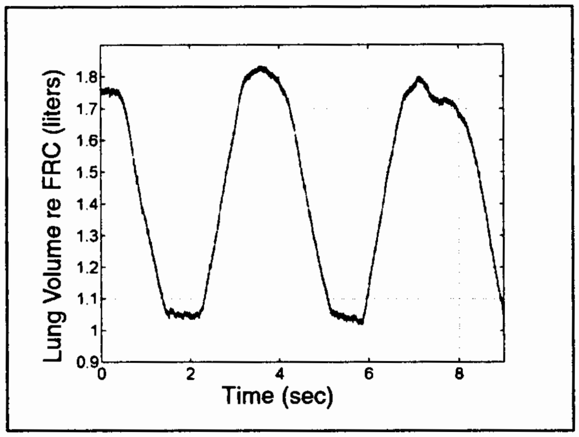


Figure 3.2: Example from a calibrated spirobag maneuver for Subject #4. The subject breathed through a mouthpiece into a 0.8 liter plastic bag. The subject was instructed to just inflate the bag fully without using forceful exhalation.

3.3.3 Alveolar Pressure

Lung pressure estimation also involved calibration maneuvers performed by the subject. As stated, esophageal pressure is a measure of pleural pressure and not alveolar pressure. In speech production modeling, subglottal pressure is used as the driving source and subglottal pressure can be considered to be equal to the pressure in the alveoli when the airflow rate is not high enough to cause a significant drop across the airway resistance. Such high rates of flow generally occur during speech only for some aspirated consonants.

Since pleural pressure is the pressure between two compliant elements (the lungs and the chest wall), there is a volume-dependent component to this pressure. At a high lung volume with the air pressure in the lungs at atmospheric, pleural pressure will have a large negative value since it must exert considerable force to expand the lungs to that volume. As lung volume decreases, pleural pressure will become less negative. This volume-dependent component of pleural pressure is called the static recoil pressure of the lungs and must be accounted for in the calibration process (as in Kunze,1964).

As outlined in Section 2.2 "Mechanics", pleural pressure is determined by the static recoil pressure of the lungs, P_L, and the pressure in the lungs P_{ALV}:

$$P_{PL} \cong P_{ALV} - P_L \tag{3.1}$$

In general, esophageal pressure, considered to equal pleural pressure 'was measured by connecting the tubing of the esophageal balloon to one port of a differential pressure transducer. The other port was left open to atmospheric pressure, P_{ATM}. However, in the calibration process, to measure the static recoil pressure of the lungs, one port of the differential pressure transducer measured pleural pressure and the other port measured oral pressure through a mouthpiece. Oral pressure measured with the glottis spread is a good estimate of alveolar pressure. This mouthpiece has an opening to the atmosphere that could be occluded by the experimenter. The subject was asked to inhale a large breath and exhale in a slow, relaxed manner. During the exhalation, the experimenter occluded the mouthpiece for brief periods of time. During those occlusions, the differential transducer measured $P_{PL} - P_{ALV}$, or $-P_L$. This pressure-volume relaxation curve for static recoil pressure was used for calibration. An example curve of lung volume versus differential pressure is shown in Figure 3.3. The moments of occlusion are the points fitted with a third-order curve. For all calibrated alveolar pressure data, the pressure value was compensated for the static recoil pressure of the lungs at the volume. For this reason, it was important that very reliable measures of lung volume be collected.

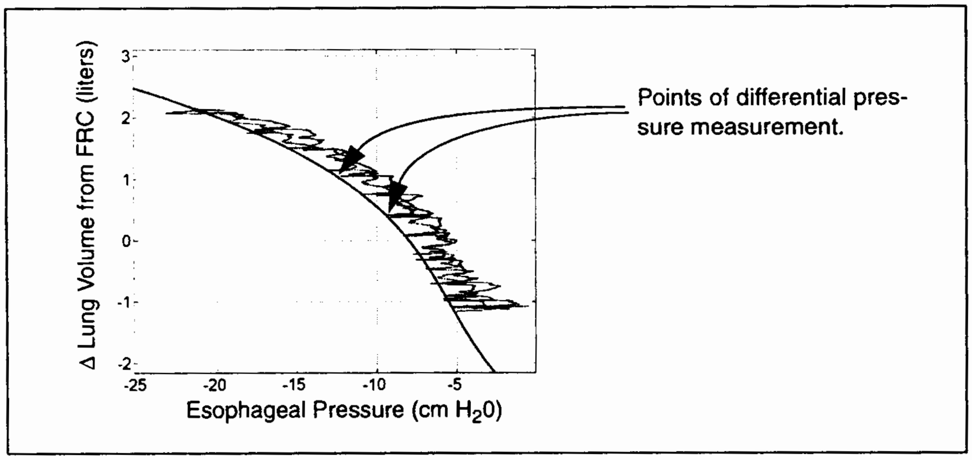


Figure 3.3: Lung volume plotted as a function of esophageal pressure in order to determine the static recoil pressure of the lungs across a range of lung volumes. This example is for Subject #7. Two different calibration maneuvers are plotted on the same axes. A third order curve is fitted to the points of differential pressure measurement, which are the leftmost points in the figure.

For verification of the calibration process, the pressure difference was measured between the esophageal pressure and intraoral pressure, while the subjects produced a series of /p ae/ syllables. During the /p/ closure, and assuming the subject preforms the task well, the intraoral pressure has been shown to be within 1 cm $H_2 0$ of the subglottal pressure as measured with a tracheal puncture (Hertegård,1994). An example of the comparison between the measured intraoral pressure and the estimated alveolar pressure is shown in Figure 3.4.

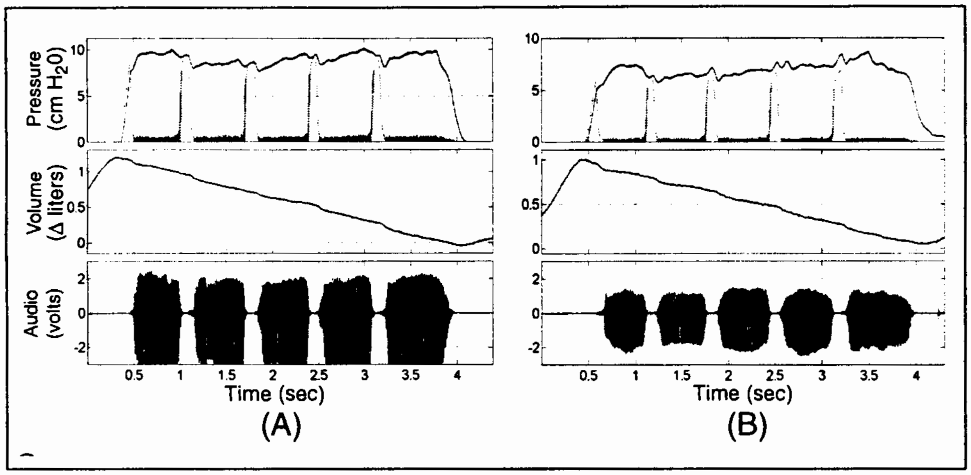


Figure 3.4: Examples of intraoral pressure and alveolar pressure during production of a series of /p ae/ syllables. The top panel shows alveolar pressure (dark line) and intraoral pressure (gray line). The middle panel is lung volume and the bottom panel is audio. During the /p/ closure, alveolar pressure should be within 1 cm H_20 of intraoral pressure. (a) Louder speaking voice for Subject #4 (b) Normal speaking voice for Subject #4.

A set of calibration figures for each subject corresponding to Figures 3.3 and 3.4 can be found in Appendix A.

3.4 Exclusion of Utterances from Analysis

While a speaker can take a breath at any point in an utterance, there is a general tendency to align utterance boundaries with breath boundaries (e.g. Henderson et al.,1965), especially during read speech (Lieberman,1967). Of the acquired utterances, only those which were initiated at the start of an exhalation and terminated at the end of an exhalation were used in the analysis. Consequently, during the following discussion, references to the start or end of an utterance will imply that such alignment with the breath is also present.

Utterances were excluded from analysis if there was evidence of a peristaltic wave during some portion of the utterance. Pressure and air flow for each utterance in the data set were the checked at the point where flow crossed from negative to positive. At this point, the pressure in the lungs should be zero. Utterances where the magnitude of this difference was greater than 1.5 cm H_20 were excluded from analysis. Also evidence of any oddity in the reading such as disfluency or stumbling over the reading of an utterance caused the utterance to be excluded. Exclusions for all of these reasons across all four subjects amounted to 16%. A perceptual test was run using the utterances, as will be discussed in the section "Perceptual Ratings of Prominences" on page 45. However, using all of the utterances in the data set would have made the length of that test unmanageable. Several utterances were arbitrarily excluded from the rating test, leaving the total number of utterances rated across all speakers as 241.

3.5 Estimation of the Area of the Constriction

Pressure in the lungs is dependent upon the actions of the chest wall and the resistance to airflow in the airways. The resistance to airflow will be largely determined by the area of the smallest constriction in the airway. This section details methods for estimating the area of the constriction and applies those methods to speech data.

3.5.1 Estimation Methods

A pressure difference across a constriction in the airway will cause airflow through that constriction. From measures of pressure and airflow, the area of the constriction can be estimated. The pressure drop across the constriction may have two components: (1) the drop due to the resistance of the constriction and (2) the drop due to the movement of the mass of air in the constriction in an oscillatory manner. The first type of pressure drop is applicable to all constrictions whether the sound source is noise-like or periodic. The second type of drop is applicable during voicing.

In the physical situation shown in Figure 3.5, a constriction in the airway is represented as having a length *l* and a cross-sectional area *A*. The pressure drop ΔP is given by:

$$\Delta P = P_1 - P_2 = RU + \frac{d}{dt}(MU) \tag{3.2}$$

where $R = \frac{\rho U}{2A^2}$ and $M = \frac{\rho l}{A}$ for most cases in speech production. The parameter ρ is the density of the air in the vocal tract and is set to 1.14 kg/m^3. The parameter U is the volume velocity through the constriction and is the low-pass filtered flow as measured with the pneumotachograph mask. The filter was a 147th order FIR filter with a cutoff of 30 Hz. The parameter P_1 is the calibrated alveolar pressure estimate. The estimate of the resistance, R, is an experimentally determined relationship. The mass element is a low-frequency approximation for a mass moving back and forth in a short tube. (For a detailed review see Stevens,1998). The area of the constriction was assumed to be rectangular with a constant length of 2.4 cm.

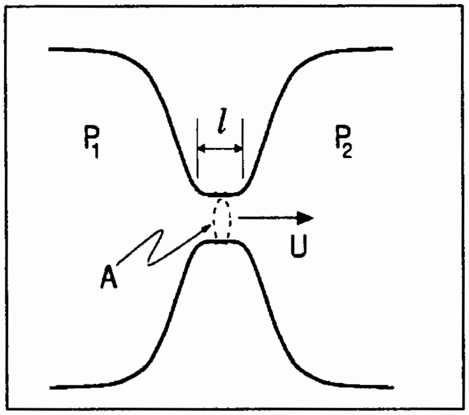


Figure 3.5: Area estimation for a single constriction with length *l* and cross-sectional area *A*.The pressure difference, $\Delta P = P_1 - P_2$ causes airflow through the constriction in the airway.

For the case where there is a single constriction in the airway and where P_2 is equal to the pressure outside the mouth (P_{ATM}), the value for P_1 is the subglottal pressure. In addition if flow rates are relatively small (< 0.3 l/sec) then there is no appreciable drop across the airways of the lungs and P_1 is equal to the alveolar pressure. While all three of these assumptions may not be present throughout every segment of a speech signal, reasonable area estimates should be obtained for most speech segments except for sound segments that may involve more than one constriction in the airway. Also, this method would not be accurate during nasals when there are two pathways for airflow from the lungs to the atmosphere.

For the actual speech data collected as a part of this study, it was not possible to determine the true area of all of the constrictions in the vocal tract. In order to compare the estimated area to actual areas, an articulatory-based speech synthesizer was used. The synthesis system, HLsyn (Sensimetrics Corp.) has among its inputs the area of the glottal constriction (A_g), the area of the constriction at the alveolar ridge (A_b), as well as the area of the constriction at the lips (A_l) and the area of the velopharyngeal opening (A_n). (For a more detailed discussion see "Synthesis of Voicing Termination Cues" on page 106.) However, for the comparison made here, it is sufficient to note that actual areas of the various constrictions are specified along with a driving pressure. The outputs of the synthesizer include the airflow through the glottis (Ug).

For comparison, the word "show" was synthesized using the measured alveolar pressure and airflow from Subject #7. Input area traces for the tongue blade constriction (+ symbols) and the glottal area (o symbols) are shown in Figure 3.6. Following synthesis, the calculated flow through the glottis U_g and the pressure P_{ALV} were used with the area estimation algorithm to produce the computed area trace (solid dark line).

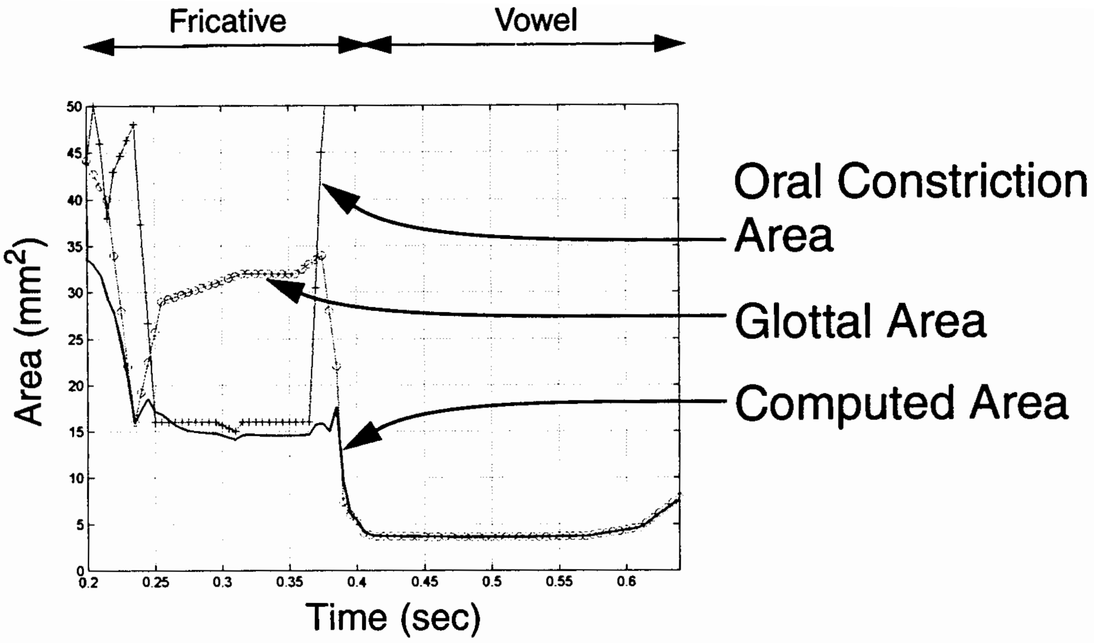


Figure 3.6: Comparison of constriction areas for the synthesized word "show." The solid dark line is the estimated area. Area of the constriction made by the tongue blade is given by the '+' symbols and glottal area is given by the 'o' symbols. Error in the estimation is present when there are two .onstrictions in the vocal tract as shown at the transition between the fricative and the vowel.

3.5.2 Examples for Speech Data

The example in Figure 3.7 shows the types of calibrated data displays used in this analysis. The F0 contour was estimated using an autocorrelation-based algorithm.

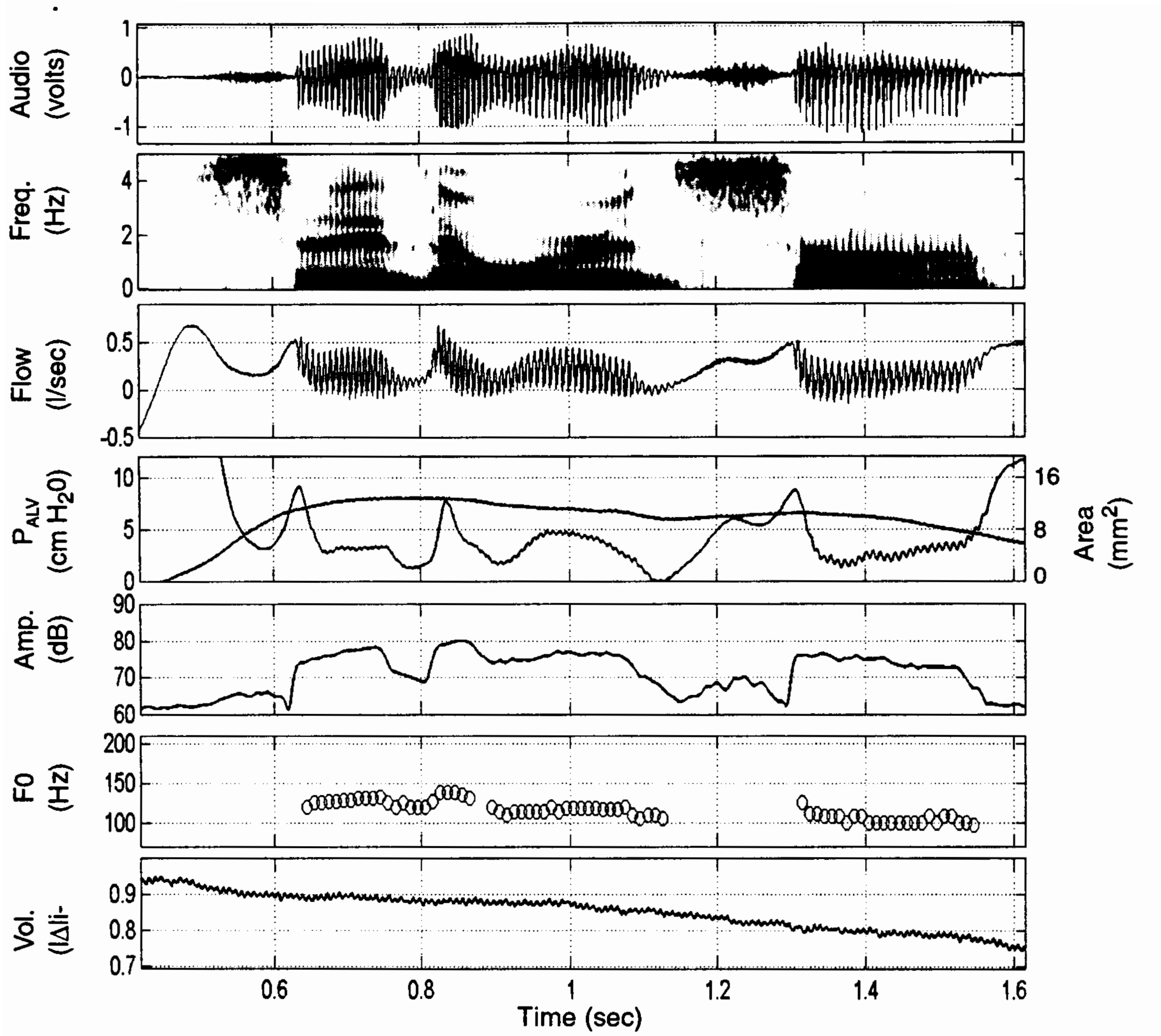


Figure 3.7: Example from Subject #4. The utterance is "Say the word saw." Panels show audio (volts), the spectrogram (kHz), airflow (liters/sec), an overlay of alveolar pressure (cm H_20, scale on the left) and area of the constriction (mm^2, scale on the right), amplitude (db), F0 (Hz) and lung volume (liters).

3.6 Perceptual Ratings of Prominences

One important concept regarding the prosody of an utterance is the prominence given to syllables within the utterance to cue the listener in regards to phrasing, emphasis and disambiguation. The current study is focused on the general prosodic cues that, in addition to F0 changes, include amplitude, duration, and source spectrum changes. In order to include all prosodic cues, it was decided to conduct a prominence rating experiment rather than execute a labeling system based on F0 changes alone. The goal of this rating experiment was to determine the initial and final perceived prominent syllables in the acquired utterance set. Such a rating-based protocol has been used by Fant and colleagues (e.g. Fant and Kruckenberg,1989). The experimental protocol asked listeners to rate every syllable in the utterance on a 0 to 30 scale of prominence. The listeners were instructed that stressed syllables should fall near 20 on the scale and unstressed syllables should fall near 10. An analysis of the ratings patterns across subjects indicated that the ratings generated were not strictly a reflection of the acoustic and physiological parameters involved in producing the speech but that the listeners also used phonetic and linguistic information (Liljencrants,1999).

The major focus of the current study has been on the boundaries of an utterance that coincide with the initiation and termination of exhalation. It has not been a study of prominences. However, as discussed in Chapter 1, there has been evidence that subglottal pressure changes occur at prosodic prominences. The use of a prominence rating test was added to gain general, essentially binary information about whether the boundary syllables in the utterance were prominent or not. For this reason and to simplify the task of the listening panel, a modified version the perceptual rating test of Fant and Kruckenberg was carried out.

3.6.1 Experimental Protocol

Listeners were instructed that they would hear a series of utterances and be asked to make a judgment about the prominence of the syllables in those utterances. They were asked to think of the prominences as being on a 0-30 scale, where prominent syllables were in the upper half of the range and all other syllables were in the lower half of half of the range. For each utterance, the listeners marked an 'X' through the first syllable in the utterance that would fall in the prominent range and marked an 'X' through the last syllable in the utterance that would fall in the prominent range. In addition, if the listeners heard a pause in the utterance, they were asked to place a slash between the words at the boundary, then place an 'X' through the first and last prominent syllables prior to the slash as well as following the slash. Each utterance could be replayed as often as the subject needed. The perceived pause information will be used in Chapter 7.

In the utterance set to be rated, 12 utterances were repeated twice (three from each speaker). These utterances and the ratings given to them were used as a consistency check on the stability of each listener's criteria for completing the task. Seven subjects were used in the listening test. One subject became fatigued about halfway through the task was unable to complete the test. Another subject only rated 2 of the 12 repeated utterances the same. Only the remaining 5 subjects were used to determine prominence labels. Those subjects had consistency measures ranging from 8/12 to 12/12 with an average value of 10/12.

It was decided that a simple majority of 3/5 votes was sufficient to label a syllable as prominent. Of the total 241 utterances in the data set, 11 onsets received less than a majority agreement (8 of those were for Subject #7) and 7 endings received less than a majority agreement (4 of those were for Subject #4). For the remaining utterances, the average vote for initial syllables was 4.35 (0.74 standard deviation) and the average vote for final syllables was 4.32 (0.79 standard deviation). In 16% of the utterances, the prominent syllable at onset was decided by a 3/5 vote. In later discussions, these ratings are considered in terms of two groups: (1) utterances where the perceived prominence was on the first or second syllable of the utterance and (2) utterances where the perceived prominence was on the third, fourth, or fifth syllable. The utterances that received a 3/5 rating made up 9% of the utterances in Group 1 and made up 37% of the utterances in Group 2. This result would suggest that there was more uncertainty in the listening panel when placing the initial perceived prominence later into the utterance. These utterances were included in the analysis to keep a wide representation of sound segments at the boundaries, utterances with pauses and without, and sentence structures in the data set.

The acoustic data were labeled for prominent syllable duration limits as well as initial and final vowel duration limits. These segmental limits were determined from combination of visual data and listening. While such limits are sometimes difficult to determine, especially in the vicinity of liquids and glides, the labels were determined as consistently as possible.

The next section will describe the speech tasks performed during the signal acquisitions.

Chapter 4

Characteristics of the Data Set

4.1 General Design of the Speech Tasks

The speech tasks used during these measurements included read isolated utterances, read paragraphs, and spontaneous responses to questions. Example material is shown in Table 4.1. One goal of this study has been to examine the interaction between respiratory or muscular habits and the realization of speech acoustics at boundaries. For these reasons, no special requirements were made regarding lung volume, syllable stress, or the loudness of the speech. The subjects were simply asked to read the utterances as naturally as possible given the array of equipment surrounding them. Each utterance was printed on a separate card and held up before the subject by the experimenter. The majority of the utterances began and ended with fricatives, vowels, voiced sonorants, or stops (See section 4.3.1 “Boundary Speech Segments” for details.)

Table 4.1: Example speech tasks.

Example isolated sentences
Ali will say the word sew.
Sally will say the word sew.
Say the word sew.
The large round rock is in the truck.
A large round rock is in the truck.
Possible routes are listed on the sheet.
Several routes are listed on the sheet.
Six will participate, even though there is room in the program for as many as eight.
Six will be asked to participate, even though there is room for eight.
It’ll be the best seat in the house.

Table 4.1: Example speech tasks.

Eight different tasks will be repeated every day.
Example paragraph
The program will begin next week. Six will be asked to participate, even though there is room for eight. Eight different tasks will be repeated every day.
Example questions for spontaneous responses.
Please describe this room.
What is the weather like today?

The utterances fall in three different general categories: (1) utterances that changed the beginning text while keeping the ending text the same, (2) utterances that changed the ending text while keeping the beginning text the same, and (3) utterances that contained a syntactic break or pause given by a comma. Comma placement occurred near the beginning of the sentence, in the middle, and near the end of the sentence.

At the time of data acquisition, subjects did produce another set of utterances with specific disambiguating stresses indicated by printing the word to be stressed in capital letters. These utterances were not used in the analysis. Subjects also gave spontaneous responses to questions and those utterances are not used in the current analysis.

4.2 Definitions and Landmarks

Several timing landmarks in the acoustic and pressure signals were defined for the initiation and termination of utterances. At initiation, times were labeled for zero flow, start of the utterance, onset of phonation, and initial alveolar pressure peak, as shown in Figure 4.1. Ideally, zero flow and zero pressure occur at the same time, but due to the possible 1.5 cm H_20 deviation in the signal, the co-occurrence of these points is not guaranteed. The zero flow point was assumed to be the more accurate determination of the beginning of the exhalation. The start of the utterance was determined from the first acoustic evidence of sound - whether that sound was noise-like or periodic. For stop consonants, the initial burst release was used as the start of the utterance. Start pressure will refer to the alveolar pressure at the start of the utterance. Onset of phonation was determined from the first acoustic indication of voicing. For utterances that began with a voiced phonetic segment, the start of the utterance was the same as the onset of phonation. However, for voiced stop onsets such as /b/, the start of the utterance was set as the burst release and may precede the start of phonation. Phonation pressure refers to the alveolar pressure at the onset of phonation. The first peak in pressure was determined automatically from an algorithm which low-pass filtered the pressure to a cutoff of 100 Hz with an FIR filter to reduce measurement variations, and found the first occurrence of two consecutive points where the slope in the pressure curve was less than or equal to zero. The pressure peak time was set to the first of those two points.

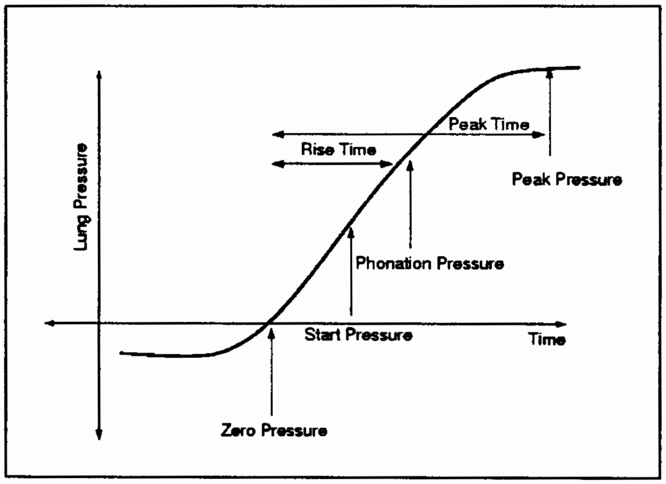
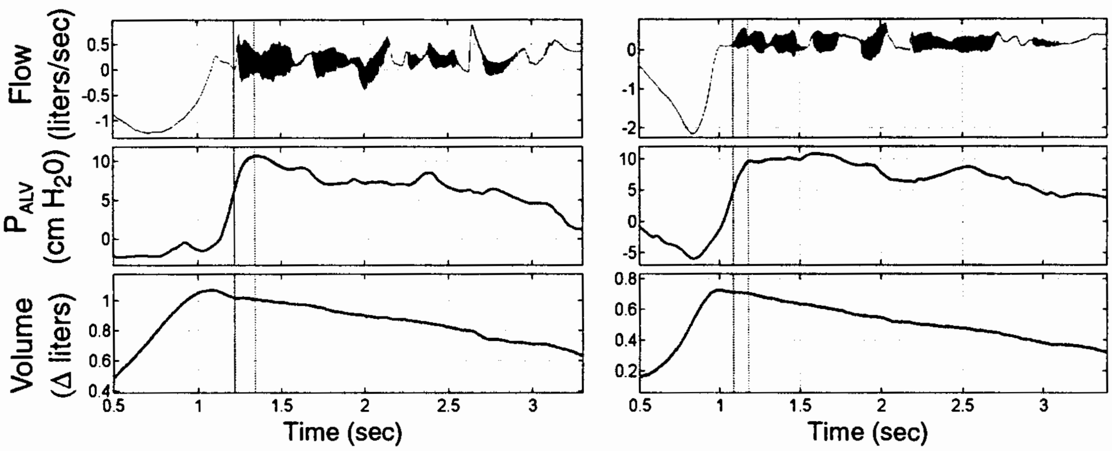


Figure 4.1: Initiation landmarks. Zero flow, utterance onset, and phonation onset are timing landmarks determined from the acoustics and flow signals. Start pressure and phonation pressure are the alveolar pressures corresponding to the start of the utterance and phonation onset, respectively. The timing landmark determined from the pressure signal is the time first peak in pressure. Rise time is defined as the time from zero flow to phonation onset. Peak time is defined as the time from zero flow to the first pressure peak.

Also seen in Figure 4.1 are the measures of rise time and peak time. The term rise time is used to refer to the time between the zero pressure instance and the onset of phonation.

The term peak time is used to refer to the time between the zero pressure instance and the time of the initial peak in alveolar pressure.

The general subglottal pressure pattern has been described as a rise, a period of level or declining pressure, and finally a fall in pressure. (See section "The role of the respiration system in speech production" on page 20.) For the data in the present study, the first peak in pressure could be a true peak with a rising slope prior to the peak and a falling slope following the peak shape (see Figure 4.2a). Or in some cases, the peak was labeled at a knee prior to a gradual slope rise or extended plateau. (See Figure 4.2b).



(a) (b)

Figure 4.2: Sample initiations for plots against time of airflow, alveolar pressure and lung volume from Subject #6, a female speaker. Vertical lines mark utterance onset and pressure peak, in sequence. (a) The utterance "The large round rock is in the path," shows an initial pressure peak. (b) The utterance "It'll be a brief walk down Willow Street," shows an initial pressure plateau.

At the end of the utterance, landmark times for end of phonation and end of utterance were determined, as shown in Figure 4.3. It was not possible to reliably determine a final pressure peak. Figure 4.4a has a clear final peak in pressure at point 1. However, as seen in Figure 4.4b, the shape of the pressure curve is affected by the resistance to the airflow. Segments with lower resistances such /s/ or /h/ can lead to dips in pressure that are not reflective of respiratory system actions. In the figure, points 2, 3, or 4 could be candidates for a final pressure peak.

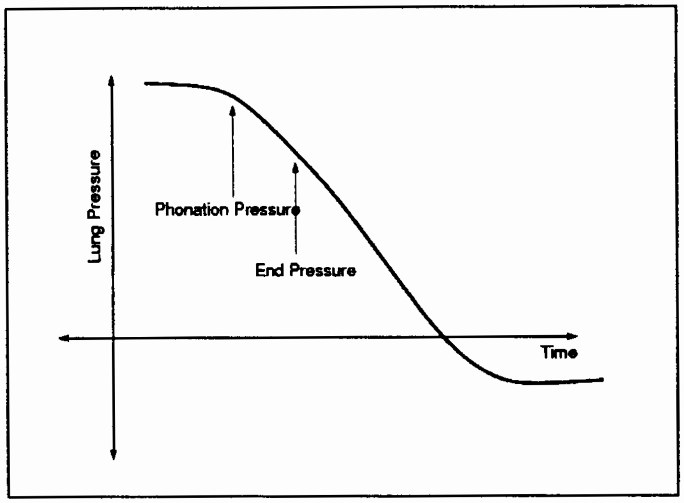


Figure 4.3: Utterance termination landmarks. Landmarks for termination of phonation and utterance termination are determined from the acoustics and the airflow.

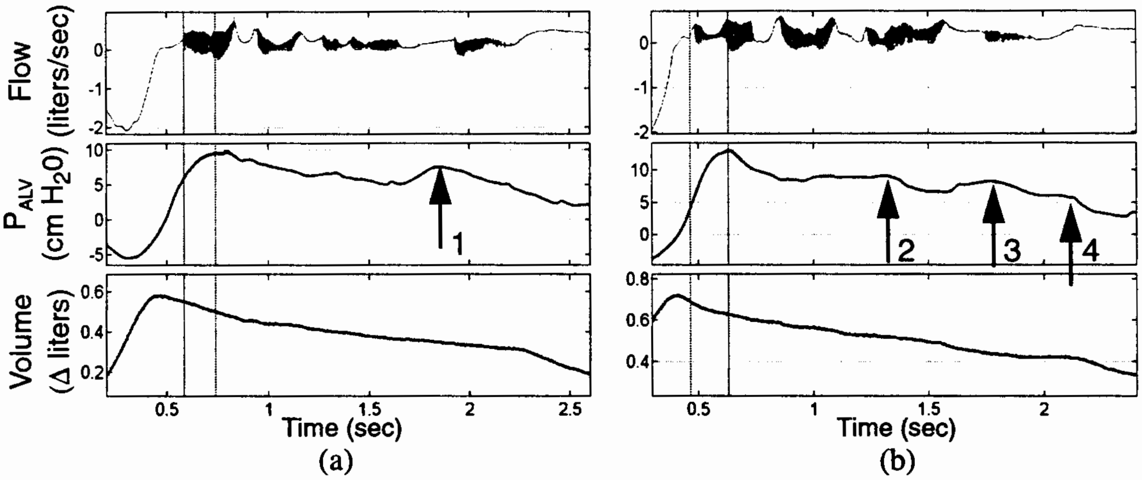


Figure 4.4: Example terminations for plots against time of airflow, alveolar pressure and lung volume from subject #6. (a) At point (1) there is a clear final peak in the alveolar pressure for the utterance "Alissa will say the word sew." (b) Points (2), (3), and (4) are all possible candidates for a final peak in alveolar pressure for the utterance "The routes are listed on the sheet." Segmental variations, such as fricatives or glottal stops, can cause pressure dips or rises that are unrelated to explicit respiratory actions. Vertical lines mark the start of the utterance and the initial alveolar pressure peak.

4.3 Characteristics of the Final Data Set

The following sections examine some of the segmental and landmark-related characteristics of the final utterance set.

4.3.1 Boundary Speech Segments

Onset and offset speech segments contain primarily voiced sonorants (in this case, vowels and the segment /y/), stop consonants, fricatives and /h/ sounds. Four of the utterances end with the segment /n/ and those are the only nasals in boundary locations. The concentrations of these segments for onsets and offsets are given in Table 4.2. In the final utterance set, 33.6% contained a syntactic boundary denoted by a comma and 20% of the utterances were read as part of a short paragraph.

Table 4.2: Segment percentages for utterance onset and offset. Reducible vowels are those vowel segments that are commonly reduced (/ʌ/ and /ɪ/).

Utterance Onset			
Vowel Onset	44%	Consonant Onset	56%
Full vowel	53%	Stop	24%
Reducible Vowel	47%	Voiceless Fricative	39%
		/ð/ as in "the"	25%
		/h/	7.5%
		/y/	4.5%
Utterance Offset			
Vowel Offset	57.3%	Consonant Offset	42.7%
Full vowel	74.6%	Stop	48.5%
Reducible vowel	25.4%	Voiceless Fricative	47.6%
		/n/	3.9%

Chapter 5

Passive and Active Respiratory Forces

5.1 Passive and Active Forces

Two major classes of force are available in the respiratory system. A speaker can apply muscular forces or make use of passive recoil forces. A model of the interaction of these forces (including active muscular forces) in the respiratory system is shown in Figure 5.1.

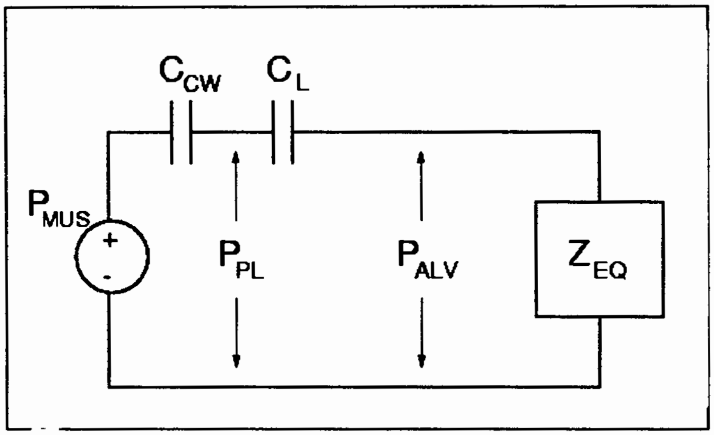


Figure 5.1: Simple model of the respiratory system using pressure as the across variable and volume velocity as the through variable. P_{PL} is the pleural pressure and P_{ALV} is the alveolar pressure. C_{CW} is the compliance of the relaxed chest wall and C_L is the compliance of the lungs. Z_{EQ} is the equivalent impedance facing air leaving the lungs. P_{MUS} is the net muscular force applied to the system.

For the analogy of pressure as the across variable and volume velocity as the through variable, the active drive to the system is muscular force (P_{MUS}). The two compliant elements represent the compliance C_{CW} of the relaxed chest wall and the compliance C_L of the lungs. The compliance of these elements is a function of lung volume. For lung volumes above or below V_{rel}, the compliant elements will exert pressure in a direction to return to rest. The pressure drops at the leftmost end of the model are related as:

$$P_{PL} = P_{MUS} + P_{CW} \tag{5.1}$$

Campbell diagrams are a graphical representation of the above equation and provide a technique for estimating regions of active muscular force and passive recoil. The following sections will give a brief review of Campbell diagrams and provide examples of the types of behavior seen in the data set. These methods will be used in Chapter 6 to examine utterance boundaries.

5.2 Campbell Diagrams and Muscular Effort

A Campbell diagram is a plot of lung volume versus pleural pressure that includes, for reference, a plot representing the pressure-volume characteristic of the relaxed chest wall. A sketch of a Campbell diagram is shown in Figure 5.2. When net inspiratory muscular force is applied to expand the chest wall, pleural pressure becomes increasingly negative. The net muscular pressure needed to depart from the relaxed chest wall is shown in the diagram as the distance between the active inhalation curve and the relaxed chest wall curve These curves only give an indication of *net* muscular force. Both inspiratory and expiratory muscles may be active simultaneously, but the position of the chest wall and the pleural pressure are a result of net forces applied to the system. For a more detailed discussion see, for example, Rahn et al. (1946), Mead et al. (1985) or Loring (1998).

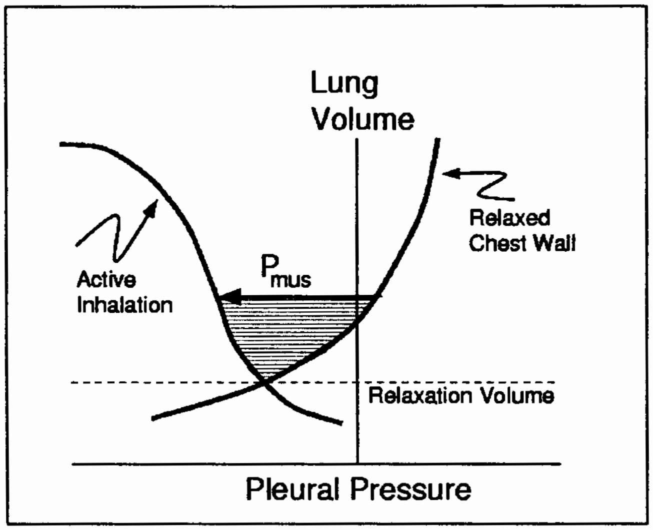


Figure 5.2: Campbell diagram. Muscular force is required for inhalation. The net magnitude of the pressure generated by the muscular effort is the pressure difference between an active inhalation curve and the relaxed chest wall curve (adapted from Loring,1998).

For example, if a person exhaled using net braking effort from a predominance of inspiratory muscle activity, the Campbell diagram would take on a shape similar to that of Figure 5.3a. The course of the exhalation falls to the left of the relaxed chest wall curve. An exhalation using net expiratory muscular force over much of the exhalation (similar to Figure 5.3b) shows the curve traveling to the right of the relaxed chest wall curve. Any portions of the exhalation that trace the relaxed chest wall curve are times of zero net muscular force. During such portions, either there is no muscular force applied or else the inspiratory muscular force balances the expiratory muscular force. The following section explains the method used to estimate the relaxed chest wall curve for the subjects in this study and uses the Campbell diagram to provide an overview of the data in this study.

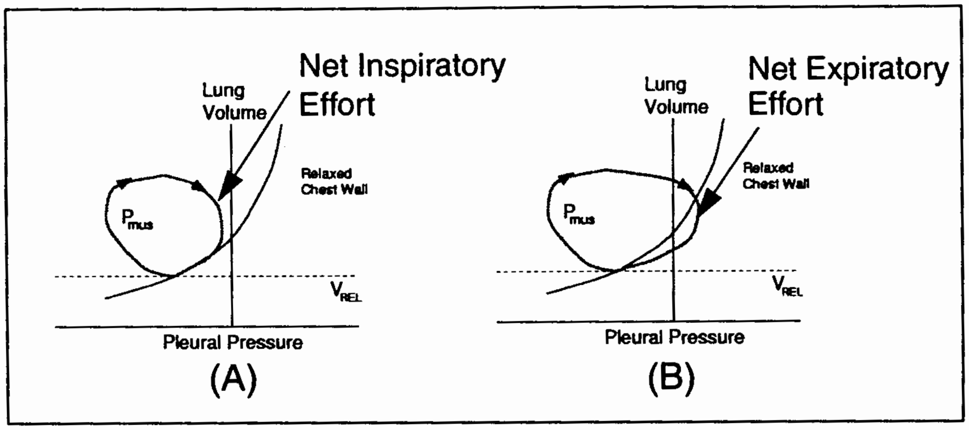


Figure 5.3: Examples of net muscular force during breathing. (a) Net inspiratory muscular pressure during exhalation (b) net expiratory muscular force during expiration. Regions of the curve that trace the relaxed chest wall curve are times of zero net muscular force.

5.3 Estimating the Relaxed Chest Wall Curve

The relaxed chest wall curve was not directly measured for the subjects in this experiment. Estimates of the curves for each subject were made using the following method.

The relaxation characteristic for the chest wall, the lungs, and the respiration system as a whole were shown in section 2.2 “Mechanics”, and are repeated here in Figure 5.4 for data measured by Knowles et al. (1959). Using this figure as a reference, estimates of four points on the relaxed chest wall curve were made. During acquisition of the physiological signals, a measure was made for the static recoil curve of the lungs in order to calibrate esophageal pressure to alveolar pressure. This is the curve corresponding to P_L in the figure. At FRC, the magnitude of the chest wall pressure is equal to the magnitude of the lung recoil pressure (P_{CW}=P_L) and this provided one of the data points for the estimated curve.

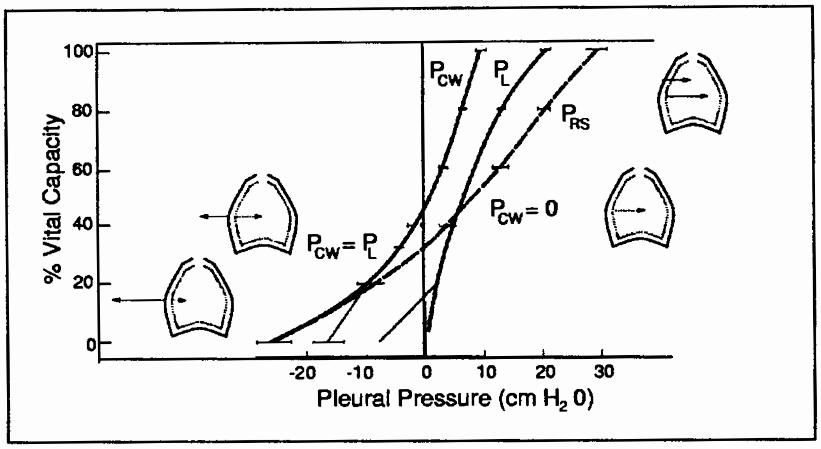


Figure 5.4: Average recoil curves. The data are from Knowles et al., 1959 and show results from four male subjects. (See legend for Figure 2.3)

. At 15-20% above FRC, the chest wall recoil is zero. For the subjects in the present study, the second point on the P_{CW} relaxation curve was set to zero at 20%VC above FRC. Two additional points were taken from the averages shown in Figure 5.4. At 65% VC above FRC, the recoil pressure was set to 6 cm $H_2$0 and at 15% VC below FRC, the recoil pressure was set to -8 cm $H_2$0. These are general landmarks that are dependent upon each subject's VC. These points were fit with a simple second order curve. An example is shown in Figure 5.5 for Subject #4. A summary of the points used in the estimation is given in Table 5.1.

Table 5.1: Data points for estimation of the relaxed chest wall curve.

Volume (%VC from FRC)	Pleural Pressure (cm $H_2$0)
-15% VC	-8
0% VC	P_L
+20 %VC	0
+65 %VC	6

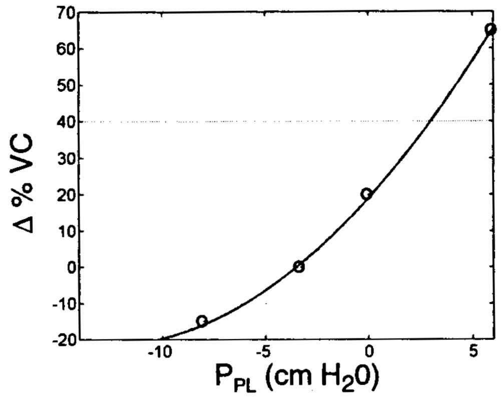


Figure 5.5: Estimated chest wall relaxation curve for Subject #4. Four points were fit with a second order curve. The points are given as black circles.

The relaxed chest wall curves are commonly used to make an estimate of C_{CW} for the linear region above FRC. For the four subjects in this study these values are given in Table 5.2. Published values for the mean and standard deviation of measured values for C_{CW} such as those found in Estenne et al. (1985) can provide a check on the validity of the values. The published values are grouped according to age and gender. However, no statistical difference was found for gender. Subjects #4, #5 and #6 fell within one standard deviation for their age, height, and weight as given in Estenne et al. (1985). The estimate for Subject #7 fell almost two standard deviations from the mean for the appropriate group and is therefore the least reliable estimate across the four subjects.

Table 5.2: Estimated compliance of the chest wall in liters/cm $H_2$0 for the linear region above FRC.

	Subject #4	Subject #5	Subject #6	Subject #7
C_{CW}	0.240	0.142	0.246	0.102

The validity of the estimated curve depends on two major factors (1) the accuracy of the FRC value as obtained from resting breathing records and (2) the accuracy of the chosen landmarks for the three other points. These are the point where P_{CW}=0, the value of

P_{CW} at +65% VC from FRC, and the value of P_{CW} at -15%VC from FRC. A graphical comparison of the types of adjustments that might be made in the estimated curve from these two factors is shown in Figure 5.6. In part (a), adjustments were made to the estimate of FRC. The upper dotted line is the estimated curve if the FRC value was raised by 0.5 liters. The lower dotted curve is the estimated curve if the FRC value was lowered by 0.5 liters. In part (b), the measured value of FRC was used but possible variations in the slope are shown using values from the appropriate group in Estenne (1985). The line with a steeper slope represents the mean C_{CW} plus one standard deviation. The line with the shallower slope represents the mean C_{CW} minus one standard deviation. A brief comparison of these types of shifts will be made in the section "Net Muscular Force" on page 69 for results concerning net muscular activity during speech.

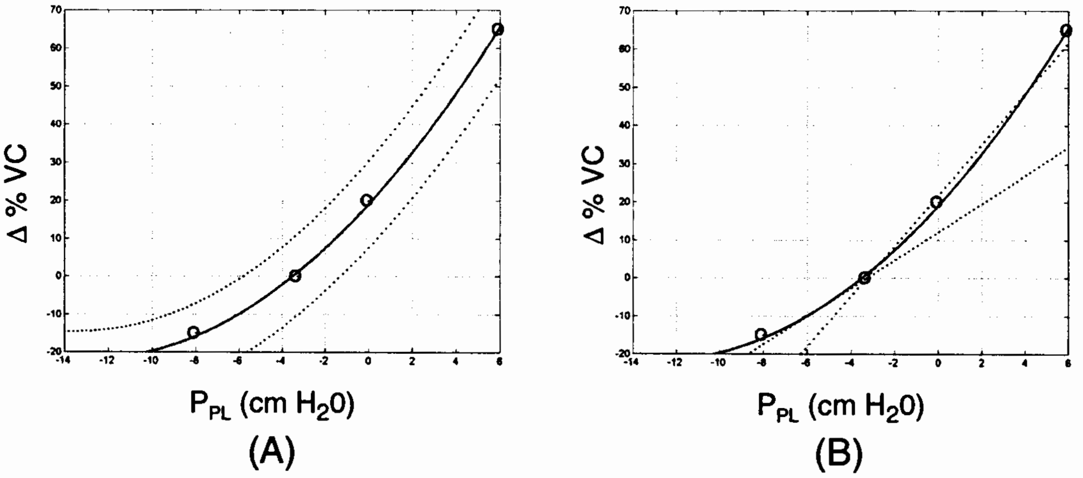


Figure 5.6: Types of possible errors in the relaxed chest wall curve estimate demonstrated on Subject #4. In both cases, the dark solid curve is the estimated curve. (a) The upper and lower dotted lines show the shift in curve location if FRC was raised by 0.5 liters and lowered by 0.5 liters respectively. (b) The dotted curves are generated using the FRC estimate obtained from resting breathing for the subject, but show the slope variation using values for the mean and standard deviation of C_{CW} as obtained from Estenne et al. (1985) for the appropriate age, height, and weight of the subject.

5.4 Campbell Diagrams for Single Utterances

Time is an important variable in speech. The timing the of articulators, the timing of transitions between sound segments, the durations of segments, and the time course of prosodic changes are all examples of the importance of the time track of information in speech. However, the pressure-volume relationship has meaningful interpretations for the dynamics of the respiratory system and it does not contain a time measure. An attempt to bring some element of time into the Campbell diagrams during speech is given in Figure 5.7. This is an example of a time-lapse series of events that makes use of the timing landmarks established in the section "Definitions and Landmarks" on page 50. Figure 5.7a shows the time tracks for audio, P_{ALV}, airflow and lung volume. Vertical lines correspond to the time-lapse images of Figure 5.7b.

Each utterance file in the data set contains some portion of time preceding and following the utterance. In Figure 5.7b, panel 1 shows that period of time prior to the onset of sound. In this case there is an inhalation and the start of an exhalation. Panel 2 extends this curve to include the start of the utterance through the first alveolar pressure peak. Panel 3 includes the rest of the utterance through the end of all sound, and Panel 4 adds on the silence after the end of the utterance.

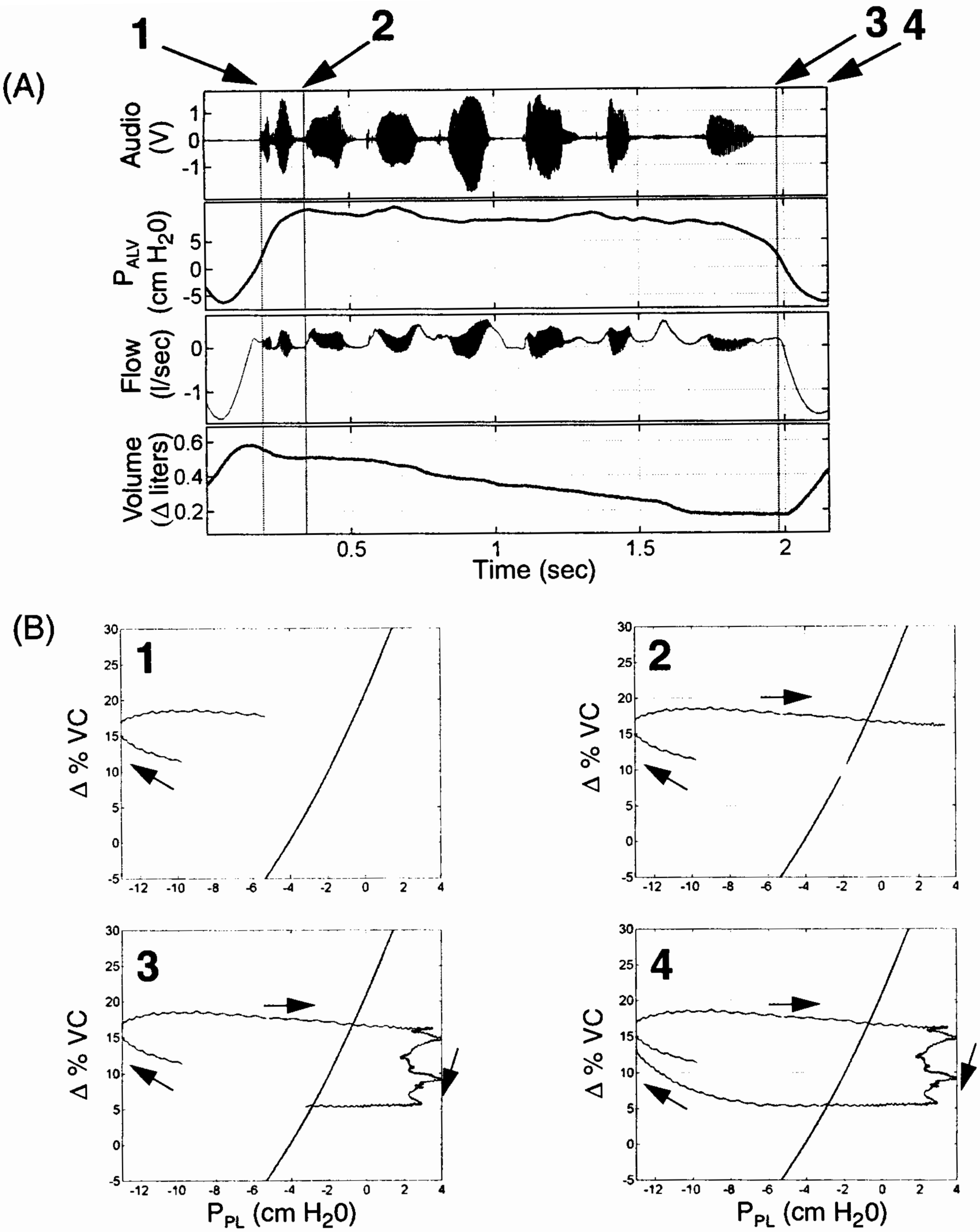


Figure 5.7: (a) Time tracks for audio, alveolar pressure, air flow, and lung volume from Subject #6. The utterance is "It'll be a brief walk down Fifth Street," and was read within a paragraph. Vertical bars correspond to the time segments shown in panels 1-4. (b) Pressure-volume curve for the example in (a). The gray line is an estimate of the relaxed chest wall curve. (1) Start of the acquisition until the onset of the utterance. (2) Start of the acquisition through the first pressure peak. (3) Start of the acquisition through the end of the utterance. (4) The entire acquisition file including a segment of time following the end of the utterance.

In this example, utterance onset occurs during a time of little volume change but rapid pressure change. The pressure peak occurs to the right of the relaxed chest wall curve and the bulk of the utterance stays to the right of the curve. Details of the timing of specific portions of the utterance are shown in the example of Figure 5.8. Each circle along the pressure-volume curve is separated, in time, by 80 milliseconds. Onset and offset timing landmarks are given by vertical bars. Regions that have large distances between the circles are times of rapid pressure changes, rapid volume changes, or both rapid pressure and volume changes. If the circles are close together, the changes are much slower by comparison. Onset is a rapid action through the first pressure peak with little change in volume. After that point (3 in Figure 5.7) the circles are more closely spaced through the offset of phonation. Between points 4 and 5, there is a quick drop in pleural pressure of about 2 cm H_20.

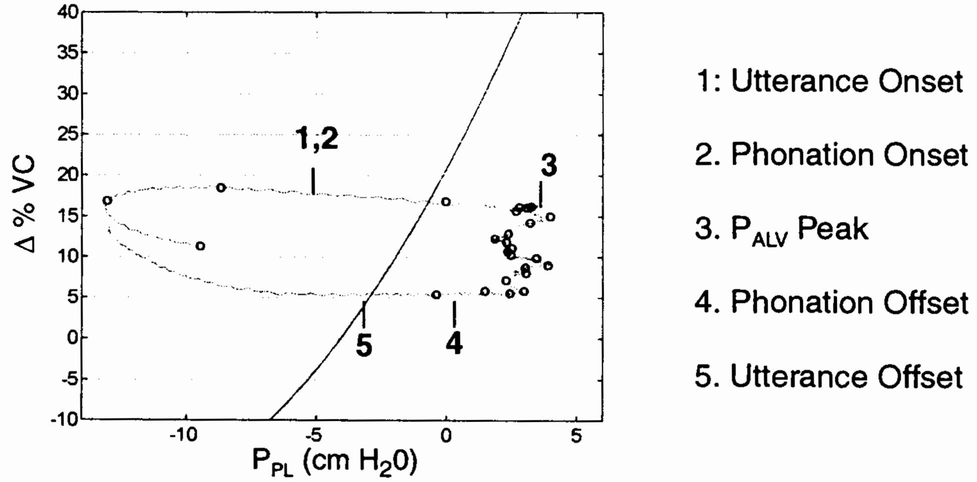


Figure 5.8: Time course of the pressure-volume curve. Each circle along the curve is separated by 80 milliseconds. The utterance is “It’ll be a brief walk down Fifth Street” as read in a paragraph by Subject #5. Landmarks are indicated by the numeric legend.

From this simple example, an initial description of the respiratory actions at boundaries can be made. Onset is rapid with little change in volume. Timing landmarks for utterance onset and phonation onset occur during this segment of rapid pressure change and the peak in alveolar pressure corresponds, roughly, to a point to the right of the relaxed chest wall curve. This general outline will be further examined with a wide range of utterances from all four speakers in Chapter 6.

In the following sections, it will be shown that the end of the utterance frequently occurs to the right of the relaxed chest wall curve, implying the presence of net expiratory muscular effort. The details of these actions, as well as the laryngeal correlations of phonation offset will also be examined in Chapter 6.

Chapter 6

Boundaries with Initiations and Terminations of Breaths

During speech, both recoil and muscular forces are manipulated to generate the drive for speech and to maintain the required gas exchange function of the respiratory system. These two aspects are reconciled with cognitive choices to create speech prosody. For example, consider the end of an utterance. As discussed in Chapter 1, regional prosodic cues at the end of an utterance may include a drop in signal amplitude, a lengthening in duration of the ending speech segment, and a drop in the fundamental frequency. However, these cues could have been *cognitively* chosen to be anything - a rise in signal amplitude or a shortening of segmental duration. It may be that the cues that *are* implemented are those cues that arise from the effort in managing the dynamics of the respiratory system. As such, their realization is influenced by changes in that system.

The questions addressed in this chapter focus on the details of the pressure dynamics, the laryngeal area dynamics, and the corresponding acoustics at utterance boundaries that are aligned with the initiation and termination of exhalations. Specifically (1) How are the dynamics of initiation and termination related to the recoil pressures of the chest wall and applied muscular pressures? (2) Are timing and/or pressure landmarks correlated to prosodically prominent syllables? (3) What are the influences, if any, of the boundary sound segments (and the airway resistance of those sound segments) on the shape of the pressure curve at utterance onset? (4) In what way are the amplitude changes influenced by coordination of the area of the vocal tract constriction and alveolar pressure at voicing termination?

6.1 Utterance Initiation

6.1.1 Utterance Initiation Landmarks

When a speaker takes a breath in preparation to speaking, the pressure in the lungs is less than atmospheric and air flows into the lungs. At the end of the inhalation, the transition is made to exhalation and speaking. During this transition, the pressure in the lungs must be raised to a level that is higher than atmospheric pressure for exhalation to occur. This pressure must also be in the proper range to create the desired sound sources for the speech. For the purposes of this study, utterance initiation has been defined to span the start of the exhalation until the initial alveolar pressure peak. Averages and standard deviations for the pressure and timing values at utterance initiation for the landmarks defined in section 4.2 are given in Table 6.1.

Table 6.1: Pressure and timing values corresponding to acoustic and physiologic landmarks at utterance initiation. Values are given as 'mean (standard deviation).' Time is given in milliseconds, pressure in cm H_20, and lung volume (LV) in %VC from FRC.

Quantity	Subject #4	Subject #5	Subject #6	Subject #7
Time (msec)				
Rise Time	120 (60)	90 (50)	150 (70)	140 (80)
Peak Time	240 (60)	200 (60)	260 (70)	280 (80)
Alveolar Pressure (cm H_20)				
Start P_{ALV}	3.33 (1.43)	3.52 (2.13)	4.62 (2.24)	2.68 (1.61)
Start Phonation P_{ALV}	5.04 (1.79)	6.71 (3.11)	6.87 (1.95)	3.77 (1.45)
Peak P_{ALV}	8.25 (1.09)	11.92 (1.82)	9.94 (1.16)	6.54 (0.92)
Lung Volume (Δ %VC from FRC)				
LV at zero flow	30.0 (4.4)	32.0 (7.7)	17.6 (3.7)	13.4 (6.5)
LV at utterance start	29.5 (4.3)	31.3 (7.8)	16.9 (3.6)	13.2 (6.6)
LV at start phonation	29.2 (4.2)	30.2 (7.9)	16.5 (3.5)	13.0 (6.6)
LV at peak P_{ALV}	28.8 (4.3)	29.0 (8.0)	16.0 (3.6)	12.6 (6.9)

The rise phase for pressure has been stated to be from 120-200 milliseconds as discussed in Chapter 1. Although t.ıe specific landmarks used to estimate that value are unclear, it may be most closely compared to the parameter peak time in the present study. Average peak times for all four subjects are longer than this stated range. The change in lung volume from the start of the exhalation to the peak in pressure is on the order to 1 to 3% VC.

The following sections will examine utterance initiation in terms of net muscular force (braking vs. assisting the recoil forces), perceptually rated prominences, and initial sound segment type. Results will be used to propose prototypes for utterance initiation.

6.1.2 Net Muscular Force

The major goal of this study is to characterize the respiratory dynamics at speech boundaries and determine the correlations to prosodic acoustic cues for those boundaries. Toward that effort, this section will examine the major trends in net muscular force, as determined through the use of modified Campbell diagrams, and the corresponding timing landmarks for utterance initiation.

As previously introduced in Chapter 5, Campbell diagrams provide an estimate of net muscular force by plotting the relationship between lung volume and pleural pressure and allowing comparison to the relaxed chest wall characteristic. Two typical examples are shown in Figure 6.1 for Subject #6. Vertical bars mark timing landmarks and the dark circles are separated in time by 80 milliseconds. Figure 6.1a is an utterance read in isolation and Figure 6.1b is an utterance read as part of a short paragraph. Utterance onset begins to the left of the relaxed chest wall characteristic and indicates some level of net inspiratory force. However, the change in pressure is rapid, taking approximately 400 milliseconds to span the peak of the inhalation (largest inspiratory P_{MUS}) to the peak in alveolar pressure at Point #3 for the utterance read in isolation, and about 240 milliseconds for the utterance read as part of a paragraph. At the end of an inhalation, whether resting breathing or speech breathing, the inspiratory muscles do not instantaneously end their action. Instead, inspiratory muscular activity tapers as the exhalation proceeds. In the case of speech breathing, expiratory muscles are generally considered to become active as the inspiratory muscles reduce in activity. (For a review see Weismer,1985) In this view of speech breathing, both inspiratory and expiratory muscles may be continuously active throughout the utterance, to greater or lesser degrees, to meet the demands of the utterance without having to initiate from complete rest.

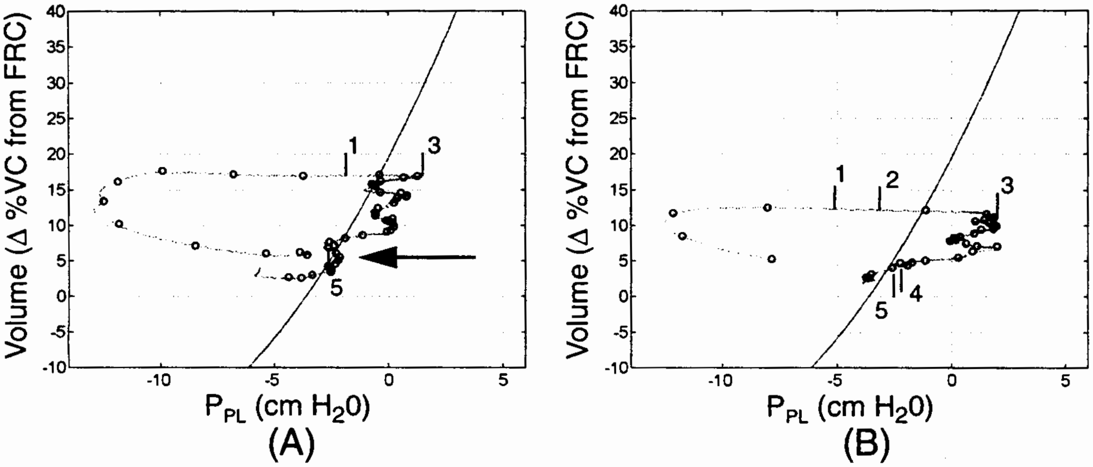


Figure 6.1: Pressure-volume examples from Subject #6. (a) The utterance, "Ali will have the best seat in the show," was read in isolation. The arrow points to the segment of time after the end of the utterance, while the speaker waits to read the next utterance. (b) The utterance, "There is a large round rock in the path," was read as part of a short paragraph. Vertical bars denote timing landmarks (1=Utterance onset, 2=Phonation onset, 3=Initial pressure peak, 4=Phonation offset, 5=Utterance offset), and dark circles are separated by 80 milliseconds.

At the very end of the utterance in Figure 6.1a, there is a brief period of time after the end of the utterance that could be interpreted as arising from the experimental protocol. The subjects read utterances from cards. The timing of the card presentation was controlled by the experimenter. As each subject waited for the next card to be presented, most postponed finishing exhalation until the next card appeared. (See Figure 6.1a around the arrow.) When the subjects saw the next card, they quickly finished the exhalation and inhaled to prepare for the next utterance.

Although the speakers were given no special instructions concerning their use of lung volume, occasionally, a speaker would produce utterances at higher lung volumes. Figure 6.2 shows two example utterances, read in isolation, from Subject #5. In Figure 6.2a, the utterance began near 37%VC above FRC and ended near 25% above FRC. In Figure 6.2b, the utterance began near 20%VC above FRC, which is closer to the lung volumes used most frequently by this subject. For the utterance of Figure 6.2a, the speaker used inspiratory muscle effort throughout. Again, this statement is subject to the accuracy of the estimate of the relaxed chest wall curve. However, the net muscular activity, whether inspiratory, expiratory or some combination, is on a different level than for the utterance of Figure 6.2b.

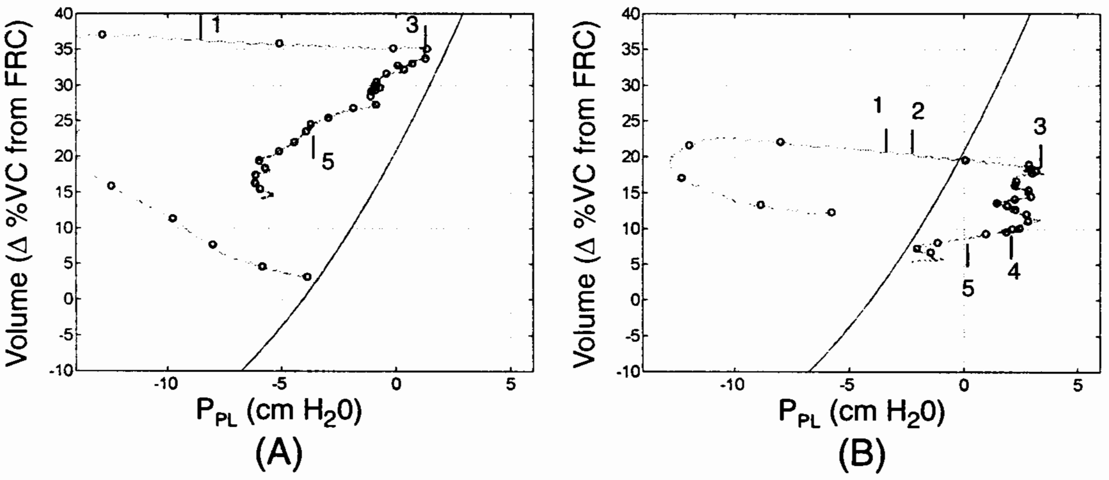


Figure 6.2: Pressure-volume examples from Subject #5. (a) The utterance, "Alissa will say the word sew," was produced at a higher lung volume and makes use of a different level of muscular force than the utterance of part (b). (b) The utterance, "The routes are listed on the sheet," was produced at a lower lung volume and makes use of expiratory muscular force. Vertical bars denote timing landmarks (1=Utterance onset, 2=Phonation onset, 3=Initial pressure peak, 4=Phonation offset, 5=Utterance offset), and dark circles are separated by 80 milliseconds.

Two of the major initiation landmarks, the time of the start of any sound (Point #1) and the time of the peak in alveolar pressure (Point #3), are shown for all utterances in the data set and for all speakers in Figure 6.3. The grey 'x' symbols mark the time of utterance onset. The dark 'o' symbols mark the peak in alveolar pressure. Subject #4 and Subject #6 produced utterances that began within a limited lung volume range. The pressure peak was clustered close to, or slightly to the right of the estimated relaxation curve. Subjects #5 and #7 used a wider range of lung volumes at utterance initiation. For the initial alveolar pressure peak data, values at higher lung volumes occur more to the left than values at lower lung volumes.

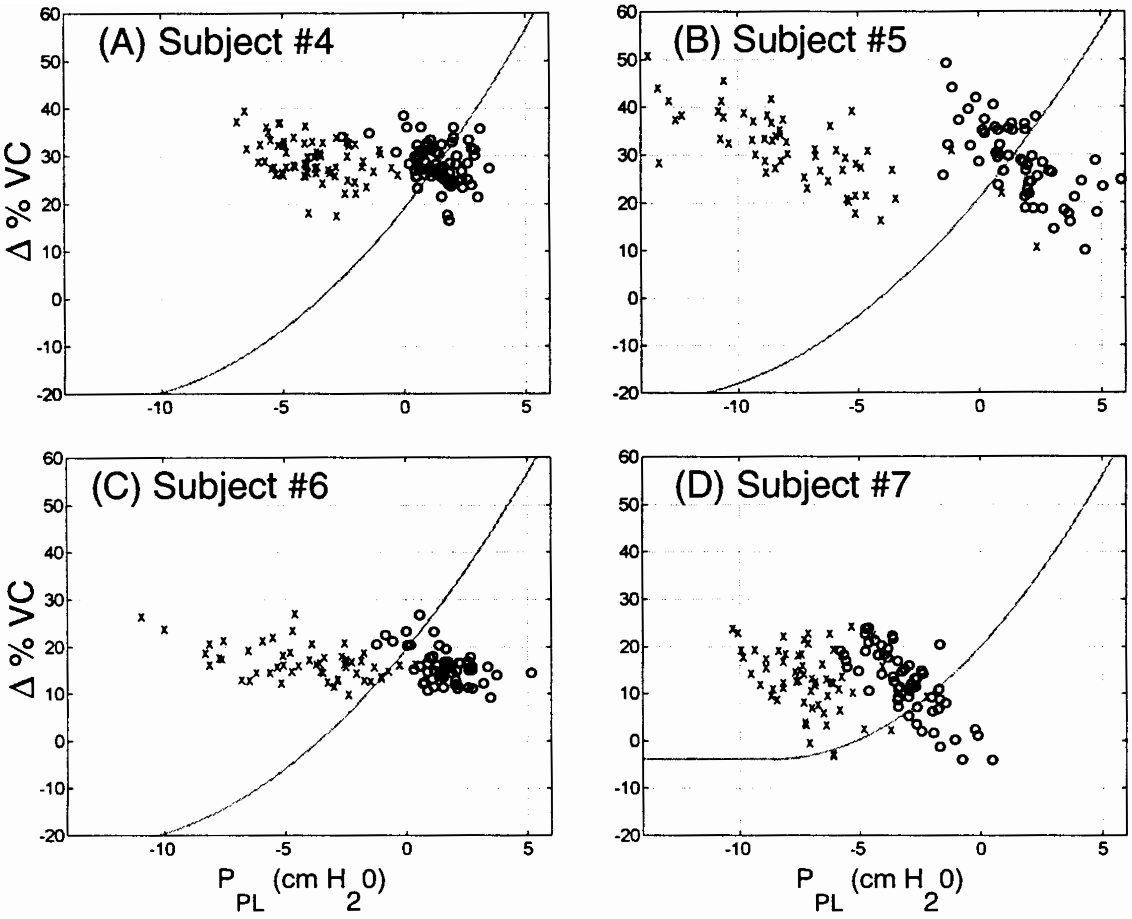


Figure 6.3: Pressure-volume onset landmarks. The 'x' symbols correspond to the landmark of utterance onset. The 'o' symbols correspond to the landmark of initial alveolar pressure peak. The estimated relaxed chest wall curve is plotted for reference. (a) Subject #4 (b) Subject #5 (c) Subject #6 (d) Subject #7.

These four subjects predominantly began utterances during net inspiratory muscular force; whether from a tapering of inspiratory activity prior to expiratory muscles use or from deliberate braking of the recoil force. As mentioned in section "Estimating the Relaxed Chest Wall Curve" on page 58, the estimated chest wall curve could be shifted due to inaccuracies in the estimate of FRC and/or it could be altered in slope due to inappropriate landmarks. A brief comparison of these types of alterations in the chest wall curve is given in Figure 6.4 for the data of panels (A) and (B) from Figure 6.3. While the variations in the estimated chest wall curve affect trends regarding the peak in alveolar pressure, the utterance onsets still occur primarily during net inspiratory muscular force.

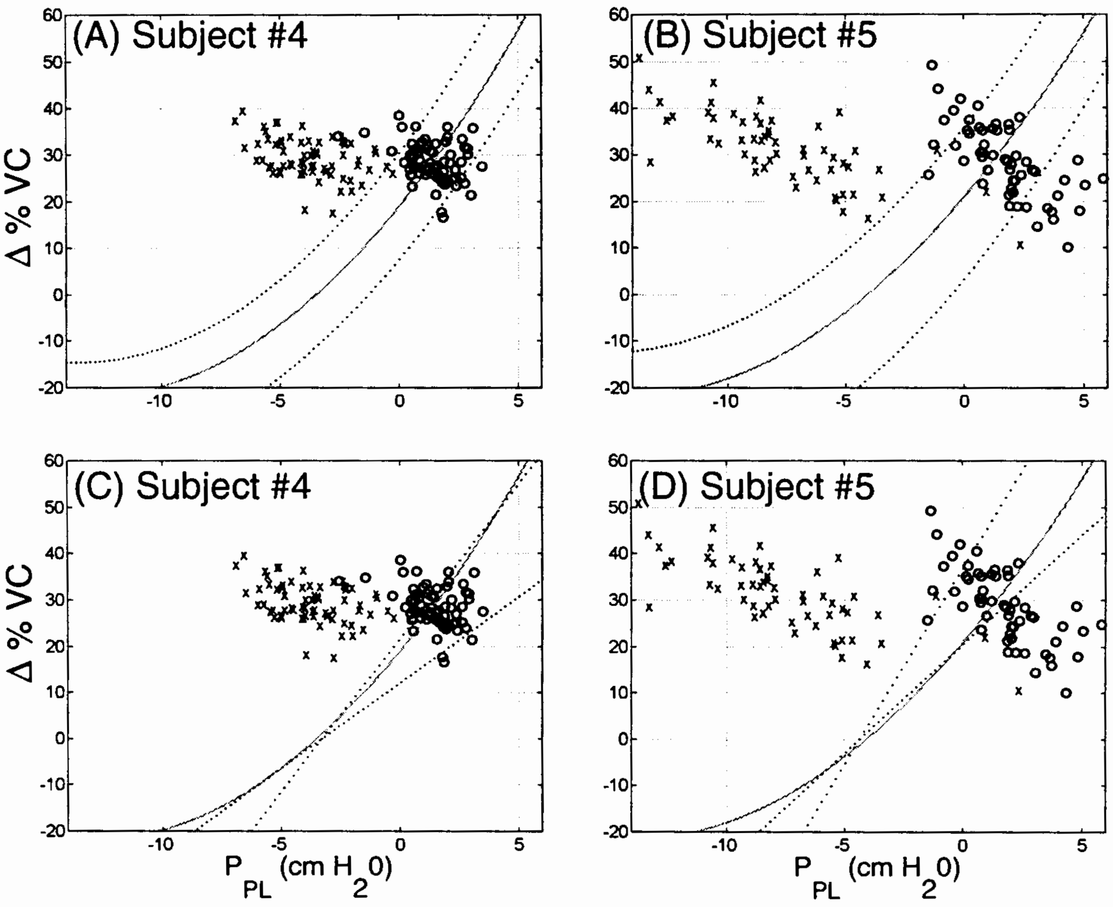


Figure 6.4: Comparison of pressure-volume onset landmarks and chest wall curve alterations for Subjects #4 and #5. The ‘x’ symbols correspond to the landmark of utterance onset. The ‘o’ symbols correspond to the landmark of initial alveolar pressure peak. The estimated relaxed chest wall curve is plotted for reference in the solid gray line. In panels (A) and (B) the relaxed chest wall curves given by the dotted lines are for an estimate of FRC that is raised by 0.5 liters and one that is lowered by 0.5 liters respectively. In panels (C) and (D) a linear estimate of C_{CW} from Estenne (1985) is used to show plus and minus one standard deviation around the mean slope. For these plots, the measured value of FRC was used as the intercept point for the line.

6.1.3 Rated Prominences

In speech production, prosodic cues such as amplitude and spectral changes are influenced by both the laryngeal configuration and the subglottal pressure. The first perceived prominence may be, in part, cued by these acoustic changes as well as by fundamental frequency and context information. Because prosodic prominences within an utterance have been linked to subglottal pressure changes (e.g. Ladefoged,1967, Lieberman,1967, Atkinson,1972), it may be that the first perceived prominence in the utterance is correlated to the initial pressure peak. One possible strategy could be to time the pressure peak with the first prominent syllable in the utterance. Another possibility is that the pressure rise is unrelated to the first prominent syllable and is guided solely by the physical requirement of generating a pressure that is in the proper range for speech as well as managing the recoil forces of the chest wall. This section compares the timing of the first perceived prominent syllable with the timing of the initial alveolar pressure peak.

Initial prominent syllables were rated by a listening test as described in the section "Perceptual Ratings of Prominences" on page 45. The time limits of these syllables were determined through visual inspection of the audio signal and the corresponding spectrogram as well as through listening to the audio signal. Some of the acoustic and physiologic landmarks are listed in Table 6.2 for the following cases: (1) the first syllable in the utterance is perceived as the initial prominence (2) the second syllable in the utterance is perceived as the initial prominence and (3) the initial perceived prominent syllable is later in the utterance than the first or second syllable. In the averages given in the table, three of the four subjects show a trend for phonation pressure to decrease and for rise time to shorten as the initial prominence is moved later into the utterance. The exception is Subject #7.

Table 6.2: Pressure and timing values corresponding to syllable prominence ratings for the initial prominent syllable in the utterance. Values are listed as 'mean (standard deviation).' Time is given in milliseconds, pressure in cm H_20, and lung volume in %VC from FRC.

Quantity	Subject #4	Subject #5	Subject #6	Subject #7
First Syllable is Prominent				
Rise Time	150 (70)	120 (50)	170 (70)	150 (80)
Start P_{ALV}	3.46 (1.66)	3.72 (2.64)	4.28 (2.53)	2.20 (1.59)
Phonation P_{ALV}	5.83 (2.1)	8.07 (2.8)	7.46 (2.2)	3.81 (1.6)
LV at utterance start	29.8 (4.6)	30.3 (6.7)	16.0 (3.2)	12.5 (6.8)
Second Syllable is Prominent				
Rise Time	100 (20)	70 (20)	150 (70)	120 (50)
Start P_{ALV}	2.73 (1.03)	3.60 (1.63)	5.20 (1.97)	2.99 (1.29)
Phonation P_{ALV}	4.49 (0.80)	6.07 (2.43)	6.76 (1.24)	3.84 (1.13)
LV at utterance start	28.8 (3.4)	31.3 (8.7)	18.6 (3.9)	14.6 (5.2)
Later Syllable is Prominent				
Rise Time	80 (30)	60 (30)	110 (30)	140 (90)
Start P_{ALV}	3.75 (1.16)	2.98 (1.38)	4.48 (1.47)	3.49 (1.66)
Phonation P_{ALV}	4.15 (1.39)	4.67 (3.21)	4.76 (1.08)	3.62 (1.49)
LV at utterance start	29.8 (4.6)	33.5 (9.0)	15.9 (3.3)	13.3 (7.6)

Figure 6.5 considers two groups of utterances: (1) those utterances where the initial rated prominence falls in the first or second syllable (gray bars) and (2) those utterances where the initial rated prominences fall later than the second syllable of the utterance (black bars). For each group of utterances, the timing of the pressure peak either preceded, fell within, or followed the limits of the perceptually prominent syllable. The percentages for each type of timing and for each utterance group are shown in Figure 6.5.

For these results, the pressure peak usually occurs in the initial prominent syllable if that syllable is the first or second syllable in the utterance. For this study, this pattern occurred in an average of 82% of the rated utterances (where 185 utterances have initial prominences rated in the first or second syllable). When the initial perceived prominence occurred later in the utterance, the pressure peak preceded that syllable in an average of 91% of the utterances (where 45 utterances had initial prominences rated in the third, fourth, or fifth syllable). For such late initial prominences, the pressure peaked in the first syllable for 29% of the utterances, in the second syllable for 58% of the utterances, and in the third syllable 13% of the utterances. As stated in Chapter 3, the level of agreement on the listening panel was less for late initial prominences. The speakers showed consistent ability to make small adjustments in the alignment of the peak timing, in that, if the prominence occurred on the first syllable, pressure peaked in the first syllable. If the prominence occurred in the second syllable, pressure peaked in the second syllable. However, there appeared to be limits to the degree to which the peak was delayed into the utterance or uncertainty over the rating of late prominences. Later prominences still had a pressure peak somewhere in the first three syllables.

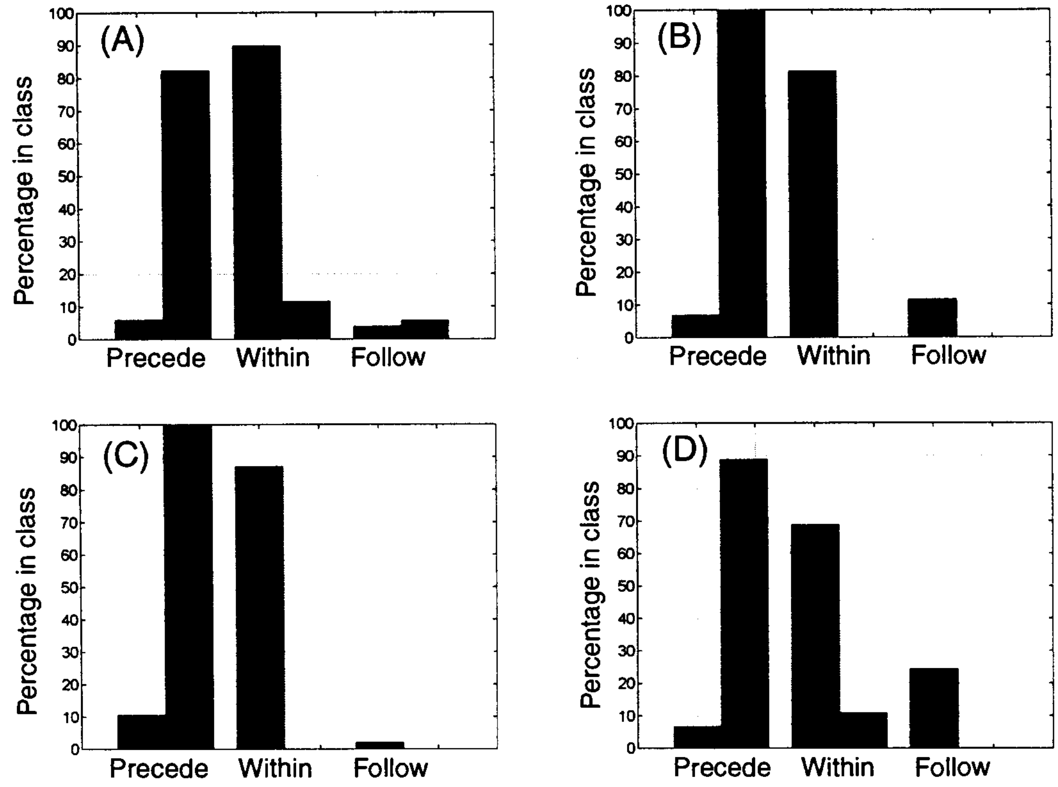


Figure 6.5: Comparison of pressure peak time to prominent syllable boundaries. The timing of the initial pressure peak either preceded, fell within, or followed the boundaries of the initial prominent syllable. That alignment is represented in the x-axis. Two groupings of perceived prominent syllables are shown for each timing alignment. Gray bars represent rated prominences that fell on the first or second syllable. Black bars represent prominences that fell on the third, fourth, or fifth syllable. (a) Subject #4 (b) Subject #5 (c) Subject #6 (d) Subject #7

The first alveolar pressure peak has been defined as the location of the first leveling off or downturn in the pressure once it begins its rise from zero pressure. Pressure may peak early, but in some cases, where the rated prominence falls late in the utterance, the pressure may continue to rise, as in Figure 6.6a. In this example, the prominence was rated from a 3/5 vote. For comparison, Figure 6.6b shows a similar utterance, with rated prominence on the same syllable (5/5 vote), but lacking such a distinct pressure plateau at the prominent syllable. Later prominences may be cued by changes that involve local pressure variations, F0 changes or changes involving control of the nature of the laryngeal closure during voicing. In both the examples the pressure appears to reach a working level or 'relatively constant level' prior to later variations at the time of the rated prominence.

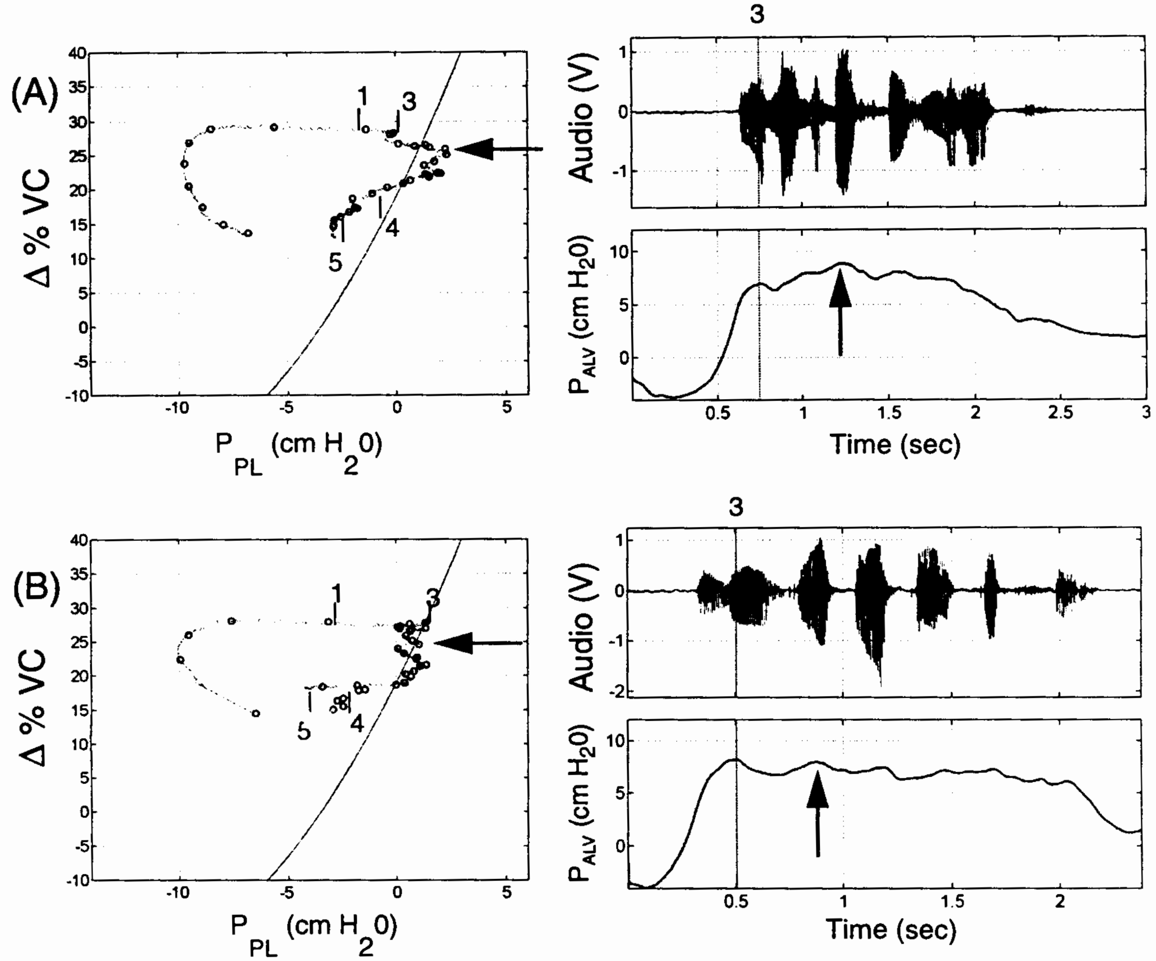


Figure 6.6: Late prominence examples for Subject #4. On the left are pressure-volume curves. Vertical bars denote timing landmarks (1=Utterance onset, 2=Phonation onset, 3=Initial pressure peak, 4=Phonation offset, 5=Utterance offset), and dark circles are separated by 80 milliseconds. On the right are the audio signal (volts) and the alveolar pressure (cm H_20) as functions of time. (a) The text "You will have the best seat in the house" was read in isolation. The initial prominence was perceived to be on "best." (b) The text "It'll be a brief walk down Fifth Street" was read as part of a paragraph. The initial prominence was perceived to be on the word "brief."

Consider the cases where the initial prominence was placed on the first or second syllable of the utterance. The speaker may have varied the rise time to place the peak within the prominent syllable, may have changed the realization of the initial syllable at the articulators (such as a very short and reduced form of "the" at utterance onset), or used some combination of those strategies. The data in Table 6.3 are the peak level and peak time means and standard deviations for the utterances as grouped by the location of the initial syllable prominence. In comparing the first and second syllable groups, except for Subject #5, the suggestion is that the peak time is slightly delayed but only on the order of 20-50 milliseconds on average. It may be possible that the timing realization of an initial syllable that is not prominent is compressed. This question would benefit from studies of articulatory dynamics.

Table 6.3: Peak pressure and timing values for the syllable prominence categories. Time is given in milliseconds and pressure in cm H_20. Values are given as 'mean (standard deviation).'

	Subject #4	Subject #5	Subject #6	Subject #7
First Syllable is Prominent				
Peak Time	230 (60)	210 (60)	240 (50)	260 (70)
Peak P_{ALV}	8.21 (1.00)	11.51 (1.64)	9.53 (1.12)	9.30 (0.93)
Second Syllable is Prominent				
Peak Time	260 (70)	190 (60)	280 (80)	280 (90)
Peak P_{ALV}	8.02 (0.73)	12.1 (1.96)	10.68 (1.03)	6.98 (0.82)
Later Syllable is Prominent				
Peak Time	240 (50)	210 (60)	260 (60)	310 (90)
Peak P_{ALV}	8.60 (1.50)	12.53 (1.94)	9.64 (0.52)	6.63 (0.85)

6.1.4 Initial Sound Segment Resistance

To create a pressure in the lungs, there must be a resistance to the airflow. At the start of the utterance, this resistance is created by the articulators and vocal folds moving into position to create the initial sound of the utterance. For initial voiced sound segments, the vocal folds may begin vibrating as the folds move together or the folds may be pressed together prior to vibrating. In the latter case, the folds are separated at the onset of phonation after pressure had reached a level sufficient to initiate phonation. The resistance to flow can be expressed as

$$R = \frac{\rho U}{2A^2} \qquad (6.1)$$

See section “Estimation of the Area of the Constriction” on page 41 for further discussion. Resistance to flow is inversely proportional to the square of the area of the constriction.

Table 6.4: Rise time and phonation pressure for a single male speaker. Subglottal pressure was estimated using a tracheal puncture. Data are from Atkinson,1973.

Initial Phoneme	Rise Time Range (msec)	Rise Time Mean (msec)	Average Phonation Pressure (cm $H_2 0$)	Number of Utterances Examined
/b/	230-360	270	6.7	8
/ð/	110-250	180	7.2	2
/w/	150-250	190	7.5	2
/h/	65-125	105	9.0	6

Lieberman (1967) characterizes the onset of an utterance in terms of the subglottal pressure estimate from an esophageal balloon for three male subjects and 19 total utterances (p. 95). He concludes that for a longer phonation delay (rise time), phonation begins at a higher subglottal pressure. Atkinson (1973) followed these results with data for a single male speaker using a tracheal puncture to estimate subglottal pressure. Average times for rise time and phonation pressure from that work are shown in Table 6.4. The speech tasks included sentences where the subject produced utterances with prominence on specific syllables. Examples are the contrast between “BEV loves Bob” (prominence on BEV) and “Bev LOVES Bob” (prominence on LOVES). At the time of data acquisition for the research of the present study, each subject produced a number of utterances with this format. However, the pressure patterns were markedly different from the utterances produced without a preset prominence pattern. While a subject is capable of a huge range of actions with the respiratory system, vocal folds, and articulators, the muscular habits were the focus of the present research. For these reasons, the utterances produced with guided prominence patterns were not included in the analysis. The current study extends the subject pool to include female subjects, uses a larger range and number of utterances, and does not require specific prosody from the speaker. Some of the acoustic and physiologic landmarks are given in Table 6.5 for the sound segments classes that begin the utterances in this task set.

Table 6.5: Pressure and timing values corresponding to acoustic and physiologic landmarks as grouped by the initial sound segment of the utterance. Time is given in milliseconds, pressure in cm H_20, and lung volume in %VC from FRC.

Quantity	Subject #4	Subject #5	Subject #6	Subject #7
Voiced Sonorant Onsets				
Rise Time	80 (30)	50 (20)	130 (70)	110 (70)
Phonation P_{ALV}	4.11 (1.19)	4.34 (2.31)	5.80 (1.61)	3.26 (1.30)
LV at start phonation	29.4 (3.3)	31.6 (7.4)	15.3 (2.8)	12.9 (6.7)
Stop Consonant Onset				
Rise Time	120 (60)	100 (50)	130 (40)	160 (80)
Start P_{ALV}	4.42 (1.51)	3.00 (1.16)	4.92 (1.60)	2.83 (2.46)
Phonation P_{ALV}	5.62 (2.31)	6.81 (1.23)	6.69 (1.11)	4.46 (1.86)
LV at utterance start	30.1 (3.1)	29.0 (5.8)	17.0 (2.7)	14.8 (6.6)
Noise-like Consonant Onsets				
Rise Time	160 (60)	120 (50)	180 (60)	180 (70)
Start P_{ALV}	2.23 (0.85)	2.67 (1.92)	3.38 (2.32)	1.72 (1.17)
Phonation P_{ALV}	5.98 (1.74)	8.69 (2.72)	8.06 (1.88)	4.36 (1.23)
LV at utterance start	29.6 (5.5)	31.7 (8.8)	18.5 (4.0)	13.1 (6.7)

In the following discussion, the initial time and pressure landmarks for phonation onset are normalized to the initial pressure peak time and level. Figure 6.7 plots the normalized phonation pressure against the normalized rise time. The ‘o’ symbols represent utterances beginning with a voiced sonorant, the ‘x’ symbols represent utterances beginning with a voiceless fricative or aspirant, and the ‘+’ symbols are used for all other utterance onset segments (voiced fricatives and stops). The voiced sonorant onsets occur in the lower left quadrant and the voiceless fricative onsets occur primarily in the upper right quadrant. For this utterance set (241 utterances total), an utterance beginning with a voiced sonorant has a phonation onset pressure that is generally 20-50% of the eventual pressure peak. However, if the pressure rise occurs during a voiceless fricative, the phonation pressure is much closer to the initial peak pressure (70-100%). The other class of segments (voiced fricatives and stops) falls in the middle range. As the pressure rises toward the peak, the amplitude of the signal and its spectral characteristics go through changes. For utterances beginning with a voiced sonorant, these changes occur during voicing. For utterances beginning with a voiceless fricative, these changes primarily occur prior to phonation.

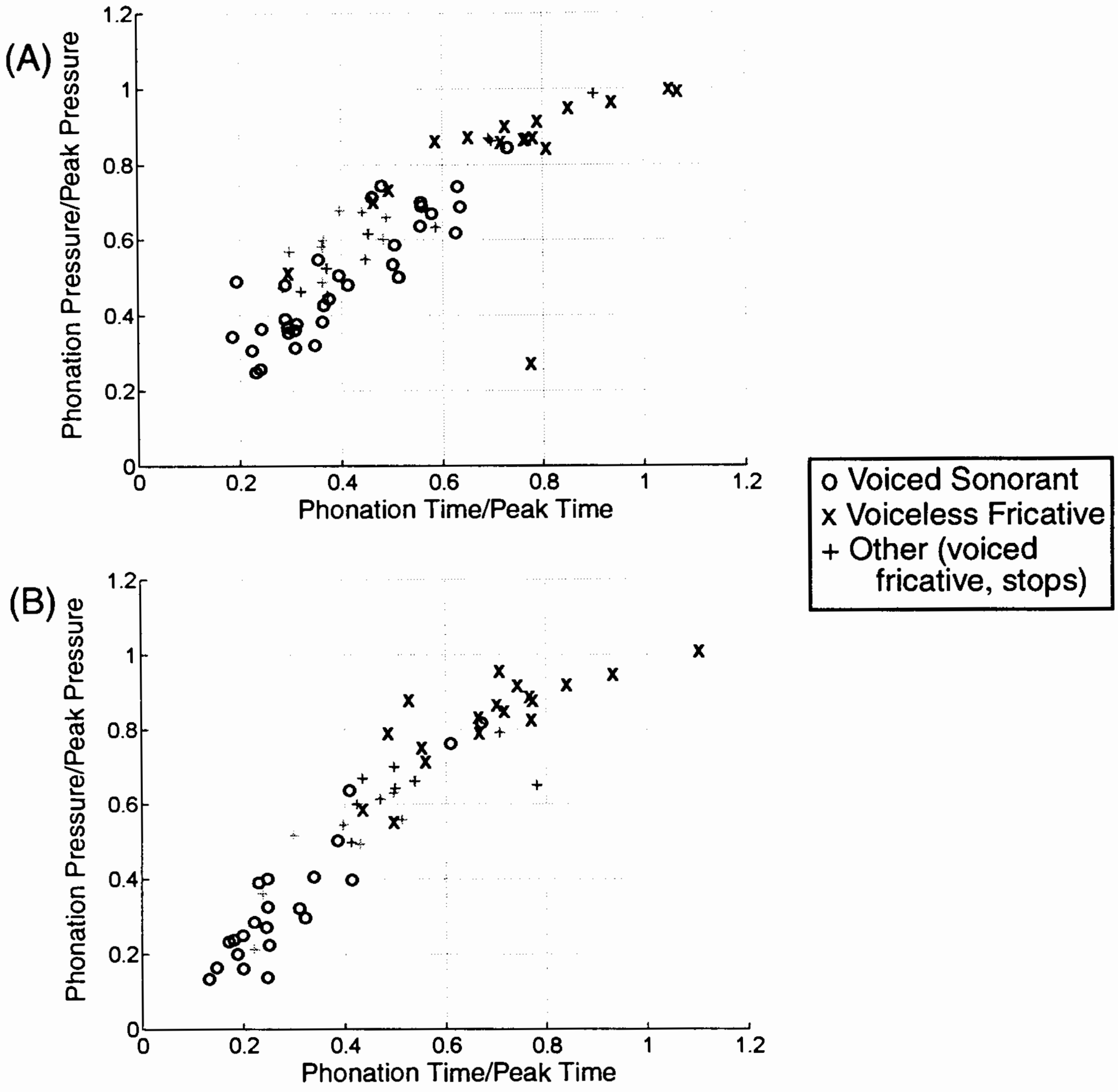


Figure 6.7: Onset timing and pressure landmarks that are normalized to the time and level of the initial pressure peak. The sound segment types in the legend refer to the class of the initial sound segment in the utterance. (a) Subject #4 (b) Subject #5.

Figure 6.8 shows an example of each type of onset from Subject #5. The example in Figure 6.8a begins with the word "say." The /s/ begins near 1 cm H_20 in pressure (Point #1) and phonation begins at a pressure of 7.8 cm H_20. The peak in pressure is 9.4 cm H_20. For the voiceless fricative onset segment, the pressure rise occurs predominantly during the /s/. The example in Figure 6.8b begins with the word "you." In this example voicing begins at 5.8 cm H_20 and rises to 14.6 cm H_20 at the peak. In general, it has been shown experimentally that the overall sound pressure level during vowel production increases approximately as

$$SPL \propto (P_{ALV})^{\frac{3}{2}} \tag{6.2}$$

by such researchers as Ladefoged (1962), Isshiki (1964) and Bouhys et al. (1968). For the example onset of Figure 6.8b, the predicted SPL would go through about a 12 dB rise at utterance onset. In the amplitude values from the calibrated audio, there is a 14 dB change over the same interval. Amplitude of voicing can be altered by changing the pressure and/ or by changing the character of the glottal closure. The result here suggests that the change in amplitude at onset is guided more by the change in pressure than by the change in the character of the glottal closure. In both of these examples, the first syllable in the utterance was rated as the initial perceived prominence.

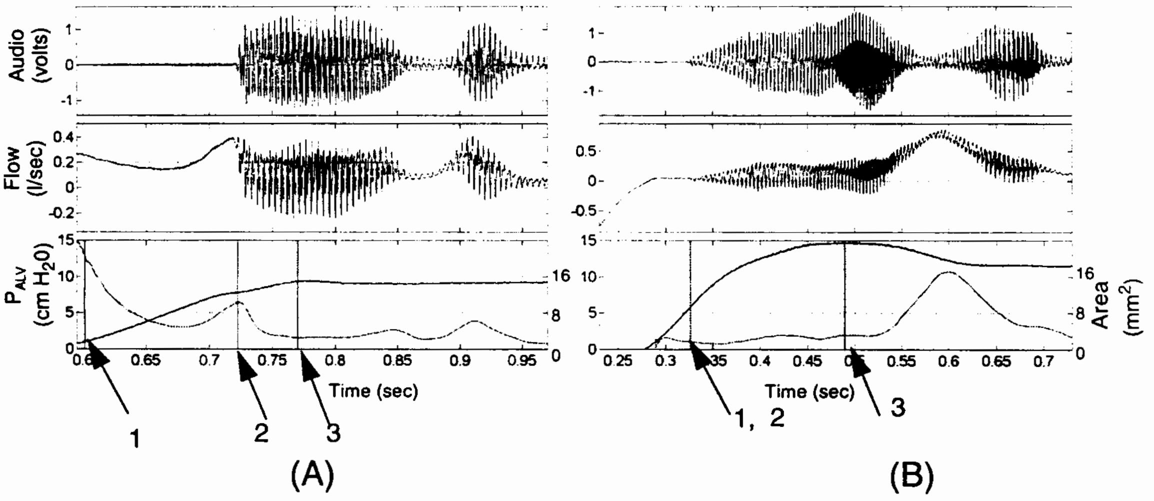


Figure 6.8: Utterance initiations for Subject #5. Panels plots audio (volts), airflow (liters/sec), and an overlay of alveolar pressure (cm H_20, scale on the left) and estimated area of the constriction (mm^2, scale on the right). (a) Utterance beginning with the words "Say the." (b) Utterance beginning with the words "You will." Vertical bars mark (1) utterance onset, (2) phonation onset, and (3) initial pressure peak.

6.1.5 Discussion

Utterance onset was shown to be rapid with changes in net muscular force and resistance of the airway that lead to alveolar pressure changes. Utterances begin while net inspiratory effort is present either from a tapering off of activity from the inhalation or a deliberate braking of recoil forces. This consistent trend was observed in modified Campbell curves where the landmark for utterance initiation fell to the left of the relaxation curve. Hixon and Mead (1976) used pressure-volume curves to assess the muscular contributions of the ribcage and abdomen during speech production for sustained vowel tasks, reading, and conversational speech. During that study it was observed that, except for cases of high initial lung volume, speech was initiated to the right of the relaxation curve. In a recent study by Johnston et al. (1999), a similar observation was made in a study using modified Campbell curves to compare the muscular effort of normal speakers with that of stutterers. Speech was initiated to the right of the relaxation curve. The current study finds that sound is initiated as soon as the conditions permit, and those conditions generally occur during the rapid transition from inhalatory effort toward a net muscular effort that is capable of generating the 'working level' of alveolar pressure. Neither of the mentioned studies focused on the details of the onset timing but were concerned with actions spanning the duration of the utterance. The implications for the acoustic realization at utterance onset is that the initial driving pressures are low and go through a substantial rise from the time of utterance onset until the initial pressure peak. As listed in Table 6.1, mean start pressures range from 2.68 to 4.62 across the subjects and mean peak pressures range from 6.54 to 11.92 across the subjects. Signal amplitude and spectral changes are present at utterance onset and could be used by the listener as cues for the boundary.

The utterances predominantly showed quick action from utterance onset to a peak in alveolar pressure. The average time from utterance onset to peak pressure, across all four speakers, was 162 milliseconds (63 milliseconds standard deviation, n=241). During this transition time the respiratory system moves from large net inspiratory muscular forces to combined coordinated effort of inspiratory and expiratory muscles during speech. The alveolar pressure peaks occur in Campbell plots to the right of the estimated relaxed chest wall curve or are clustered close to the curve. For two subjects, utterances initiated at higher lung volumes began, on the pressure-volume charts, to the left of utterances with peaks at lower lung volumes. The other two subjects used a narrow range of lung volumes to initiate an utterance and also showed a narrow range of lung volumes at the initial peak in alveolar pressure.

The low pressure at onset goes through a rapid rise at the start of the utterance and leads to characteristic changes in the acoustic signal at the onset boundary. These changes occur largely during voicing if the onset segment is voiced. However, if the onset is a voiceless fricative, these changes occur prior to voicing. From these data on the relationship of timing landmarks at utterance initiation, some basic prototypes can be proposed. Examples are given in Figure 6.9 using landmark averages from Subject #5. Such prototypes could be used in applications such as articulatory-based text-to-speech synthesis. In such a system, alignment of appropriate landmarks at utterance onset with changes in subglottal pressure would be required to produce natural sounding prosodic boundary cues.

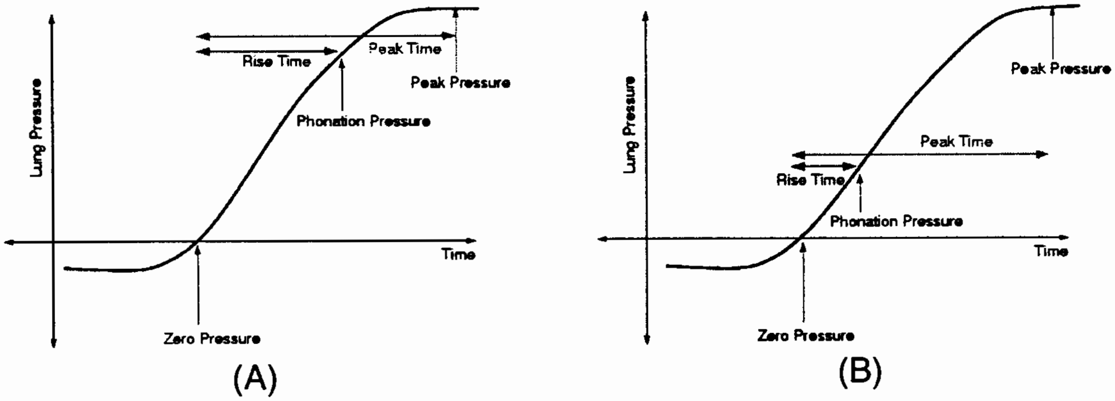


Figure 6.9: Prototypes for initial sound segments for Subject #5. (a) Typical landmarks timing for initiation of an utterance beginning with a voiceless fricative. (b) Typical landmarks timing for initiation of an utterance beginnings with a voiced sonorant.

There appears to be a limited correlation of the timing of the initial pressure peak and the location of the first perceived prominence in the utterance. The initial pressure peak occurs within the initial prominent syllable if that syllable is early in the utterance (first or second syllable in the utterance.) Otherwise, the pressure rises to its working level within the first three syllables for this data set. However, in some cases, a separate pressure rise may be associated with the late prominent syllables.

6.2 Utterance Termination

The ends of utterances have received extensive study as boundaries with robust cues. As discussed in the section "Prosody" on page 17, these cues include: lengthening of segmental durations, changes in the tone contour, amplitude decrease, and changes in spectral content. This array of cues could be considered to be a cognitively chosen set of cues added to the speech to signal the boundary. If that is the case, then the set of changes could have been chosen to be anything - a rise in signal amplitude, a shortening of duration, or even some complicated pitch change. However, it could be that the set of cues that are present are there because they are caused by the actions of the physical production system as it ends or suspends a muscular action - whether that ending is the termination an utterance, the end a breath, a pause in an utterance, or a reduction in the drive for the sound source.

An utterance showing some of these cues is given in Figure 6.10. Fundamental frequency falls approximately 50 Hz in the final syllable. Signal amplitude falls over 10 dB SPL as the vowel ends. Changes in the spectrum are consistent with an increase in subglottal coupling. This could occur if the vocal folds were spreading at the end of the vowel. The possible changes in the spectrum include the effects of the resonances of the subglottal space (possible extra poles and zeros in the transfer function) as well as increased losses that can widen the bandwidths and change the level of the formant peaks. The current research will begin by using both methods. Incomplete closure of the vocal folds, as in a glottal chink, will alter the bandwidth of the resonances and introduce resonances from the subglottal space (e.g. Rothenberg,1983, Klatt and Klatt,1990).

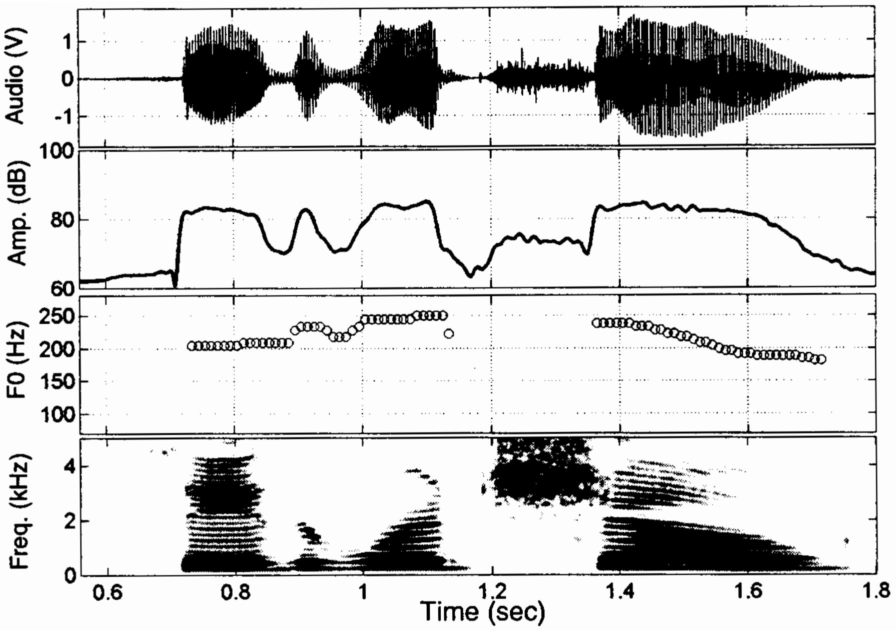


Figure 6.10: Acoustic cues example for Subject #5. The utterance is "Say the word show." The panels show audio (volts), signal amplitude (dB SPL), F0 (Hz), and the audio spectrogram (kHz).

Some of the acoustic and physiologic landmarks for utterance termination are given in Table 6.6 for the utterance set as a whole as well as subsets for the type of sound segment that ends the utterance.

Table 6.6: Pressure and timing values corresponding to acoustic and physiologic landmarks. Values are given as 'mean (standard deviation).' Pressure is given in cm H_20 and lung volume in %VC from FRC.

Quantity	Subject #4	Subject #5	Subject #6	Subject #7
End Phonation P_{ALV}	4.78 (0.87)	5.34 (1.44)	4.09 (1.32)	3.79 (0.98)
End P_{ALV}	4.12 (1.13)	4.59 (1.65)	3.58 (1.28)	2.83 (1.32)
LV at end phonation	17.9 (5.7)	14.9 (8.9)	7.2 (4.8)	2.7 (7.4)
LV at utterance end	17.3 (5.6)	14.2 (8.8)	6.9 (4.7)	2.0 (7.2)
Voiced Sonorant Offsets				
End Phonation P_{ALV}	4.65 (0.84)	5.10 (1.60)	3.70 (1.30)	3.64 (0.98)
LV at end phonation	17.0 (6.4)	14.8 (8.3)	5.9 (4.4)	1.9 (7.6)

Table 6.6: Pressure and timing values corresponding to acoustic and physiologic landmarks. Values are given as 'mean (standard deviation).' Pressure is given in cm H_20 and lung volume in %VC from FRC.

Quantity	Subject #4	Subject #5	Subject #6	Subject #7
Stop Consonant Offsets				
End phonation P_{ALV}	4.99 (0.92)	5.94 (1.01)	5.24 (1.08)	4.19 (0.68)
End P_{ALV}	3.38 (1.14)	3.57 (1.71)	3.94 (1.26)	1.75 (0.84)
LV at utterance end	18.7 (4.4)	13.8 (12.3)	8.5 (4.8)	3.0 (5.6)
Noise-like Consonant Offsets				
End phonation P_{ALV}	4.90 (0.91)	5.46 (1.12)	4.31 (0.90)	3.76 (1.14)
End P_{ALV}	3.57 (1.09)	4.07 (0.96)	2.79 (1.00)	1.92 (1.06)
LV at utterance end	16.7 (4.6)	12.5 (6.1)	8.6 (5.0)	1.4 (7.9)

In the perceptual rating experiment, the last prominent syllable was chosen to fall on the last syllable of the utterance 75% of the time across all read utterances and all speakers. This was true even though 38% of the utterances ended in a prepositional phrase such as "in the show" or "in the truck." Also, 16% of the utterances had a distinctive modifier such as "Hanna's house" or "Willow street" which again, might have lead to a final prominence that fell earlier in the utterance. The averages and standard deviations for the acoustic and physiologic landmarks as grouped by the location of the final perceived prominent syllable in the utterance are given in Table 6.7.

Table 6.7: Pressure and timing values corresponding to syllable prominence ratings at utterance termination. Values are given as 'mean (standard deviation).' Pressure is given in cm H_20 and lung volume in %VC from FRC.

Quantity	Subject #4	Subject #5	Subject #6	Subject #7
Last Syllable is Prominent				
End phonation P_{ALV}	4.89 (0.87)	5.40 (1.50)	4.04 (1.14)	3.85 (0.94)
End P_{ALV}	4.25 (1.14)	4.60 (1.7)	3.42 (0.99)	3.10 (6.63)
LV at utterance end	17.9 (5.7)	15.3 (9.2)	6.9 (5.1)	2.8 (1.3)
Second to Last Syllable is Prominent				
End phonation P_{ALV}	4.41 (1.03)	6.33 (1.51)	4.07 (2.34)	3.58 (1.18)
End P_{ALV}	3.81 (1.35)	5.62 (2.49)	4.07 (2.34)	3.19 (1.07)
LV at utterance end	16.3 (4.4)	11.0 (7.9)	7.2 (3.2)	-0.4 (7.7)
Early Final Prominence				
Phonation P_{ALV}	4.52 (0.71)	4.67 (0.83)	4.54 (0.59)	3.65 (1.04)
End P_{ALV}	3.68 (0.77)	4.05 (0.60)	4.11 (1.16)	2.61 (1.76)
LV at utterance end	14.8 (5.1)	11.7 (7.7)	6.4 (3.3)	-3.8 (9.0)

6.2.1 Phonation Offset

Phonation requires at least a certain level of pressure across the vocal folds (Titze,1988, Titze,1992). The pressure required is related to the vocal-fold stiffness, and hence the fundamental frequency and is different for the onset and the termination of phonation. For phonation, the vocal folds must also be brought into proximity and away from the spread position that they occupy during resting breathing. This section examines the pressure and glottal area changes that lead to the termination of phonation and the termination of utterances. The corresponding acoustic changes are discussed in terms of their cues for the boundary.

Figure 6.11 shows the pressure-volume relationship at the termination of phonation for all utterances in the data set. The ‘o’ symbols represent utterances that have a voiced sonorant sound segment at the end of the utterance. The ‘x’ symbols represent utterances that end in any of the other sound segments in the data set (mostly fricatives and stops). These points generally cluster near the relaxed chest wall characteristic, at a lung volume level that is near the level of FRC. There is no apparent separation based solely on the type of sound segment at the end of the utterance. For this protocol, there are utterances that end phonation to the left of the chest wall characteristic. As previously stated, these points may arise from the method of presenting the printed utterances to the subject for reading and may also reflect the estimated relaxation characteristic for the chest wall rather than an explicitly measured characteristic. The important observation from these plots is that the points cluster near the curve and near the functional residual capacity. This implies that as phonation ends, the speaker is ending net muscular effort by moving toward the relaxation characteristic in the Campbell diagram.

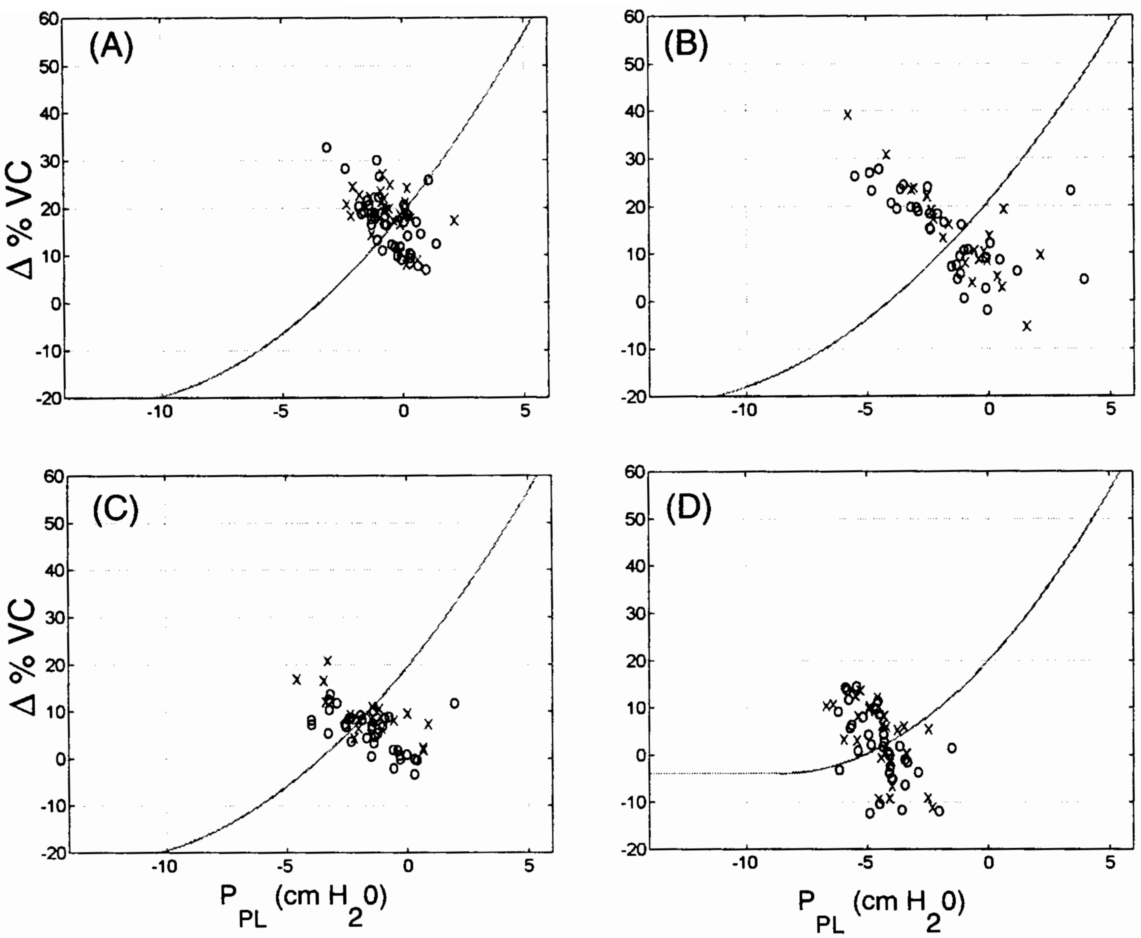


Figure 6.11: Pressure-volume relationship at the termination of phonation. The 'o' symbols are utterances that end with a voiced sonorant. The 'x' symbols represent all other ending segments (fricatives, stops). The estimated relaxed chest wall curve is plotted for reference. (a) Subject #4 (b) Subject #5 (c) Subject #6 (d) Subject #7.

6.2.2 Pressure and Glottal Area

To end voicing, a speaker can spread the vocal folds too far apart to sustain oscillation, reduce the pressure drop across the vocal folds so that it is too low to sustain oscillation, press the vocal folds together to prevent airflow, or make a combination of the above actions. This section examines the termination of voicing at the end of an utterance from the point at which the sound pressure level begins to fall until the end of voicing.

At the end of voicing the signal amplitude tapers off rather than ending in a step function. The point at which signal amplitude began to decay was visually determined from the calibrated dB SPL signal. The alveolar pressure was measured and the area of the constriction was estimated at the point where the audio amplitude began to fall and at the point of termination of phonation. An example is shown in Figure 6.12. This utterance is the same utterance as in Figure 6.10, but the additional data provided are the airflow, estimated alveolar pressure and area of the constriction. The termination of phonation is marked by the vertical bar. The start of the fall in signal amplitude is marked by the arrow. Pairs of such points were collected for all utterances in the data set and are shown in the scatter plots of Figure 6.13. The gray 'o' symbols represent the data at the start of the fall in signal amplitude. The black 'x' symbols represent the data at the termination of phonation. Plots in the left column compare the area of the constriction during the final amplitude fall. Plots in the right column compare the alveolar pressure during the final amplitude fall. The general trend is that area rises and pressure falls. The numerical values for the means and standard deviations associated with these scatter plots are given in Table 6.8. While the noise in the measurements themselves as well as the use of a visual criteria to select the start of the fall in signal amplitude should be considered when viewing these numbers, these data indicate that glottal area increases during the end voicing by a factor in the range of 2.4 to 4 across the four speakers. Alveolar pressure falls during the same period to a value that is in the range of 65-77% of its value at the start of the fall in amplitude. The change in dB SPL is a decrease by 9.9-15.4 dB SPL. As stated earlier, it has been shown that the overall sound pressure level during vowel production increases approximately as

$$SPL \propto (P_{ALV})^{\frac{3}{2}} \qquad (6.3)$$

by such researchers as Ladefoged (1962), Isshiki (1964) and Bouhys et al. (1968).

For a 70% decrease in alveolar pressure, the approximate drop in SPL is 4.6 dB. In these cases, this means that the change in area also contributes to the decrease in amplitude observed. While F0 changes are not a major focus of this work, the example in Figure 6.12 shows that the F0 drop at the end of this utterance was largely completed prior to the fall in signal amplitude. F0 has been shown to vary with changes in alveolar pressure on the order of 3 to 5 Hz/cm H_20 (Ladefoged,1962, Ohala and Ladefoged, 1970). The relationship of change in F0 to change in subglottal pressure depends on where the F0 falls within a speaker's range. Results from Titze (1989) show a value of 6.1Hz/cm H_20 for slack vocal folds and 0.54 Hz/cm H_20 for tense vocal folds. In this example, F0 falls about 40 Hz in the span of time (from 1.4 to 1.6 seconds in Figure 6.12) that pressure falls about 2 cm H_20. This is consistent with a pitch change that is created primarily through adjustment of the tension of the vocal folds.

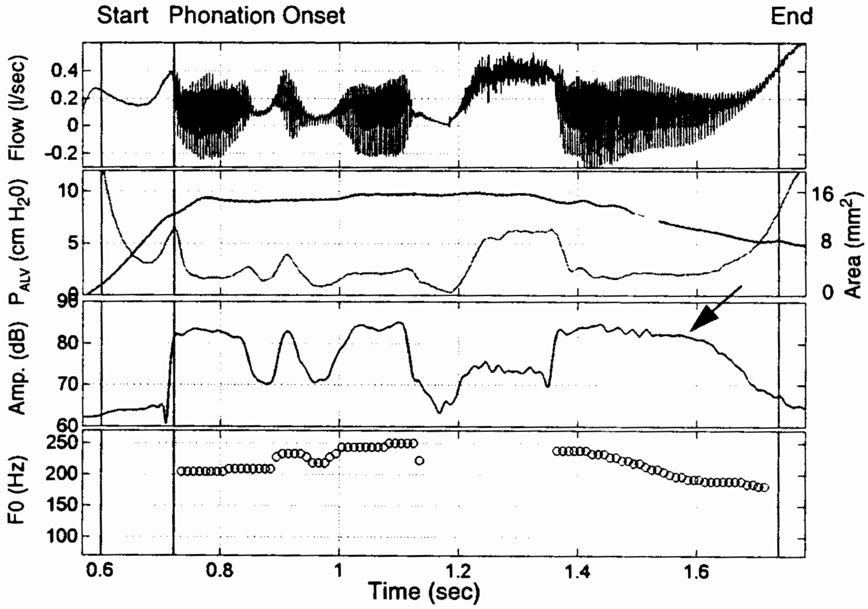


Figure 6.12: Example amplitude, area, and alveolar pressure variations at utterance termination. The utterance is "Say the word show,“ as produced by Subject #5. (a) Airflow (liters/sec) (b) an overlay of the alveolar pressure (cm H_20, scale on the left) and estimated area of the constriction (mm^2, scale on the right) (c) signal amplitude (dB SPL) (d) estimated fundamental frequency (Hz). The vertical bar denotes the landmarks for utterance onset, phonation onset and termination of voicing. The arrow marks the beginning of the final fall in signal amplitude.

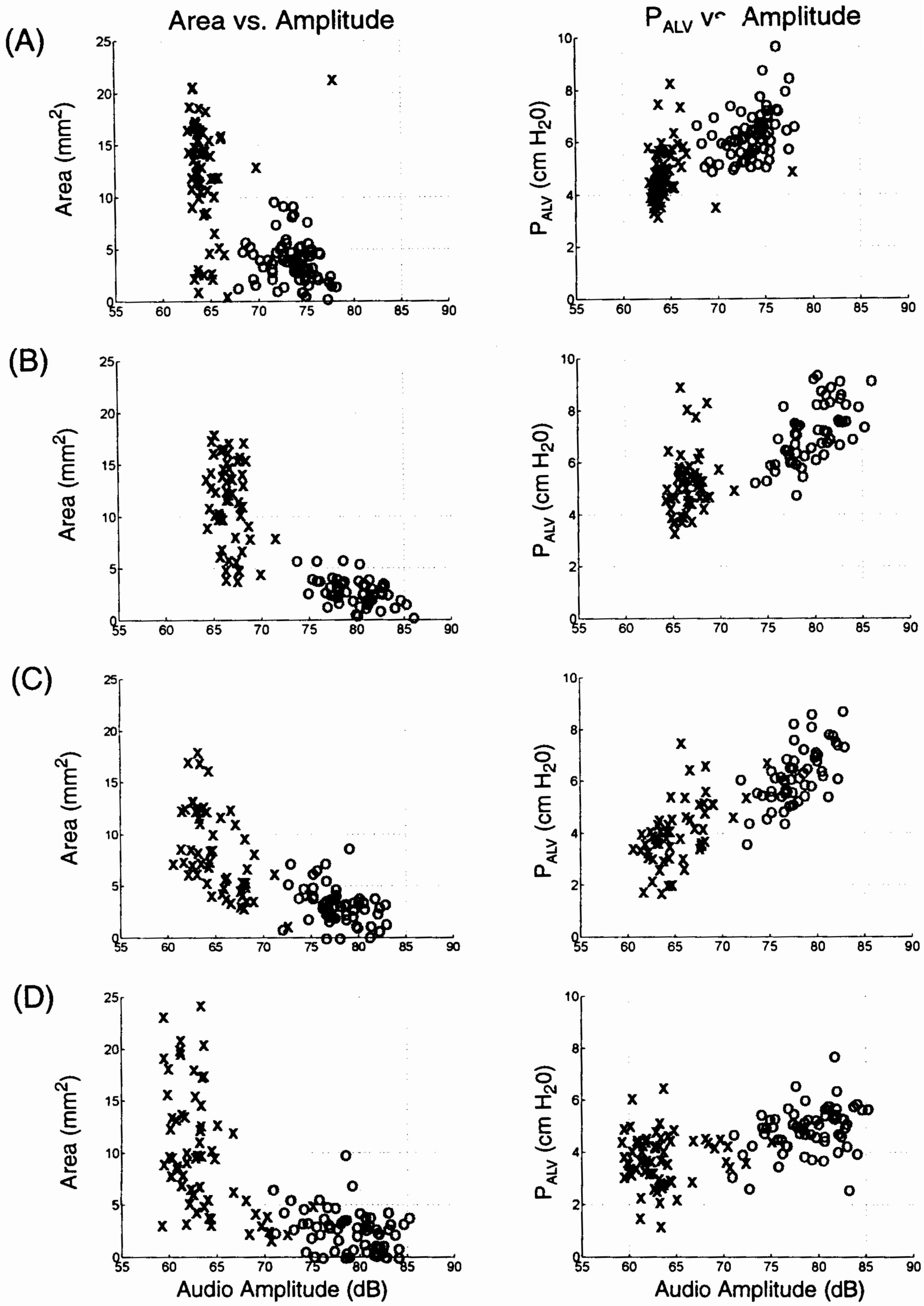


Figure 6.13: Amplitude comparisons for area (left column) and alveolar pressure (right column) where 'o' indicates the start of the amplitude fall and 'x' represents the termination of phonation. (a) Subject #4 (b) Subject #5 (c) Subject #6 (d) Subject #7.

The fact that the area increases and the pressure falls implies a combination of the two methods proposed for terminating phonation: both abduction of the vocal folds and a decrease in pressure. The pressure decrease could arise from a decrease in vocal tract resistance alone or it could result from both the abduction and a decrease in muscular effort. The pressure threshold for phonation is in the range of 3-4 cm H_20 (as in Titze,1992). The majority of the pressure values in this study at the end of phonation are above this threshold. Phonation appears to be terminated largely by abduction of the vocal folds rather than simply removing the drive for oscillation. It may be that the speaker had greater and finer control over the smaller structures of the vocal folds than the larger chest wall structures that have longer time constants.

Table 6.8: Mean and standard deviation for the scatter plots of Figure 6.13.

	Subject #4	Subject #5	Subject #6	Subject #7
dB SPL Start	72.9 (2.4)	79.3 (2.7)	77.2 (2.6)	78.3 (3.6)
Area Start (mm^2)	4.0 (2.1)	2.8 (1.3)	3.2 (1.8)	2.6 (1.9)
Palv Start (cm H_20)	6.3 (0.9)	7.4 (1.8)	6.3 (1.3)	4.9 (0.9)
dB SPL End	63.0 (7.2)	66.1 (1.4)	63.7 (8.3)	62.9 (3.4)
Area End (mm^2)	12.6 (5.1)	11.4 (4.0)	8.2 (4.0)	10.2 (5.8)
Palv End (cm H_20)	4.8 (0.9)	5.3 (1.4)	4.1 (1.4)	3.8 (1.0)

A closer look at the ending sound segment, as shown in Figure 6.14, shows that most of the cases that showed little increase in the area of the constriction as voicing ends were those cases which end in a stop consonant such as /t/. In some cases, /t/ may be created by a glottal stop (pressing the vocal folds together) rather than a constriction by the tongue blade. Also, during the transition from a vowel to the /t/, there are two constrictions in the vocal tract. The two constrictions occur as an overlap as the vocal folds spread and the tongue blade moves to create an obstruction of the airway. As stated in section "Estimation of the Area of the Constriction" on page 41, that is a condition for which the estimated area of the constriction is not as accurate. Both of these considerations could place these /t/ segments at the low end of area increase. It also implies that the average area increase factor may be biased to a lower value.

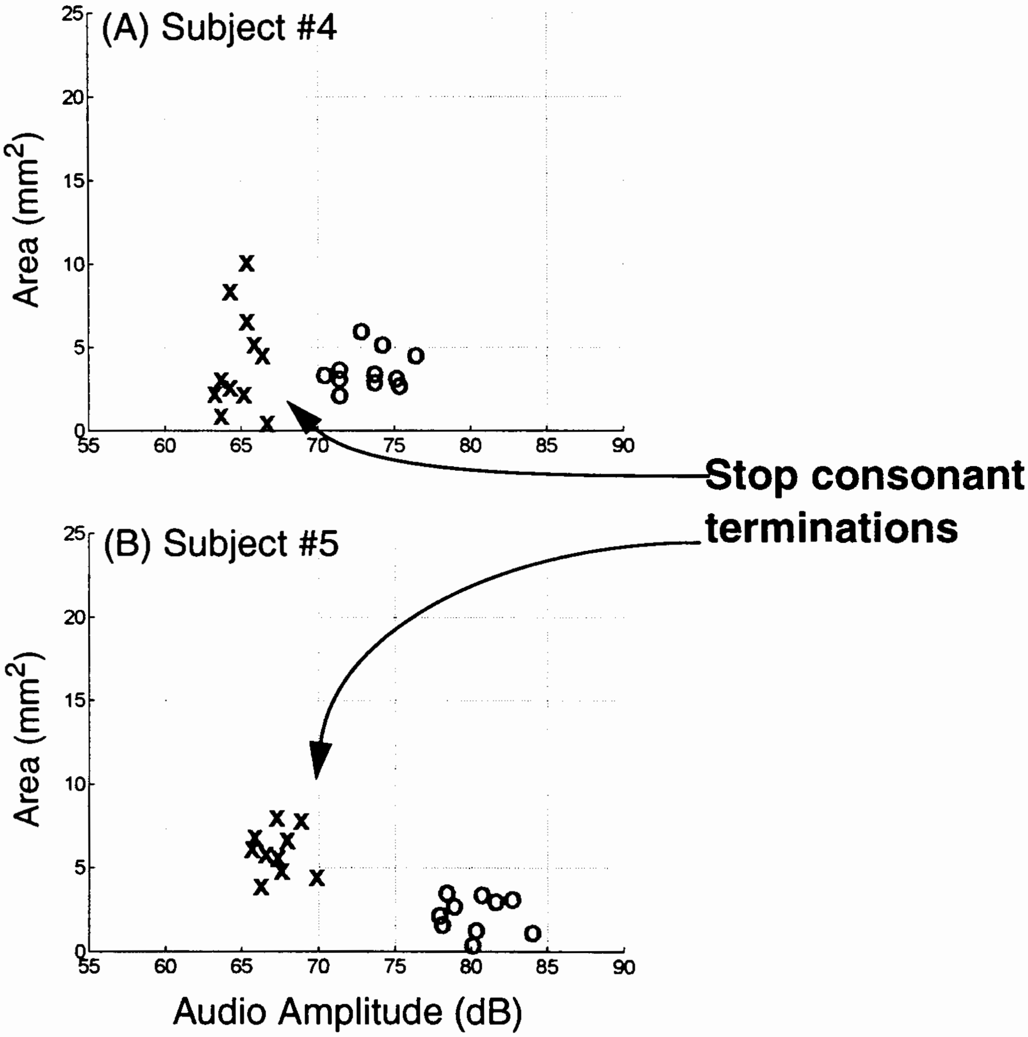


Figure 6.14: Detail in ending area of the constriction vs. audio amplitude. These plots are the same as Figure 6.13 (a) and (b) except that the only sound segment represented is /t/. Gray letters represent the time at which audio amplitude begins to fall. Black letters represent the time at which phonation ends. (a) Subject #4 (b) Subject #5.

6.2.3 Irregular Glottal Vibration

As the drive to the vocal folds (subglottal pressure) and the tension and/or degree of abduction of the vocal folds changes at the end of an utterance, there can be instabilities in the oscillation. This results in irregular glottal vibration that is audible to the listener, and can be interpreted as cues to the boundary (for example, Henton and Bladon,1988, Dilley et al.,1996). One type of irregular occurrence has been labeled as a glottalization. A glottalization is generally considered to be an instance where, in general, the vocal folds are adducted accompanied by short intervals of glottal opening during a vibration cycle (Ladefoged,1982, Stevens,1994). The airflow pulses have considerable high frequency components due to the sharpness of the vocal fold closure. Pierrehumbert and Talkin (1992) surmise, from examination of acoustic and EGG signals, that braced partially-abducted vocal folds could also generate an irregular fundamental frequency.

The speakers in this study terminated voicing with three basic types of trends in the fundamental frequency: (1) regular voicing and a smooth decrease in signal amplitude (2) a period of irregular voicing followed by a regular oscillation showing a smooth taper in signal amplitude and (3) an irregular glottal waveform through the end of voicing. This section examines these three types of endings and possible correlations with respiratory measures.

Each utterance in the data set was classified into one of these three categories by visual inspection. This task is demonstrated in the examples shown in Figure 6.15 for Subject #4. Three people served as evaluators. Each was shown examples of the types of endings in the data set and asked to classify each ending as one of those types. Evaluators were instructed to examine from the end of voicing, using plots of the raw audio signal as well as the airflow signal, and backwards into the utterance no more than 0.2 seconds. A simple majority determined the type of ending. Average agreement between the evaluators was

92%. The percentage of time that each speaker terminated an utterance with each class is listed in Table 6.9.

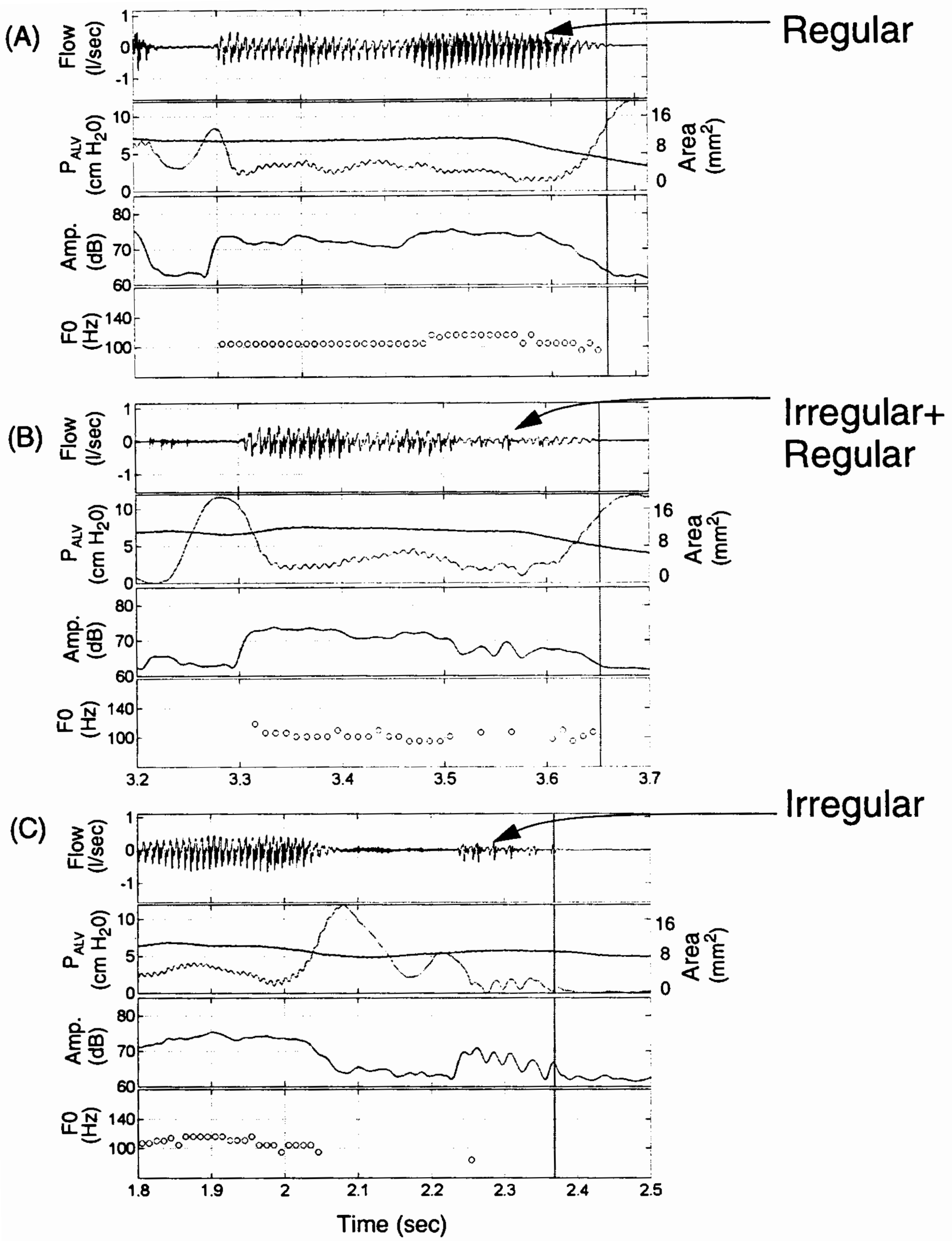


Figure 6.15: Examples of termination of voicing from Subject #4. Panels show airflow (liters/sec), an overlay of alveolar pressure (cm H_20, scale on the left) and area of the constriction (mm^2, scale on the right), amplitude (dB), and F0 (Hz). (a) Example of a smooth decrease in signal amplitude with regular voicing. The utterance ends with the word "umbrella." (b) Example of a period of irregular voicing followed by a regular tapered ending. The utterance ends with the word "Canada." (c) Example of irregular voicing. Vertical bars mark the end of phonation.The utterance ends with the word "street."

Table 6.9: Voicing termination types. Percentage of each ending type for each speaker.

Speaker	Percentage of Voicing Termination Type		
	Regular	Irregular+Regular	Irregular
Subject #4	63	27	10
Subject #5	94.6	3.6	1.8
Subject #6	5	33	62
Subject #7	6.6	69	24.4

Each subject appeared to have certain habits for terminating voicing. Subject #4, a male speaker, produced mostly a mix of regular fundamental frequency and mixed irregular+regular endings. Of his irregular voicing endings, 72% preceded a voiceless stop consonant and as stated in the previous section, did not show the rise in area typically present at the ends of the other utterances.

Subject #5, a female speaker, almost exclusively produced endings with a regular fundamental frequency that was accompanied by a rise in glottal area and a sloping fall in alveolar pressure. A typical ending for Subject #5 is shown in Figure 6.16

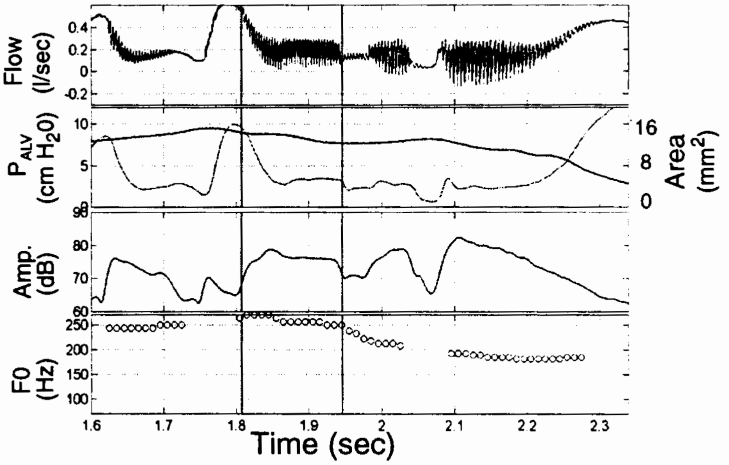


Figure 6.16: Typical termination of voicing for Subject #5. Panels are plots of airflow (liters/sec), an overlay of alveolar pressure (cm H_20, scale on the left) and area of the constriction (mm^2, scale on the right), sound amplitude (dB), and fundamental frequency (Hz.) Subject #5 generally shows regular F0 at the termination of voicing. The utterance ends with the word "Canada."

Subjects #6 and #7 produced very few endings that had a regular fundamental frequency. However, each subject had different conditions for generating an irregular fundamental frequency. Subject #7 produced a mixed voicing termination (irregular+regular) for 69% of his utterances. Two examples are shown in Figure 6.17. For these utterances, the resumption of regular oscillation for the tapered ending corresponded to a rapidly rising area. The possible co-occurrence of changes in vocal fold stiffness or length cannot be determined from these data. However, it can be concluded that the vocal folds are spreading as regular oscillation resumed.

Subject #6 had some segment of irregular F0 during 95% of the utterance terminations. Examples are shown in Figure 6.18. In these cases, the irregular F0 is not consistent with a classical definition of glottalization, nor do they suggest a braced configuration for the vocal folds. Rather the folds are in the process of being further abducted.

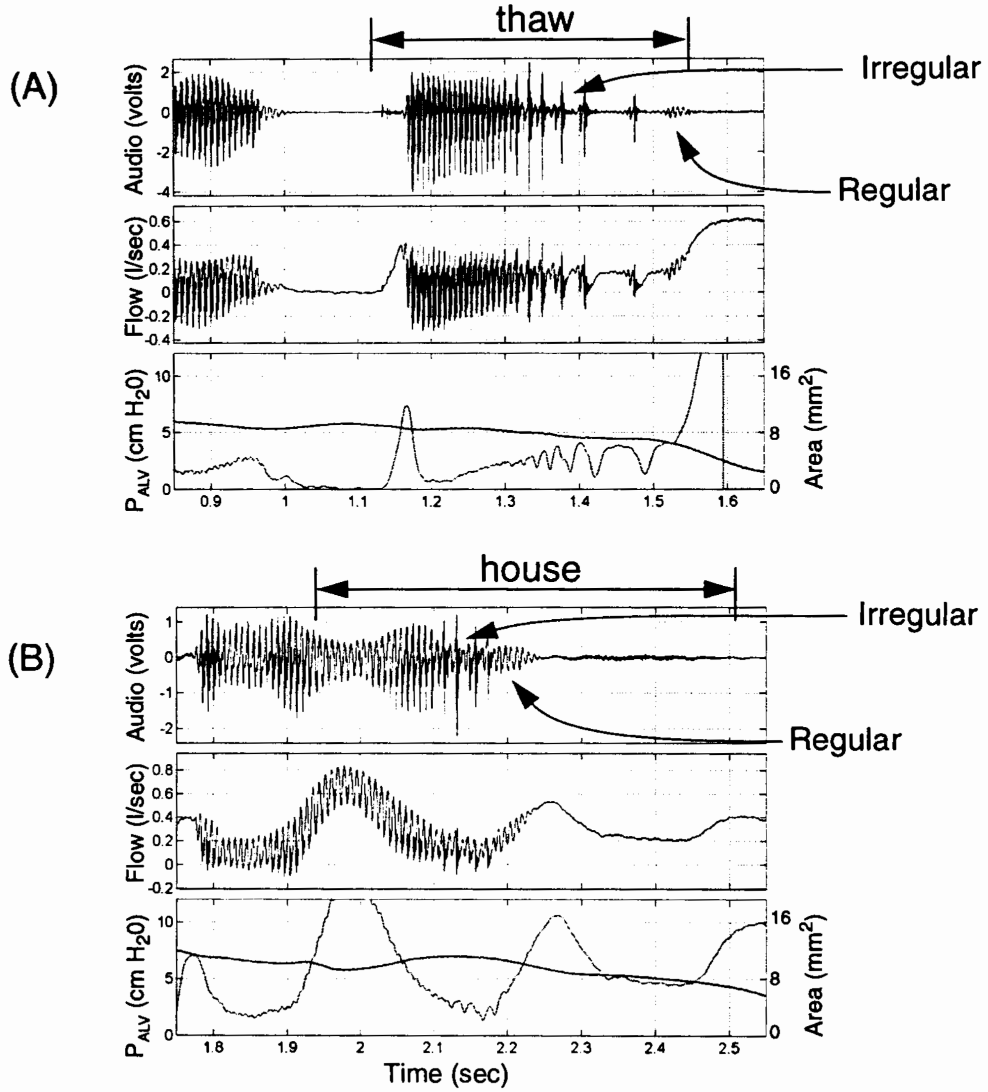


Figure 6.17: Voicing terminations for Subject #7 show a mixed termination: a period of irregular oscillation followed by regular oscillation as voicing tapers off. Panels show audio (volts), airflow (l/sec) and an overlay of alveolar pressure (cm H_20, scale on the left) and estimated area of the constriction (mm^2, scale on the right). (a) Utterance ending with the word "thaw." (b) Utterance ending with the word "house."

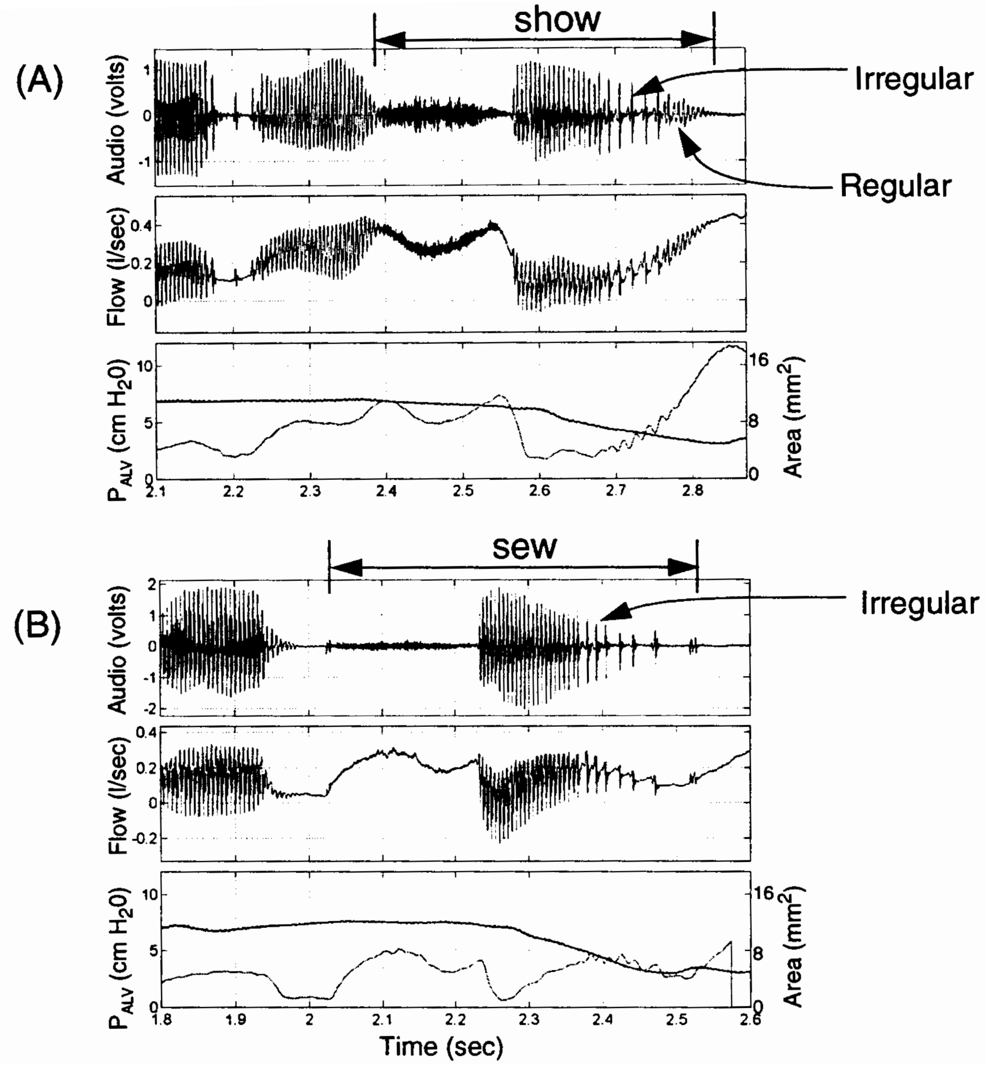


Figure 6.18: Voicing terminations for Subject #6. Panels show audio (volts), airflow (l/sec) and an overlay of alveolar pressure (cm H_20, scale on the left) and estimated area of the constriction (mm^2, scale on the right).(a) Utterance ending with the word "show" that has an irregular glottal waveform followed by a resumption of regular oscillation. (b) Utterance ending with the word "sew" that has an irregular glottal waveform through the end of voicing.

6.2.4 Synthesis of Voicing Termination Cues

Listeners are used to hearing speech that is produced from a coordinated physical gesture. The types of cues are related to this coordination. For example, it was shown in this chapter that the glottal area increases as phonation ends. This leads to not only a drop in signal amplitude but also to a change in spectral tilt. Synthesis of an ending to an utterance that did not take into account the correlated presence of changes could sound unnatural to the listener.

The synthesis system HLsyn (Sensimetrics Inc.) combines a formant synthesizer with articulatory and aerodynamic models for sound source generation. It is based on the source-filter theory of speech production (Fant,1960). This system uses as inputs some parameters that specify the area of the constriction in the airway. These parameters include the area of the glottis at the vocal folds (Ag), the area of the constriction made by the tongue blade (Ab), the area of the constriction made by the lips, and the area (Ap) of the shunt path between the arytenoid cartilages. The system uses another input parameter, subglottal pressure (Ps), and computes the sound sources for the speech. Predominantly, in development of the system and in general use, the pressure P_S, has been held constant at 8 cm H_20 or has been set to a trapezoidal rise and fall to the 8 cm H_20 level. The trapezoid ramps up from 0 to 8 cm H_20 in 25 milliseconds and falls in 25 milliseconds relative to the start and end of the utterance. The following section provides an example of the types of acoustic changes at utterance termination for changing pressure and area inputs versus constant pressure and area inputs.

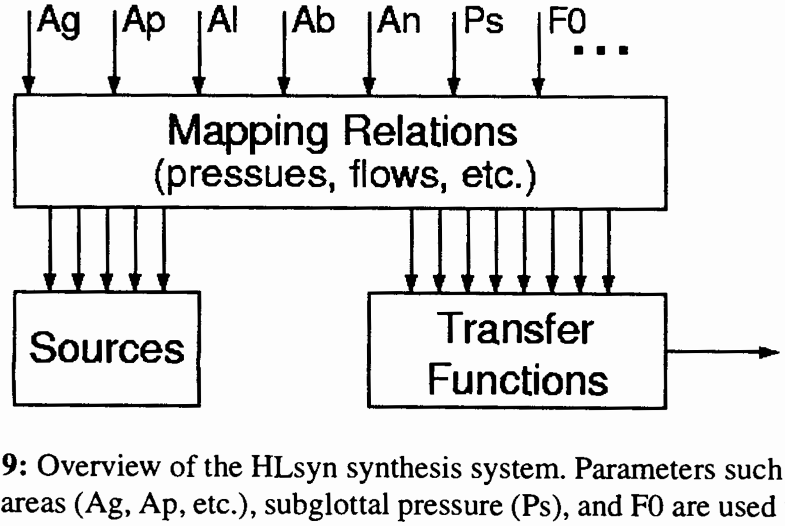


Figure 6.19: Overview of the HLsyn synthesis system. Parameters such as vocal tract constriction areas (Ag, Ap, etc.), subglottal pressure (Ps), and F0 are used to compute the sound sources for the synthesis system. (Adapted from Bickley et al.,1997).

The input parameters were estimated from analysis of the recorded acoustics, the measured flow, and the estimated alveolar pressure. The estimated area at the level of the glottis (Ag plus Ap) is plotted in Figure 6.20a for the phrase "in the way" as produced by speaker #7 at the end of an utterance. Overlaid on the estimated area track is a constant area of 4 mm^2. Figure 6.20b shows the same segment of time and compares a constant pressure level of 5.2 cm H_20 with the measured alveolar pressure. Two segments of synthesized speech are shown here. The audio and spectrogram for the segment created from varying area and pressure information are shown in Figure 6.21b, and Figure 6.21c shows the corresponding plots for constant area and pressure.

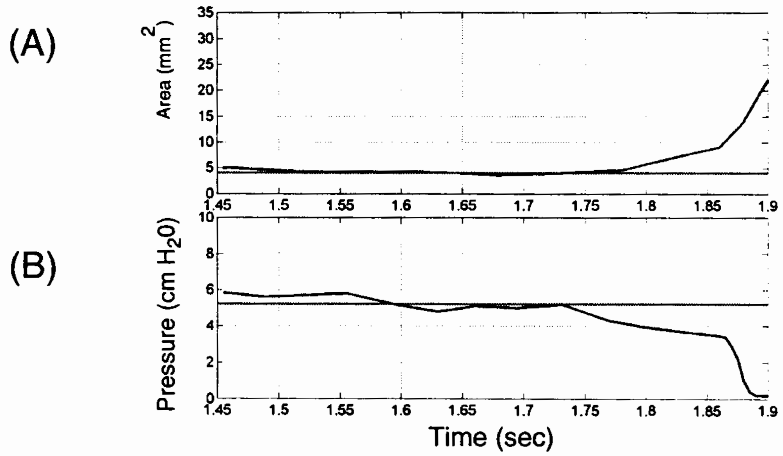


Figure 6.20: Comparison of ending synthesis parameters. Synthesis parameters were created for the phrase "in the way." This figure plots the final word "way." (a) Area of the glottis. Synthesis examples compare the level parameter track (gray line) with the rising parameter track (black line). (b) Subglottal pressure. Synthesis examples compare the level parameter track (gray line) with the falling parameter track (black line).

For the recorded audio, voicing terminates with a change in high frequency content (as seen in Figure 6.21a) and a decrease in signal amplitude, which is more easily seen in the raw audio. Also, the actual recorded audio shows several other effects at the termination of voicing. One of these is the tendency of this speaker to have a period of irregular glottal waveform that is followed by regular glottal waveform. The synthesis system does not currently have a means for including such phenomena. This example points out one possible area of application for the results of the present study. That application is to improve the naturalness of speech synthesis by providing data to refine the mapping relations of the synthesis system as well as contributing to the refinement of rules for text-to-speech generation of the input parameter tracks.

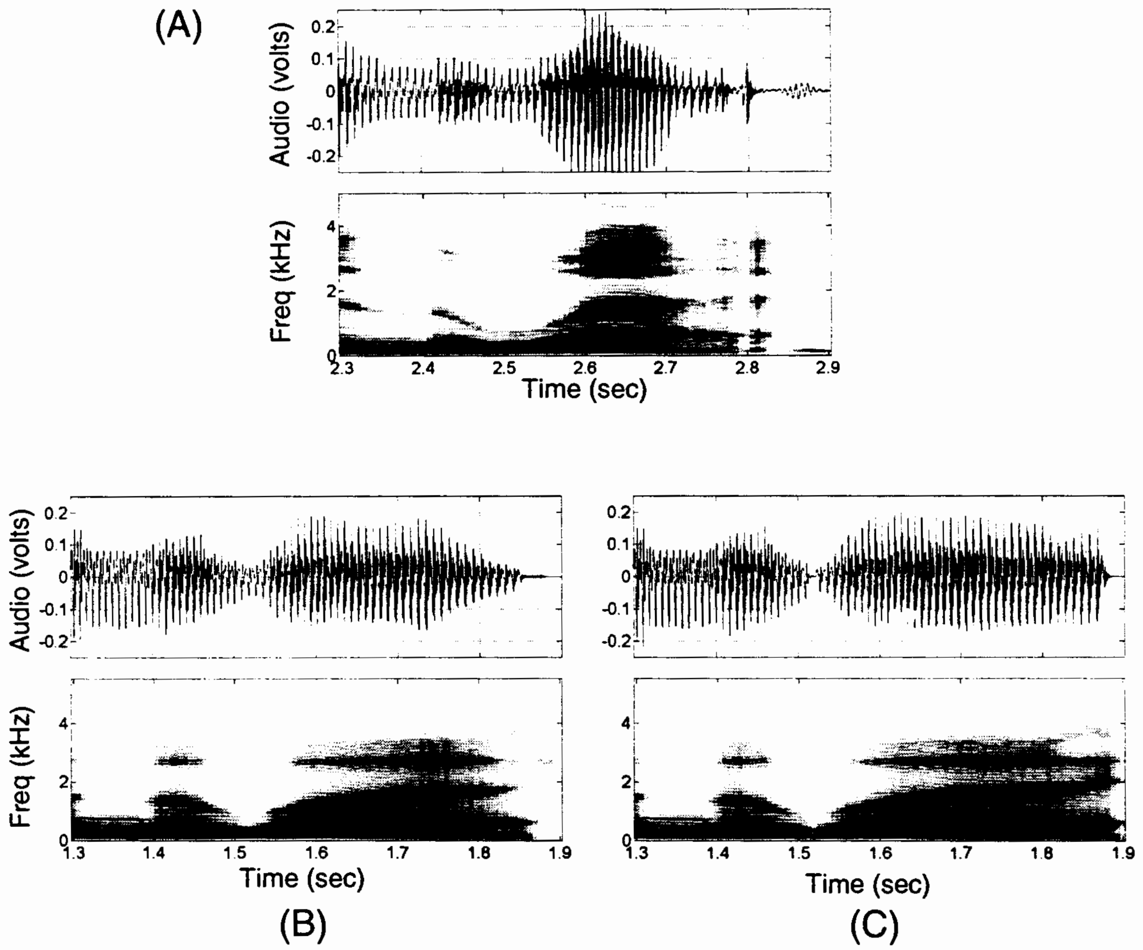


Figure 6.21: Comparison of endings in the acoustic signal. The panels are audio (volts) and the spectrogram (kHz). The phrase is "in the way." (a) Recorded audio from Subject #7. (b) Synthesized speech using falling pressure and rising area at termination of voicing in the word "way." (c) Synthesized speech using constant pressure and constant area at the termination of voicing in the word "way."

Informal listening tests with two different utterance terminations, "in the show" and "in the way", for various combinations of constant and varying pressure and area, showed that listeners never preferred the synthesis versions using constant pressure and constant area of the constriction.

6.3 Summary

This chapter focused on the details of the pressure dynamics, the laryngeal area changes, and the corresponding acoustics at utterance boundaries that coincided with the initiation and termination of exhalation. Questions stated at the start of the chapter included: (1) How are the dynamics of initiation and termination related to the recoil pressures of the chest wall and applied muscular pressures? (2) Are timing and/or pressure landmarks correlated to prosodically prominent syllables? (3) What are the influences, if any, of the boundary sound segments (and the airway resistance of those sound segments) on the shape of the pressure curve at utterance onset? (4) In what way are the amplitude changes influenced by coordination of the area of the vocal tract constriction and alveolar pressure at voicing termination? The following section briefly summaries some of the results that address these questions.

At utterance initiation, boundary sound segments were seen to influence the timing of the onset landmarks. For utterances beginning with voiceless fricatives, the pressure at the onset of phonation was very near to the initial peak in alveolar pressure (on the order of 90% of the peak pressure level). The corresponding rise time fell in the range of 60-80% of the peak time. By contrast, a voiced sonorant at the boundary segment was shown to have a lower phonation pressure (20-70% of the peak pressure) and a shorter rise time (20-60% of the peak time).

At the end of the utterance, termination of voicing was found to result from an approximately three-fold increase in area and a 1-3 cm H_20 drop in alveolar pressure. However, if the boundary segment was /t/, the area rise did not generally occur. This trend may be the result of inadequate estimation of the area of the constriction (two constrictions in the airway as the glottis widens and the tongue blade constriction narrows) or of the speaker's actions in realizing the /t/ segment as a glottal stop.

At utterance initiation, the speakers aligned the timing of the pressure peak to fall within the syllable that was perceived as the initial prominent syllable of the utterance, if that syllable was the first or second syllable of the utterance. If the initial prominence was perceived to fall on the third, fourth, or fifth syllable in the utterance, the trend was for the pressure to reach a peak prior to that syllable. Consider the case where the second syllable in the utterance was judged to be the initial prominence of the utterance. In that case, the peak time may be longer or the first syllable in the utterance may be compressed or reduced in duration. Some combination of the two methods may also be used.

Utterance initiation was a rapid action with a large pressure change and very little volume change in the pressure-volume curves. Utterance onset began during net inspiratory muscular force. The muscular force at the pressure peak could only be estimated due to uncertainties in the chest wall relaxation curve estimation. However, the alveolar pressure peaked close to or to the right of the relaxation curve, except when utterances were produced at the high end of a speaker's range of lung volumes for utterance initiation.

Irregular glottal waveform segments at the termination of utterances that were aligned with the termination of exhalations were seen to occur mainly during periods of increasing area of the constriction in the airway, indicating that abduction of the vocal folds was increasing. This pattern was in contrast to the classical definition of glottalizations which occur as periods of largely adducted vocal folds with brief irregular open segments. Irregular glottal waveforms that occur during periods of increasing abduction will have acoustic cues in spectral tilt that are a result of the increased influence of glottal and subglottal impedances.

Chapter 7

Boundaries within a Breath

Prosodic boundaries coinciding with a breath boundary have been shown in the preceding chapter to have acoustic cues - in amplitude, spectral content, and F0 irregularities - that are correlated to the actions of the respiratory system as it manages the acts of renewing the pressure drive for speech and breathing for gas exchange purposes. However, prosodic boundaries are also acoustically cued even when a breath is not taken at the boundary. It could be that the actions of the respiratory system are not involved in generating the acoustic cues for boundaries internal to a breath. In that case, the changes in the sound source would be controlled by laryngeal or area of the constriction changes. However, given that the basic set of acoustic cues is the same, it could be reasonable to expect that there are similar physical dynamics generating the cues. This would imply that there should be evidence of deliberate actions in the respiratory system at those prosodic boundaries. This section extends the study of prosodic boundaries within an utterance beyond examination of the acoustics alone to include physiological signals as well.

7.1 Pauses and Boundaries

A pause, in terms of an acoustic definition, is a brief period of silence within the speech signal. There are periods of silence within speech that arise from actions of the articulators in generating some of the sound segments, such as a voiceless stop consonant like /p/, and there can be pauses between words to make lexical distinctions. Other pauses are associated with a phrase boundary within an utterance. The suspension of sound sources requires that vocal-fold vibration and/or noise-like sources be inhibited. The source terminations can be achieved by increasing the area of the constriction, decreasing the pressure drive across the constriction, obstructing the airflow (as might be done by pressing the vocal folds together) or some combination of the above actions. However, a boundary may be perceived even when there is no suspension of sound. These cases are sometimes referred to as “filled pauses.”

Locations of major boundaries in the utterance set were determined by a listening panel. (See “Perceptual Ratings of Prominences” on page 45.)

7.2 Boundary Gestures

The Campbell diagram provides a means for examining the net muscular activity involved during actions of the respiratory system. The series of figures, Figure 7.1 through Figure 7.3, shows both airflow and alveolar pressure as functions of time for segments of speech around a perceived boundary within an utterance as well as the corresponding Campbell diagram for the entire utterance. In each figure in the series, the same three utterances are displayed as produced by three different speakers (Subjects #4, #5 and #6). Locations of perceived boundaries are marked with arrows. Three examples of sound segment pairs on either side of the boundary are given: a fricative to a stop (panel A), a vowel to a vowel (panel B), and a stop to a vowel (panel C).

In all cases, there is a distinct gesture by the respiratory system at the prosodic boundary. This gesture involves either a reduction in the net expiratory muscular force, a switch from net expiratory force to net inspiratory force or an increase in the net inspiratory force. The exact details are subject to the accuracy of the estimate of the relaxed chest wall curve. The action appears in the Campbell diagram as an almost horizontal inset in the pressure-volume curve. An almost horizontal segment implies little volume change during a larger pressure change. However, as seen in the utterances in Figure 7.1a through Figure 7.3a, a relatively larger volume change occurs for the fricative to stop sound segments at the boundary (Panel A) as compared to the utterances on Panels B and C.

In some cases, there is a cessation of air flow at the boundary. Examples are in Figure 7.1c, Figure 7.2b, and Figure /.3c. As seen in the corresponding Campbell diagrams, the pressure drops quickly and then rises again during the period that the airflow is obstructed. Changes in the pressure during this time period do not influence the acoustics as no sound is being produced. However, the pressure is changing prior to and after the airflow obstruction. In Figure 7.1c, the cessation of airflow occurs from about 3.2 to 3.4 seconds. Prior to the 3.2 seconds mark, the alveolar pressure falls about 2 cm H_20 across the voiced part of the syllable "pate" in "participate." After the 3.4 seconds mark, the pressure rises about 2.5 cm H_20 at the start of the word "even." For the previously discussed relationship of SPL proportional to $(P_{ALV})^{3/2}$, the change in SPL due to pressure is roughly 5.3 dB at the start of the word "even."

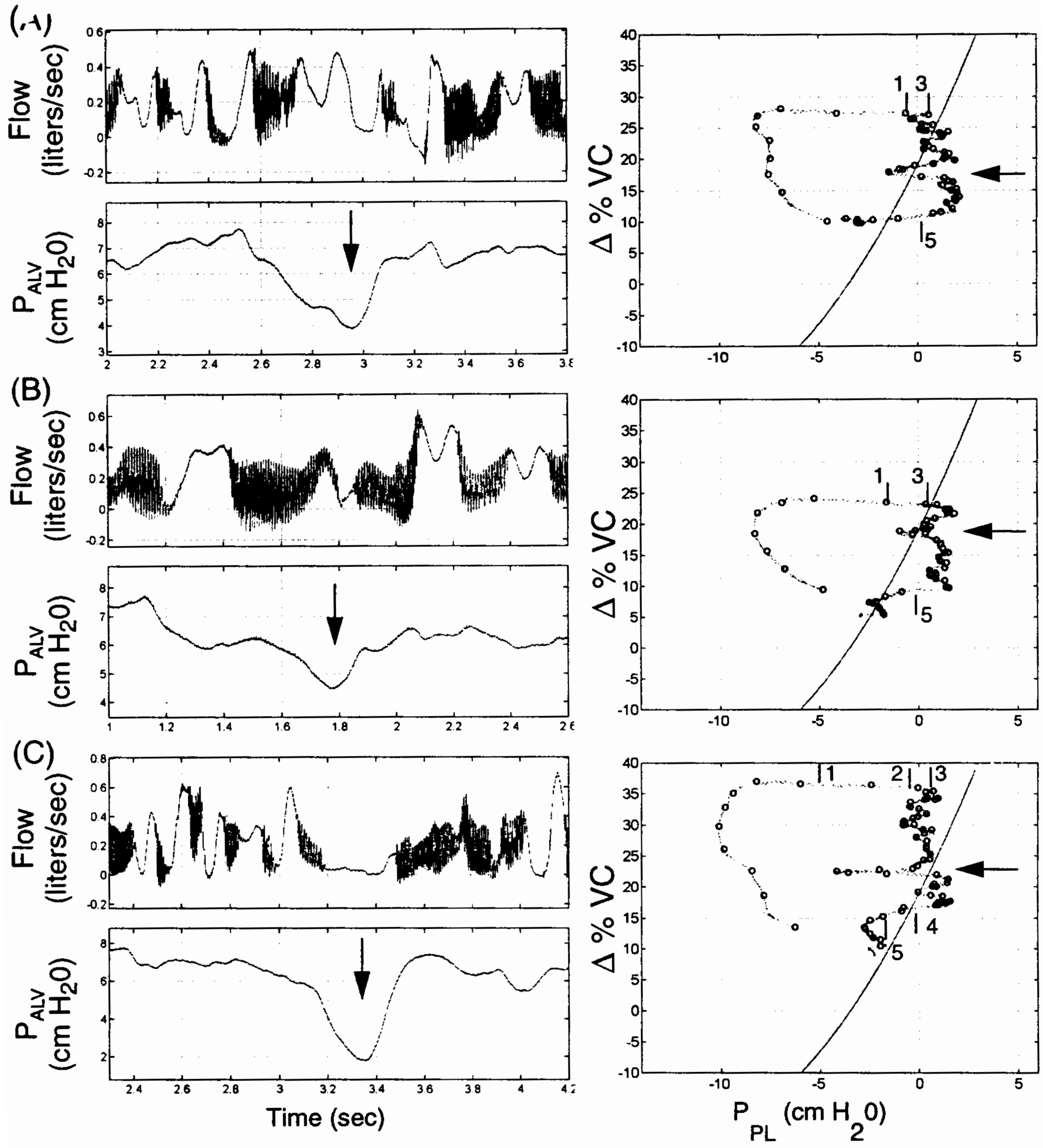


Figure 7.1: Utterances with pauses for Subject #4. On the left are the time segments around the boundary for airflow and alveolar pressure. On the right is the Campbell diagram for the entire utterance. Arrows indicate the perceived boundary. Vertical bars denote timing landmarks (1=Utterance onset, 2=Phonation onset, 3=Initial pressure peak, 4=Phonation offset, 5=Utterance offset), and dark circles are separated by 80 milliseconds.(a) Text: "Ask the operator if you can speak to Jeff, but call before the show." (b) Text: "If you say the word show, Alissa will say the word sew." (c) Text: "Six will be asked if they would be willing to participate, even though there could be eight."

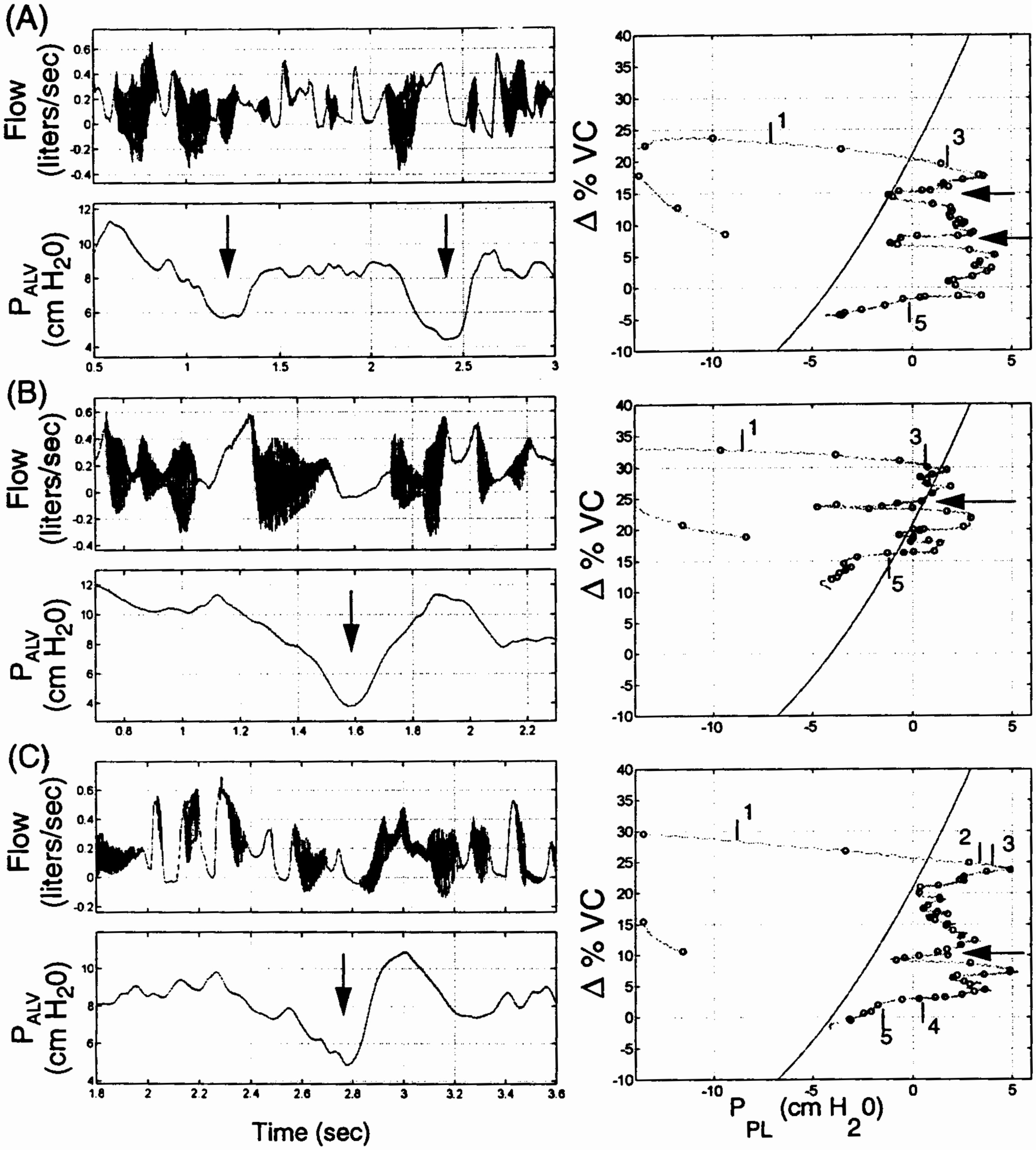


Figure 7.2: Utterances with pauses for Subject #5. On the left are the time segments around the boundary for airflow and alveolar pressure. On the right is the Campbell diagram for the entire utterance. Arrows indicate the perceived boundary. Vertical bars denote timing landmarks (1=Utterance onset, 2=Phonation onset, 3=Initial pressure peak, 4=phonation offset, 5=utterance offset), and dark circles are separated by 80 milliseconds. (a) Text: "Ask the operator if you can speak to Jeff, but call before the show." (b) Text: "If you say the word show, Alissa will say the word sew." (c) Text: "Six will be asked if they would be willing to participate, even though there could be eight."

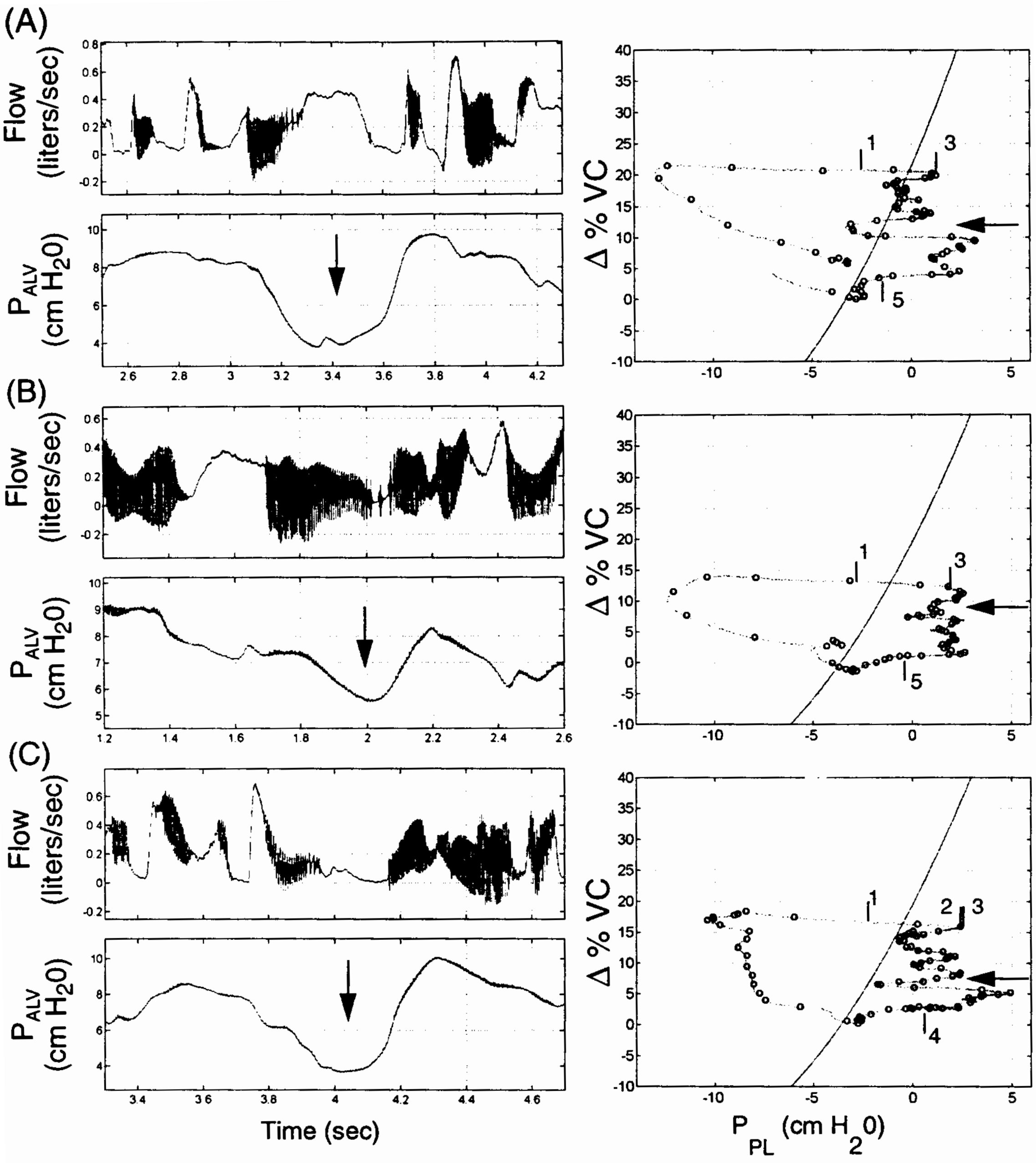


Figure 7.3: Utterances with pauses for Subject #6 On the left are the time segments around the boundary for airflow and alveolar pressure. On the right is the Campbell diagram for the entire utterance. Arrows indicate the perceived boundary. Vertical bars denote timing landmarks (1=Utterance onset, 2=Phonation onset, 3=Initial pressure peak, 4=Phonation offset, 5=Utterance offset), and dark circles are separated by 80 milliseconds. (a) Text: "Ask the operator if you can speak to Jeff, but call before the show." (b) Text: "If you say the word show, Alissa will say the word sew." (c) Text: "Six will be asked if they would be willing to participate, even though there could be eight." (The final /t/ was not released.)

The major focus of this study was on simple read utterances that, at least syntactically, were *not* expected to produce a pause or major boundary within the utterance. However, each subject did read, in isolation, 14 different utterances with a syntactic break indicated by a comma. Eight of those utterances were read twice at different points in the recording session, giving a total of 22 utterances. For all of these utterances, a boundary was perceived by the listening panel at the syntactic break indicated by the comma. In some cases, additional pauses, not aligned with a comma, were also perceived by the panel.

The task set is not exhaustive and as such is not useful for a complete specification of the conditions determining the details of this respiratory gesture. However, some general observations can be made from the data available.

In two series of utterances, there was an attempt to keep some variables constant in the utterance while moving the location of the syntactic break within the utterance. One series was as follows:

(1) Ask Jeff, but speak with him well before the start of the show.

(2) Ask to speak to Jeff, but call him well before the show.

(3) Ask the operator if you can speak to Jeff, but call before the show.

These utterances were chosen to make a first assessment of whether the respiratory system would be involved to varying degrees depending on the recoil forces available at that point in the utterance. An example of this series of utterances, in terms of Campbell curves, is shown in Figure 7.4 for Subject #4. These three utterances were all initiated near 27% VC above FRC. Recall that subject #4 used a narrow range of lung volumes to initiate utterances and showed little volume change while moving to the initial alveolar pressure peak. (See Figure 6.3 on page 73.) In all three utterances, a respiratory gesture is made that has an excursion of approximately 3 cm $H_2 0$ in pleural pressure and 3% VC in lung volume. The duration of the segment is on the order of 0.56 seconds.

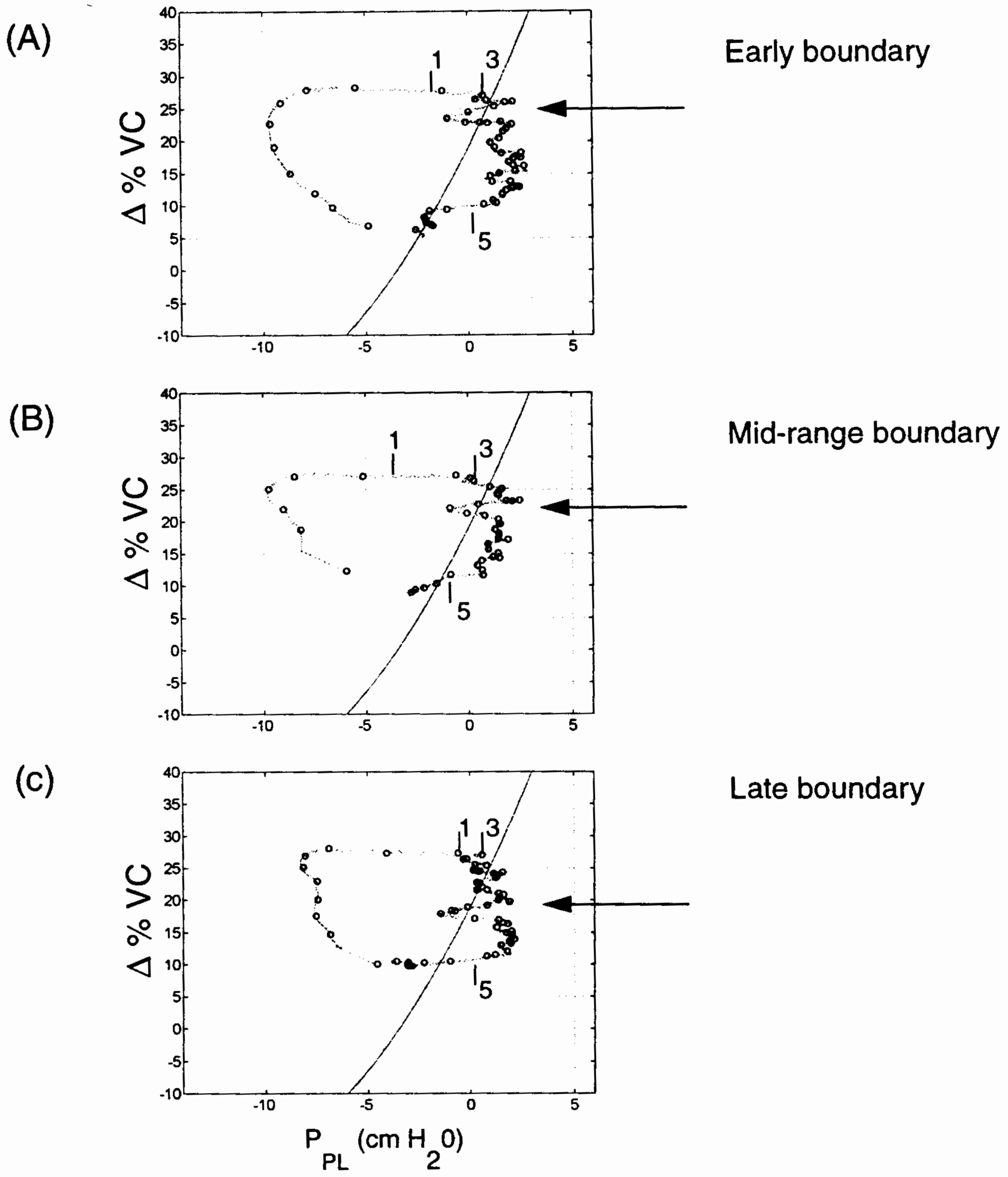


Figure 7.4: Boundary location timing within an utterance. Three utterances from Subject #4. Arrows indicate the perceived boundary. Vertical bars denote timing landmarks (1=Utterance onset, 2=Phonation onset, 3=Initial pressure peak, 4=Phonation offset, 5=Utterance offset), and dark circles are separated by 80 milliseconds. (a) Text: "Ask Jeff, but speak with him well before the start of the show." (b) Text: "Ask to speak to Jeff, but call him well before the show." (c) Text: "Ask the operator if you can speak to Jeff, but call before the show."

In Figure 7.4a through c, the leftmost excursion of the pressure-volume relationship occurs, respectively, near 24.5 % VC, 22 %VC, and 18% VC above FRC. As a whole, the utterances begin near 27% VC above FRC and end near 10%VC above FRC. Similar patterns were found for all four speakers for the utterance series used in Figure 7.4 as well as one other series of utterances with a progressive pause within the utterance. For comparison, the Campbell curves for a similar series of utterances for the same subject (#4) are shown in Figure 7.5. These sentences read for this series of utterances also had a syntactic break at different locations. The sentences are:

(1) Six will participate, even though there is room in the program for as many as eight.

(2) Six will be asked to participate, even though there is room for eight.

(3) Six will be asked if they would be willing to participate, even though there could be eight.

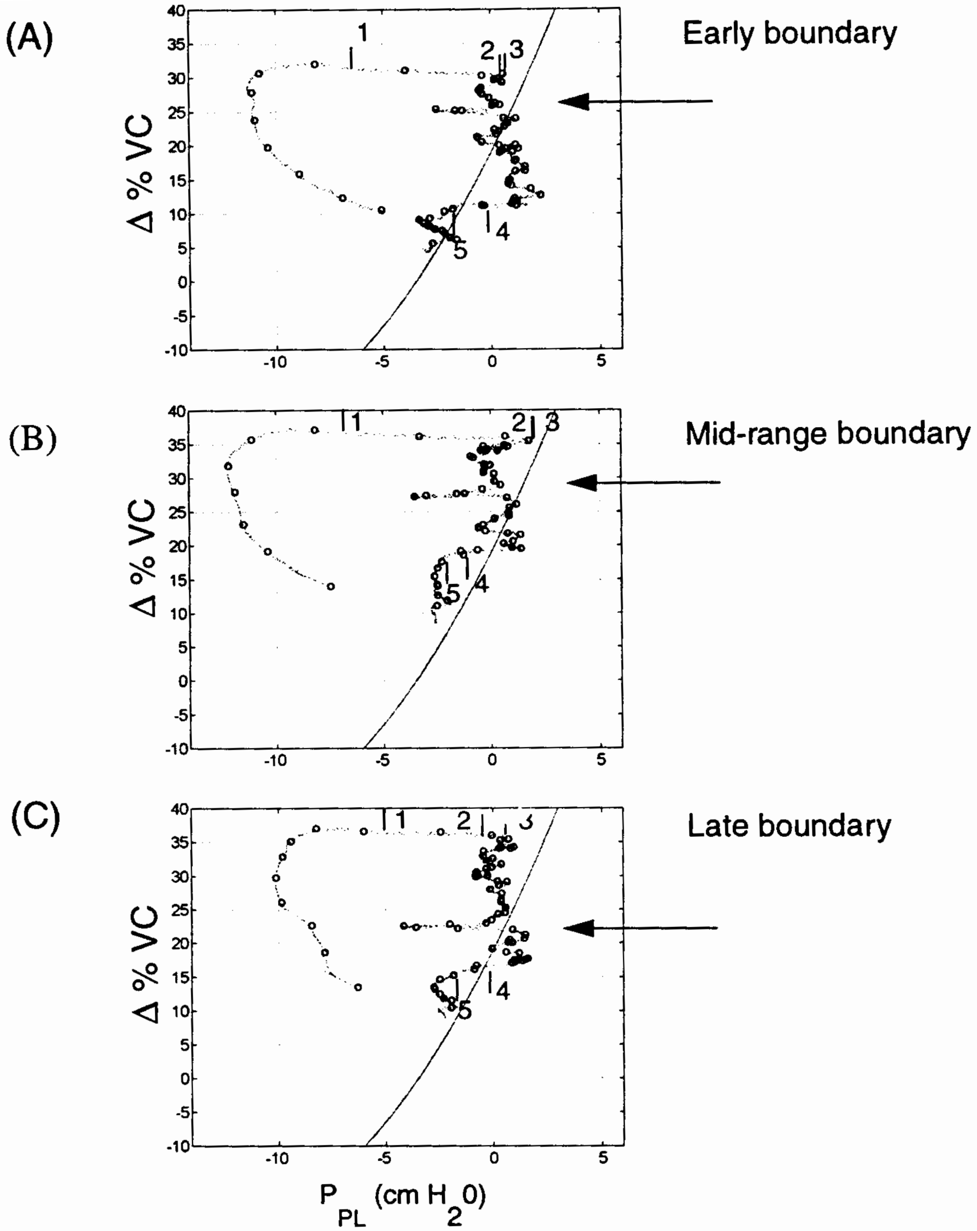


Figure 7.5: Boundary location timing within an utterance. Three utterances from Subject #4. Arrows indicate the perceived boundary. Vertical bars denote timing landmarks (1=Utterance onset, 2=Phonation onset, 3=Initial pressure peak, 4=Phonation offset, 5=Utterance offset), and dark circles are separated by 80 milliseconds. (a) Text: "Six will participate, even though there is room in the program for as many as eight." (b) Text: "Six will be asked to participate, even though there is room for eight." (c) Text: "Six will be asked if they would be willing to participate, even tough there could be eight."

During data acquisition, the set of utterances read by each subject included a few utterances that were repeated. One of these utterances is the utterance in Panel B of Figures 7.1 through 7.3. The second reading of this utterance for Subject #6 is shown in Figure 7.6. In this case, the voicing through the boundary is fairly continuous. The fall in F0 does not include irregular F0 as in Figure 7.3b, but rather shows a fall to a comparatively low F0. By comparison, the minimum pressure at the boundary is about 7.5 cm H_20 for the reading in Figure 7.6a where the F0 lowers but does not have an irregular glottal waveform. For Figure 7.6b, the minimum pressure at the boundary is about 5.5 cm H_20 and there is an irregular glottal waveform. This difference suggests that future work should examine the types of mechanisms that lead to in irregular glottal waveform, and pressure at the boundary could be a factor. Again, there is a respiratory gesture at the boundary even though there is clearly not a period of silence in the speech signal.

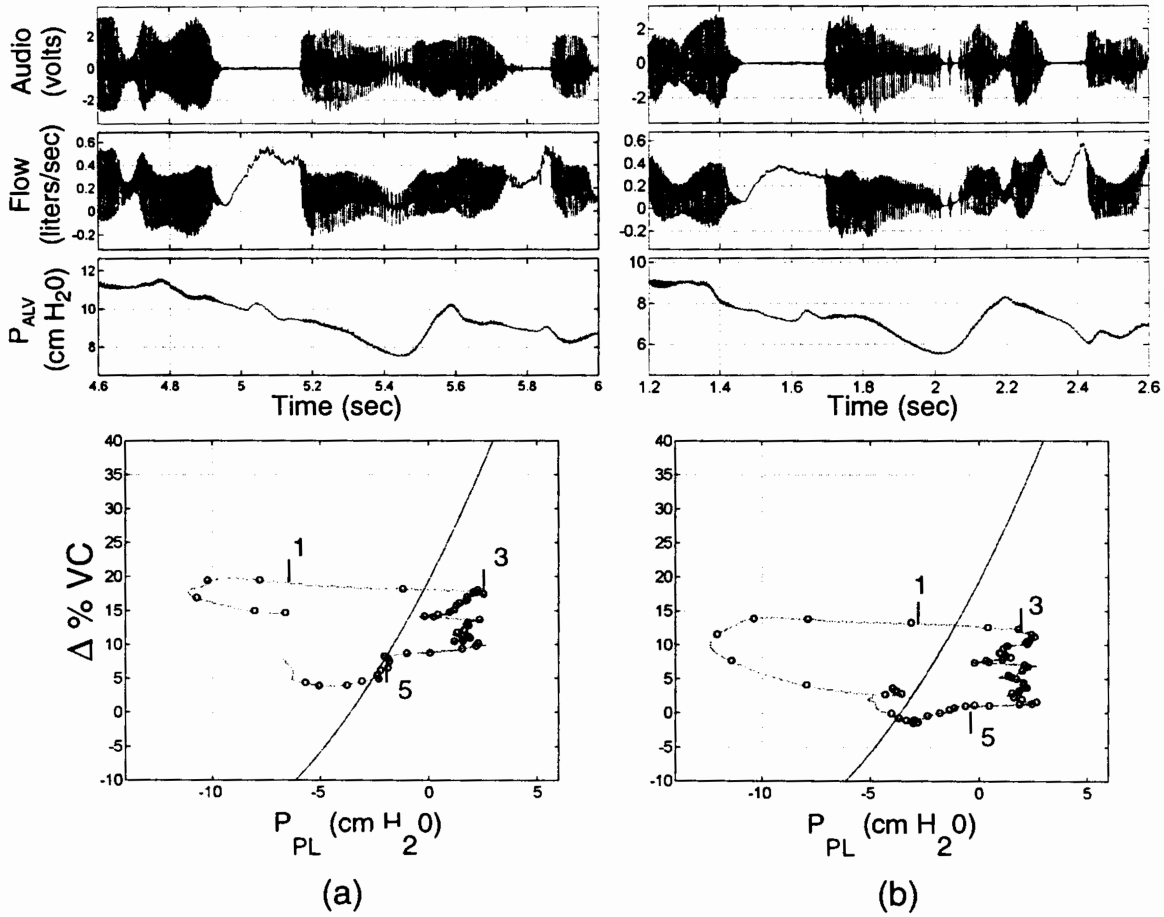


Figure 7.6: Voiced boundary from Subject #6. The top three panels in are plots of audio (volts), airflow (liters/sec) and P_{ALV} (cm H_20) for the time segment bracketing the rated boundary. The bottom panel is the modified Campbell diagram for the entire utterance. Vertical bars denote timing landmarks (1=Utterance onset, 3=Initial pressure peak, 5=Utterance offset), and dark circles are separated by 80 milliseconds. Both panels represent the utterance "If you say the word show, Alissa will say the word sew." (a) Regular F0 continues across the boundary. (b) Irregular F0 occurs at the boundary.

7.3 Discussion

Researchers found that during read speech, subjects created the same pause structure (placement of pauses in relation to the words in the sentence) independent of the number of breaths used to generate the speech (e.g. Grosjean et al.,1979). The speaker fits the breaths in when they are able to do so, and that usually occurs at major constituent boundaries such as utterance boundaries. Presumably, then, between breaths, the standard role of the respiratory system is in effect: that of providing a relatively constant subglottal pressure. However, the research presented here shows that the respiratory system is still involved in the generation of the cues for the boundary or pause.

During speech, the respiratory system is not operating near its limits of performance. The lung volume range is not that much beyond that used for resting breathing. The range of pressures, 5-10 cm H_20, is much less than the maximal 100 cm H_20 that the respiration system can produce. As, such, it is not an activity that requires gross muscular manipulation, but rather can employ a finer, more subtle control. The speaker is able to execute the actions needed to produce a gesture in the respiratory system at the boundary. This respiratory system gesture appears to be one in which airflow and the pressure are reduced, as seen in the Campbell diagrams.

The respiratory gesture was apparent at both silent and filled pauses associated with a phrase boundary. The gesture may be motivated by a need to produce the appropriate acoustic cues to signal the boundary prior to and following the boundary. At the end of the phrase, the falling pressure reduces the signal amplitude but the speaker may also spread the area of the constriction to change the spectral content. If so, an increase in airflow is averted by decreasing the pressure. Or, it could be that in order to create a boundary using an obstruction of airflow, the pressure is decreased to create the signal cues while at the same time avoiding an inappropriate build-up of pressure against the obstructed airway during the pause. The example in this study showed a relatively large rise in alveolar pressure at the start of the next phrase after a boundary. Such a change leads to an audible amplitude change for the listener. In some cases, voicing through the boundary showed regular but low F0 or an irregular glottal waveform. In this particular example, the pressure level was higher at the boundary that showed a low but regular F0 than the pressure at the boundary that showed irregular oscillation. This finding suggests that future studies might benefit from examining the specific instances when irregular F0 occurs at a boundary with the alveolar pressure levels.

The presence of this respiratory gesture adds another layer to the presumed function of the respiratory system during speech. This function has been stated to be the generation of a relatively constant subglottal pressure in the required range for speech. (See "The role of the respiration system in speech production" on page 20.) The results of this study indicate that the respiratory system is actively involved in the generation of pauses within an utterance.

Chapter 8

Summary

The major goal of this study has been to characterize the respiratory dynamics at speech boundaries and determine the correlations to prosodic acoustic cues for those boundaries. Simultaneous recordings of the acoustic signal and several physiologically-related signals were collected, analyzed and used to estimate alveolar pressure and area of the constriction in the airway. Perceptual ratings were used to provide information on the prominences that were closest to the boundaries.

The speech data consisted of read utterances and many were read in isolation. The environment was a laboratory setting where the subjects were wearing a variety of measuring devices, including an airflow mask and an esophageal balloon that passed through their nasal passages. Laboratory conditions, measurement noise, and biological noise are all factors influencing the data, 'ɔut consistent and quantitative trends were observed at prosodic boundaries.

8.1 Boundary Cues

This research has shown that utterances begin prior to the point at which the respiratory system has generated a 'relatively constant working level' for alveolar pressure during speech. Rather than the general trend being to initiate utterances with primarily net expiratory force (as in Hixon,1976 and Johnston et al.,1999) or with primarily a net inspiratory force acting to brake the recoil (as in Ladefoged,1962), this work has shown that most utterances begin during a rapid transition in muscular force from a strong inhalatory force to the level of muscular force necessary to produce pressure in the speech range. This consistent trend was observed in modified Campbell curves where the onset of the utterance occurred primarily during net inspiratory force. Prior studies were not focused on the details of the onset portion of the speech.

This quick rise in pressure has been further shown to have timing landmarks for phonation onset where phonation begins at relatively low pressures if the initial sound segment is voiced. Pressure rises through the voiced segment leading to changes in signal amplitude and spectrum. Phonation onset pressure is relatively high for an unvoiced fricative onset compared with a voiced sonorant onset. These pressures are taken as relative to the initial pressure peak. The current research agrees with prior studies (such as Atkinson,1973) in terms of such landmarks at rise time, but used an expanded utterance set, male and female subjects, and normal read speech without demands on prominence placement, lung volume, or loudness. The timing of the initial pressure peak fell within the initial perceived prominent syllable if that syllable was early in the utterance. However, it appears that because of the rapid nature of the transition to the alveolar pressure peak and the apparent use of the recoil forces of the chest wall, the timing of that rise is not delayed to coincide with initial perceived prominences that occur later in the utterance. The specific actions in aligning the pressure peak with the initial prominence could involve the respiratory system or could involve the articulators and the specific realization of the initial syllables in the utterance.

At the end of an utterance, the tendency is to apply little net muscular force. While alveolar pressure fell on the order of 1-3 cm H_20 during the final amplitude fall, the termination of voicing was largely achieved through an increase in glottal area. The corresponding acoustic cues from apparent increased coupling to the subglottal space indicate that the spectral changes and amplitude fall are correlated cues that should co-vary, and this covariance should be realized in a speech synthesis scheme. Several types of endings in phonation occurred within the data set. A smooth taper to voicing amplitude with a regular F0 throughout most commonly occurred with a falling pressure and rising area. In many cases, ending F0 irregularities were present and were linked to an increase in glottal area. As area continued to increase, voicing returned to a regular F0. This pattern was in contrast to the classical definition of glottalization as a largely adducted gesture.

Boundaries within an utterance, such as pauses, that are not linked to the initiation or termination of a breath, were seen to be marked with a muscular gesture from the respiratory system. This gesture was present for perceived pauses in which there was a period of silence in the utterances as well as perceived pauses where voicing continued through the boundary. The respiration system involvement was present for various types of sound segments at the boundary as well as various locations of the pauses within the utterance. Evidence of this gesture was even seen when the airway was completely obstructed during the pause. This gesture adds another layer to the role of the respiration system during speech. The respiratory system is active in generating acoustic cues for some prosodic boundaries that are not aligned with the initiation or termination of exhalation.

8.2 Future Directions

While the current research did collect segments of spontaneous speech, in the form of responses to questions, those data were not included in the analysis. The primary goal of the research involved the establishment of a baseline of behavior at the initiations and terminations of respiratory activity and speech boundaries. Future analysis should examine the types of modifications that are made during spontaneous speech as well as during conversation.

The data provided here can be used to aid the development of a model of the respiratory system during speech that is capable of producing accurate subglottal pressure variations at major speech boundaries (such as utterance and pause boundaries). Some future experiments may attempt to expand measures of the details of the respiratory muscular gesture within a breath. Such studies should include a measured value for the chest wall relaxation curve and would also benefit from electromyography of some of the muscles of inspiration and expiration.

There appear to be many types of conditions that lead to irregularities in the fundamental frequency during voicing. Many of these irregularities occur at major prosodic boundaries. Attempts to acquire data about the physiological actions during these instances can lead to models for the synthesis of such irregularities and can supplement methods for predicting prosodic cues from text.

This research has provided a baseline for some of the parameters related to the respiration system during speech. In some cases of speech disorders or speech produced by people with little or no sense of hearing, such a baseline could provide a starting point for studies determining the departure from the baseline. The results of this study could contribute significantly to the development of an articulatory-based speech synthesis system, particularly for the synthesis of acoustic events at the initiation and termination of utterances and at pauses within an utterance.

References

[1] Agostoni, E. and Mead, J. (1964) Statics of the respiratory system. In: W.O. Fenn and H. Rahn (eds.) *Handbook of Physiology, Section 3, Respiration. Vol. I,* Chapter 13. Washington, D.C.: American Physiological Society, 387-409.

[2] Annoni, J., Cot, F., Ryalls, J., and Lecours, A. (1993) Profile of the aphasic population in a Montreal geriatric hospital: a 6-year study. *Aphasiology,* 7(3), 271-284.

[3] Atkinson, James E. (1973) Aspects of Intonation in Speech: Implications from an Experimental Study of Fundamental Frequency, Ph.D. Thesis, The University of Connecticut.

[4] Baydur, A., Behrakis, P.K., Zin, W.A., Jaeger, M., and Milic-Emili, J. (1982) A simple method for assessing the validity of the esophageal balloon technique. *Am Rev Respir Dis*, 126(5), 788-791.

[5] Beckman, M.E. and Pierrehumbert, J. (1986) Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255-309.

[6] Bickley, C.A., Stevens, K.N., and Williams, D.R. (1997) A framework for synthesis of segments based on pseudoarticulatory parameters. In: J.P.H. van Santen, R.W. Sproat, J.P. Olive, and J. Hirschberg (eds.) *Progress in Speech Synthesis*. New York: Springer. 211-220.

[7] Bolinger, D.L. (1958) A theory of pitch accent in English, *Word*, 14, 109-149.

[8] Bouhys, A., Proctor, D., and Mead, J. (1966) Kinetic aspects of singing. *Journal of Applied Physiology,* 21, 483-496.

[9] Bouhys, A., Mead, J., Proctor, D.F., and Stevens, K.N. (1968) Pressure-flow events during singing. *Annals of the New York Academy of Sciences*, 55, 165-176.

[10] Byrd, D. and Saltzman, E. (1998) Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics,* 26, 173-199.

[11] Childers, D.K. and Krishnamurthy, A.K. (1985) A critical review of electroglottography, *CRC Critical Reviews in Biomedical Engineering*, 2(2), 131-161.

[12] Dilley, L., Shattuck-Hufnagel, S. and Ostendorf, M. (1996) Glottalization of word-initial vowels as a function of prosodic structure, *Journal of Phonetics*, 24, 423-444.

[13] Draper, M., Ladefoged, P., and Whitteridge, D. (1959) Respiratory muscles in speech, *Journal of Speech and Hearing Research*, 2, 16-27.

[14] Estenne, M., Yernault, J-C., and De Troyer, A. (1985) Rib cage and diaphragm-abdomen compliance in humans: effects of age and posture. *Journal of Applied Physiology,* 59, 1842-1848.

[15] Fant, G. (1960) *Acoustic Theory of Speech Production.* 's-Gravenhage: Mouton and Co.

[16] Fant, G. (1987) Interactive phenomena in speech production, in *Proceedings of XIth International Congress of Phonetic Sciences (ICPhS),* Tallin, USSR, 3, 376-381.

[17] Fant, G., Hertegård S., Kruckenberg, A., and Liljencrants, J. (1997a) Covariation of subglottal pressure, F0 and glottal parameters. *Proc. Eurospeech '97*, Rhodes, 453-456.

[18] Fant, G., Hertegård S., Kruckenberg, A., and Liljencrants, J. (1997b) Accentuation and subglottal pressure in Swedish. *FCSA Workshop on Intonation*, Athena.

[19] Fant, G. and Kruckenberg, A. (1989) Preliminaries to the study of Swedish prose reading and reading style, *STL-QPSR*, 2/1989, 1-83.

[20] Fant, G. and Kruckenberg, A. (1994) Notes on stress and word accent in Swedish, *STL-QPSR*, 2-3/1994, 125-144.

[21] Fougeron, C. and Keating, P.A. (1997) Articulatory strengthening in prosodic domain-initial position, *UCLA Working Papers in Phonetics*, 92, 61-87.

[22] Fry, D.B. (1958) Experiments in the perception of stress, *Lang. Speech*, 1, 126-152.

[23] Gold, W. M. (1994) Pulmonary function testing. In: J. Murray and J. Nadel (eds.) *Textbook of Respiratory Medicine, Volume I, 2nd Edition*, Philedelphia: W.B. Saunders Co., 817-833.

[24] Grosjean, F., Grosjean, L. and Lane, H. (1979) The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, 11, 58-81.

[25] Hammen, V.L. and Yorkston, K.M. (1996) Respiratory Patterning and Variability in Dysarthric Speech. In: D.A. Robin, K.M. Yorkston, and D.R. Beukelman (eds.) *Disorders of Motor Speech: Assessment, Treatment, and Clinical Characterization*, Baltimore: Paul H. Brookes Publishing Co., 181-192.

[26] Hayes, B. (1989) The prosodic hierarchy in meter. In: P. Kiparsky and G. Youmans (eds.), *Phonetics and Phonology, Vol 1: Rhythm and Meter.* San Diego: Academic Press. 201-260.

[27] Henderson, A., Goldman-Eisler, F. and Skarbek, A. (1965) The common value of pausing time in spontaneous speech, *Quantitative Journal of Experimental Psychology*, 17, 343-345.

[28] Henton, C. and Bladon, A. (1988) Creak as a sociophonetic marker. In: L.M. Hyman and C.N. Li (eds.) *Language, Speech, and Mind: Studies in Honour of A. Fromkin*, London: Routledge. 2-29.

[29] Hertegård, S. (1994) Vocal fold vibration as studied with flow inverse filtering, Med. Dr. Thesis, Karolinska Institute, Sweden.

[30] Hixon, T.J., Mead, J., and Goldman, M.D. (1976) Dynamics of the chest wall during speech production: Function of the thorax, rib cage, diaphragm, and abdomen. *Journal of Speech and Hearing Research*, 19, 297-356.

[31] Hoole, P. and Ziegler, W. (1997) A comparison of normals' and aphasics' ability to plan respiratory activity in overt and covert speech. In: W. Hulstijn, H.F.M. Peters, and P.H.H.M. Van Lieshout (eds.) *Speech Production: Motor Control, Brain Research and Fluency Disorders.* Amsterdam: Elsevier. 77-80.

[32] Isshiki, N. (1964) Regulatory mechanism of voice source intensity variation. *Journal of Speech and Hearing Research*, 7, 233-244.

[33] Johnston, S., Yan, S., Sliwinski, P., and Macklem, P.T. (1999) Modified Campbell diagram to assess respiratory muscle action in speech. *Respirology*, 4(3), 213-222.

[34] Klatt, D. and Klatt L. (1990) Analysis, synthesis, and perception of voice quality variations among female and male talkers, *Journal of the Acoustical Society of America*, 87(2), 820-57.

[35] Knowles, J.H., Hong, S.K. and Rahn, H. (1959) Possible errors using esophageal balloons in determination of pressure-volume characteristics of the lung and thoracic cage, *Journal of Applied Physiology*, 14(4), 525-530.

[36] Konno, K. and Mead, J. (1967) Measurement of the separate volume changes of rib-cage and abdomen during breathing. *Journal of Applied Physiology*, 22, 407-422.

[37] Kunze, L. (1964) Evaluation of methods of estimating sub-glottal air pressure. *Journal of Speech and Hearing Research*, 7, 151-164.

[38] Ladefoged, P. (1962) Subglottal activity during speech. In: *Proceedings of the Fourth International Congress of Phonetic Sciences.* The Hague, Netherlands: Mouton. 73-91.

[39] Ladefoged, P. (1967) *Three areas of experimental phonetics*. London: Oxford University Press.

[40] Ladefoged, P. (1982) A Course in Phonetics, Second Edition. Jovanovich, NY: Hartcourt Brace.

[41] Lieberman, P. (1967) *Intonation, perception, and language*. Cambridge, MA: MIT Press.

[42] Liljencrants, J. (1999) Judges of prominence in *Gothenburg Papers in Theoretical Linguistics, Proceedings of Fonetik 99*, Gothenburg, Sweden, 81, 57-60.

[43] Lehiste, I., Olive, J.P. and Streeter, L.A. (1976) The role of duration in disambiguating syntactically ambiguous sentences, *Journal of the Acoustical Society of America,* 60, 1199-1202.

[44] Loring, S. H. (1998) Mechanics of the lungs and chest wall. In: J. J. Marini and A. S. Slutsky (eds.) *Physiological Basis of Ventilatory Support.* New York: Marcel Dekker. 177-208.

[45] Loring, S. H. and Mead, J. (1982) Abdominal muscle use during quiet breathing and hyperpnea in uninformed subjects. *Journal of Applied Physiology: Respirat. Environ. Exercise Physiol.*, 52(3), 700-704.

[46] Mead, J. and Agostoni, E. (1964) Dynamics of breathing. In: W.O. Fenn and H. Rahn, (eds.) *Handbook of Physiology, Section 3, Respiration. Vol. I, Chapter 14*. Washington, D.C.: American Physiological Society. 411-427.

[47] Nespor, M. and Vogel, I. (1986) *Prosodic Phonology.* Dordrecht: Foris Publications.

[48] Netsell, R. (1973) Speech physiology. In: F. D. Minifie, T. J. Hixon and F. Williams (eds.) *Normal Aspects of Speech, Hearing, and Language*, Englewood Cliffs, NJ: Prentice-Hall. 211-234.

[49] Ohala, J.J. and Ladefoged, P. (1970) Further investigation of pitch regulation in speech. *UCLA Working Papers in Phonetics*, 14, 12-24.

[50] Ohala, J.J. (1990) Respiratory activity in speech, In: W.J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modelling, NATO ASI Series D: Vol. 55*. Boston: Kluwer Academic Publishers. 23-51.

[51] Oller, D.K. (1973) The effect of the position in utterance on speech segment duration in English, *Journal of the Acoustical Society of America*, 54, 1235-1247.

[52] Peters, H.F.M. and Boves, L. (1988) Coordination of aerodynamic and phonatory processes in fluent speech utterances of stutterers, *Journal of Speech and Hearing Research,* 31, 352-361.

[53] Pierrehumbert, J. and Beckman, M. B. (1988) *Japanese Tone Structure.* cambridge, MA: MIT Press.

[54] Pierrehumbert, J. and Talkin, D. (1992) Lenition of /h/ and glottal stop. In: G. Doherty and D.R. Ladd (eds.) *Papers in laboratory phonology II: gesture segment prosody.* Cambridge: Cambrdige Uniceristy Press. 90-117.

[55] Rahn, H., A. B. Otis, L. E. Chadwick, and W. O. Fenn. (1946) The pressure-volume diagram of the thorax and lung. *Am. J. Physiol.* 146: 161-178.

[56] Rothenberg, M. (1973) A new inverse-filtering technique for deriving the glottal airflow waveform during voicing. *Journal of the Acoustical Society of America*, 89, 1777-1781.

[57] Rothenberg, M. (1983) Source-tract acoustic interaction in breathy voice. In I.R. Titze and R.C. Scherer (eds.) *Vocal fold physiology, biomechanics, acoustics, and phonatory control.* Denver: The Denver Center for the Performing Arts. 465-481.

[58] Selkirk, E. O. (1978) On prosodic structure and its relation to syntactic structure. In T. Fretheim (ed.) *Nordic Prosody II*, Trondheim: TAPIR.

[59] Selkirk, E. O. (1984) *Phonology and Syntax: The Relation Between Sound and Structure*, Cambridge, MA: MIT Press.

[60] Selkirk, E. O. (1986) On derived domains in sentence phonology. *Phonology Yearbook 3,* 371-405.

[61] Sharp, J.T., Johnson, F.N., Goldberg, N.B., and van Lith, P. (1967) Hysteresis and stress adaptation in the human respiratory stsrem, *Journal of Applied Physiology*, 23(4), 487-496.

[62] Shattuck-Hufnagel, S. and Turk, A. (1996) A prosody tutorial for investigators of auditory sentence processing, *Journal of Psycholinguistic Research*, 25(2), 193-247.

[63] Silverman, K., Beckman, M.B., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (1992) ToBI: A standard for labeling English prosody. In: *Proceedings of the International Conference on Spoken Language Processing*, Banff, II, 867-870.

[64] Stevens, K. N. (1994) Prosodic influences on glottal waveform: preliminary data. In *Proceedings of the International Symposium on Prosody*, Yokohama, Japan, 53-63.

[65] Stevens, K. N. (1998) *Acoustic Phonetics*. Cambridge, MA: MIT Press.

[66] Stevens, K. and Bickley, C. (1991) Constraints among parameters simplify control of Klatt formant synthesizer, *Journal of Phonetics*, 19(1), 161-174.

[67] Streeter, L.A. (1978) Acoustic determinants of phrase boundary perception, Journal of the Acoustical Society of America, 64(6), 1582-1592.

[68] Titze, I.R. (1988) The physics of small-amplitude oscillation of the vocal folds, *Journal of the Acoustical Society of America*, 83(4), 1536-1552

[69] Titze, I.R. (1989) On the relation between subglottal pressure and fundamental frequency in phonation. *Journal of the Acoustical Society of America*, 85(2), 901-906.

[70] Titze, I.R. (1992) Phonation threshold pressure: A missing link in glottal aerodynamics, *Journal of the Acoustical Society of America*, 91(5), 2926-2935.

[71] Van den Berg, Jw. (1956) Direct and indirect determination of mean subglottal pressure, *Folia Phoniatrica*, 8, 1-24.

[72] Weismer, G. (1985) Speech breathing: Contemporary views and findings. In: R. G. Daniloff (ed.) *Speech Science*, San Diego: College Hill Press. 47-72.

[73] Wohl, M., Turner, J. and Mead, J. (1968) Static volume-pressure curves of dog lungs - in vivo and in vitro. *Journal of Applied Physiology*, 24(3), 348-354.

[74] Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M. and Price, P. (1992) Segmental durations in the vicinity of prosodic phrase boundaries, *Journal of the Acoustical Society of America*, 91, 1707-1717.

Appendix A

Alveolar Pressure Calibration Figures

The calibration procedure for deriving alveolar pressure from measures of esophageal pressure was discussed in Chapter 3. This appendix provides plots (Figure A.1) of the lung recoil curves as determined from calibration maneuvers. Once the calibrations were completed, the procedure was checked through a comparison of intraoral pressure and alveolar pressure while subjects produced a series of /p ae/ syllables. Figure A.2 shows such a series for each subject in the study.

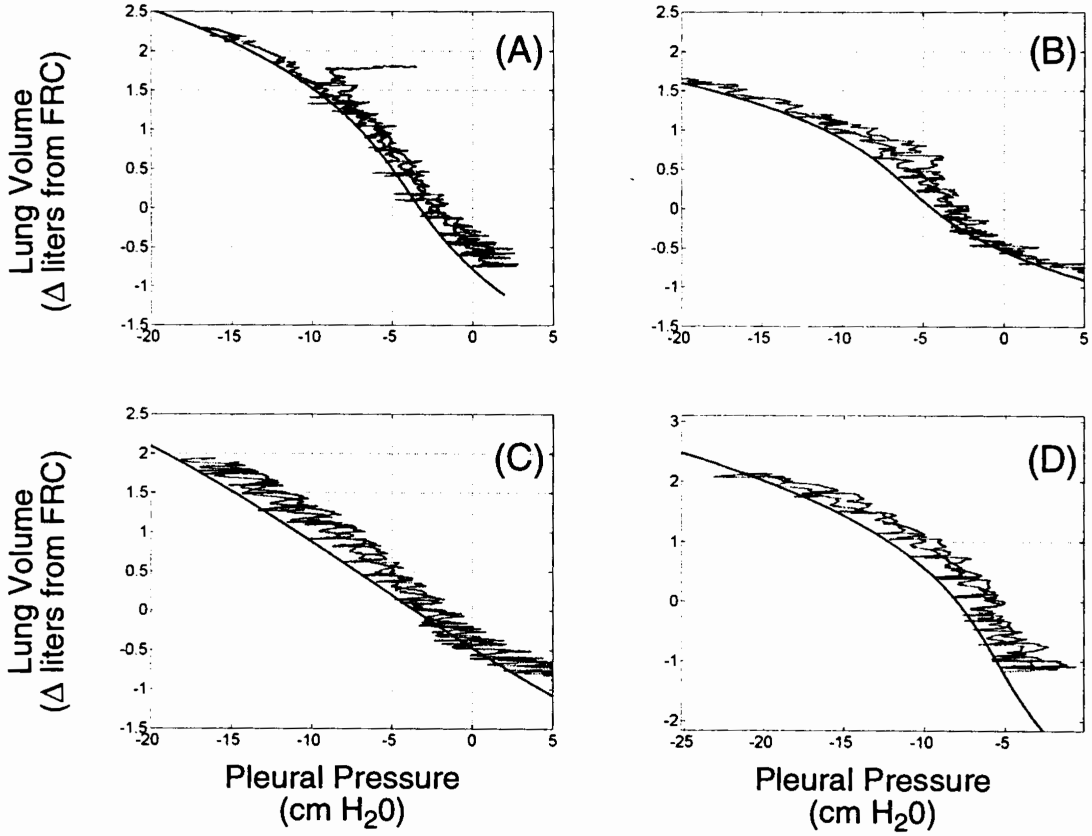


Figure A.1: Lung recoil pressure curves. Each panel shows the overlay of two calibration maneuvers and the third-order curve that was fit to the points of differential pressure measurement. (a) Subject #4 (b) Subject #5 (c) Subject #6 (d) Subject #7

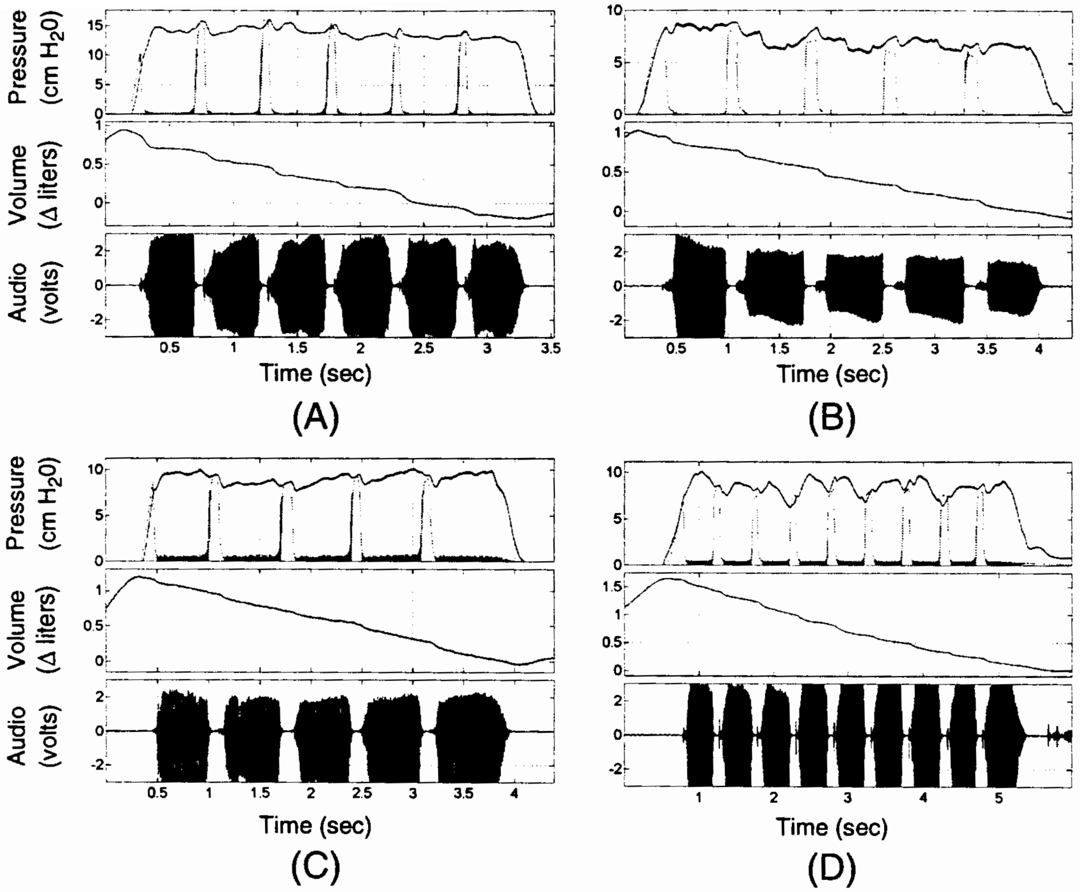


Figure A.2: Examples of intraoral pressure and alveolar pressure during production of a series of /p ae/ syllables. The top panel shows alveolar pressure (dark line) and intraoral pressure (gray line) in cm H_20. The middle panel is lung volume (liters from FRC) and the bottom panel is audio (volts). During the /p/ closure, alveolar pressure should be within 1 cm H_20 of intraoral pressure. (a) Subject #4 (b) Subject #5 (c) Subject #6 (d) Subject #7.