Analysis and Synthesis of Fricative Consonants

by

Lorin F. Wilde

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1995

Author .....
Department of Electrical Engineering and Computer Science
February 11, 1995

Certified by .....
Kenneth N. Stevens
Clarence J. Lebel Professor of Electrical Engineering
Thesis Supervisor

Accepted by .....
Frederic R. Morgenthaler
Chairman, Departmental Committee on Graduate Students

Analysis and Synthesis of Fricative Consonants

by

Lorin F. Wilde

Submitted to the Department of Electrical Engineering and Computer Science
on February 12, 1995, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis studies and interprets the inventory of acoustic events associated with the changing vocal-tract configurations that characterize fricatives preceding vowels. Theoretical considerations of the articulatory, aerodynamic and acoustic aspects of the production of fricatives provide the foundation for interpreting the acoustic data. Fricative characteristics are considered with respect to the adjacent vowel. All eight English fricatives in six vowel contexts were studied using a controlled database of consonant-vowel sequences recorded by four speakers. There are two separate acoustic analysis components: 1) analysis of formant patterns and 2) analysis of time-varying noise characteristics. In addition, listening tests with utterances containing synthetic fricative consonants are used to determine which acoustic aspects are perceptually important. The acoustic data are applied to modify the theory, where appropriate, and to set the time-varying control parameters in a speech synthesizer.

Detailed acoustic analysis begins by first identifying events or landmarks in the speech signal, such as at consonant-vowel (CV) and vowel-consonant (VC) boundaries. Once landmarks are located, signal processing and analysis can be focused in their vicinity, where change is occurring and information is concentrated. The strategy is to extract appropriate acoustic information in the vicinity of these landmarks, and also in the intervals when the vocal tract is most and least constricted. This thesis takes a kinematic view of the acoustics of fricative consonants, rather than making the more common, stationary assumptions.

In the study of formant onset patterns we focus on the transitions of the second and third formants. Formant transitions at the release from a fricative into a vowel allow the positions of the major articulators to be inferred. It is shown that variability in the acoustic signal at the release of a fricative decreases when production of the fricative places greater constraints on the position of the tongue body.

In the series of studies on the time-varying noise characteristics of fricatives, we further quantify the acoustic attributes of turbulence noise generated in fricatives with respect to adjacent vowels. We develop an analysis system, using averaging methods, to examine how fricative spectra change with time. The attribute of stridency, signaled by greater energy in the high frequencies in the consonant relative to the vowel, was examined. The weak and strong fricatives were well-separated:

the maximum amplitude above 2 kHz in the fricative, normalized relative to vowel amplitude, is 15-20 dB more for /s/ and /š/ than for /f/ and /θ/.

In addition, we use speech-copying techniques to investigate the effects of varying the consonant-vowel ratio and mimicking observed time variations over the duration of the fricative. The objective is to determine the acoustic characteristics that have perceptual importance for speech understanding systems. A significant difference in the production constraints between nonstrident and strident fricatives was compatible with perceptual results using synthetic stimuli. For example, listeners were more tolerant of the manipulations in the consonant-vowel ratio for /f/ than /s/. We failed to demonstrate that listeners are sensitive to time variations in fricative noise when judging naturalness. However, we have shown that listeners reject stimuli with energy at unexpected times and frequencies.

Finally, we conclude by relating what we learned about the acoustics of fricative consonants to the models of the relative timing of the changing vocal-tract configurations. For example, the spectral differences between strident and nonstrident fricatives suggested that source-filter models, i.e., models of the filtering of the noise source by the cavity in front of the constriction, might be improved if the losses in the vocal tract and location of the noise source were better represented.

Thesis Supervisor: Kenneth N. Stevens
Title: Clarence J. Lebel Professor of Electrical Engineering

Acknowledgments

It is with great pleasure that I end the writing of this thesis by recognizing the support I've received from the beginning of this endeavor.

Deep gratitude goes to my thesis advisor, Ken Stevens, who supported my continuing education and told me from the outset that graduate school should be fun. It is a privilege and pleasure to be among the diverse set of researchers who have been supervised by Ken. He generously shares his extraordinary insights and enthusiasm, while encouraging independence of ideas.

The contribution of each of my readers cannot be underestimated. Victor Zue provided clarity and perspective, and I always benefitted from his thoughtful advice and questions. Charlotte Reed was always available for discussing how to formulate ideas. And I was fortunate to have Cam Searle as a trustworthy presence on all my committees; he always kept me honest.

Thanks go to my Oral Qualifying Exam committee, Al Grodzinsky, Tom Knight and Cam, who listened to the earliest design of my first speech analysis and synthesis system, and then recommended that I get back in the lab and build it. Thanks also go to my Area Exam committee, Ken, Cam and Larry Frishkopf, who participated in my opportunity to explore auditory modelling.

Bill Peake was inspirational, and an integral part of the learning process. Formative discussions were ongoing with Jim Glass, Li Deng, Mark Randolph, Yi Xu, and Alice Turk, among other colleagues. Melanie Matthies certainly made statistics comprehensible and enjoyable. Software tool development and collaboration with Mark Johnson, Dan Ellis, Ellen Eide, Nabil Bitar, Hugh Secker-Walker, Noel Massey, Giulia Arman-Nassi and Reiner Wilhems-Tricarico made data analysis possible.

I appreciated the feedback from additional reviewers, including Corine Bickley, Caroline Huang, Marilyn Chen, Sharlene Liu and Suzanne Boyce, and additional listeners, including Stefanie Shattuck-Hufnagel, Joe Perkell, Sharon Manuel and Kelly Poort. Help with formatting came from Jeff Kuo, Karen Chenausky, and from Jane Wozniak for the "figure that wouldn't die". Help above and beyond administrative support came from Arlene Wint, Janice Balzer and Deborah Manning.

I'm grateful for all my friends at and away from MIT. Special recognition goes to my dear comrades Ellen Eide, Dan Ellis, Mark Johnson, Marilyn Chen and Hwa-Ping Chang, who were always there to help, even at the oddest hours. Amy Kronenberg and Sandy Fruean provided their support from a distance, which kept them close.

This task would have been impossible without the unfailing love and support of my parents Edith and Irwin Wilde. Bob and Bobbi Kross and Bubby also brought love and balance to our lives.

Heartfelt thanks go to my husband, Mark Kross, for his humor, music and improvisional talent at turning my figures into reality. Our mutual support makes our lives together a never-ending gift. And loving thanks go to our daughter Joelle, our finest collaboration, who helped to keep life's priorities straight.

All the collaborative efforts made this thesis more than just a sum of its parts. This research was partially supported by a grant from the National Institute of Health. This work remains in remembrance of Dennis Klatt and Keith North.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Audible speech is a combination of buzzes, hisses, and clicks. All sounds generated in the vocal tract involve the modulation of air flow. Better models still need to be developed for the relative timing and contribution of the sources of vocal sound: voicing, the quasi-periodic buzz generated by the vibration of the vocal folds, and the noise sources that involve random fluctuations in airflow at a constriction in the vocal tract. This thesis will focus on the relative contribution and timing of the vocal sources involved in fricative sound production, and the filtering of those sources by the vocal tract.

Fricative consonants are distinguished from other speech sounds by their manner of production. Fricatives are produced by forming a narrow constriction in some region along the length of the vocal tract. Air blown through this constriction becomes turbulent in flow, typically near an obstacle in the airstream or at the walls of the vocal tract. The acoustic result of this turbulence is the generation of noise. This noise is then filtered by the vocal tract, with the acoustic cavity in front of the constriction contributing the greatest influence on the filtering.

The eight fricatives in English, shown in Table 1.1, are distinguished by the location of the consonantal constriction: labiodentals /f, v/, dentals /θ, ð/, alveolars /s, z/, and palatals /š, ž/. The first of each pair is voiceless and the second is the voiced cognate, in which the voicing source is superimposed on the noise. One approach for characterizing these sounds is to describe them by a set of distinctive features. These

features specify the articulators, such as the tongue, lips and larynx, that are used in producing the sounds, and describe how these articulators are adjusted in forming constrictions in the vocal tract (Jakobson, Fant and Halle, 1965; Chomsky and Halle, 1968; Halle and Stevens, 1991; Stevens and Keyser, 1992).

Table 1.1: The following words and symbols illustrate the sounds of English fricatives.

Phonetic Symbol	Word Example	Arpabet Notation
/f/	fat	f
/v/	vat	v
/θ/	thesis	th
/ð/	the	dh
/s/	sip	s
/z/	zip	z
/š/	sure	sh
/ž/	azure	zh

This thesis takes a kinematic view of the acoustics of fricative consonants, rather than making the more common, stationary assumptions. During spoken communication, the vocal tract alternately opens and closes. The major features that subdivide speech sounds into classes focus on the articulatory and acoustic events that occur at different phases of this opening and closing. Vowels are produced when the vocal tract is least constricted and the vocal folds are positioned so that spontaneous voicing occurs. The following acoustic events are associated with fricatives produced in an intervocalic context:

- An interval of frication noise with a spectrum that is shaped by the location of the constriction.
- Formant transitions into adjacent vowels that provide additional place of articulation information.
- Detailed events, which occur during the transition from noise production to voicing onset, that signal the distinction between voiced and voiceless fricatives and contribute to naturalness (Klatt, Chapter 6, unpublished manuscript).

One convenient way to visualize these acoustic events is to use a spectrographic display. In Figure 1.1 spectrograms of the utterances /isi/ and /izi/ display differences that exist between acoustic realizations of a voiced and voiceless fricative.

The objective of this thesis is to study and interpret the inventory of acoustic events associated with the movement between a fricative and a vowel, and evaluate their perceptual importance. Detailed acoustic analysis begins by first identifying events or landmarks in the speech signal, such as at consonant-vowel (CV) and vowel-consonant (VC) boundaries. The strategy is to extract appropriate acoustic information in the vicinity of the landmarks. Despite a rich history of research (e.g., Hughes and Halle, 1956; Strevens, 1960; and Heinz and Stevens, 1961), the unique characterization of each fricative according to its physical quantities remains elusive.

1.1 Summary of the Problem

Shadle (1989) postulates three reasons for current limitations in our characterization of the acoustic mechanism of fricatives: 1) the theory of sound generation due to turbulence is incomplete; 2) the primary sound generation process, unlike vowels, does not include a mechanical vibration that is clearly correlated with the speech signal; and 3) an intrinsically noisy speech signal must be described statistically, rather than analytically. These limitations, which exist even for static vocal-tract configurations, become even more interesting when considered in terms of the kinematics of moving between a consonant and a vowel.

Vowels are characterized by the predomination of low frequencies and periodic voicing. Fricatives are characterized by high frequency aperiodic excitation. What glues together such acoustically different sounds? Formant transitions, which reflect the changing vocal-tract resonances, help capture the movement of the vocal tract from one configuration to another. For fricatives, not only does the shape of formant trajectory depend on place of articulation, but there is a place-dependent interaction with the vowel context. For example, Wilde and Huang (1991) demonstrated that labiodentals, which normally have the lowest formant onset frequencies, show a higher

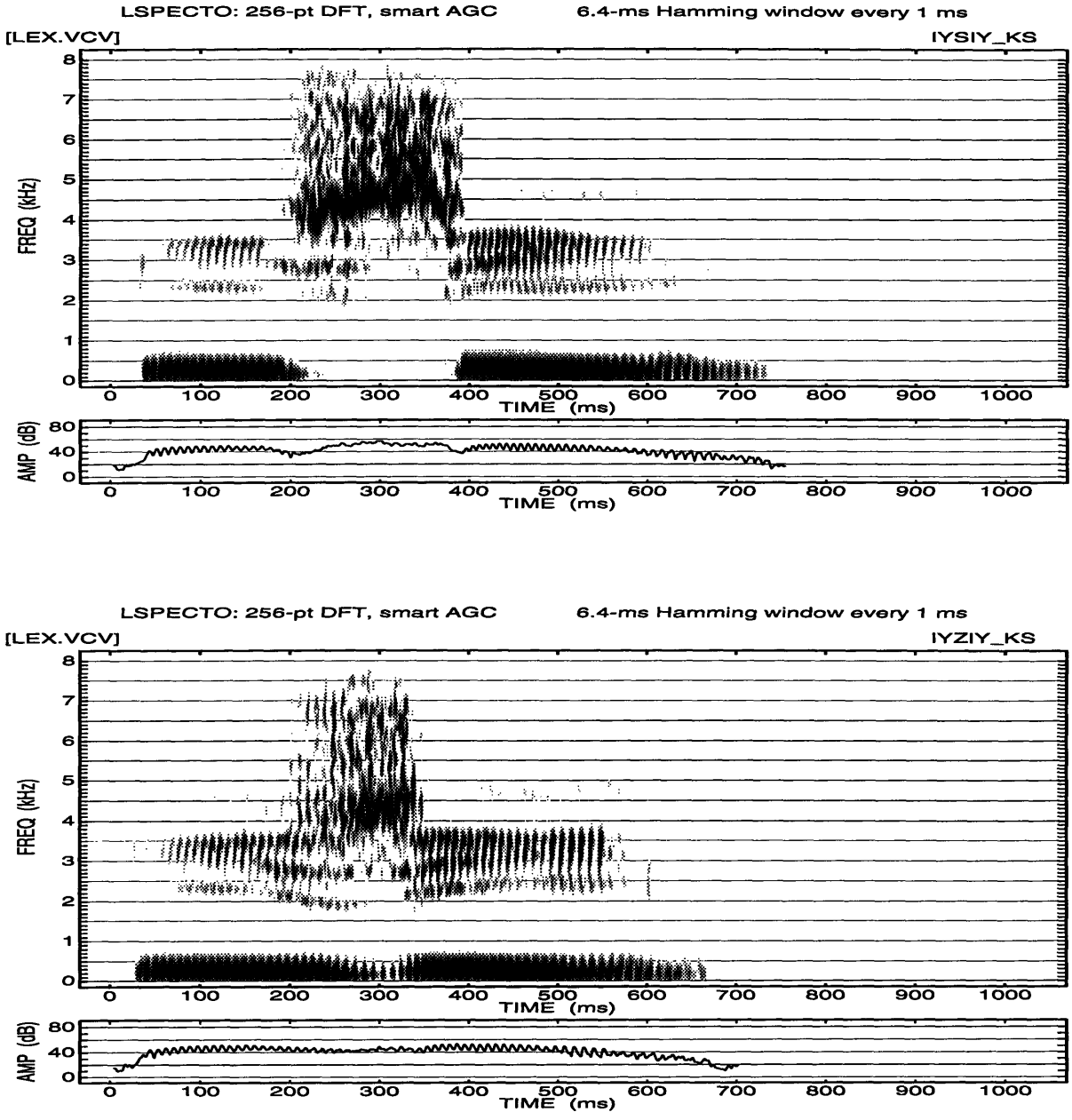

Figure 1.1: Spectrograms of the utterances /isi/ (top) and /izi/ (bottom).

second formant onset than dentals when followed by a high, front vowel. Formant transition information has been shown to be important for discriminating among fricatives when strong spectral cues are absent (Harris, 1958).

Fricatives seem to naturally cluster into two basic groups. Sibilant sounds, such as alveolars and palato-alveolars, have more acoustic energy at higher frequencies than do labiodentals and dentals. Sibilant refers to an acoustic property, "the amount of hissing noise in a sound" and is a primary way to distinguish between alveolar fricatives and dentals (Ladefoged, 1993, p. 43). Strident sounds, such as the alveolar and palatal fricatives, are also "marked acoustically by greater noisiness than their nonstrident counterparts" (Chomsky and Halle, 1968, p. 329). The feature stridency appears to be quite robust for automatic speech recognition (Deng and Sun, unpublished data). That is, under quiet conditions, there are relatively few errors that evidence confusion between segments labeled as strident and nonstrident. However, to date, the range of normal variation of stridency has not been systematically quantified and there is no accepted threshold that automatically separates sounds along the stridency dimension. While the alveolar and palatal fricatives are unequivocably classified as strident, there has been considerable debate about whether the labiodental fricatives (f, v) should be considered strident or be classified as nonstrident along with the dental fricatives (θ, ð) (McCasland, 1979).

Part of the problem of quantifying the normal variation of stridency is to clarify the definition, as that will determine which acoustic measures are appropriate. Stridency has been used to refer to both acoustic and articulatory properties. It has been suggested that the strident quality is achieved by directing a concentrated jet of air against an obstacle. Catford (1977) distinguished between a "channel turbulence" which results from flow through an articulatory channel, and a "wake turbulence" that is generated downstream from an obstacle and contributes high frequencies to the noise spectrum. Shadle (1989) also distinguished two source mechanisms for fricative sound generation: 1) an obstacle source, where sound is primarily generated at a rigid body approximately perpendicular to the airflow, and 2) a wall source, where sound is primarily generated along a rigid wall, which is parallel to the flow.

For /s, z/ and /š, ž/, the teeth are considered to be the obstacles. The upper lip may form an obstacle for /f, v/.

Models of the articulation of fricatives provide one way to study their acoustic mechanism. If we could model these sounds better, we could also improve the performance of speech synthesizers. Fricative consonants have been difficult to optimize in articulatory synthesis and analysis (Sondhi and Schroeter, 1992). The need for further study is also apparent in rule-based speech synthesis where fricatives continue to present intelligibility, as well as naturalness, problems. The results for natural and synthetic fricatives in a study comparing high-quality text-to-speech synthesis with natural speech (Wilde and Huang, 1991) confirmed the need to model better the source changes at fricative-vowel boundaries. The intelligiblity of synthetic labiodental and dental fricatives was found to be poorer than natural, even when formant transitions appeared to be reproduced accurately. The stronger, more distinctive noise spectra of alveolar and palatal fricatives /s, z, š, ž/ allows them to be synthesized more intelligibly. However, these synthetic fricatives often still sound unnatural.

1.2 Thesis Motivation

It is the speech signal that provides the acoustic link between the production and perception systems. The acoustic signal for fricative consonants has been shown to be rich in cues that distinguish these sounds. One objective of this thesis is to contribute to the quantification of acoustic cues available for the recognition of fricative consonants. These cues will be interpreted in terms of the constraints of the human production and perception systems. The motivation for this work is to evaluate the acoustic characteristics of fricative consonants in the context of the speech chain of spoken communication (Denes and Pinson, 1973).

1.3 Thesis Overview

The following is a roadmap to the remainder of this thesis. Theoretical considerations of the articulatory, aerodynamic and acoustic aspects of the production of fricatives in intervocalic position (Chapter 2) form the basis for predicting the acoustic patterns expected from the relative timing of the glottal and supraglottal constrictions. The modelling in Chapter 2 provides a foundation for interpreting the acoustic data on fricatives and for specifying strategies for synthesizing these sounds. There are two separate acoustic analysis components (Chapters 3 and 4). Chapter 3 presents a study of formant onset patterns and will serve to provide tables of values for setting relevant parameters in a formant synthesizer. In Chapter 3, we focus on the transitions of the second and third formants. That is, we are mainly looking at what is occurring behind the constriction, since it is the back cavity shape that primarily determines the second and third formant frequencies for the fricatives of English. Chapter 4 presents a study of the time-varying noise characteristics of fricatives. A system, using averaging methods, is developed to further quantify the acoustic attributes of fricatives with respect to adjacent vowels. We examine how fricative spectra change with time and we quantify the feature [strident]. Results of Chapter 4 are applied to modify the theory, where appropriate, and to set the time-varying control parameters in a speech synthesizer. In view of the different domains covered in the two acoustic chapters, further description of the relevant literature is presented separately in Chapters 3 and 4. Next, listening tests with utterances containing synthetic fricative consonants are used to determine which acoustic aspects are perceptually important (Chapter 5). In Chapter 5, we use speech-copying techniques to investigate the effects of varying the consonant-vowel ratio and mimicking observed time variations over the duration of the fricative. Finally, we conclude (Chapter 6) by relating what we learned about the acoustics of fricative consonants to the models of speech production and perception.

Chapter 2

Background: Production Modelling

During speech production, humans perform rapid vocal gymnastics. Speech production models are attempts to capture the aerodynamic and acoustic consequences of the naturally occurring movements of our articulators. The models are approximate representations which attempt to explain observed behavior in a systematic way. In modelling the production of fricative consonants, our goal is to understand how humans generate these noisy sounds and how computers can simulate the acoustics of natural fricatives. This chapter will present an acoustic theory of fricative production.

One standard way to determine the expected acoustics for a speech sound is in terms of sound sources and filter functions. In the source-filter description of the acoustic theory of speech production (Fant, 1960), the speech wave is the response of a vocal-tract filter to one or more excitation sources. The acoustic theory of speech production represents the vocal tract as an acoustic tube with varying cross-sectional area. An acoustic source can form the excitation of this tube either at the glottal end or at points along its length, and the shape of the tube determines how the source is to be filtered.

We begin with a static description of fricative production and then consider the rapidly changing geometry of the vocal tract and larynx that is associated with producing a fricative in a controlled phonetic context: vowel-fricative-vowel (VCV). We

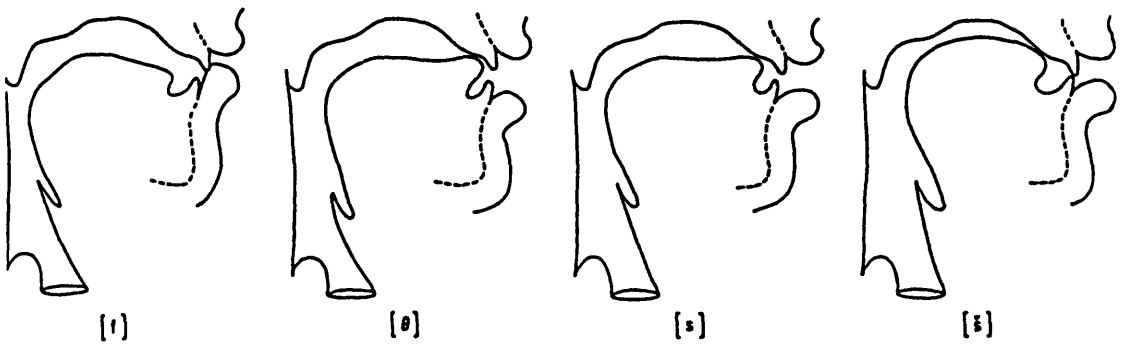

Figure 2.1: Midsagittal cross-sections of a vocal tract configured to distinguish between English fricatives according to the location of the consonantal constriction: labiodental /f/, dental /θ/, alveolar /s/ and palato-alveolar /š/. Voiced fricatives /v, ð, z, ž/ would be expected to have similar configurations to those of the voiceless cognates. (These schematics were traced by Dennis Klatt from film described in Perkell (1969).)

then review the aerodynamic conditions that arise when constrictions are made along the vocal tract. Finally, we examine existing models of the acoustic effects governed by the articulatory and aerodynamic processes of speech production.

2.1 Articulation and Aerodynamics

Fricative consonants are formed by making a narrow constriction in the vocal tract at a point above the larynx. The major articulator that forms this narrowing is usually the lips, the tongue blade or the tongue body. Midsagittal views of the vocal tract during the constricted portion are shown in Figure 2.1, and contrast the four places of articulation that distinguish fricatives in English.

At the boundary region between a fricative and vowel, the three most important

articulatory adjustments are the narrowing or widening of the vocal tract for the consonantal constriction, abduction or adduction of the vocal folds for voicing, and changes in the enclosed volume of the vocal tract (Scully et al., 1992).

Data on vocal-tract shape and dimensions are necessary for more complete understanding of the articulatory processes of speech production. However, direct evidence, especially needed three-dimensional data, is notably lacking. Baer et al.(1991) provide a review of speech production models and measurements that have been made to test existing theory in their paper on the use of magnetic resonance imaging (MRI) to study vocal tract shape and dimensions during vowel production. Much hope has been expressed for MRI to improve our models of speech production.

However, a major drawback of current MRI techniques is the time required to perform the image processing, i.e., tens of seconds to minutes, which restricts study to sustained speech sounds. Additional limitations involve the resolution difficulties, e.g., identifying the teeth, as calcified structures may be indistinguishable from the airway. Trade-offs between temporal and spatial resolution are inherent in the contemporary choice of non-invasive techniques. At present, therefore, MRI can not be used to directly examine the kinematics involved in moving from one sound to another. Fortunately, the continuant nature of fricatives makes study of sustained production possible.

A recent MRI study of fricative consonants (Narayanan et al., 1994) sustained by four speakers has provided measurements of vocal tract lengths and area functions, as well as descriptions of tongue shape. Greater inter-speaker differences were noted for nonstrident, as compared to strident, fricatives. Voiced fricatives were observed to have larger pharyngeal volumes, as compared to unvoiced, due to the advancing of the tongue root. Observed ranges of supraglottal minimum cross-sectional areas for the voiceless fricatives are shown in Table 2.1. These cross-sectional areas can be used to make estimates of airflows and pressures during fricative consonant production.

A model of fricative consonant production that permits calculation of airflows and pressures is shown in Figure 2.2. Figure 2.2(a) shows an outline derived from an x-ray of the midsagittal cross-section of a vocal tract configured for /s/. For

Table 2.1: Area ranges of supraglottal constriction for sustained fricatives produced by four speakers in MRI study by Narayanan et al. (1994).

Fricative	Cross-sectional Area (cm^2)
/f/	0.25 - 0.40
/θ/	0.15 - 0.35
/s/	0.10 - 0.30
/š/	0.10 - 0.30

/s/ the tongue tip is right behind the upper teeth, forming a constriction at the alveolar ridge. Figure 2.2(b) depicts the schematized vocal tract, according to the simplifying assumptions of a concatenated tube model with two constrictions: one at the glottis and one at the constriction formed by the supraglottal articulators. The parameters identified allow calculation of the pressure drop across the glottis ΔP_g from the subglottal pressure P_{sub} minus the intraoral pressure in the mouth, $P_m = \Delta P_c$, assuming zero atmospheric pressure (P_{atm}),

$$P_{sub} = \Delta P_c + \Delta P_g \tag{2.1}$$

The pressure drop ΔP (in dynes/cm^2 or cm H_2O) across a constriction, or resistance to airflow, can be determined from the Orifice Equation (Stevens, 1971):

$$\Delta P = \frac{k\rho U^2}{2A^2} + viscosity\ term \tag{2.2}$$

where

ρ is the density of air (.00114 g/cm^3)

U is the volume velocity in cm^3/sec

A is the area of the constriction in cm^2

k is a constant that depends on cross-sectional shape of the constriction and on the degree of discontinuity at the inlet and outlet of the constriction.

When the cross-sectional area A is greater than 0.05 cm^2 the effects of viscosity can be shown to be less than 10 % of the overall resistance to flow, and therefore can be neglected as a first approximation. When viscosity effects are negligible, as is characteristic for the subglottal pressures being considered (6-8 cm H_2O), Equation 2.2 can

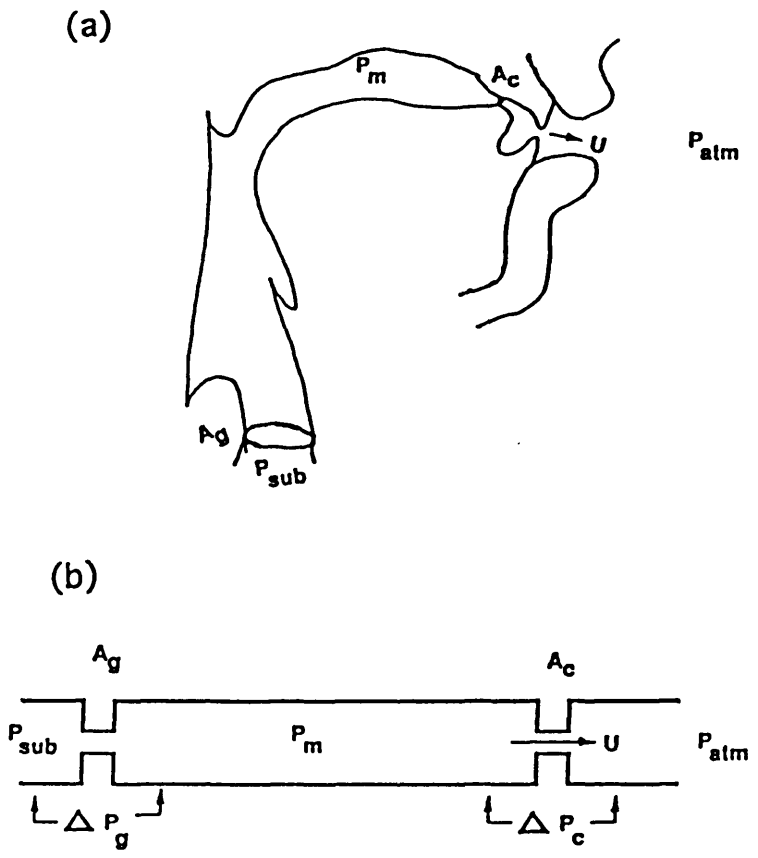

Figure 2.2: (a) Midsagittal cross-section of a vocal tract as configured for /s/. (b) This vocal tract idealized as a uniform tube with constrictions at the glottis and in the oral region.

be interpreted as a dynamic pressure drop across a nonlinear resistance.

When a fricative consonant is produced in intervocalic context, the cross-sectional area of the constriction A_c during the consonant is significantly smaller than the minimum cross-sectional area of the vocal tract in the adjacent vowels. While there is no easily obtainable method for physically measuring these areas directly in connected speech, the values of A_c and A_g can be inferred from the airflow data (Klatt et al., 1968; Scully et al., 1992). For fricatives in natural speech, the vocal tract is rarely completely constricted, and complete closure may be viewed as inadvertent overshoot. The range of the area of narrowest vocal-tract constriction for consonants is approximately 0-0.4 cm^2. The range for fricatives, as estimated from the MRI data shown in Table 2.1, may span 0.1-0.4 cm^2. However, estimates from airflow studies are generally smaller and suggest that a range of 0.05-0.3 cm^2 for the minimum constriction areas during fricatives production may be more typical. For example, data from Scully et al (1992) for /s/ are in the range of 0.05-0.2 cm^2. The average glottal opening A_g may range from about 0.1-0.4 cm^2 during voiceless fricative production, and may be somewhat smaller during voiced fricative production (Stevens, 1971). The average area of the glottis, usually about 0.03-0.05 cm^2 for modal voicing during vowel production, is generally determined from airflow and also fiberoptic studies.

It is useful to represent the concatenated-tube model of Figure 2.2 in its circuit analog, as depicted in Figure 2.3, where pressure is analogous to voltage, and volume velocity is analogous to current. This circuit was discussed by Stevens (1993) for modelling affricates, and is similar to circuits proposed earlier by Rothenberg (1968) and others for modelling the breath stream dynamics of simple stop consonants. It represents an aerodynamic model of the average pressures and flows in the vocal tract, given a subglottal pressure P_{sub}. The resistances due to glottal and superglottal constrictions are R_g and R_c, respectively, and the behavior of the vocal-tract walls at low frequencies is modelled as a resistance R_w in series with an acoustic compliance C_w. As a first approximation, C_w and R_w, which play a role in the modelling of rapidly released stop-like consonants, can be neglected when the rates of change of the parameters are slow enough. For example, when a complete closure is made in

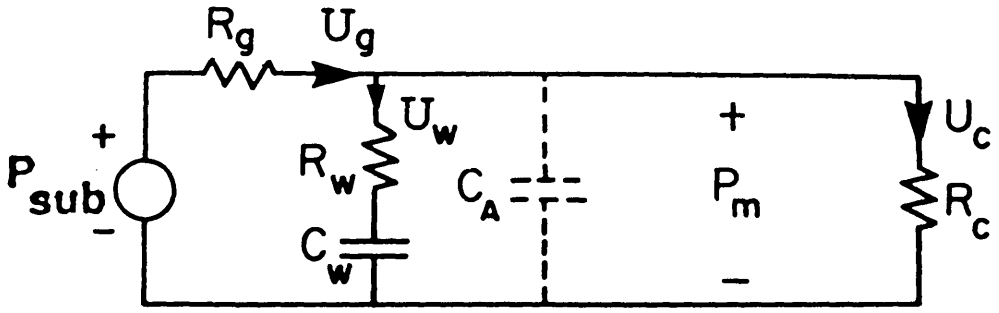

Figure 2.3: Low-frequency equivalent circuit of vocal-tract model, for estimating airflows and pressures. The component C_A shown by dashed lines is neglected in calculations of airflow. See text. (From Stevens, 1993, p. 36)

the vocal tract as is the case during production of the stop consonant /p/, there is no airflow out of the mouth; therefore, the value of R_c is infinite and U_c is zero. During the interval of complete closure for a stop consonant, when U_c is zero, the airflow U_w expands the walls of the vocal tract. In contrast, for a fricative such as /f/, the supraglottal constriction is not complete, and the flow U_c through the constriction and out of the mouth is significantly greater than the flow U_w which expands the vocal-tract walls.

A quasi-static solution can be obtained when the rates of change of the time-varying resistances, R_g and R_c, are sufficiently slow that $U_w \ll U_c$, and the capacitor C_w can be considered an open circuit. For this quasi-static solution of the circuit, the flow out of the mouth U_c equals the flow through the glottis U_g. The values of the resistances R_g and R_c have been empirically determined to be proportional to the inverse of the square of the area of the corresponding constriction. The form of the equation for these kinetic resistances

$$R_{kinetic} = \frac{k\rho U}{2A^2} \tag{2.3}$$

is related to Equation 2.2. Further, when the viscosity term is negligible and k is assumed to be the same for both constrictions, then there is the following voltage divider relationship between the voltages (or pressures) and relative resistances of the

constrictions:

$$\Delta P_c = \frac{R_c}{R_c + R_g} P_{sub} = \frac{\frac{U}{A_c^2}}{\frac{U}{A_c^2} + \frac{U}{A_g^2}} P_{sub} = \frac{1}{1 + (\frac{A_c}{A_g})^2} P_{sub} \tag{2.4}$$

Assuming a constant subglottal pressure, we can compute the build-up of pressure in the mouth $P_m = \Delta P_c$, according to the relative sizes of the time-varying adjustments of the glottal and supraglottal constrictions.

It is generally observed that the intraoral pressure P_m is higher in the voiceless case, often approaching the subglottal pressure (Klatt et al., 1968). Production of voiced fricatives requires a delicate balance between maintaining enough intraoral pressure to generate frication and maintaining enough pressure drop across the glottis to sustain vocal fold vibration. Devoicing occurs when the pressure drop across the glottis becomes less than a critical amount of approximately 2-3 cm H_2O (Titze, 1992). The data presented by Stevens et al. (1992) on devoicing of voiced fricatives are in general agreement with an earlier study by Haggard (1978).

The vocal-tract pressures and flows can be estimated from the low-frequency equivalent-circuit model of Figure 2.3. In this circuit, the resistances R_g and R_c vary in time as the areas A_g and A_c change. We used a numerical simulation of this equivalent circuit that was developed at Sensimetrics (Stevens et al., in preparation), which incorporates vocal-tract dimensions and tissue properties. The solution is found in terms of the air flow through the vocal tract, the pressure drop (ΔP) across a kinetic resistance and the cross-sectional area of the constriction, as seen from solving for U in Equation 2.2:

$$U = \sqrt{\frac{2\Delta P}{k\rho}} A \tag{2.5}$$

Schematized trajectories of the glottal and supraglottal constrictions as hypothesized for an intervocalic voiceless and voiced labiodental fricative are shown in the top left and right panels of Figure 2.4 from Stevens et al. (in preparation). The simplifying assumption for these plots is that the trajectory for the supraglottal opening A_c is the same for the voiced and voiceless cognates, while the glottal area A_g remains

wider for the voiceless fricative. The calculated flows and pressures that correspond to the time-varying area functions for intervocalic labiodental fricatives are given in the middle and bottom panels. The buildup of intraoral pressure P_m is shown to be greater for intervocalic voiceless fricatives (bottom left) as compared to voiced fricatives (bottom right).

The plots of airflow in Figure 2.4 illustrate the double peaks and large volume for intervocalic fricatives (Klatt, et al., 1968) relative to the adjacent vowels. The double peak in airflow is determined by the relative timing of the laryngeal and articulator movements. Consider the consequences of these relative movements for the calculated flows for the following example: /əfɑ/. During the /ə/, the glottis begins to open, even as the vocal folds continue to vibrate. The consequence of opening the glottal area is that airflow through the glottis A_g increases. Then when the upper teeth begin to come in contact with the lower lip, the following consequences occur: a constriction is formed resulting in a rise in the pressure in the mouth P_m causing an abrupt cessation of vocal-fold vibration. The supraglottal articulators continue to constrict during the /f/ until flow resistance reaches a local maximum and flow reaches a local minimum. Then the articulators begin to move apart in anticipation of the following /ɑ/: flow resistance lowers and the airflow through the glottis increases. As a consequence of the reduction in intraoral air pressure and the action of adducting muscles, the vocal folds begin to approximate and vocal-fold vibration begins. These adducting glottal movements increase the total flow resistance, and the flow drops to the expected value for the vowel.

A similar flow pattern is expected for intervocalic voiced fricatives. However, during voicing, there is increased laryngeal resistance with the result that airflow is reduced and the peaks are less pronounced for voiced fricatives than for their voiceless cognates. Still when pressure conditions allow vocal fold vibration to continue throughout a fricative, the flow remains higher for the voiced fricative than it would for a vowel. As U_w was found to be relatively small in these outputs, we will ignore it in future calculations.

The cross-sectional area trajectories shown in Figure 2.4 are idealized. In practice,

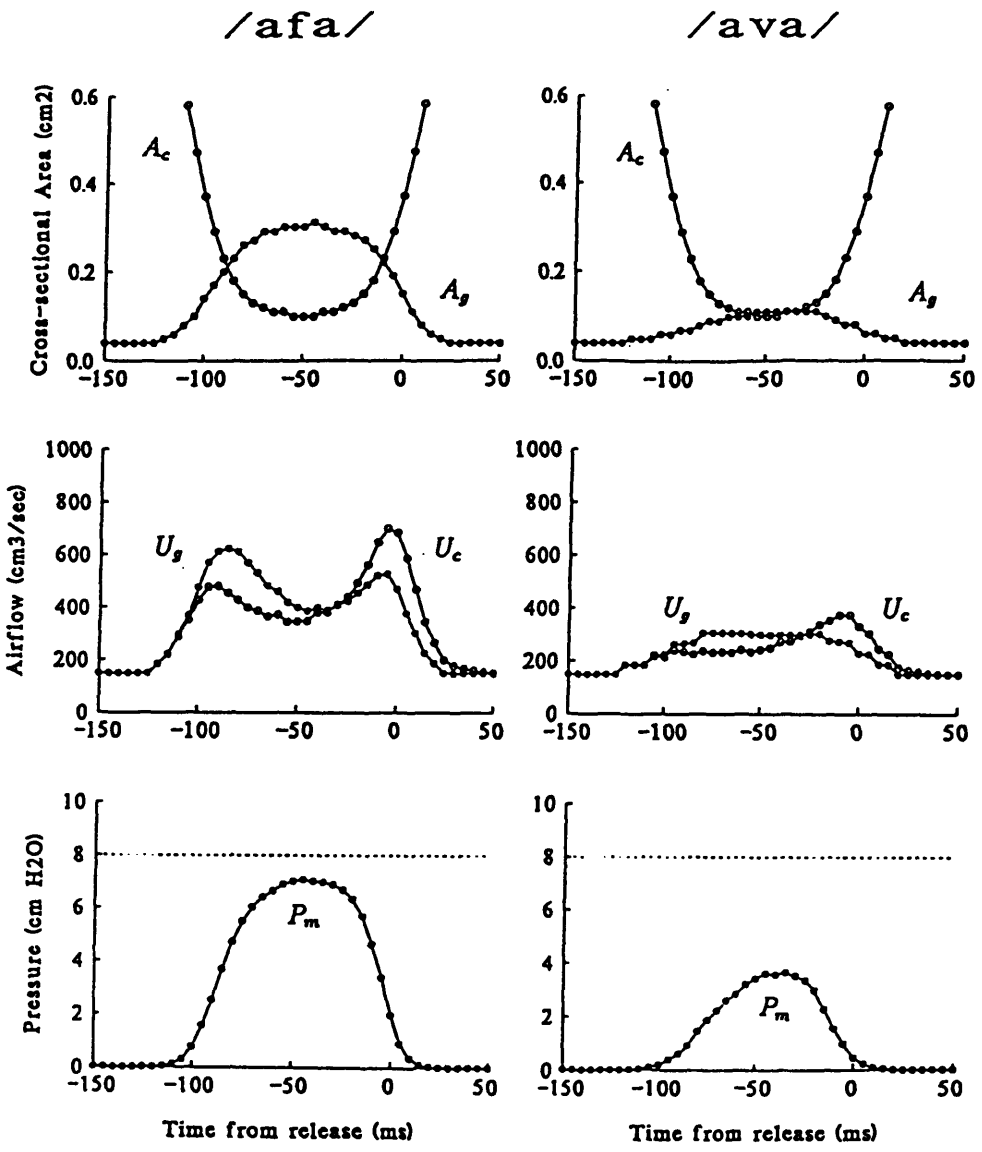

Figure 2.4: Airflows and pressures are shown with associated trajectories of the glottal (solid circles) and supraglottal (open circles) constrictions, A_g and A_c, as hypothesized for intervocalic voiceless (left) and voiced (right) labiodental fricatives. The difference between the flow through the glottis U_g (solid circles) and flow through the supraglottal constriction U_c (open circles) is the contribution of the wall flow U_w. Intraoral pressure P_m (solid circles) and a fixed subglottal pressure (dashed line) are shown in the bottom panel (From Stevens et al., in preparation).

the relative timing, or mistiming, of the glottal and oral articulator movements during VCV production can result in area trajectories that are quite asymmetrical; these asymmetries would be reflected in the resulting pressures and flows. The effects of the area trajectories on the relative amplitudes of the sound sources will be discussed in the following section.

2.2 Acoustics

There are three primary sources of sound that may be varied to produce fricatives. Two are at or near the glottis: 1) the quasi-periodic vocal fold vibration for voicing and 2) the generation of aspiration noise; and one is at the consonantal constriction: 3) the generation of frication noise. In some cases, a transient source is observed at the release of the constriction.

2.2.1 Estimating sources from aerodynamics

The turbulent noise sources are generated by random fluctuations in the velocity of air flow, superimposed on an average flow. Shadle (1985, 1990) related results of experiments with mechanical models to improving descriptions of the noise source characteristics. Pastel (1987), motivated by Shadle's work, reviewed a transmission line model of the vocal tract and explained how the airstream for aspiration or frication can be modelled as a turbulent jet. The Reynolds number, a dimensionless constant used in flow analysis, gives a measure of the turbulence of the airflow. A noise source generated by air impinging on an obstacle (i.e., zero velocity) can be modelled as a sound pressure source (called a dipole source). Obstacles often extend into the main jet stream, where velocity is higher and the source magnitude is proportional to the cross-sectional area interacting with the jet. In addition to the pressure source, there can be fluctuations in flow through the constriction resulting from irregularities further upstream. Turbulence noise used to signal a phonetic distinction in speech primarily involves turbulence generated at an obstacle or surface (Stevens, in preparation).

Estimates of the time-variations of the amplitudes of the voicing and noise sources can be made if the pressures across the constrictions are known. It has been shown that the amplitude of the source from vocal-fold vibration is proportional to $(P_m - P_{sub})^{1.5}$ (Isshiki, 1964) and the amplitude of the frication noise source is proportional to $P_m^{1.5} A_c^{0.5}$ (Stevens, 1971; Shadle, 1985). This relation for frication noise has recently been confirmed for one subject in a study by Badin et al. (1994) using enhanced Electropalatography to measure constriction size and shape during production of sustained fricatives and whispered vowels. The equation for amplitude of the aspiration noise is assumed to have the same form as frication noise. However, as discussed above, the overall amplitude of the noise source may depend on the presence of obstacles in the airstream and the effect of obstacles may be different for the two kinds of turbulent noise sources.

Asymmetries in the noise amplitude of fricatives have been observed in our acoustic analysis (and by Behrens and Blumstein, 1988 and Shadle et al., 1992). One factor in observed asymmetries for intervocalic fricatives is the relative timing of the glottal and supraglottal trajectories, and consequently, how much aspiration noise is superimposed on the frication noise. Two hypothetical time courses relating the area of the glottal opening A_g and the area of the supraglottal constriction A_c are shown in the top panels of Figure 2.5 (Stevens, in preparation). Asymmetries in the breathiness observed in the vowels at vowel-fricative and fricative-vowel boundaries have suggested that the schematized area trajectories shown in the top left panel of Figure 2.5 can be modified to show a more abrupt increase in A_g at the vowel offset and more gradual decrease in A_g at the vowel onset, as shown in the right panel. These movements are slow enough to warrant consideration of the quasi-static solution to the circuit shown in Figure 2.3. The calculated amplitudes of the noise sources N_g and N_c in the vicinity of the glottal and supraglottal constrictions, respectively, are shown in the bottom panels. The relative contributions of the noise source are shown to change as the relative movement of the articulators is varied. One characteristic result is that the frication noise source amplitude remains relatively constant during the consonantal interval and another is that another is that aspiration noise persists

for a longer time interval at the fricative-vowel boundary.

2.2.2 Source spectra

As discussed above, one way to examine the characteristics of a turbulent sound source is to model it mechanically. Stevens (1971) reviews studies of the sound source generation by turbulent flow at an obstruction, called a spoiler, in a pipe. The frequency distribution f from the fluctuation in the force is centered on a frequency that is proportional to flow velocity V divided by the cross-dimension d of the spoiler. Shadle (1985) used a mechanical model to derive source functions for turbulence, including far-field sound measurements produced when airflow through a constriction hits an obstacle. The measured spectrum of the sound-pressure source that results from turbulence at an obstacle tends to have a broad peak centered at a frequency that is proportional to V/d, and then falls off slowly above this frequency. The relative amplitudes of measured spectra associated with two different flow velocities are shown in Figure 2.6 for sound-pressure sources used to model the obstacle case. In these spectra the frequency of the broad peak is shown to increase as the rate of flow increases. During fricative production, typical ranges for rates of flow are 300-600 cm^3/sec.

2.2.3 Vocal tract filtering of sources

The spectrum of the sound radiated from the lips during fricative production can be considered to be the product of three spectra: 1) the spectrum of the noise source, 2) the transfer function from the sound pressure source to the volume velocity at the lips, which is determined by the vocal tract shape, and 3) the radiation characteristic at the lips. Therefore, another way to study the noise source characteristics of fricatives is to start with natural speech and account for the filtering effects of the vocal tract. In this approach, inverse filtering, the noise source is modelled as a series pressure source located at a single point in the vocal tract with characteristics that are independent of the vocal tract shape (Badin, 1991).

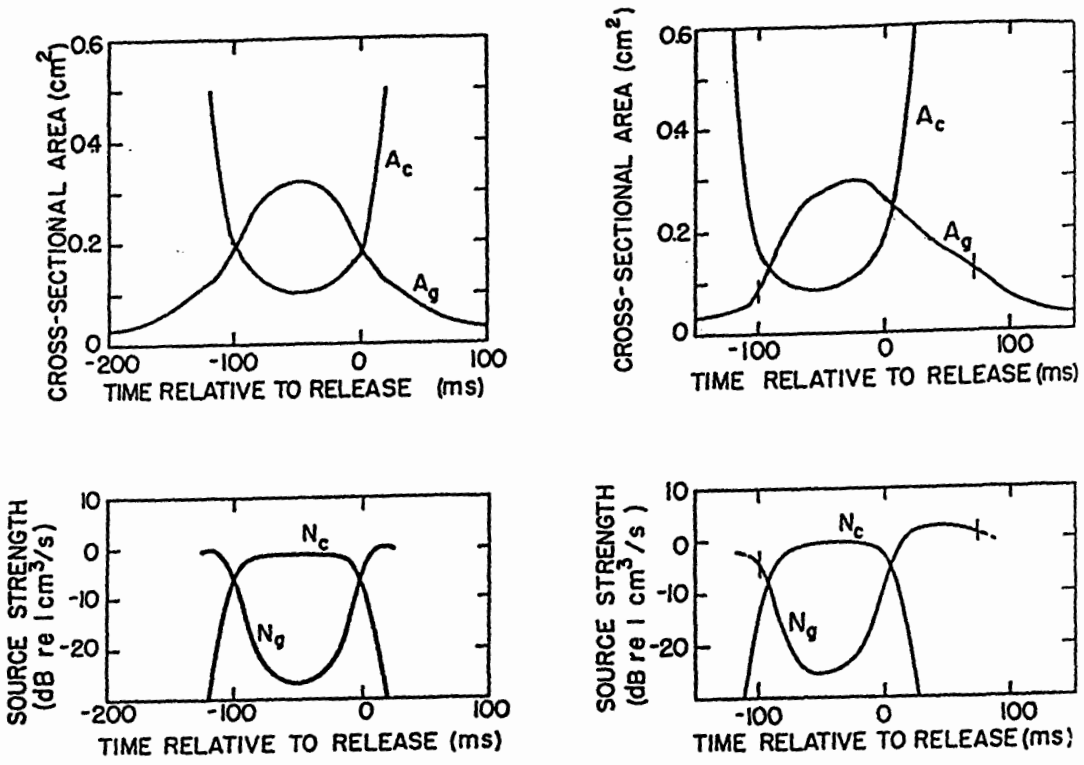

Figure 2.5: (Top) Schematized trajectories of the glottal and supraglottal constrictions (A_g and A_c) as hypothesized for an intervocalic fricative, during which the movement of A_c in the right panel is advanced 50 ms with respect to the movement of A_g as shown in the left panel. (Bottom) The amplitudes of the noise sources N_g and N_c near the glottal and supraglottal constrictions, respectively, are calculated assuming quasi-static conditions. The subglottal pressure is assumed to be 8 cm H_2O. The vertical marks in the curves A_g and N_g show estimated times of offset and onset of vibration of the vocal folds (From Stevens, in preparation).

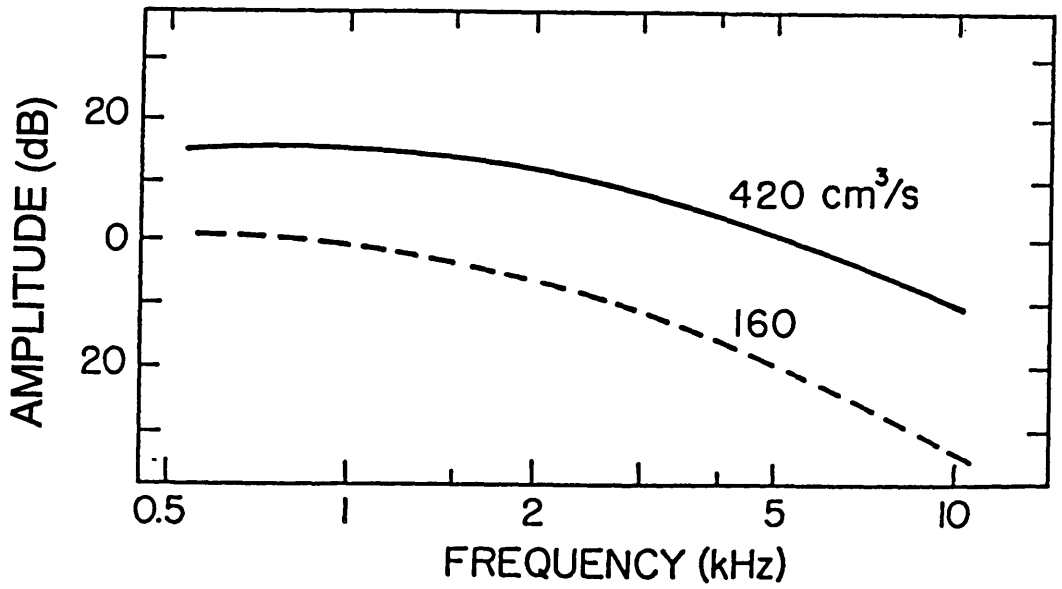

Figure 2.6: Measured spectra of sound pressure source for two different flow velocities. From Stevens (in preparation).

The filtering function of the vocal tract can be considered in terms of its pole-zero decomposition. A pole frequency is a frequency for which a small excitation at the input produces an infinite response at the output. That is, the poles are the natural resonances of the entire system under consideration. A zero frequency is a frequency for which even a large input to the system produces zero output.

A series of models for the voiceless alveolar fricative /s/, as shown in Figure 2.7, will be used to introduce the effect of the vocal-tract filtering function. When the constriction A_c is very narrow compared to the cross-sectional area of the concatenated tubes shown at the top left, it is assumed that the coupling between the cavity behind the constriction and the cavity in front of the constriction is negligible. That is, the constriction can be replaced by a rigid wall, as shown on the top right. For this approximation, the only resonances that appear in the output spectrum are those due to the cavity in front of the constriction. This model of decoupling the front and back cavities is one way to represent the absence of back cavity resonances in the output.

Another description is that the zeroes of the back cavity cancel the corresponding system poles. In the following calculations, we will also assume that there is a single pressure source at a fixed location, rather than a spatially distributed source, and we will neglect losses. We can then specify the transfer function T_n of the system shown at the bottom of Figure 2.7:

$$T_n = \frac{U_m}{P_s} = \frac{U_1}{P_s} \times \frac{U_m}{U_1} \ . \tag{2.6}$$

We can use the solution to the one-dimensional wave equation (e.g., as derived in Beranek (1954)) in order to solve the right-hand side of Equation 2.6. The first factor can be shown in terms of the impedances Z_1 and Z_2 seen to the right and the left of the pressure source:

$$\frac{U_1}{P_s} = \frac{1}{Z_1 + Z_2} \tag{2.7}$$

where

$$\begin{aligned} Z_1 &= j\frac{\rho c}{A}\tan kl_1 \ , \\ Z_2 &= -j\frac{\rho c}{A}\cot kl_2 \end{aligned}$$

and

$$k = \frac{\omega}{c} = \frac{2\pi f}{c} \ .$$

The second factor is given by

$$\frac{U_m}{U_1} = \frac{1}{\cos kl_1} \ . \tag{2.8}$$

Equation 2.6 reduces to

$$T_n = j\frac{A}{\rho c}\frac{\sin kl_2}{\cos kl} \tag{2.9}$$

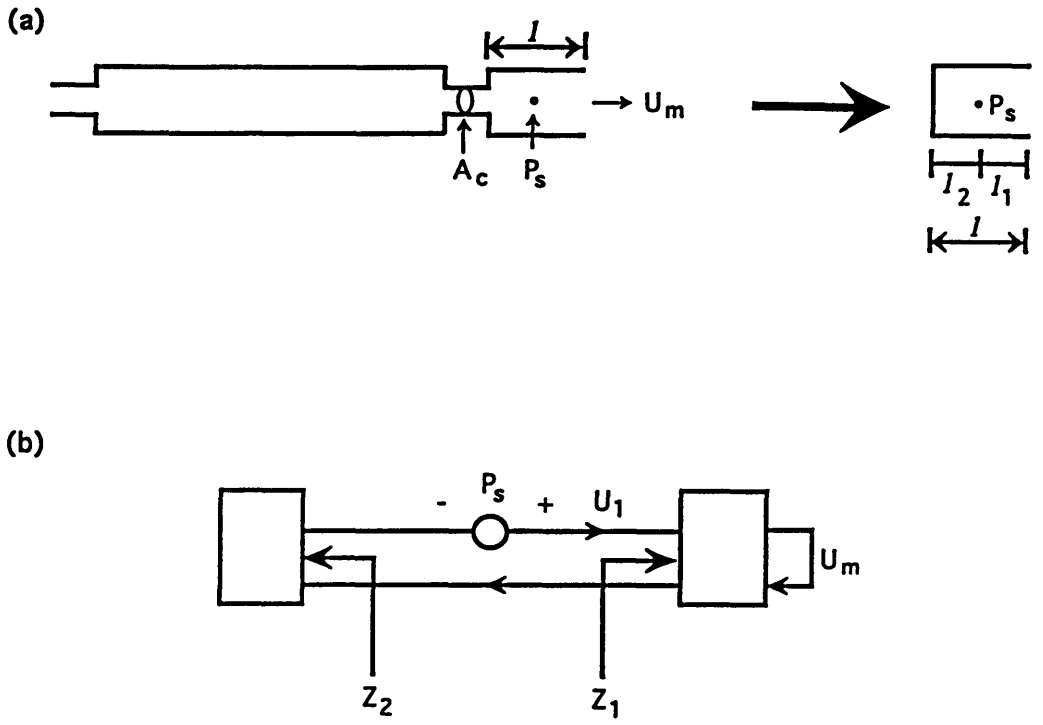

Figure 2.7: Models of the vocal-tract for an alveolar fricative, in which a pressure source P_s is located at a fixed position in front of the constriction. (a) When A_c is sufficiently small, the constriction is modelled as a rigid wall. (b) The impedances Z_1 and Z_2 seen to the right and the left of the pressure source are shown.

where

$$l = l_1 + l_2 \ .$$

To find the pole frequencies, f_{pn}, the denominator of T_n is set to zero. Therefore, the poles of the system occur when

$$\cos kl = 0$$

which yields

$$f_{pn} = \frac{c}{4l}, \frac{3c}{4l}, \frac{5c}{4l}, ... \tag{2.10}$$

To find the zero frequencies, f_{zn}, the numerator of T_n is set to zero. Therefore, the zeroes of the system occur when

$$\sin kl_2 = 0$$

which yields

$$f_{zn} = 0, \frac{c}{2l_2}, \frac{c}{l_2}, ... \tag{2.11}$$

In summary, the system shown in Figure 2.7 can be decomposed into a quarter-wave resonator representing the cavity in front of the constriction, which contributes the system poles, and a half-wave resonator representing the portion of the vocal tract between the constriction and the pressure source, which contributes the zeroes. That is, the zeroes depend on the location of the source.

For example, we can calculate the poles and zeroes of the following configuration: a front cavity of length 2 cm with the pressure source located 1 cm in front of the constriction. The first pole of the system in front of the constriction is the first resonance of the quarter-wave resonator: 4425 Hz ($c/4l$ where c, the speech of sound, is 35,400 cm/sec and length l is 2 cm). The zeroes are contributed by the half-

wave resonator: the zero frequencies for the whole system are the same as the pole frequencies for the half-wave resonator in this simulation. There is a zero at 0 Hz, and one at 17,700 Hz ($c/2l_2$, where l_2= 1 cm).

In order to determine the contribution of the transfer function T_n of this vocal-tract configuration, we derive the magnitude of the transfer function at the formant frequency given by the above pole at 4425 Hz. The peak-to-valley ratio (in dB) of T_n is given by

$$\frac{2S}{\pi B} \tag{2.12}$$

where

S is the spacing between the formants in Hz

B is the formant bandwidth in Hz.

The bandwidths at the frequency ranges we are considering for the frication noise are mostly due to radiation losses. Given a spacing of 8850 Hz (i.e, twice the lowest natural frequency) and bandwidths in the range of 500-1000 Hz, we get a peak-to-valley ratio due to the system poles of 15-21 dB. In addition, the combined effect of the first two system zeroes, as discussed above, is to bring down the spectrum about 3 dB at a frequency of 4425 Hz, and to further reduce the spectrum amplitude at lower frequencies.

The calculated spectrum of the fricative output is shown as the curve labeled "FRIC." in Figure 2.8. The frication spectra is shown for an alveolar consonant, such as /s/. There is a high frequency peak in the frication spectrum which reflects the filtering of the pressure source by the short cavity in front of the constriction at the alveolar ridge, i.e. a tube with a first resonance of approximately 4500 Hz. As discussed above, there may also be a contribution from a volume velocity source, but this small, predominantly low-frequency effect is neglected in our model of production. The radiation resistance increases with frequency; i.e., there are increased losses as frequency increases. Therefore, the effect on the overall spectrum of speech is to widen the bandwidths and decrease the amplitude at higher frequencies, as can be seen in the formant structure for curve labeled "VOWEL". The vowel curve is shown for a

neutral vowel, in which formants are regularly spaced in frequency. Recall that the noise source for aspiration is at the glottis, i.e., at one end of the tube model of the vocal tract. Therefore, resonance peaks in the aspiration spectrum correspond to the formants in the adjacent vowel. However, because of the losses at the glottis which are maximum at low frequencies, the lower formants, especially the first formant is considerably damped, and may not be visible in the aspiration output.

Our research will focus on the acoustic consequences of fricative production. Our approach to modelling fricatives will be to be to set the time-varying controls in a speech synthesizer. When we synthesize speech sounds with computers, we are modelling production. With an analysis-by-synthesis approach (e.g., Scully et al., 1992), model parameters are set to estimated values based on acoustic analysis of naturally-produced fricatives, and then are iteratively adjusted to improved the match. In rule-based synthesis, we seek a set of simple rules, i.e., a model, to gain better understanding of how natural and intelligible fricatives are generated. We will consider the synthesis of intervocalic fricatives (Chapter 5) after a discussion on the acoustic analysis of natural fricatives (Chapters 3 and 4). The modelling in this chapter provides a basis for interpreting the acoustic data on fricatives and for synthesizing fricatives.

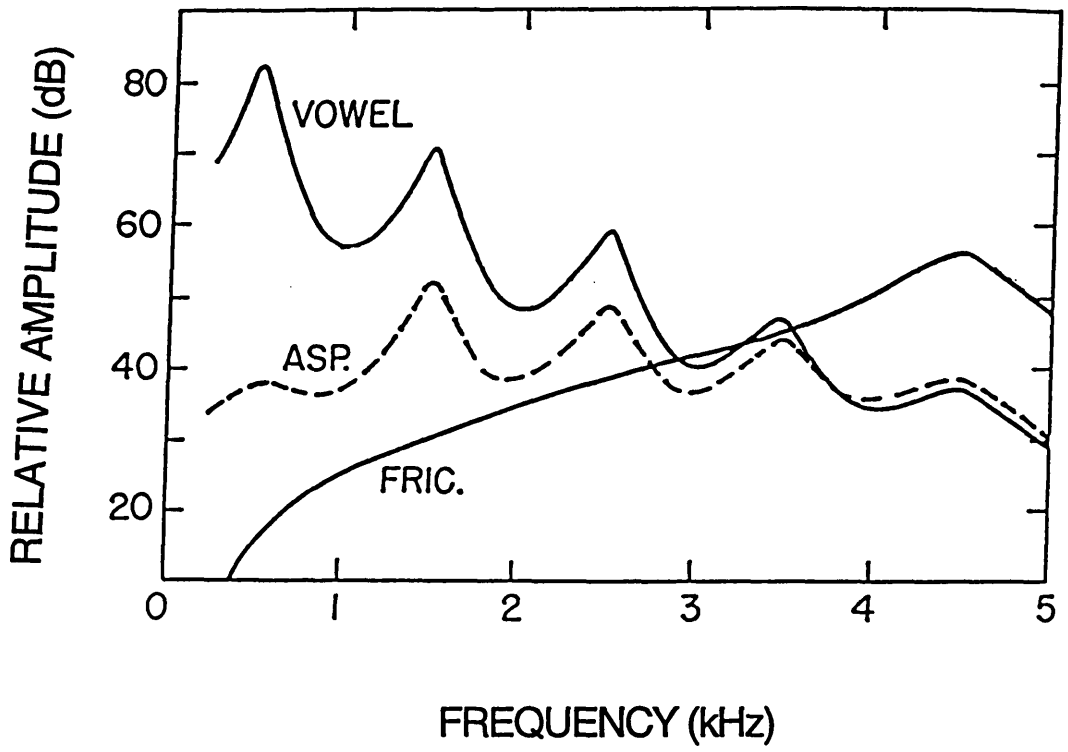

Figure 2.8: Calculated frication noise spectrum (FRIC) from a pressure source as might be seen for an alveolar fricative consonant, such as /s/ compared to the spectra for a neutral vowel (VOWEL) and aspiration (ASP). Spectra represent relative sound-pressure levels that might be measured at a distance in front of the mouth opening. Calculations are based on source models and assumptions about vocal-tract configurations for an adult male speaker. From Stevens (in preparation).

Chapter 3

Formant Patterns

The outcome of perceptual experiments on fricative consonants (Harris, 1958) showed that formant transition information is important for discriminating among fricatives when strong spectral cues are absent. Previous studies have focused on examining the acoustic correlates of place of articulation in formant transitions in natural speech (Kewley-Port, 1982) and synthetic speech signals (dating from classical Haskins studies in the 1950's documenting the role of the second-formant transition in signaling stop place of articulation).

The early work describing the articulator position from acoustic properties available at consonant-vowel (CV) syllables was in response to the locus theory advanced by Delattre et al. (1955), which proposes the presence of a place on the frequency scale to which a formant transition is moving. More recently, slope and y-intercept parameters of locus equations have been suggested for classifying stop place categories (Sussman et al., 1991; Sussman et al., 1993). Fowler (1994) argues that it is coarticulation resistance, i.e., the degree of coarticulatory overlap between consonant and vowel, that is reflected by study of the relationship between formant onset frequencies at a CV boundary and the formant values in the steady portion of the vowel.

In this chapter, it is argued that formant transitions at the release from a fricative into a vowel allow inferring the positions of the major articulators. It is shown that variability in the acoustic signal at the release of a fricative decreases when production of the fricative places greater constraints on the position of the tongue body.

3.1 Speech Corpus, Recording Procedure and Equipment

A database was collected in order to examine in detail the acoustic attributes of fricative consonants in the front, back and back-rounded vowel contexts. Three normal speakers of American English, one male and two female, recorded 'CVCV'CVC nonsense syllables. The consonant was one of the eight English fricatives: /f, v, θ, ð, s, z, š, ž/ and the stressed vowels were /i, ɛ, ɑ, ʌ, o, u/. The first and third vowels in an utterance were the same. The second vowel was always the reduced vowel /ə/. Two repetitions of each fricative in pre-stressed position were studied.

The speech was recorded in a sound-treated room using an Altec ominidirectional microphone (Model 684A), which was amplified by a Shure Microphone Preamplifier (Model M67) feeding a Nakamichi tape recorder (Model LX-5). The microphone was located approximately 25 cm in front of the speaker at approximately 5 cm above the speaker's mouth. This placement was chosen so that the airstream was not directed against the microphone diaphragm and also to avoid problems of picking up incorrect relative amplitudes of radiation from the lips and neck, if too close, or reflection of low frequencies from the walls of the room, if too far away.

The recordings were digitized at 16 kHz after being passed through an anti-aliasing filter which had a cut-off frequency of 7.5 kHz. This cut-off frequency was the highest available at the Speech Communication Laboratory at the time of this study.

One additional male speaker, previously low-pass filtered at 4.8 kHz and digitized at 10 kHz by Klatt (unpublished manuscript), was also studied. The combined database of four speakers allowed examination of intra-speaker variability for fricatives in intervocalic position and was adequate for setting synthesis parameters for stimuli used in the subsequent perceptual experiments.

3.2 Measurement of Formant Onset Frequencies

Formant frequencies were measured for the eight English fricatives preceding the six vowels. (In the later analysis, the data for the different tokens are grouped according to whether the vowels are front /i ɛ/, back unrounded /ɑ ʌ/ or back rounded /o u/.) Measurements were made at an identified landmark between the fricative and the vowel, during the vowel and, when possible, in the consonantal interval. Discrete Fourier transforms were computed with a 6.4 ms Hamming window. The window was carefully placed in order to maximize inclusion of the closed portion of the glottal cycle; this placement avoids the widening of formant bandwidths, which may be associated with introducing acoustic losses at an open glottis, and enhances identification of spectral peaks.

Formant onset frequencies are identified as **F1onset**, **F2onset**, and **F3onset**. Formant onset times at the CV boundary designate the point when the amplitude of the first formant increases most rapidly. This energy increase could often be detected from the time waveform, especially at the landmark between voiceless fricatives and the following vowels. The first formant amplitude at each pitch period in the vicinity of the boundary was examined with a short (6.4 msec) window to confirm the landmark, and to resolve ambiguity, as needed, especially for cases with voiced fricatives. In addition, wide-band spectrograms were used to further confirm the location of the CV boundary. The formant onset frequencies were measured at the first pitch pulse after the CV boundary.

Midvowel formant frequencies are identified as **F1vowel**, **F2vowel**, and **F3vowel**. Formant frequencies were measured and recorded at the first pitch pulse 70 ms after the CV boundary. In Figure 3.1 example spectra illustrate the reported formant measures for the utterance /ði/ spoken by one of the male speakers. For this particular utterance, formants can be easily tracked throughout the entire CV. However, the difficulty of assigning a single point as the CV boundary between voiced fricatives and vowels is also evident.

Formant onset frequencies and values in the middle of the following vowel for

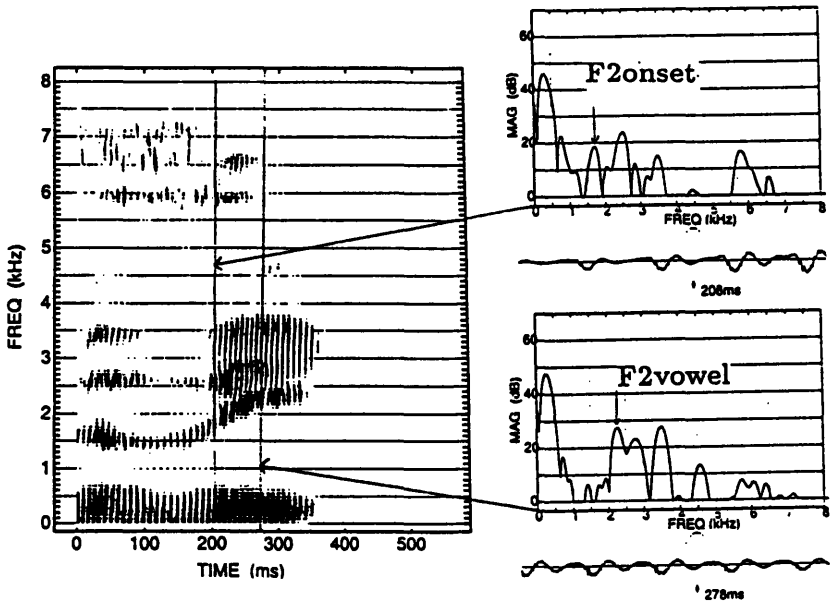

Figure 3.1: Examples of spectra for /ði/ spoken by a male speaker illustrate formant frequency measures at formant onset and in the vowel 70 ms later. The relative timing of these measurements can be visualized on the corresponding spectrogram.

all four speakers are presented in the appendices (Female speakers, F1 and F2, in Appendix A and male speakers, M1 and M2, in Appendix B). Formant onset data were first analyzed separately for each speaker. In the initial analysis, averaged values from the two instances of each pre-stressed CV token excised from each 'CVCV'CVC utterance were used. Results of this initial analysis were used to suggest further combination of tokens for subsequent analysis.

3.3 Results

Fricative data were compared with measurements reported from investigations of place categorization of stop consonants. Locus equation scatterplots (**F2onset** vs. **F2vowel**) were generated, as reported in Sussman et al. (1991) and Sussman et al. (1993). In addition, **F2onset** and **F3onset** values were combined in a two-dimensional (F2xF3) space as reported by Kewley-Port (1982). The range of **F2onset**, as discussed below, was one measure that was found to show solid group differences according to vowel context, place of articulation and fricative-vowel interaction.

3.3.1 Effects of Vowel Context

The constraints placed on consonant onset frequencies are different for different vowels. This can be seen from Table 3.1, which compares the ranges of **F2onset** frequencies. **F2onset** range is calculated separately for each subject by subtracting the lowest obtained frequency value of **F2onset** from the highest frequency value for a particular context, in this case for each vowel context. These minimum and maximum values are taken from the averages of measures from tokens containing voiced and voiceless cognates (i.e., the average of four tokens, two voiced and two voiceless) to provide one measure for each distinct place of articulation. Averaging the entire fricative data set allows examination of the variability in **F2onset** values according to vowel context. Initial examination of the data suggested that the formant measures for the non-high back vowels, /ɑ/, /ʌ/, and /o/, could be combined.

Results in Table 3.1 show the same trend for each speaker according to vowel

context. The different fricative consonants have similar starting frequency when preceding /i/. In contrast, there is a wide range of onsets when preceding /u/, and ranges within these two extremes are found for /ɛ/ and the non-high back vowels. This progression is evident in the averages of **F2onset** ranges across all four speakers as a function of vowel context. For stops, Kewley-Port (1982) also found that there is a large separation for /b/ and /d/ in the back vowel context, as compared to their similarity in the front vowel context.

Table 3.1: F2 ranges obtained for each speaker, averaged over data obtained for the same place of articulation, are compared to examine the effect of the following vowel context. F2 range is calculated by subtracting the minimum value of **F2onset** from the maximum value of **F2onset**, as described in the text.

Range of F2 Onset Values (Hz): Effect of Vowel Context				
Speaker	/i/	/ɛ/	non-high back vowels	/u/
F1	184	376	526	891
F2	166	276	687	693
M1	217	388	580	606
M2	419	464	527	796
Group	247	376	580	747

A limitation in analyzing the range in this simple manner is that the effects of outliers cannot be determined. Therefore, additional analysis of the interquartile range (IQR), a simple and robust measure of data dispersion, was undertaken. The IQR is defined to be the 75th percentile (Q3) minus the twenty-fifth percentile (Q1). That is, the IQR is defined as Q3-Q1, where Q2 is the median **F2onset** value. Therefore, the IQR includes the middle 50 % of the data, as opposed to the previous range analysis which looks at the difference between two data extremes. The medians and interquartile ranges of **F2onset** values for each vowel context for the individual speakers and the group data are shown in the boxplots in Figure 3.2. The horizontal line in the interior of the box is located at the median of the data. The height of the box is the IQR. The horizontal lines are the whiskers; for data described by a Gaussian distribution, approximately 99.3 % of the data falls within the whiskers (S-PLUS, 1991). The IQR results confirms the previous trend for each condition for three out of four

of the individual speakers and the group IQR for /u/ (412 Hz) is twice as large as the group mean IQR for /i/ (221 Hz).

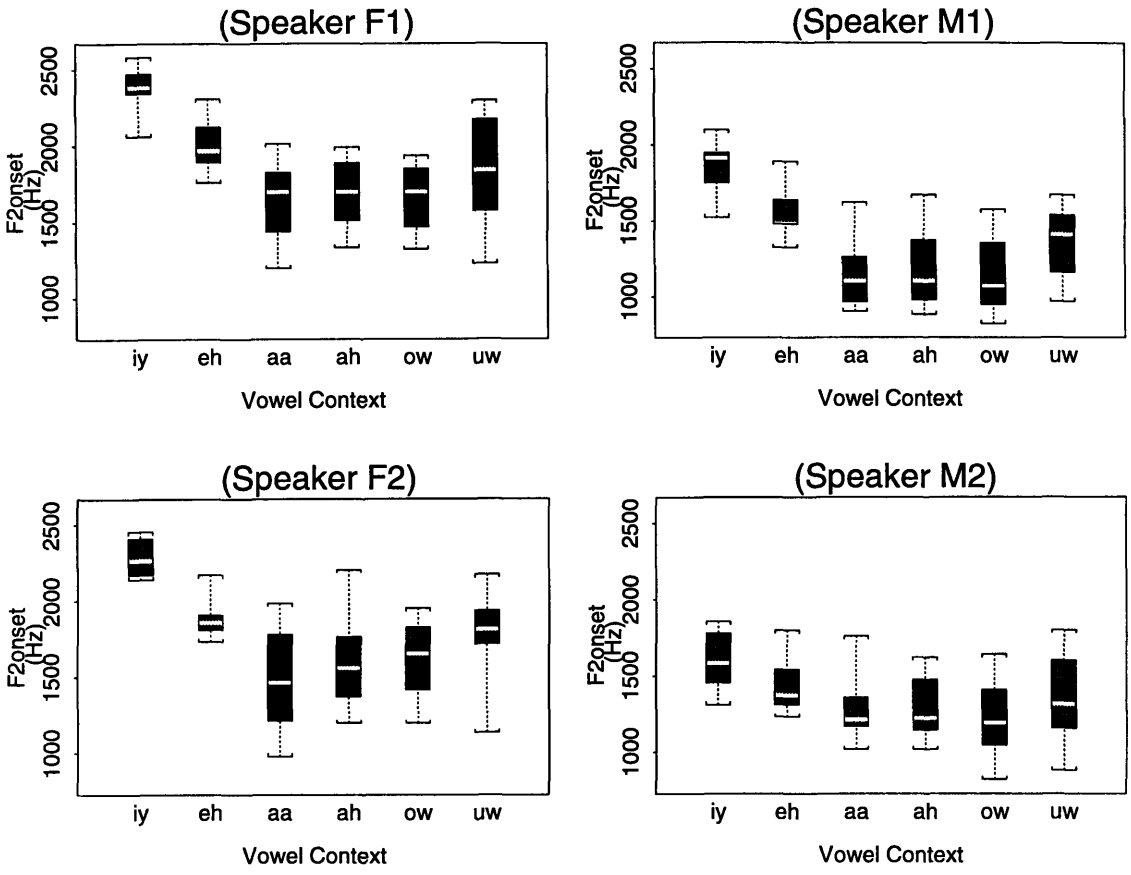

Figure 3.2: The medians (white lines) and interquartiles ranges (IQR=box height) illustrating the effect of vowel context on **F2onset** are shown for each speaker. Each box corresponds to a different vowel and represents 16 data points (2 repetitions of all 8 fricatives). The whiskers (the dotted lines extending from the top and bottom of the box) extend to the extreme values of the data or a distance 1.5xIQR from the center, whichever is less.

3.3.2 Effects of Place of Articulation

For fricatives, in general, **F2onset** values are lowest for labiodentals and are successively higher for dentals, alveolars, and palato-alveolars. This result is observed particularly for back vowels for which the F2 range is greatest and is consistent with the second resonance being affiliated with the cavity behind the primary constriction.

There is variability of consonant locus when averaging over the vowel contexts. Figure 3.3 plots **F2onset** vs. **F2vowel**, where each line represents a separate locus equation of the form:

$$F2onset = m * F2vowel + b \tag{3.1}$$

where m and b represent the slope and y intercept, respectively.
Calculations are from straight-line regression fits to data points for **F2onset** plotted against corresponding **F2vowel**. The locus equation slope shows the extent to which consonant locus changes with the following vowel. A slope equal to zero means that **F2onset** loci are constant across all vowel contexts (i.e., no accommodation). A slope of one means that **F2onset** is equal to **F2vowel** or is displaced by a fixed amount from **F2vowel** (i.e., total accommodation).

Table 3.2 shows the slopes for each of the fricative places of articulation and also the slopes for stops (Sussman et al., 1991). The y intercept data, which cannot be interpreted with respect to an articulatory correlate, are not discussed here.

Table 3.2: Slopes for the locus equations fit to the fricative data, which include both voiced and voiceless cognates, are compared with values for labial and alveolar voiced stop consonants reported in Sussman et al. (1991). Calculations are from straight-line regression fits to data points for **F2onset** plotted against corresponding **F2vowel** for data from female and male speakers. Each point represents data from one CV utterance.

	Labiodental f v	Dental θ ð	Alveolar s z	Palato-alveolar š ž	Labial b	Alveolar d
Female	0.72	0.45	0.43	0.30	0.90	0.40
Male	0.77	0.53	0.52	0.42	0.87	0.43

These results show that slope of the locus equation is systematically different according to place of the supraglottal constriction. Sounds made at the lips show the greatest accommodation to vowels. The palato-alveolars show the least accommodation, and alveolars and dentals are intermediate. Stops (Sussman et al., 1991) and fricatives show similar trends, with slope decreasing as the point of constriction is

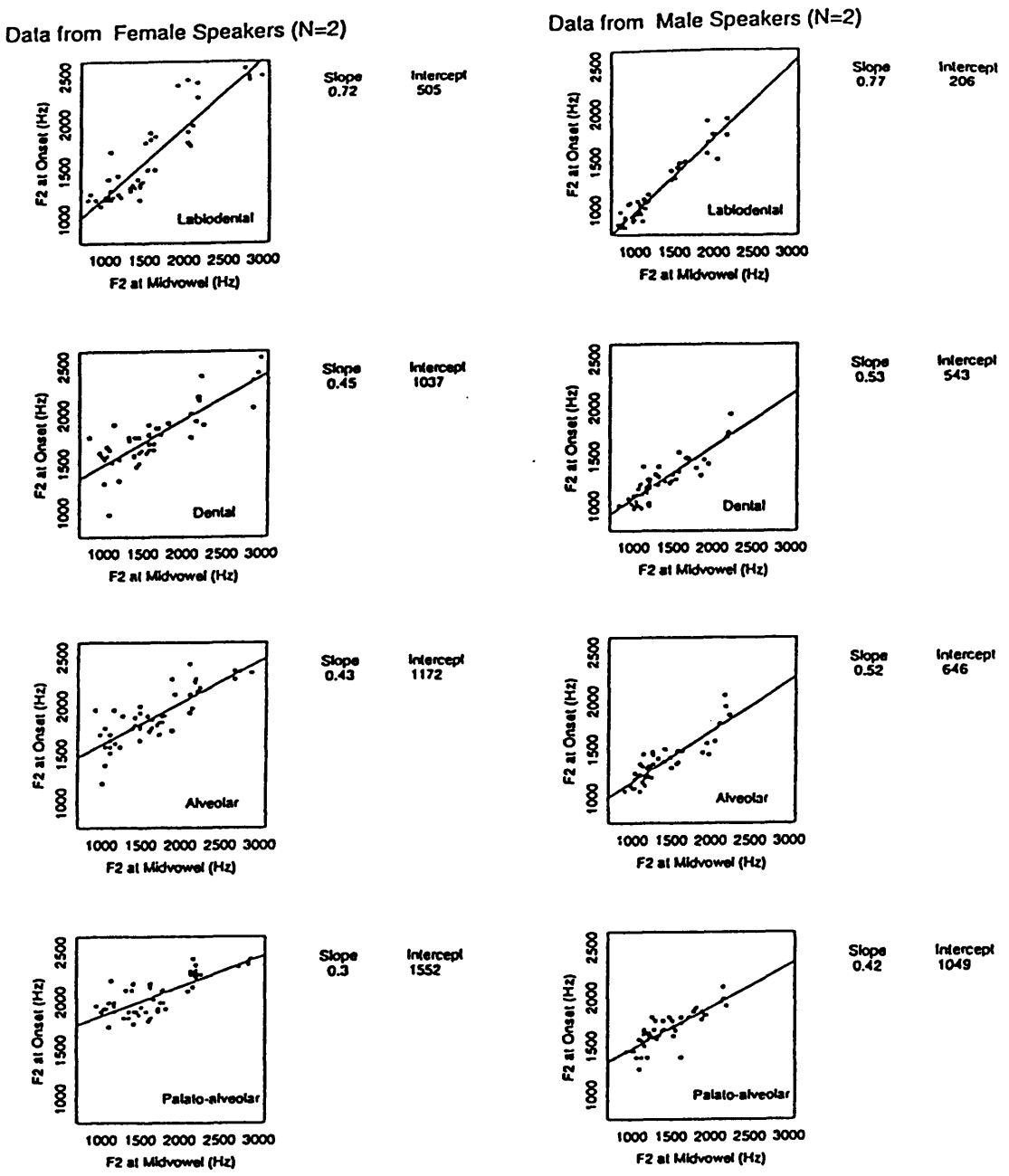

Figure 3.3: Straight-line regressions are fit to data points for **F2onset** plotted against corresponding **F2vowel**. These locus equations are calculated and plotted for data, from female and male speakers, that were shown in Table 3.2.

moved further back in the vocal tract. It is interesting to note that the slopes for the labial stops are even greater than those for the labiodental fricatives.

F2onset range is another way of illustrating the relationship captured by the slope of the locus equations. Table 3.3 compares the ranges of **F2onset**, as averaged over vowel context, to examine the differences in variability of F2 onset values according to place of articulation. The method for calculating ranges is the same as that described previously for different vowels, except that the ranges here are for different consonants.

Table 3.3: F2 range, averaged over data obtained for the same vowel context, is compared to examine the effect of place of articulation. The same F2 range calculations described previously are used here.

Range of F2 Onset Values (Hz): Effect of Place of Articulation				
	Labiodental	Dental	Alveolar	Palato-alveolar
F1	1201	757	519	491
F2	1150	719	686	297
M1	880	738	718	501
M2	727	205	269	114
Group	990	605	548	351

F2onset range was found to be systematically different according to place of articulation. There is a progressively smaller range as the location of the constriction moves further back in the vocal tract. The medians and interquartile ranges of **F2onset** for each place of articulation for the individual speakers are shown in the boxplots in Figure 3.4 and show a clear separation between labiodental and palatals. For labiodentals, vocal tract shape is the least constrained and **F2onset** range is widest. For palato-alveolars, vocal-tract shape is most constrained for tongue blade and body and **F2onset** range is narrowest. These results are consistent with Kewley-Port's results (1982), in which there is a wide range of F2 loci for /b/ across vowel contexts, while the range for /d/ is much narrower.

F2xF3 plots at onset compared to the vowel centers showed a tendency for voiceless tokens to be closer to the vowel target than voiced cognate tokens. However, this trend was not present in all cases; it was necessary to have a sufficient range of formant

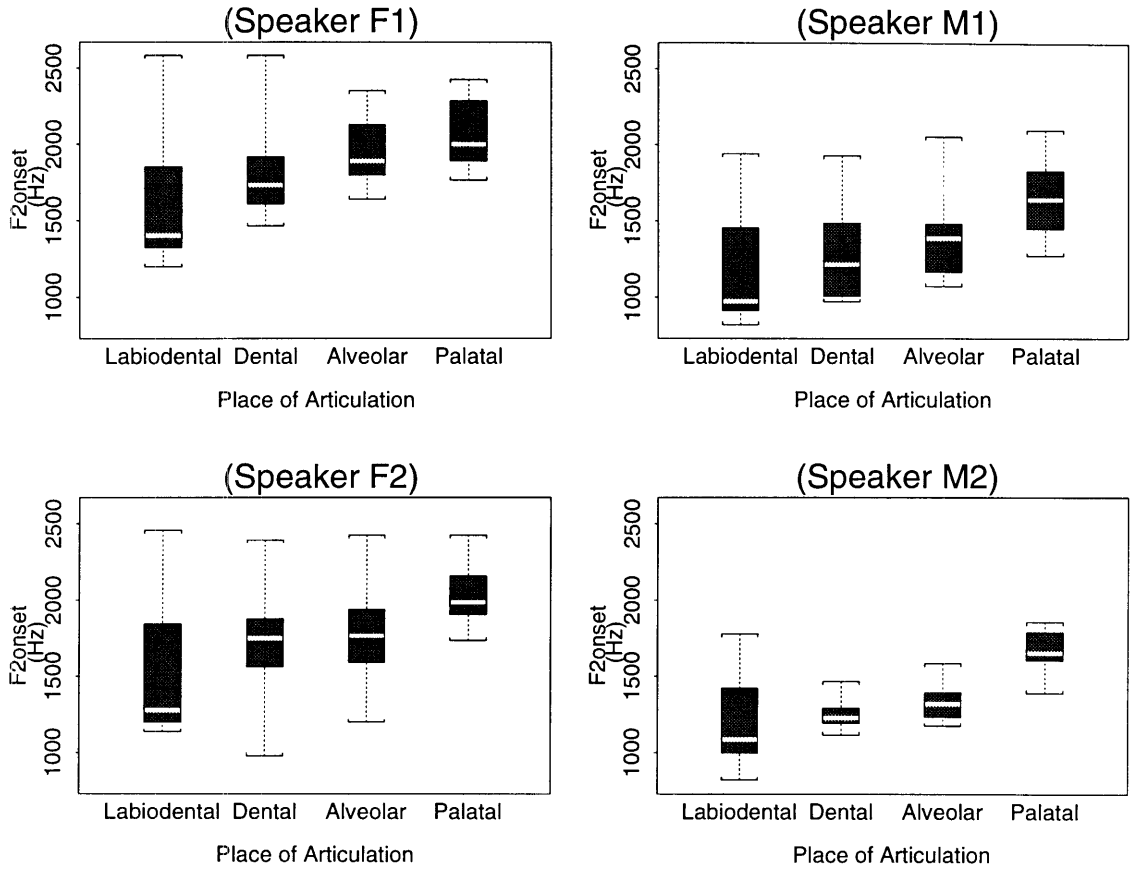

Figure 3.4: The medians (white lines) and interquartiles ranges (IQR=box height) illustrating the effect of place of articulation on **F2onset** are shown for each speaker. Each box corresponds to one place of articulation and represents 24 data points (2 repetitions of voiced and voiceless fricatives before each of six vowels). The whiskers (the dotted lines extending from the top and bottom of the box) extend to the extreme values of the data or a distance 1.5xIQR from the center, whichever is less.

values in order to discern a consistent difference between voiced and voiceless cognates. Figure 3.5 shows an example of an F2xF3 plot for one female speaker, in which the F2 and F3 values of the voiceless tokens tend to be closer (on average) to those of the vowel center for /u/, while all tokens in the /i/ context are tightly packed.

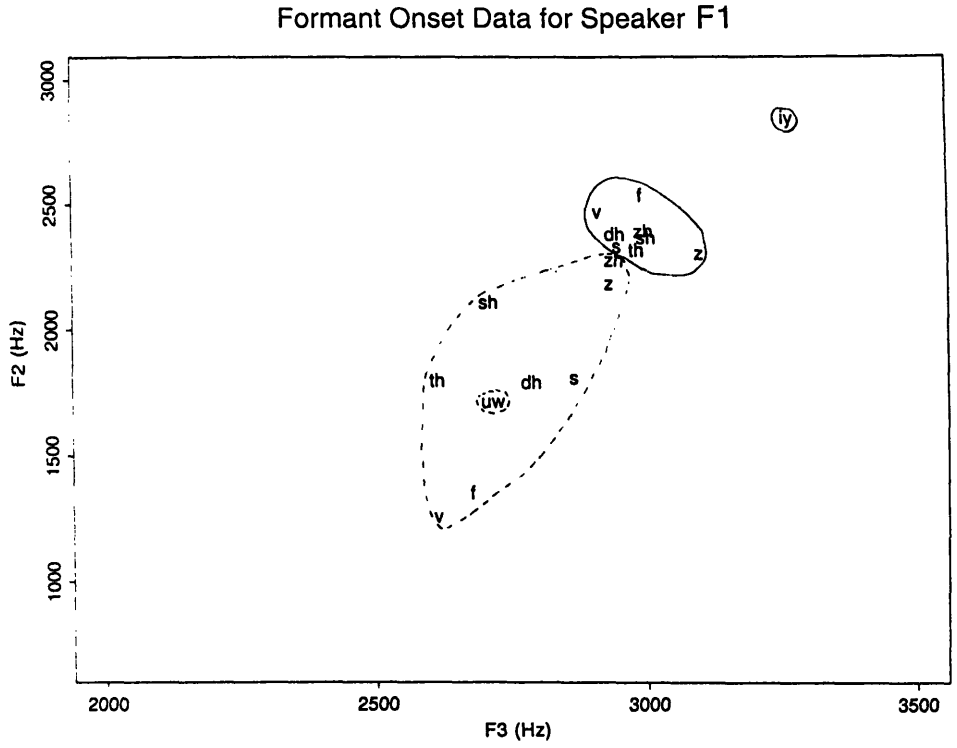

Figure 3.5: Onset values of F2xF3 for one female speaker are shown for fricatives before the vowels /i/ (grouped by solid line) and /u/ (grouped by dashed line). The vowel centers for each context (noted as iy and uw) is established by averaging midvowel values over all tokens with the same vowel context. Each fricative point represents the average between two instances of the same CV token.

3.3.3 Effects of Fricative-Vowel Interaction

The previous results showed that there is very tight clustering of **F2onset** data in the /i/ context, and the widest separation in the /u/ context. To explore the fricative-vowel interaction, the means and standard errors of the **F2onset** were calculated for each place of articulation separately for each vowel. Each point plotted in Figure 3.6 represents the mean values of four repetitions for each of the four subjects. The standard errors range from 38-100 Hz. Again, the biggest contrast in behavior is

noted between the labiodental and palato-alveolar context. The labiodental values are noted to drop sharply from the high-front (/i/) to low-back (/ɑ/) vowel context, and remains constant in the back vowel context. While the mean values for the palato-alveolars drop slightly from /i/ to /ɑ/, they remain relatively insensitive to the vowel context.

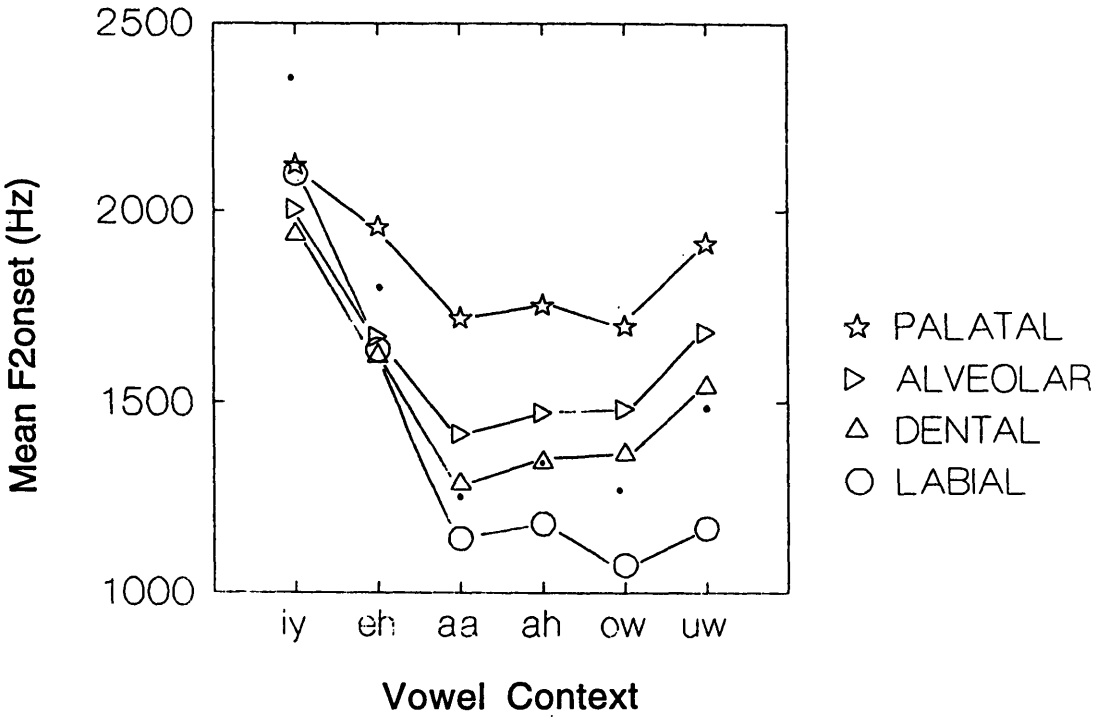

Figure 3.6: Mean **F2onset** values are plotted as a function of vowel context for each place of articulation. Each onset value represents the average of 16 tokens (four repetitions for each of the four subjects). The dots represent the average value of the second formant in the center of the vowel.

3.4 Inferring Articulatory Movements from Formant Patterns

The trends that were observed can be interpreted with respect to the existing articulatory data. As discussed in the previous chapter, there is still more data needed, especially for study of the kinematics during speech production. Midsagittal views of the vocal tract during consonant and vowel production in sentences generated by French speakers, obtained with cineradiography (Bothorel, 1986), are shown in Figure 3.7. Examples were chosen among the existing data for labiodental, alveolar and palato-alveolar fricatives in front and back vowel context.

3.4.1 Effects of Vowel Context

Each vowel imposes different amounts of constraint on **F2onset** values and, by implication, on the degree to which the vocal-tract shape behind the consonant constriction accommodates to the following vowel. One explanation for less acoustic variability in fricatives in the high, front vowel context /i/ is that the tongue blade position is high. Therefore, the distance to any of the fricative consonants is relatively small. In contrast, there is more acoustic variability in the back vowel context; the tongue blade must move a relatively large distance, and the jaw must be raised, to make the constriction for fricatives. This situation can be visualized from the x-ray tracings shown in Figure 3.7. The vocal-tract configurations for fricatives preceding front vowels (solid lines) and back vowels (dashed lines) in Figure 3.7(a), especially /fü/ vs. /fu/, illustrate a big effect of vowel context.

3.4.2 Effects of Place of Articulation

The vocal-tract configurations for /fu/ and /šu/ in Figure 3.7(a) illustrate a large variation in shape between different fricatives, for a given vowel /u/. Different consonant places of articulation show different amounts of acoustic variability. The wide range of acoustic variability for labiodentals implies that the tongue body and blade

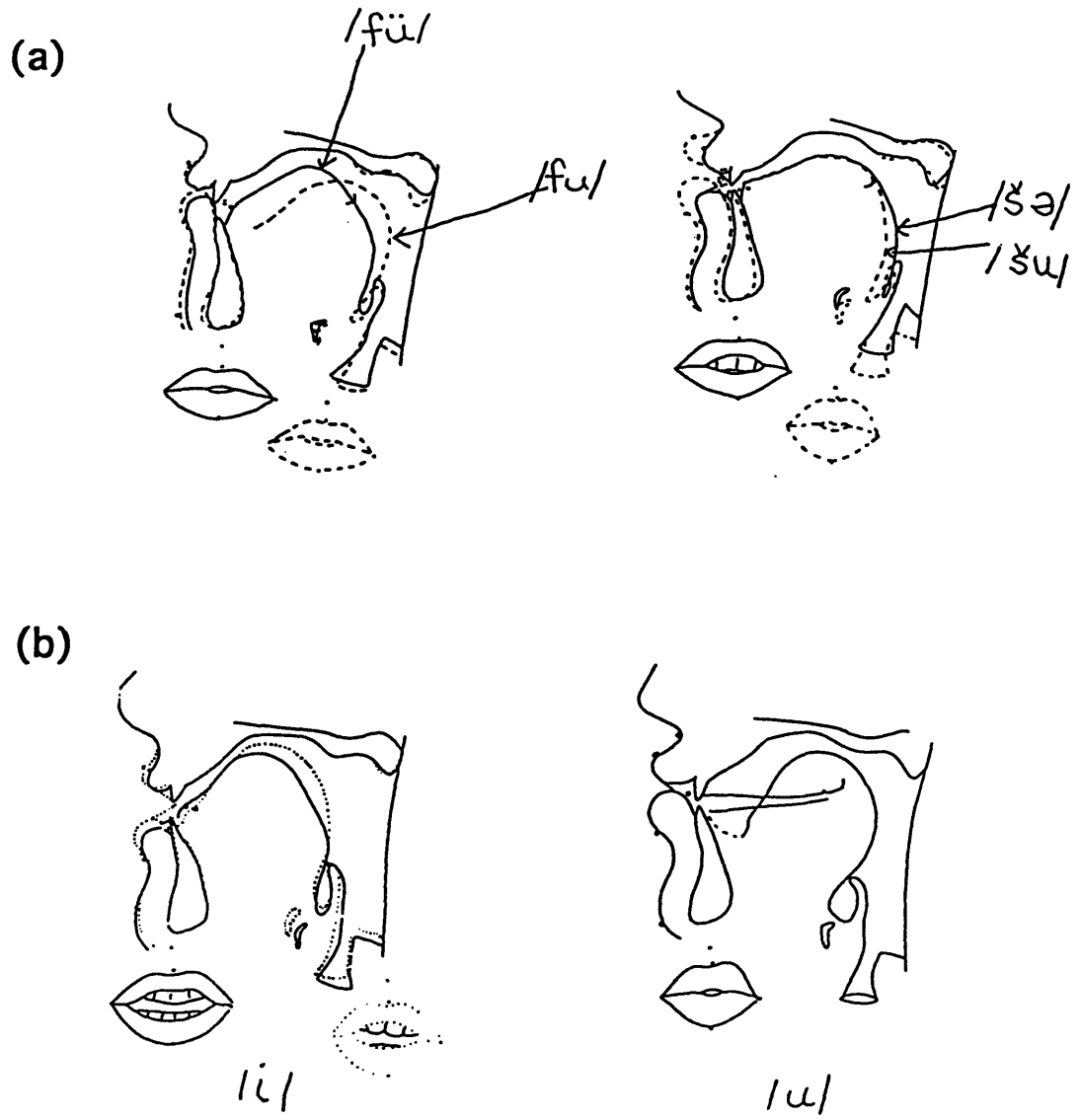

Figure 3.7: (a) Tracings from the x-rays of the same speaker during the first fricative in the following French phrases: Left: *Il fume son tabac* (solid line) and *Un fourré touffu* (dashed line). Right: *Ma chemise est roussie* (solid line) and *Une pâte à choux* (dashed line). These contrast /f/ before a high front rounded vowel /ü/ and before /u/ with /š/ before /ə/ and before /u/. (b) Two different subjects saying /i/ (left) and /u/ (right).

are free to anticipate the following vowel. The least amount of acoustic variability is observed for the palato-alveolar fricatives, for which the front part of the tongue body is quite constrained.

Voiceless fricative measures can be considered to be at a point closer to the vowel because voicing begins later relative to the oral release gesture. Onset frequencies for voiced fricatives may therefore reflect a more extreme supraglottal posture. It should be noted that more variability is inherent in establishing a precise location for a CV boundary in voiced as compared to voiceless fricatives. These observations suggest that it is more appropriate to consider a region, rather than a specific point, when locating the acoustic landmarks between the consonant and the vowel.

Overall, the present findings support the trends for fricative-vowel coarticulation reported in Wilde (1993) and are consistent with Fowler's (1994) concept of coarticulation resistance. The findings concerning the different amount of constraint that adjacent vowels and fricatives impose on each other are consistent with the notion of a "vocalic frame" as an integral part of consonantal identity (Sussman et al., 1991). Findings which link consonant identification to vowel context have been used to support a syllable-based view of speech perception.

Chapter 4

Time-varying Noise Characteristics

The acoustic consequences during fricative production, as discussed in Chapter 2, include continuous spectral variations over time. First, the articulation and aerodynamics in noise generation during a particular fricative are continuous. In addition, the acoustics of fricatives produced in connected speech, as opposed to sustained fricatives, are influenced by concurrent coarticulatory movements.

In this chapter, we will begin by describing the problems inherent in analyzing the noise in fricative consonants. In our search for solutions, we will review studies which seek to characterize the the time-varying noise in fricative consonants and present new research on the description and quantification of fricative properties that vary with time.

4.1 Statement of the Problem: Difficulty in Frication Analysis

In the processing of random noise signals, and therefore the analysis of fricatives, care must be taken to do appropriate averaging in order to observe the broad spectral properties. For a general discussion on the analysis and processing of random signals,

the reader is referred to Leon-Garcia (1994). In addition, for fricatives, the spectrum is changing over time as the supraglottal constriction is made and released. Therefore, characterization of noise properties, such as stridency, are highly dependent on the measurement techniques and the averaging time that is used.

What is common among the different approaches that previous researchers have applied, with various data reduction techniques, is finding the most critical aspects of the fricative that must be represented for listeners to correctly perceive them. The spectral moments measure, a statistical procedure successfully applied by Forrest et al. (1988) and Jassem (1965), is used to classify voiceless obstruents. This procedure treats a small slice of the consonant noise as a frequency distribution, and calculate its moments (mean, variance, skewness and kurtosis); these moments describe the signal on the basis of its central tendency, its shape and its symmetry. Tomiak (1990), comprehensively reviewed this literature in her dissertation, which presented a series of experiments to test the robustness of the spectral moments metric, including the most perceptually appropriate window for prototype derivation. Forrest et al. (1988) found the Bark scale enhanced their classification accuracy; however, Tomiak (1990) found very similar performance between the Bark and linear-based spectral moment profiles. Tomiak demonstrated the potential and limitations of the moments profiles, including that "metric information will be most representative when based upon extended frication sections, with window placement dependent on the place-of-articulation category under consideration" (p. 191).

Recent studies (Behrens and Blumstein, 1988a; Scully et al., 1992, Shadle et al., 1993; Xu and Wilde, 1994) have provided additional evidence that the kinematics of fricative articulation create an acoustic signal that is inherently non-static. Until recently, however, descriptions of the noise characteristics of fricative consonants using traditional methodology for speech analysis, i.e., use of the short-time Fourier transform (STFT), relied upon the assumption that the statistics of the signal do not change within the time interval specified by the analysis window.

An additional difficulty in the analysis of fricatives, compared to general speech analysis issues encountered for other sounds, arises from the nature of random noise

generation in fricative production. The nature of a noise source complicates the accurate measurement of spectral properties associated with the articulatory movement.

This difficulty in obtaining detailed spectral measurements of fricatives is illustrated in Figure 4.1, an example borrowed from a study of voiceless fricatives in Mandarin (Xu and Wilde, 1994). When a long analysis window is used, as shown in Figure 4.1(a), there are numerous local fluctuations, which reflect the randomness of the source, which are not characteristic of the vocal tract resonances. In comparison, when a short analysis window is used, as in Figure 4.1(b), there is considerable inconsistency among spectra that are very close together in time, as well as the statistical uncertainty attributable to noise source.

Fant (in press) explains that additional averaging of spectral data is necessary in order to reduce statistical uncertainties in the amplitude versus frequency analysis of unvoiced speech. In order to obtain an acceptable random error, the spectral section must be averaged over a time interval T_a which is substantially longer than T, the length of the FFT window. Fant cites that the following expression for the error in dB has been empirically validated:

$$20log_{10}(\sigma_e/A_e) = 4(BT_a)^{-0.5}(dB) \tag{4.1}$$

where

A_e is the estimate of the amplitude of a spectral component

σ_e is the standard deviation and

B=1/T is the reciprocal of the spacing between spectral samples.

That is, without averaging, the uncertainty in the amplitude estimates for fricatives is on the order of 4 dB, and the longer the averaging interval relative to the bandwidth, the more the error can be reduced. As we will discuss, the type of averaging, as well as parameters such as the length of the averaging interval, must be carefully chosen.

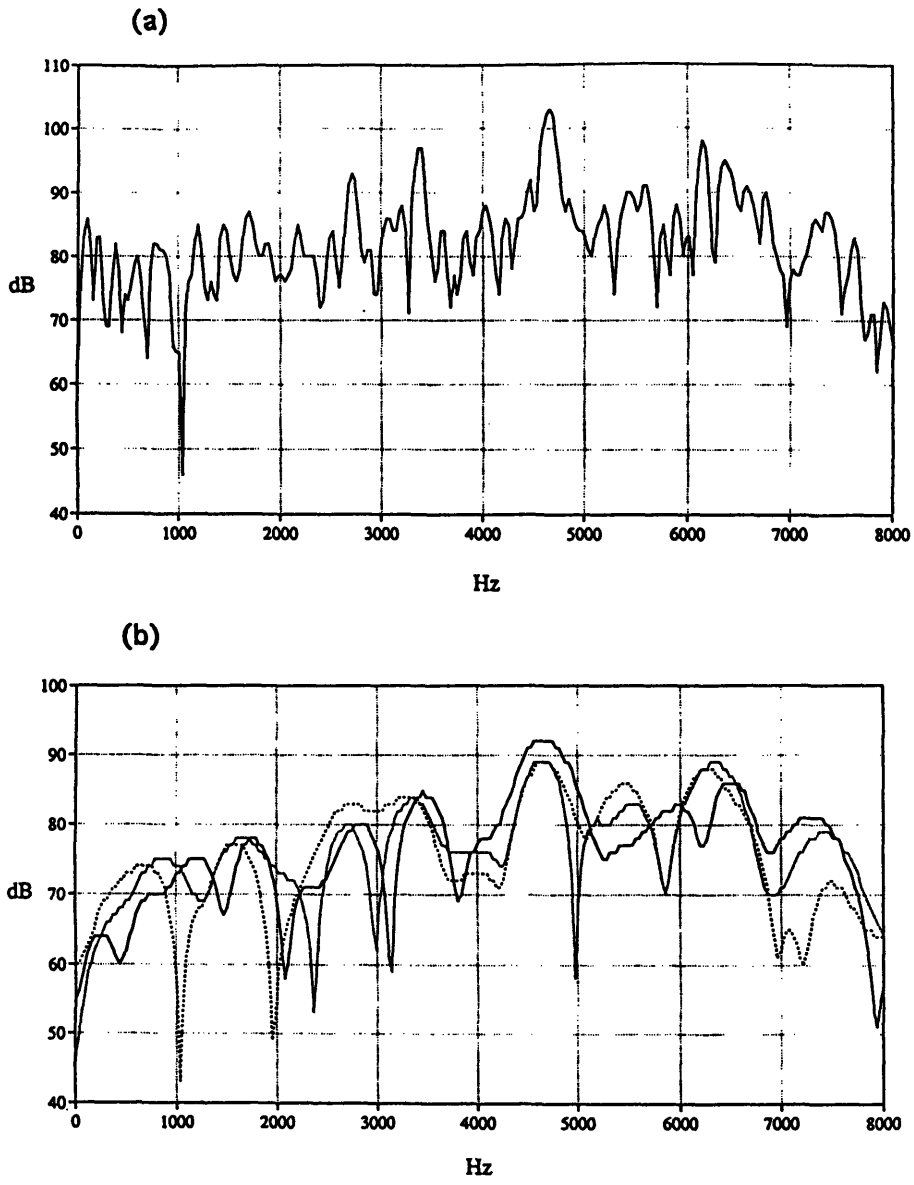

Figure 4.1: Spectra of the palatal fricative /ɕ/ illustrating the effect of analysis with different sizes of FFT windows. (a) Spectrum taken in the middle of frication with an analysis window size of 20 ms. (b) Three consecutive spectra of the same fricative obtained with an analysis window size of 4 ms. The windows are positioned 1 ms apart from one another. From Xu and Wilde (1994).

4.2 Literature Review

There is a growing body of literature on how to analyze fricative consonants, as the need remains for development of appropriate techniques. Shadle et al. (1992) compared time-averaging through the stationary portion of one fricative token with ensemble-averaging across intervocalic fricative tokens at specific events. They showed that ensemble-averaging captured the changes in spectrum over relatively short periods of time. However, both approaches involve assumptions that may not always hold. For time-averaging, it is assumed that the fricative token is stationary over the interval that it is averaged. For ensemble-averaging, it is assumed that all tokens are produced in identical fashion.

Of course, the methodology chosen is dictated by the questions that are being asked, as well as the applications that are being addressed. Klatt (unpublished manuscript) suggested that ensemble averaging over tokens with the same fricative, but different vowels, could yield representative templates of the eight English fricatives. He computed these separately for front, back and back-rounded vowel contexts, called the results "grand-average spectra" and used them as a basis for speech synthesis.

Coarticulation effects in fricative spectra in prevocalic position have been well-documented. For example, Soli (1981), using mean LPC spectra, revealed anticipatory vowel coarticulation effects in initial sibilant fricatives, as evidenced by spectral peaks affiliated with the second formant of the following vowel (/ɑ, i, u/). This evidence was consistent with results from a companion perceptual study (Yeni-Komshian and Soli, 1981) which showed that listeners could identify the following high vowels /i, u/ from fricatives excised from fricative-vowel syllables. For both studies, the effects for the low vowel /ɑ/ were weakest, presumably because the opposing articulatory gestures yielded the least assimilation. In addition, in these fricative-vowel tokens, more coarticulatory information was observed to be present in the latter portion of the fricative.

Encouraging results were obtained in a study by Xu and Wilde (1994) that com-

bined time averaging and ensemble averaging into a method for analyzing anticipatory coarticulation of lip-rounding on voiceless fricatives in Mandarin. The results indicated that, once the numerous local fluctuations due to the noise source were smoothed, prominent peaks could be identified as resonances characteristic of the vocal tract. The nature of the corpus also allowed the relations between peaks in frication and formants in adjacent vowels to be inferred, the time course of anticipatory coarticulation to be revealed, and the spreading effect of vowel rounding to be examined. We will consider the analysis method employed by Xu and Wilde (1994) as one extension of the methodology used in the present study of English fricatives.

Context effects in broad phonetic environments have also been examined. Spectrogram reading experience and studies on the acoustic regularities in speech have demonstrated that fricatives can exhibit distinct spectral changes depending on the phonetic context. For example, Zue (1985) observed that while /ð/ is often realized in a stop-like manner, when preceded by a consonant, its acoustic realization is weaker in amplitude and higher in frequency than a true /d/ burst. Glass (1988) used cluster analysis to classify two acoustic patterns for the voiced dental fricative. Acoustic classification of 473 instances of /ð/, taken from two 100 talker databases, were found to cluster into different classes: one dominated by low-frequency energy, and the other by high-frequency energy. The latter, a more stop-like configuration, was observed when /ð/ was preceded by silence or an obstruent, as opposed to the other pattern associated with a preceding sonorant. A similar result was also obtained for the other weak voiced fricative /v/.

One objective of the present research was to examine the differences in the nature of the observed time variations in fricatives that are classified as strident vs. nonstrident. A further goal was to quantify the normal range of amplitudes and spectra that characterize stridents as opposed to nonstridents. During the course of this ongoing study, several other investigators have shared this goal (Utman and Blumstein, 1994; Bitar, 1993).

The objective of the study by Bitar (1993) was to automatically classify fricative sounds as strident or nonstrident. The metric for correct classification was chosen to

be the phonetic labeling of the TIMIT database, in which the nonstrident fricatives were considered to be /f, v, θ, ð/ and the strident fricatives were /s, z, š, ž/. The training set consisted of all SI sentences (N=192) in the TIMIT database. The test set allowed comparing TIMIT labeling of 1004 fricatives spoken by ten male and ten female speakers from eight geographical regions. The strategy was to compute the energy in different frequency bands and reference this to a measure of the total energy taken over the entire utterance, which was sampled at 16 kHz. First the boundaries of "steady-state" region(s) of the fricative were estimated, by computing the overall energy and then searching for the fastest transition showing an increase or decrease. Then the energy was computed in each of the chosen frequency bands: 2-4 KHz, 4-6 kHz and 6-8 kHz. Presumably, the alveolar fricatives would be identified as strident due to the spectral peak expected in the 4-6 kHz band and the palato-alveolar fricatives would be expected to have a broad spectral prominence that falls within the 2-4 kHz band. Then a normalized energy was calculated by referencing the obtained band energies to the total energy for the entire utterance.

Bitar's error analysis showed that the approximately three percent classification error was systematic. Strident fricatives appeared to be misclassified as nonstrident when found in a nonstressed vowel context. In the few cases in which nonstrident segments were labeled as strident, the phonetic environments were as follows:

1. retroflexed (N=3)
2. adjacent to stridents (N=2)
3. in stressed syllable (N=5) as defined by listening and identifying the phrase level stress pattern of the utterance

While Bitar used a simple convention for determining the label as strident or nonstrident, the present study adopts the definition that stridency is "evidenced by greater energy in the high frequencies in the consonant, relative to the energy in the corresponding regions in the vowel" (Ohde, 1992, p. 112). Therefore, typically weak fricatives, such as /f/ that are strengthened in certain phonetic contexts, such as those

observed in the above classification "errors", would indeed be considered strident if they met our acoustic criteria.

4.3 Methodology

In the present series of studies, the speech corpus described in Chapter 3, consisting of the 'CVCV'CVC utterances of two males and two females, was used. Averaged spectra were calculated for utterances containing all eight English fricatives in six vowel environments. Characteristics of the noise spectra were considered with respect to the adjacent vowel spectra.

The first step in signal processing was to reduce the error in the measurement of the spectrum through temporal averaging. The averaged spectra were obtained by stepping a relatively short Hamming window of 6.4 ms every 1 ms and then averaging 15 overlapping windows. This method is schematized in Figure 4.2. The averaging interval of 20 ms was chosen to be long enough to minimize the measurement error and short enough to allow for quantifying the time variations present in individual tokens. According to Equation 4.1, the uncertainty in the amplitude measurements made with the chosen parameter values (T_a = 20.4 ms for 15 frames of a window of effective length 6.4 ms) would be reduced to 2.2 dB. Since a Hamming window has about half the effective length of the same size rectangular window, we can further calculate that the amplitude values we measure are within about $\pm$ 1.5 dB of the expected amplitude values.

An averaged spectrogram, created with the time-averaging process applied to the entire fricative-vowel utterance of /si/ spoken by one male subject (M1) is shown in Figure 4.3. We will look at the spectral characteristics that are revealed from this representation. First, we will characterize the spectra in the steady part of the noise, if there is one. Next, we will focus on the evidence of time-varying characteristics, involving (a) amplitude of noise for back cavity resonances that shows up more strongly at the edges, where they are possibly due to aspiration, and (b) changes in the higher frequencies that occur over the course of the consonant. In the example

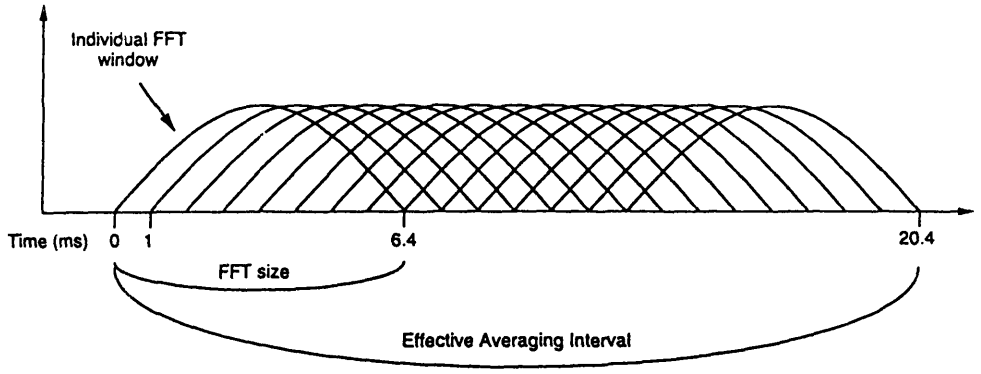

Figure 4.2: A schematic of time averaging, in which 15 spectra obtained with 6.4 ms Hamming windows are averaged over a 20.4 ms interval in order to generate a single averaged spectrum. The windows are advanced in 1 ms time steps.

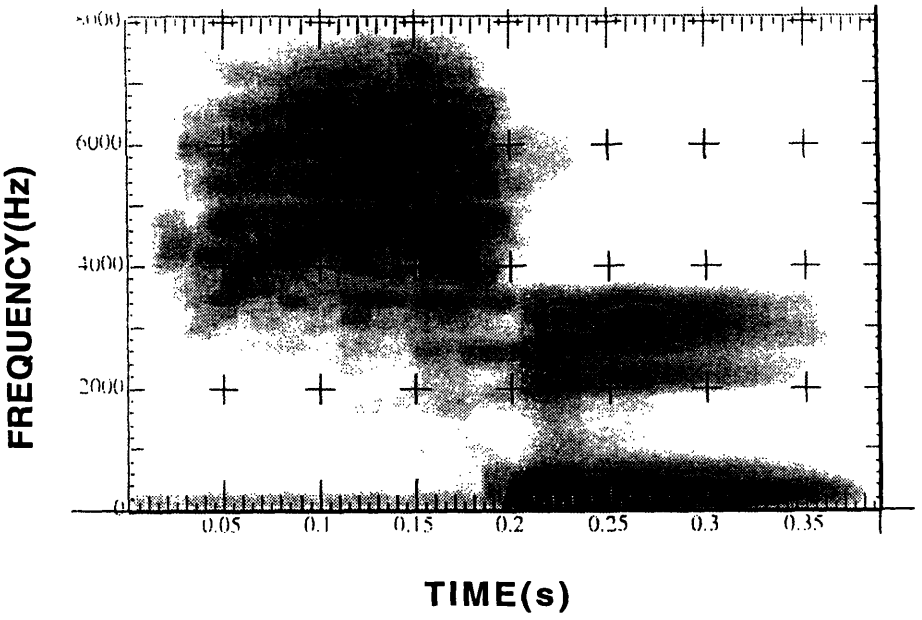

Figure 4.3: A time-averaged spectrogram created with the time-averaging process shown in the previous figure of the utterance /si/ spoken by one male subject (M1) is shown.

shown in Figure 4.3, the third and fourth formants show up very close to the boundaries between the fricative and the vowel, where they are possibly due to aspiration. For this study, the presence of aspiration is identified qualitatively as an interval of noise which includes an energy prominence continuous with the second and higher formants in the adjacent vowel. It should be noted that this measure by itself does not rule out the possibility of incomplete pole-zero cancellation with a frication noise source near the constriction, as has also been proposed (Badin, 1991). We also see changes in the frequency of the major frication peak, which in this example shifts from about 4.0 to 4.5 kHz from beginning to end (as well as another concentration of energy change around 6.5 kHz). In this example, these higher frequency resonances must be attributable to the front cavity.

Time-averaged spectral slices from the same utterance as in Figure 4.3 are shown in Figure 4.4. The fricative /s/ is taken at its midpoint and the following vowel /i/ is taken 20 ms after the fricative-vowel boundary. The concentration of energy in the low frequencies for the vowel and high frequencies for this strident fricative are readily apparent. All spectra are calculated without preemphasis, unless otherwise indicated.

Time-averaged spectra taken about the middle of the fricative in the utterances for /əfɑ/, /əθɑ/, /əsɑ/, and /əšɑ/ are shown in Figure 4.5. The vertical positioning of these fricative spectra represent relative amplitudes. The amplitudes were corrected by up to 3 dB to compensate for differences in the amplitude of the first formant in the following vowel. These spectral shapes are consistent with the general findings in the literature for fricatives at the four places of articulation in English. Spectra for the labiodental /f/ and dental /θ/ are relatively flat and weak in intensity. In contrast, prominent spectral peaks are noted for the stronger fricatives. The primary peak for /s/ is just above 4 kHz. The spectrum for /š/ shows several peaks, with maximum amplitude between 2 and 4 kHz.

In the present analysis, the amplitudes in restricted frequency regions of the noise were examined with respect to the neighboring vowel. It was hypothesized that relative measures could be found to capture important characteristics of the gross spectral

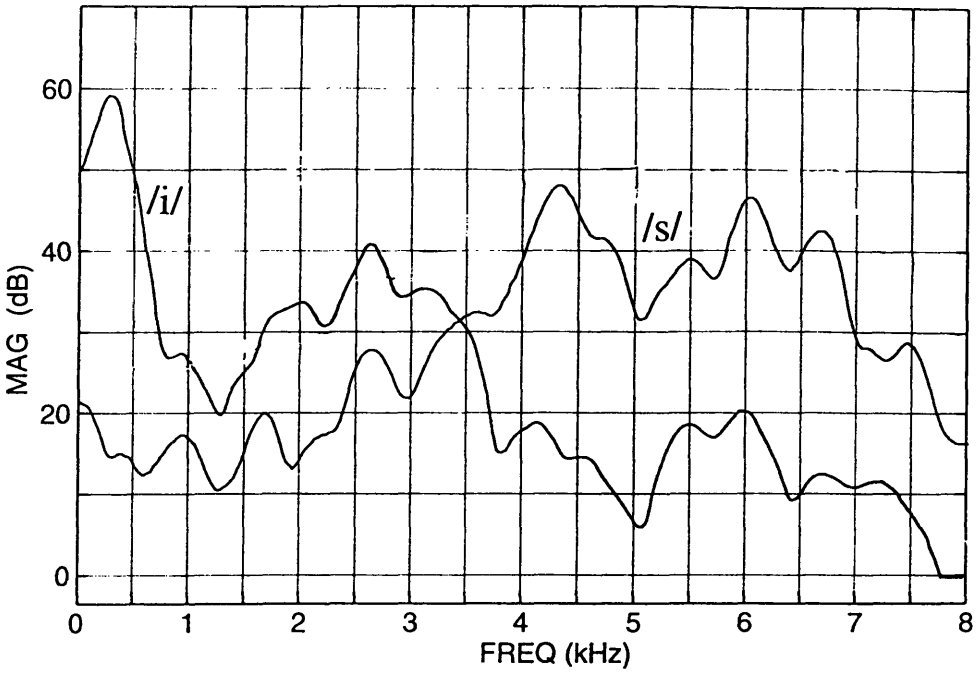

Figure 4.4: Comparison of averaged spectra taken in the fricative /s/ and following vowel /i/ spoken by one male speaker (M1). The spectra are calculated by moving a 6.4 ms Hamming window in 1 ms time steps and averaging over a 20 ms interval.

shape and reduce the dimensionality of the data. In addition, it was hoped that individual variability could be examined and thresholds that capture ranges of variability, e.g. for stridency, could be found. Octave bands were chosen in consideration of preliminary findings and with the additional motivation that they maintain an essential feature of the tuning of the auditory system (i.e., poorer spectral resolution at the higher frequencies with maintenance of constant bandwidth with log frequency above about 1 kHz). Therefore, given the bandwidth of the speech under analysis, five bands were examined this study: 1) 0-500 Hz, 2) 500-1000 Hz, 3) 1000-2000 Hz, 4) 2000-4000 Hz and 5) 4000-8000 Hz. These frequency bands are indicated below the x-axis of Figure 4.5.

We looked at peak amplitudes in these discrete bands. It is important to document the sources of variability that could arise from this procedure. First, the sharp frequency boundaries used would be expected to introduce predictable behavior analogous to those in signal processing using non-overlapping rectangular windows in time. And, of course, quantization errors are inherent in any methodology in which

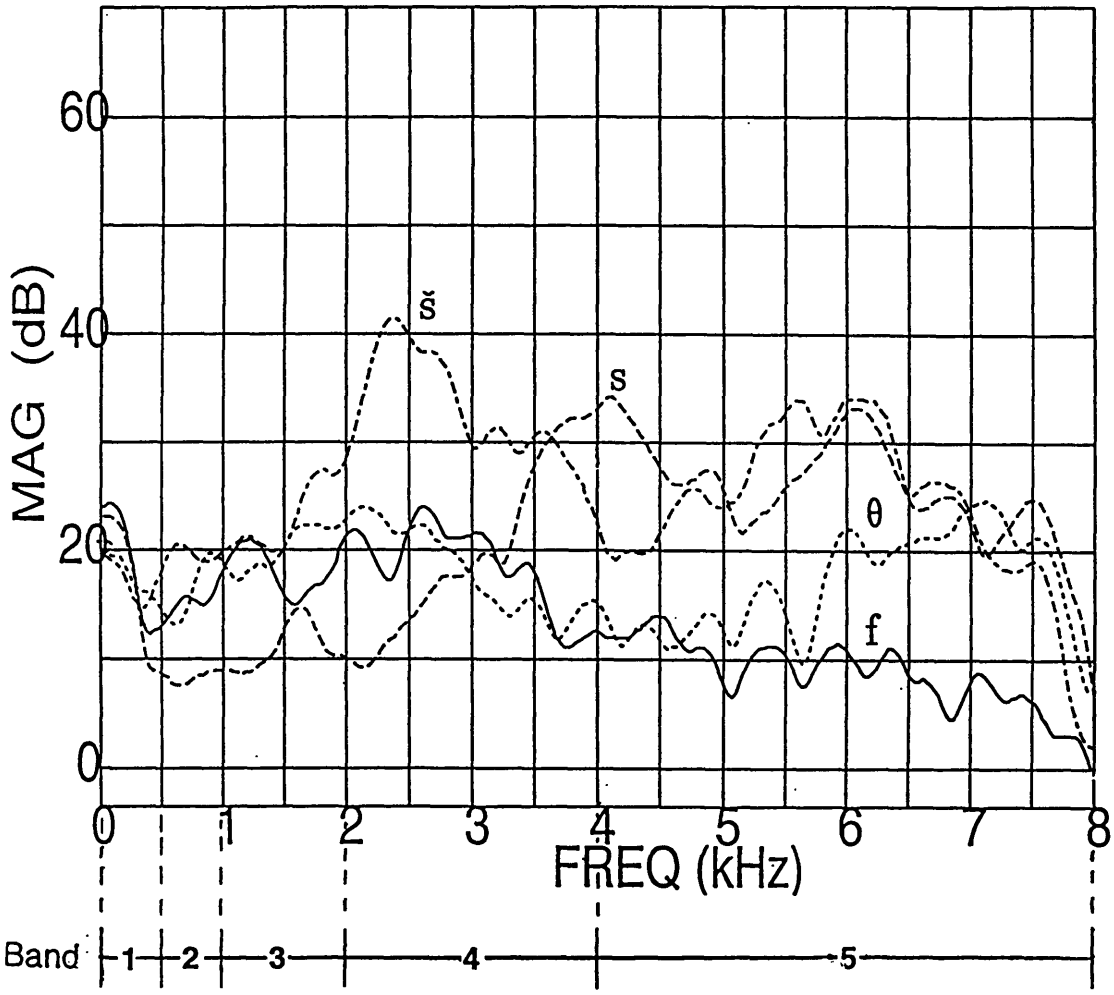

Figure 4.5: Time-averaged spectra of the four voiceless fricative consonants, preceding /ɑ/, spoken by a male speaker (M1). The amplitudes have been corrected by up to 3 dB to compensate for differences in the F1 amplitude of the following vowel.

a continuous range is partitioned into discrete units. If the amplitude measures are expected to indicate amplitudes of spectral prominences, then errors could result from two primary conditions: 1) if formants moved across frequency band boundaries or 2) if more than one formant fell within a single band.

The approach of picking out the amplitude maximum can also yield a reasonable estimate of the total energy in dB, especially when there is no dominant peak. Preliminary results of applying this method using Klattools programs (Klatt, 1984) confirmed that the analysis process could be further simulated in software. Standard signal processing routines from ESPS (1992), such as **sgram** to create a spectrogram and **pplain** to translate files to a text format, were used when possible. Additional signal processing routines, **avg-sgram** to create a customized time-averaged spectrogram and **track-peak** to pick out amplitude peaks, had already been developed by colleagues to work with ESPS routines.

A family of programs, **read-amp.c**, was written to concentrate on finding acoustic cues in the vicinity of acoustic events called landmarks (Stevens, 1991). For the purposes of this study, the landmarks consisted of the fricative-vowel and vowel-fricative boundaries. Stevens (1985) acknowledged that there is a growing body of evidence that information about phonetic features is concentrated in the interval within 10 to 30 ms of these boundaries. Once acoustic landmarks are located, signal processing and analysis can be focused in their vicinity, where change is occurring and information is concentrated. Our approach to describing time-varying noise in fricatives with respect to landmarks, and especially to quantifying the stridency feature, is consistent with a model for lexical access based on features (Stevens et al., 1992). In this more general context, landmarks include more than just phonetic boundaries and have a particular role in the production and perception of speech sounds: they are events rich in acoustic cues about phonetic features.

The fricative-vowel and vowel-fricative boundary times used in this study were the carefully marked events described in Chapter 3 for this database. In order to design a framework that could also work on larger databases such as TIMIT (Zue et al., 1990), label files were created to document these times. The function **phn-cls.c**

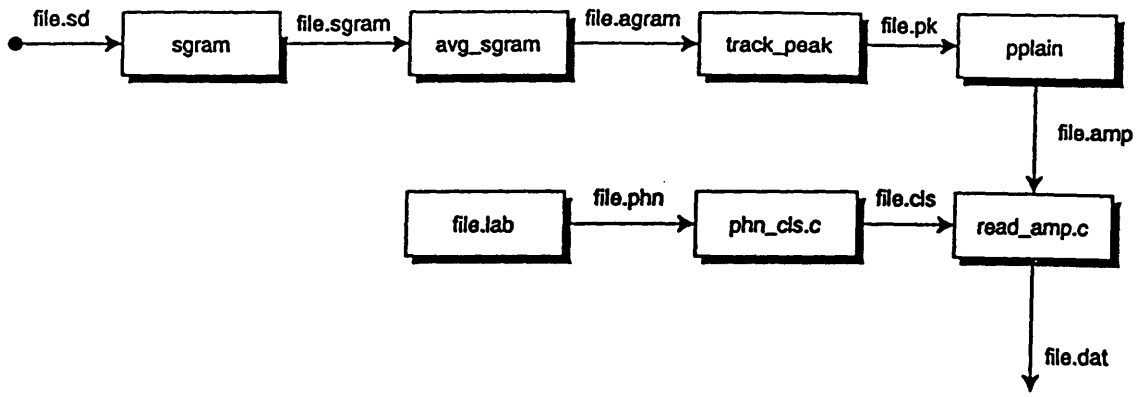

Figure 4.6: Overview of the system developed for quantifying the time-varying noise spectra of fricative consonants.

already existed; it classifies the labels into broad phonetic categories and allows these methods to be adapted for speech sounds besides fricatives.

The system developed for automatically creating and analyzing averaged spectra is displayed in Figure 4.6. This system inputs the digitized waveform file (file.sd) and a file containing identified times of the fricative-vowel boundary (file.lab) for each utterance. The function **track-peak** takes the time-averaged spectrogram (file.agram) and picks out the spectral peaks, i.e., the maximum value in any frequency bin found within the pre-specified frequency bands. Finally, the system outputs the amplitudes and frequencies of those spectral peaks occurring at times of interest relative to the identified landmark(s) (file.dat). Plots of the peak amplitude variations in the five frequency bands aligned to the corresponding spectrogram for the utterance /əfɛ/ are shown in Figure 4.7.

The following questions were asked in order to confirm that important spectral characteristics of the fricatives being studied were maintained in the output of the system. First, what is the gross spectral shape at each place of articulation? Next, what do the spectra look like in relation to the vowel?

We will always be talking about relative (not absolute) amplitudes. Several ways

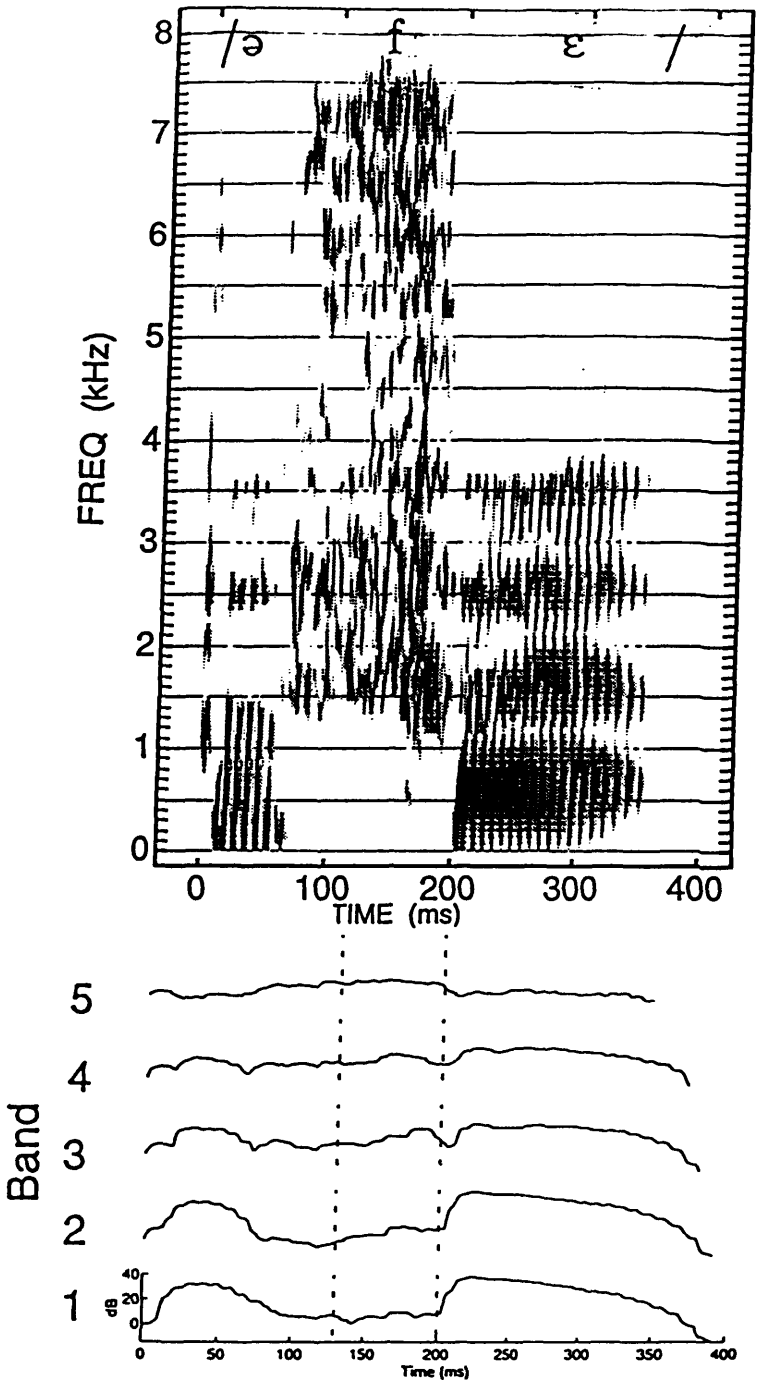

Figure 4.7: Amplitude variations for the five frequency bands aligned with the spectrogram of the utterance /əfɛ/.

of normalizing the noise amplitudes were considered. One simple method was to normalize peak amplitudes in the specific frequency bands in the noise by subtracting the peak amplitudes in the corresponding region in the vowel. Another was to normalize peak amplitudes in the noise with respect to the amplitude of the first formant in the vowel. The overall level of the vowel almost always reflects the amplitude of the first formant, as used in Utman and Blumstein (1994). In our system the amplitude of F1 in the vowel would be reflected by finding the maximum between the peak amplitude in Band 1 (0-500 Hz) and the peak amplitude in Band 2 (500-1000 Hz). That is, F1 amplitude is represented by the maximum peak below 1000 Hz in the vowel. Unless otherwise specified, the time of this reference measurement was chosen to be 20 ms after the consonant-vowel boundary ($CV + 20$). When looking for acoustic cues on either side of our acoustic landmark, the fricative-vowel boundary, we have the benefit of looking at speech output when the source is changing most rapidly, yet the vocal tract shape has not changed very much.

4.4 Results

We begin by describing the gross spectral characteristics of fricatives. We start with measures made at one time relative to the consonant-vowel boundary (CV), in order to compare our results with those of traditional studies on the acoustics of fricatives. For example, $relative\ time = CV - duration/2$ denotes the midpoint of the fricative. We then consider the variation in the noise over the duration of the consonant. Finally, we review results of applying our methodology to the quantification of the feature [strident].

4.4.1 Gross Spectral Characteristics

In this section we examine the gross spectral charcteristics of the fricatives in relation to the vowel. Results for two methods of amplitude normalization in relation to the vowel are discussed below.

Effects of Place of Articulation

Results for the amplitude differences obtained for the voiceless and voiced fricatives in the five frequency bands, normalized by simply subtracting the peak amplitude in the corresponding frequency band in the following vowel, are shown for one male speaker (M1) and one female speaker (F1) in Figures 4.8 and 4.9, respectively. The lines between the data points in each frequency band were included to give a rough indication of overall spectral shape; there was no actual interpolation between the five data points per plot. Each anchor point represents the average of 12 values (1 fricative x 6 vowels x 2 repetitions) each of which represents the peak amplitude found in a particular frequency band (Band1: 0-500 Hz, Band2: 500-1000 Hz, Band3: 1000-2000 Hz, Band4: 2000-4000 Hz and Band5: 4000-8000 Hz) at a specified time in the fricative relative to the CV boundary.

The plots shown in Figures 4.8 and 4.9 give an indication of the relative amplitude in corresponding octave frequency bands between English fricatives and vowels. Some expected characteristics, which could be predicted from averaged spectra such as those in Figures 4.4 and 4.5, are evident. Negative values indicate that the amplitude peaks in the vowel are stronger than those in the fricative. Positive values are observed only for Band 5 for the alveolar fricatives and for Bands 4 and 5 for the palato-alveolar fricatives. This result is consistent with amplitude predictions from models for these stronger fricatives. Fricative spectra, according to the acoustical theory of speech production and simplified electrical models (Heinz and Stevens, 1961), can be characterized by poles and zeroes, which depend on the location of the constriction in the vocal tract and the location of the source of excitation. That is, the resonance frequencies for fricatives are inversely related to the size of the front cavity. In Chapter 2, we calculated the first spectral peak for an alveolar fricative to be about 4.5 kHz for an average adult male. Prior studies have reported major frequency peaks within the 3.5-5 kHz range for /s/ and 2.5-3.5 kHz range for /š/ and essentially flat spectra for /f/ and /θ/ with diffuse spread of energy from about 1.8-8.5 kHz (Hughes and Halle, 1956; Strevens, 1960). We would expect the first resonance for the weak fricatives to occur at frequencies above the cutoff frequency used in this study.

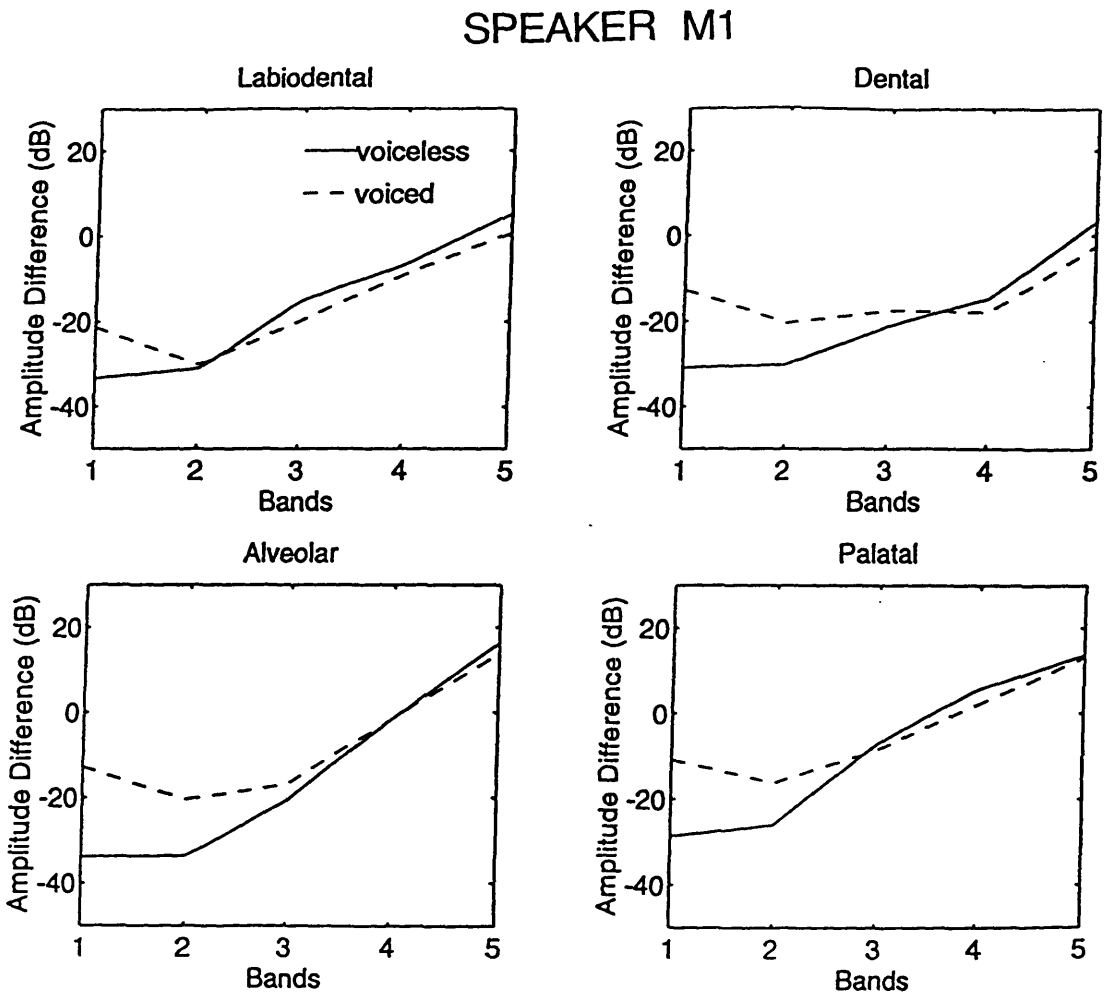

Figure 4.8: Results of band-by-band amplitude normalization comparing voiceless (solid lines) and voiced (dashed lines) fricatives are shown for one male speaker (M1). In this plot, each anchor point represents an average of 12 values, which are normalized with respect to the amplitude of the peak in the same band in the following vowel (Band1: 0-500 Hz, Band2: 500-1000 Hz, Band3: 1000-2000 Hz, Band4: 2000-4000 Hz and Band5: 4000-8000 Hz). All measures were made in the vicinity of the fricative-vowel boundary: *relative time* $= CV - 20$ for the fricatives and $CV + 20$ for the vowels.

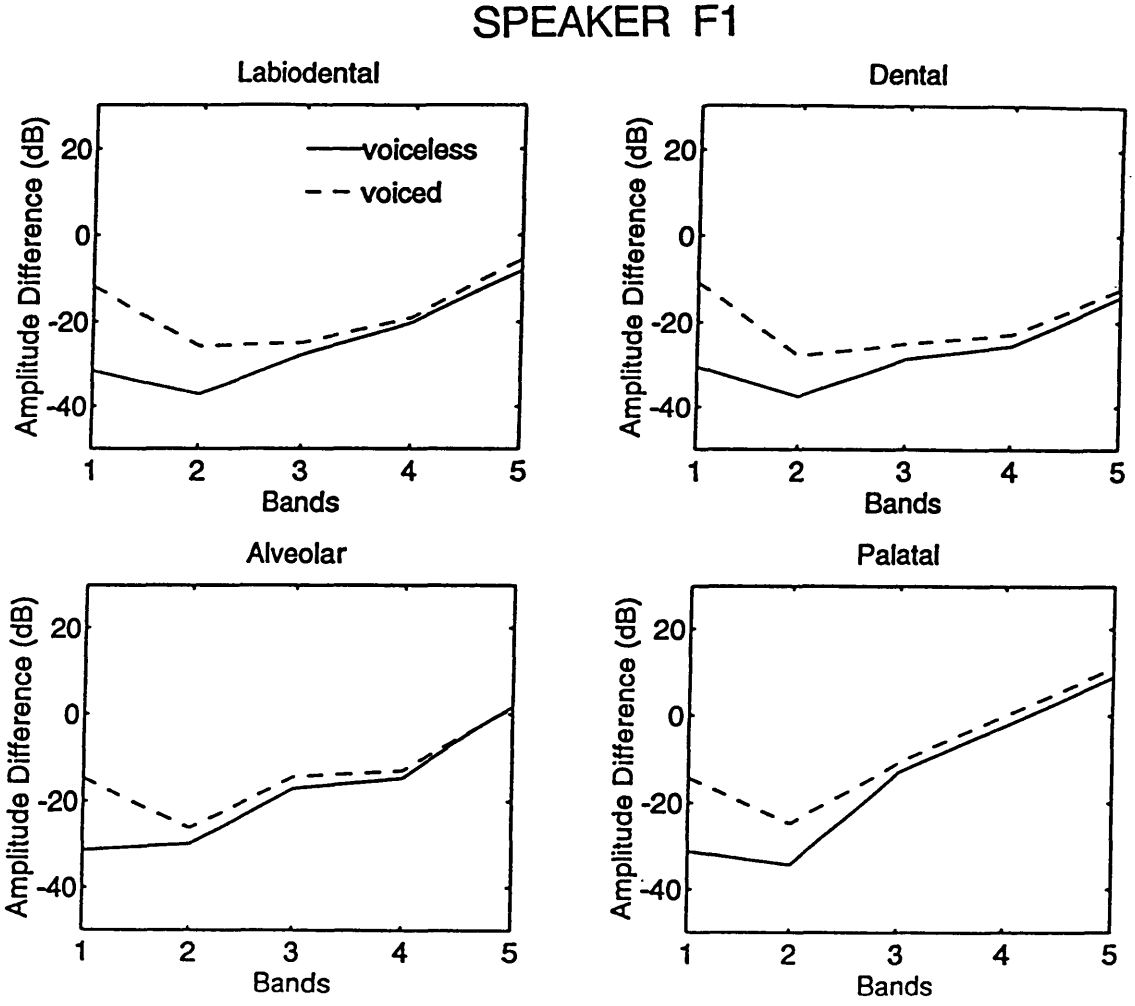

Figure 4.9: Results of band-by-band amplitude normalization comparing voiceless (solid lines) and voiced (dashed lines) fricatives separately for one female speaker (F1). In this plot, each anchor point represents an average of 12 values, which are normalized with respect to the amplitude of the peak in the same band in the following vowel (Band1: 0-500 Hz, Band2: 500-1000 Hz, Band3: 1000-2000 Hz, Band4: 2000-4000 Hz and Band5: 4000-8000 Hz). All measures were made in the vicinity of the fricative-vowel boundary: *relative time* $= CV - 20$ for the fricatives and $CV + 20$ for the vowels.

The average frequency values for shorter vocal tract, e.g. from women and children, would be shifted upwards, but the relations would remain essentially the same. In our results, the highest normalized amplitude value for the female speaker (F1), as compared to the male speaker (M1), for the palatal falls in Band 5 rather than Band 4. Frequency peaks for /š/ above 4 kHz have even been found by Behrens and Blumstein (1988a) in their data on the speech of three male speakers, in contrast to values below 4 kHz reported in the earlier studies (Hughes and Halle, 1956; Strevens, 1960).

Trends in voicing can be discerned. The low-frequency energy for the voiced fricatives is clearly greater than for the voiceless fricatives; this is attributable to the addition of the voicing source. Voiceless fricatives, which are expected to have greater intraoral pressure, evidence the expected trend for greater amplitude noise in the higher bands for the male speaker, but not the female speaker.

Normalizing the peak amplitudes in this simple band-by-band manner has at least two inherent drawbacks that can especially affect the higher frequency bands. First, the results are highly dependent on the peak amplitude values in the vowel, which depend on the frequency and bandwidth of those peaks. Band-by-band normalization does not insure the same frequency of the peaks between the fricative and the vowel. In addition, depending on the timing of the glottal and supraglottal constrictions, frication noise may extend past fricative-vowel boundary and affect the normalization factor obtained at 20 ms into the vowel. Still, band-by-band normalization might have perceptual relevance.

Normalization with respect to the amplitude of the first formant of the following vowel was shown to register the relative amplitudes as well as maintains the gross spectral shape of the fricatives observed in Figure 4.5. Normalized amplitudes, obtained by subtracting first formant amplitude (A_1) at time $CV + 20$, are shown for voiced and voiceless fricatives in Figures 4.10 and 4.11 for subject M1 and Figures 4.12 and 4.13 for subjects F1. The normalized amplitudes were obtained by subtracting the maximum amplitude peak below 1000 Hz in the vowel (i.e., maximum (Band1 amplitude, Band2 amplitude), which is taken to reflect the amplitude of F1 (A_1),

from the peak amplitudes in the noise. An offset value of 64 dB has been added to the plots. Differences between voiced and voiceless fricatives are evident, and the Band 1 amplitude for the voiced fricatives is 10-15 dB less than the vowel.

These barplots show an alternative way of schematizing the spectral output of the analysis system, and also include information about the variation in the data that was not included in Figures 4.8 and 4.9 by displaying both the means and standard deviations. In Figures 4.10, 4.11, 4.12 and 4.13, the normalized amplitudes are averaged separately for tokens preceding front /i ɛ/ (solid bars), back /ɑ ʌ/ (dashed bars) and back-rounded vowels /o u/ (dotted bars). The effect of the following vowel context on this measure will be further discussed below. In these barplots, each bar represents an average of 4 tokens.

These barplots provide a useful view of schematized spectra, which maintain known characteristics of fricative spectral shape. These essential characteristics include the relatively weak and flat amplitude values for the labiodental and dental fricatives; i.e., results show similar amplitude values among the highest three bands. The results for the clearly strident alveolar and palato-alveolar fricatives are consistent with the presence of high-frequency resonances, that were predicted when modelling front cavity dimensions and that were shown for the male speaker (M1) in Figure 4.5. The trend is for the amplitudes of alveolars to peak in Band 5 (above 4 kHz) and the amplitudes of palato-alveolars to have a broad maximum that falls in the middle frequency region (approximately 2-4 kHz). We would expect that results for strident fricatives spoken by individuals with vocal tracts that are shorter than those of the average male to show higher frequency peaks. Female speakers, for example, might show a broad maximum for palato-alveolars that spans the 4 kHz band edge; this is reflected in the outputs for both Band 4 and 5 for the female speaker (F1) in Figures 4.12 and 4.13. Similarly, individuals with longer than average male vocal tracts might have a peak for the alveolars that would cross into Band 4. We expect these types of speaker variability, and given the known limitations of quantizing frequencies into absolute bands, we revisit this issue.

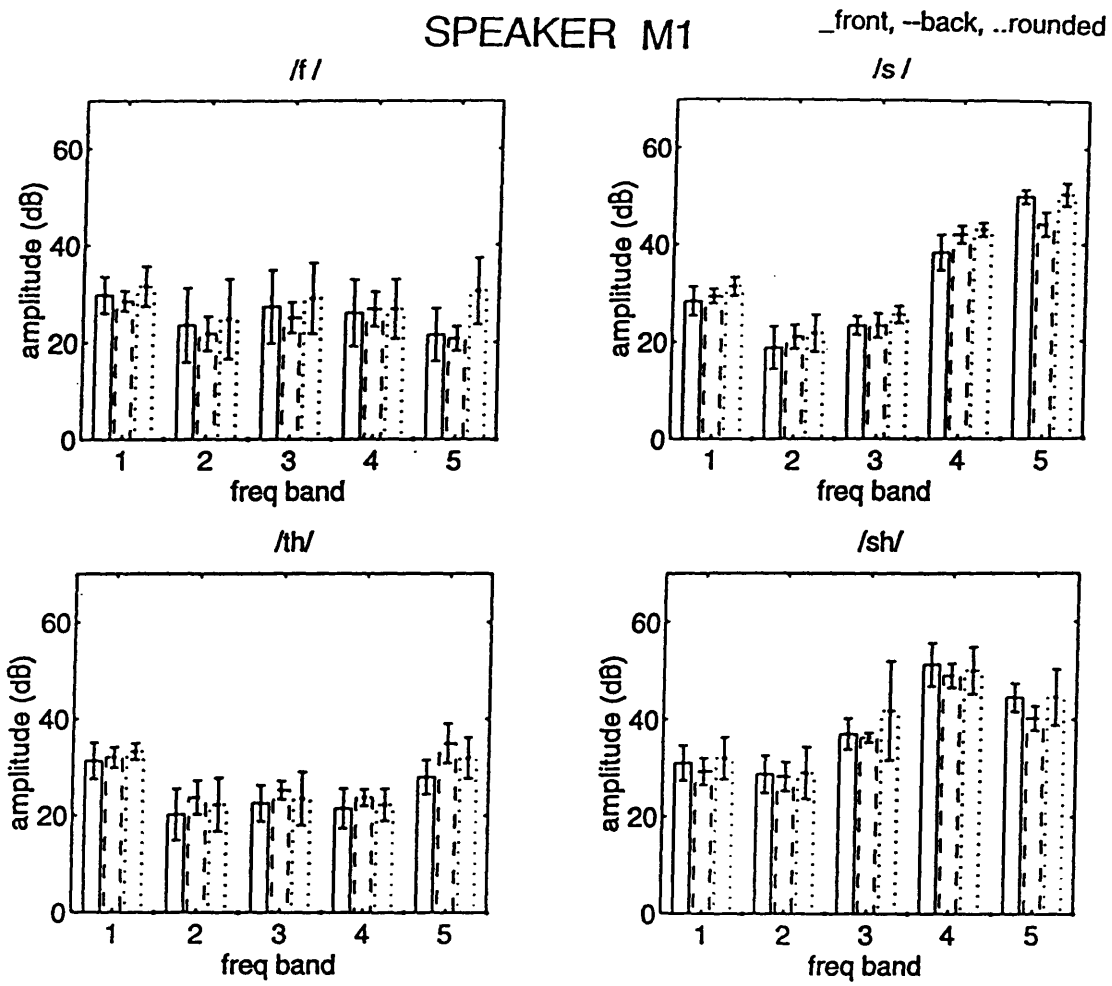

Figure 4.10: An example of plotting the output of the system for the voiceless fricatives for one male subject (M1). In this plot, each bar represents an average of 4 tokens, which are each normalized by subtracting the amplitude of the first formant in the following vowel from the maximum noise amplitude in each band. Error bars indicate standard deviations. These schematized spectra are made from measures at the middle of the fricative and are averaged separately for front (solid bars), back (dashed bars) and back-rounded (dotted bars) vowel contexts. In these plots, an offset of +64 dB has been added.

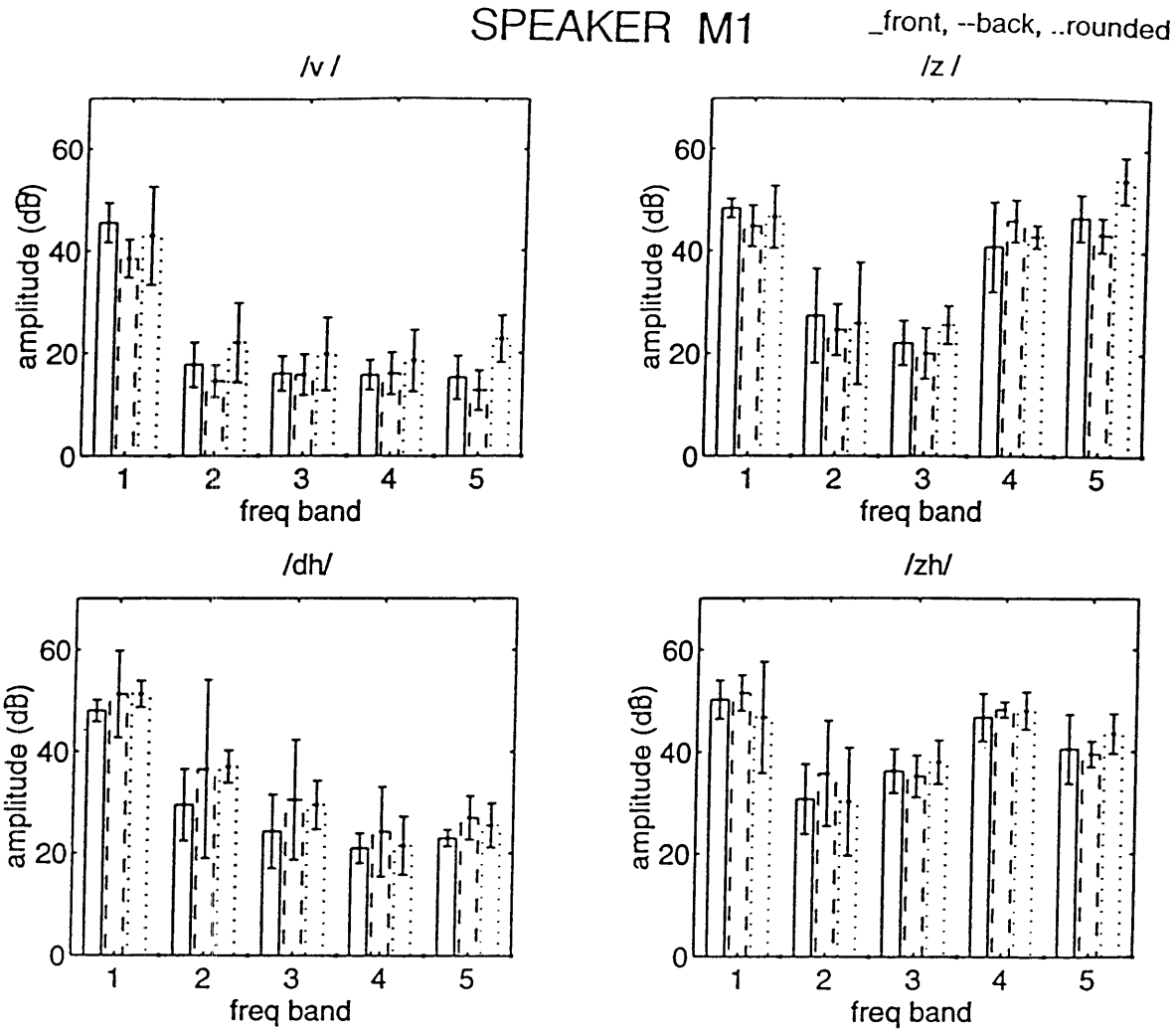

Figure 4.11: An example of plotting the output of the system for the voiced fricatives for one male subject (M1). In this plot, each bar represents an average of 4 tokens, which are each normalized by subtracting the amplitude of the first formant in the following vowel from the maximum noise amplitude in each band. Error bars indicate standard deviations. These schematized spectra are made from measures at the middle of the fricative and are averaged separately for front (solid bars), back (dashed bars) and back-rounded (dotted bars) vowel contexts. In these plots, an offset of +64 dB has been added.

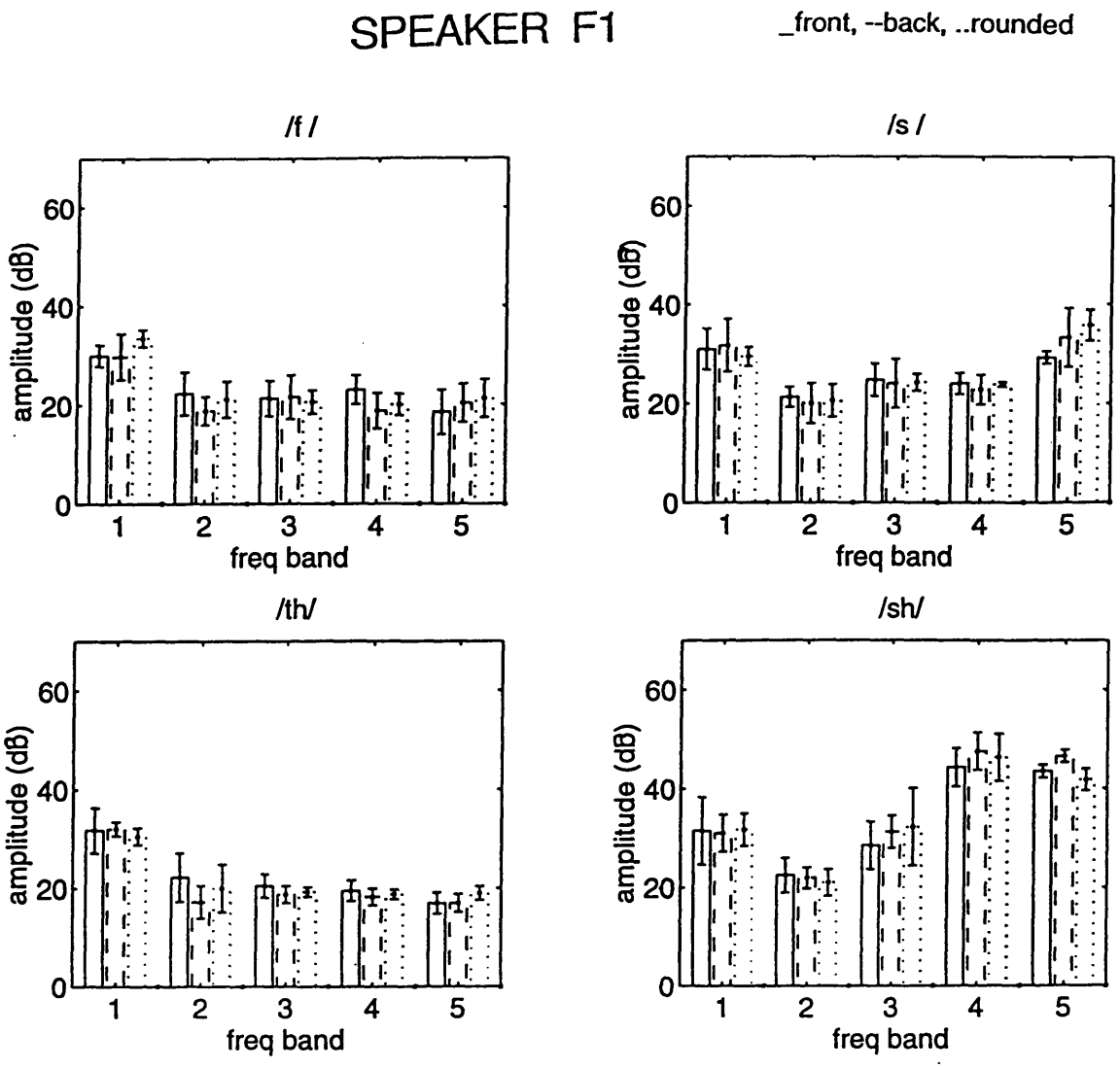

Figure 4.12: An example of plotting the output of the system for the voiceless fricatives for one female subject (F1). In this plot, each bar represents an average of 4 tokens, which are each normalized by subtracting the amplitude of the first formant in the following vowel from the maximum amplitude noise in each band. Error bars indicate standard deviations. These schematized spectra are made from measures at the middle of the fricative and are averaged separately for front (solid bars), back (dashed bars) and back-rounded (dotted bars) vowel contexts. In these plots, an offset of +64 dB has been added.

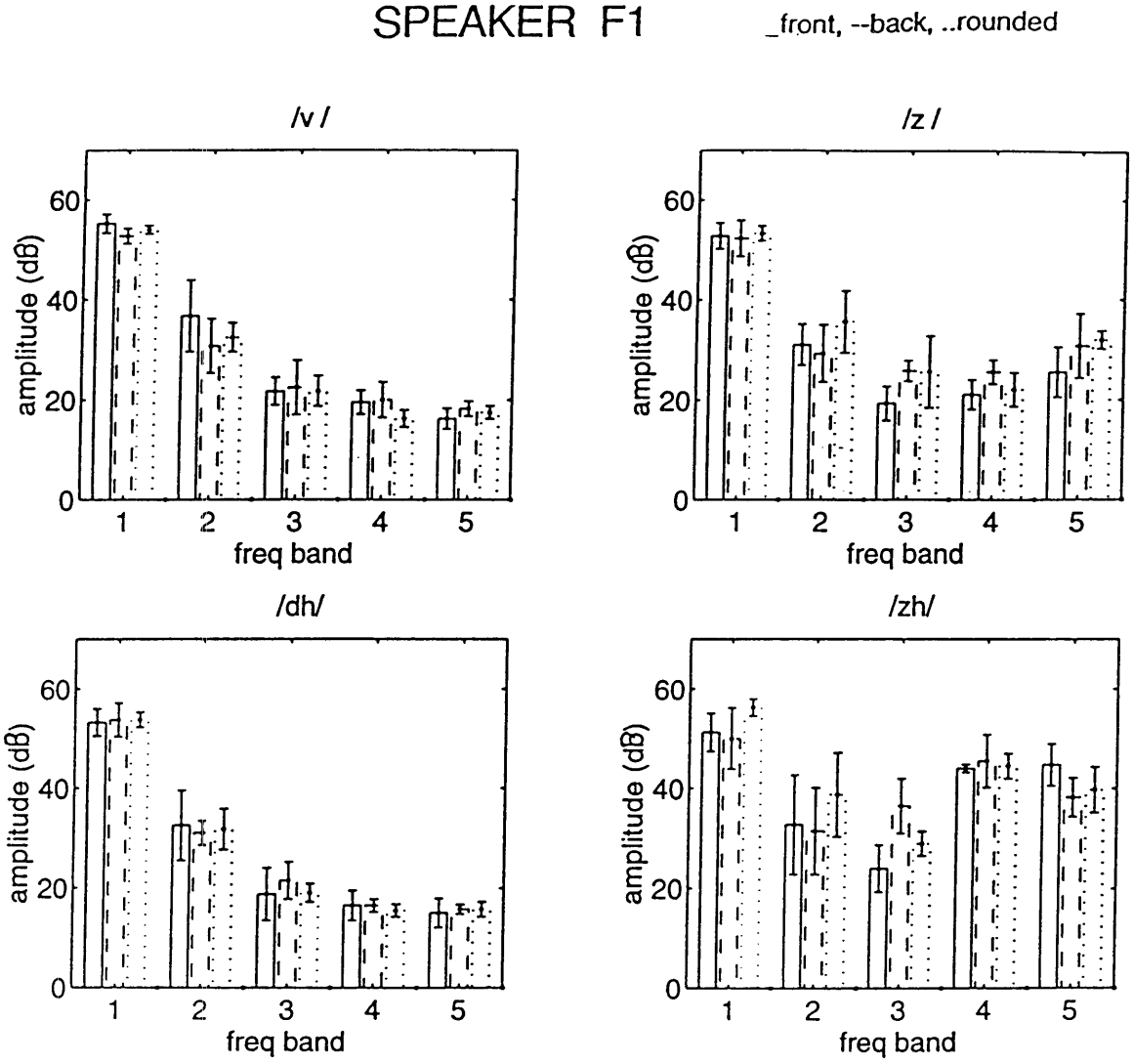

Figure 4.13: An example of plotting the output of the system for the voiced fricatives for one female subject (F1). In this plot, each bar represents an average of 4 tokens, which are each normalized by subtracting the amplitude of the first formant in the following vowel from the maximum amplitude noise in each band. Error bars indicate standard deviations. These schematized spectra are made from measures at the middle of the fricative and are averaged separately for front (solid bars), back (dashed bars) and back-rounded (dotted bars) vowel contexts. In these plots, an offset of +64 dB has been added.

Effects of Vowel Context

One effect of following vowel context that previous studies (Soli, 1981; Xu and Wilde, 1994) and the acoustic theory of fricative production would lead us to expect is the anticipatory coarticulation of lip-rounding in the back-rounded vowel context. Protruding the lips while rounding has the effect of lengthening the cavity in front of the constriction; the longer the front cavity, the lower its resonances. Therefore, we could expect lowering of front cavity resonances in fricatives that precede rounded, as opposed to unrounded, vowels. Minimal effects of lip-rounding on spectral shape were observed by Klatt (in preparation) when there was no prominent front cavity resonance. Klatt observed the following effects for strident fricative before rounded vowels: (1) high frequency intensity in /s/ was increased by approximately 6 dB and (2) the frequency of the third formant peak in palato-alveolars was slightly lowered.

Strong differences for back-rounded vowel contexts are only occasionally observed in the normalized amplitudes in Figures 4.10 and 4.11. This rounding effect shows up for speaker M1 for the labiodental and alveolar voiced and voiceless fricatives (/f, v, s, z/) as higher amplitudes in Band 5 for back-rounded as compared to back-unrounded contexts. Note that the weakest normalized Band 5 amplitudes for /s/ are observed for the back-unrounded context. The strength of coarticulatory lip-rounding appears to be speaker-dependent. Speaker F2 (not shown in the figures) was the only other speaker to show lip-rounding effects in the middle of the fricative, and then only for /z/.

The lack of robustness of vowel context effects in the above results is not altogether surprising, given the nature of the present analysis discussed thus far. The methodology of picking out amplitude peaks in broad frequency bands blurs the frequency distinction, and we cannot expect to track a within-band decrease in frequency. What we do see is occasional evidence of an increase in peak amplitude; this is consistent with the expected increase in amplitude associated with lowering the frequency of the front cavity resonance.

Xu and Wilde (1994) showed striking, preliminary results for average spectra of Mandarin palatal fricatives before rounded and unrounded vowels repeated 10 times

by five speakers. That study, unlike this one, was specifically designed to investigate the effect of anticipatory coarticulation of rounding. The effect of coarticulation with the following rounded vowel /y/ (vs. unrounded /i/) could be clearly visualized: (1) peaks started at lower frequencies before rounded vowels and (2) peak frequencies decreased over time.

Finally, we have begun by concentrating on these spectral characteristics in the middle of the fricative. We would expect coarticulatory effects of following vowel context to become more evident as we sample at times closer to the fricative-vowel boundary. We will now look at the results of measuring how the noise amplitude varies over the fricative duration.

4.4.2 Quantifying Time-varying Noise

The magnitude of amplitude peaks in the five bands over time was used to quantify changes in the noise spectra within the consonant. The relevant questions are similar to those that will be posed for quantifying stridency: How big a change and in which frequency regions? We can compare the amplitude variations for the octave bands by revisiting plots of amplitude peak output in the five bands that were shown in Figure 4.7, in which the outputs were aligned with the spectrogram of the corresponding utterance (/əfɛ/). Amplitude variations for utterances containing the consonants /f/ and /s/ in the contexts /əCɛ/ and /əCɑ/ are shown in Figure 4.14. It is evident that the amplitude of these labiodental fricatives increases from the middle (left vertical line) to the fricative-vowel boundary (right vertical line) in the middle frequency bands, and the amplitude of the alveolar fricatives is weaker at the edges than in the middle in the highest frequency bands. It is not uncommon to find brief energy minimums in several bands in the vicinity of the fricative-vowel boundary. These minimums, which can be observed for all four utterances in Figure 4.14, presumably reflect the relative timing, or mistiming, between the vocal sources. One clear trend, reflected in Band 1 trajectories, is the tendency for the voicing source to turn off relatively gradually (reflected by the decline before 100 ms) and turn on more abruptly (reflected by the increase after 200 ms).

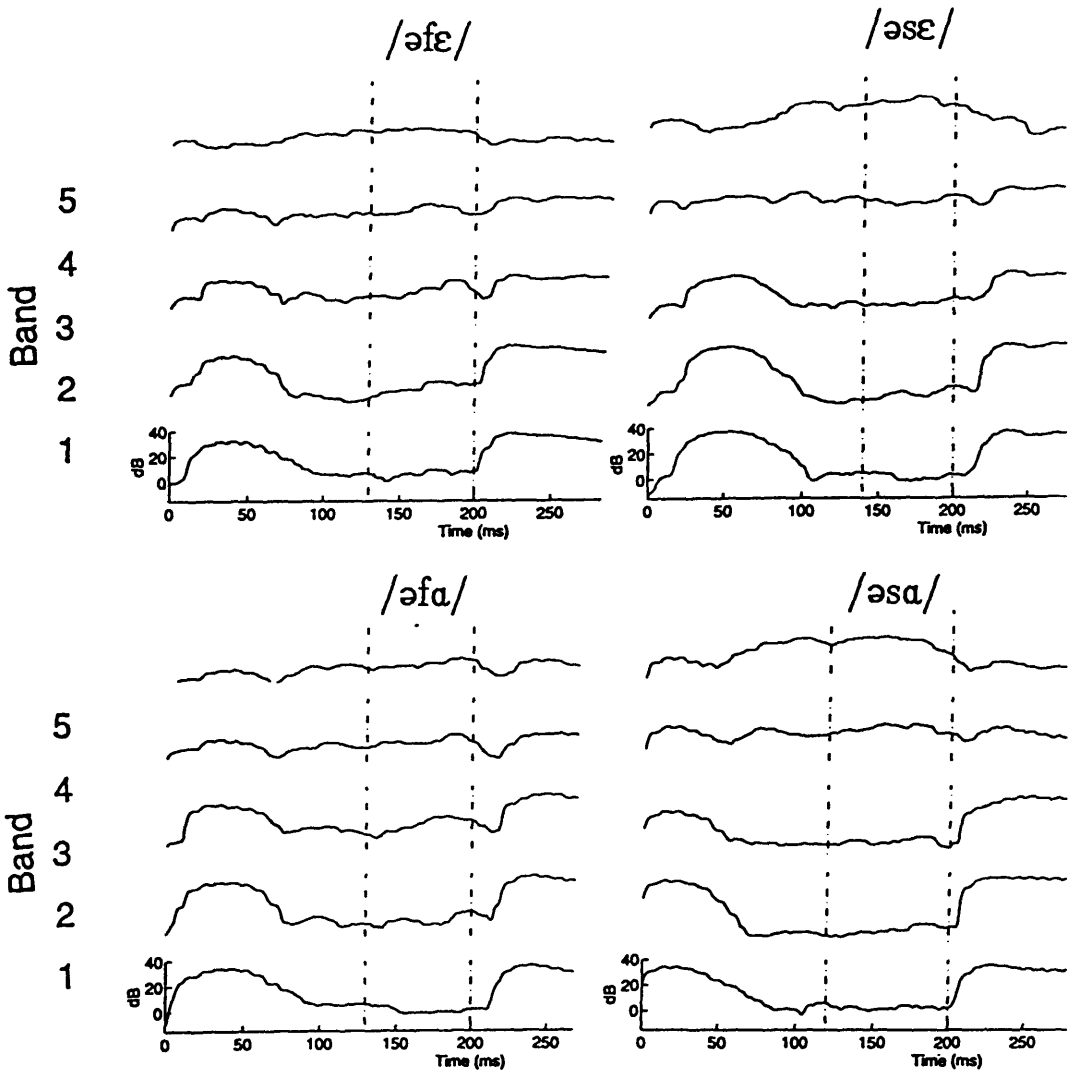

Figure 4.14: The amplitude variations for the five frequency bands, shown on the same relative dB scale. The right vertical line marks the fricative-vowel boundary; the left vertical line marks the temporal midpoint of the fricative. These tracings are the system output for utterances containing the consonants /f/ and /s/ in the contexts /əCɛ/ and /əCɑ/ produced by speaker M1.

A measure of spectral variation over time was calculated by subtracting the amplitude value at the right edge of the fricative ($relative\ time = CV - 20$) from the amplitude value at the temporal center of the fricative ($relative\ time = CV - duration/2$). For example, a negative difference means that the amplitude at the edge is greater than the amplitude at the midpoint. The following results are reported for the voiceless fricatives, in order to restrict our discussion to utterances for which the CV landmark could be accurately identified to within 4 ms. The results for the three highest frequency bands are shown for all four speakers in Figures 4.15, 4.16 and 4.17. It should be recalled that all results for the highest band for speaker M2 are limited to frequencies up to 4.8 kHz, as that was the cutoff filter used by Klatt (in preparation). The main, not unexpected finding is that there is considerable variation in noise spectra over time. That is, the noise amplitude is not constant and, from the interquartile ranges of all subjects, appears to vary from about -13 to +8 dB over the interval from the fricative midpoint to just before the fricative-vowel boundary. The individual results for each band suggest a trend for differences between the nonstrident and strident fricatives. For Band 3 (1-2 kHz) the clear trend is that there is greater amplitude difference for the nonstrident fricatives. Band 4 (2-4 kHz) shows the same trend, although the ranges are more similar. For all fricatives in these frequency ranges (1-4 kHz), the differences are negative, i.e., the edge is stronger than the middle. However, for the highest frequencies in Band 5 (4-8 kHz) there is a contrast in trends between the nonstridents, which are clearly negative, and the stridents, which are clearly positive. The tendency for the strident consonants is that the highest frequencies are more intense in the middle of the fricative. This finding is consistent with the LPC analysis of voiceless fricatives preceding five vowels by Behrens and Blumstein (1988a), who found a tendency for high-frequency peaks to appear more often in the midpoint of a fricative than in the initial or final 15 ms. The temporal midpoint of the fricative is when we might expect cross-sectional area of the constriction to be at its minimum.

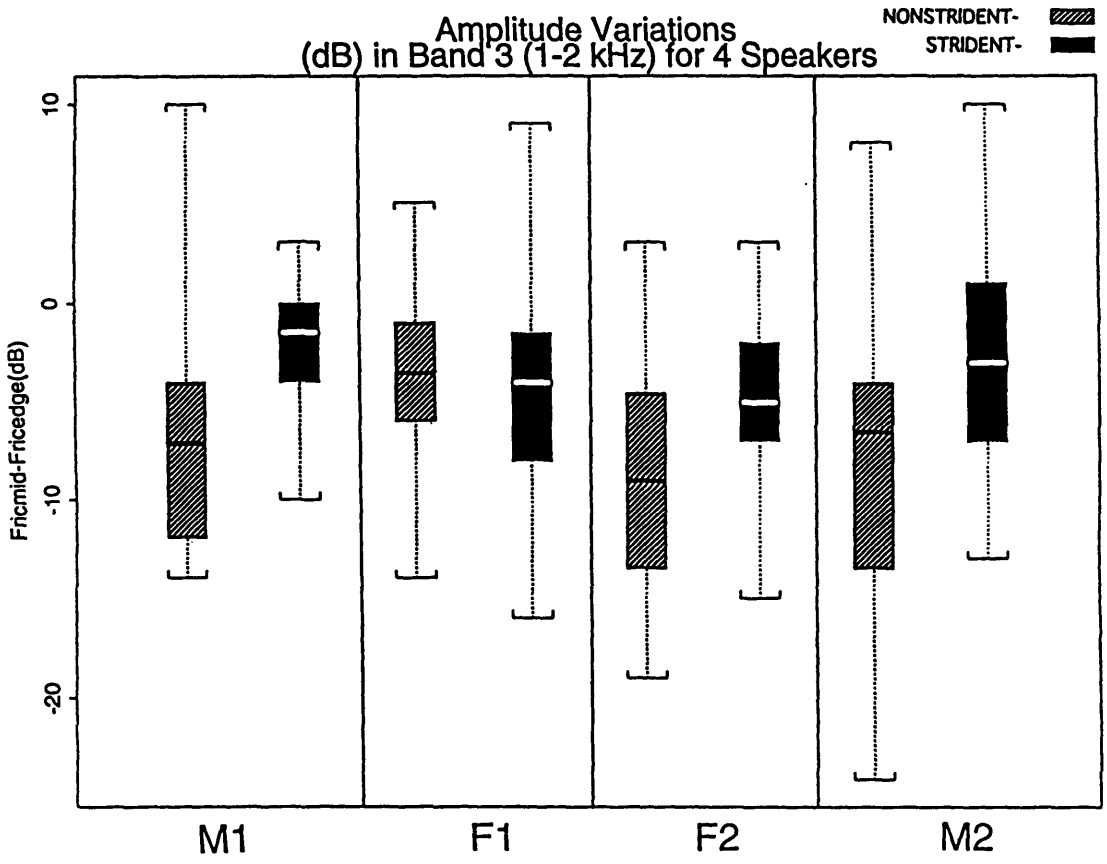

Figure 4.15: The medians (lines) and interquartiles ranges (IQR=box height) illustrate the magnitude of amplitude variations (dB) in Band 3. Each box represents 24 data points (two repetitions of two voiceless fricatives before each of six vowels), calculated by subtracting the amplitude at the right edge ($CV - 20$) from the amplitude in the middle ($CV - duration/2$), contrasted for the nonstrident vs. strident voiceless fricatives and shown separately for each speaker. The whiskers (the dotted lines extending from the top and bottom of the box) extend to the extreme values of the data or a distance 1.5xIQR from the center, whichever is less.

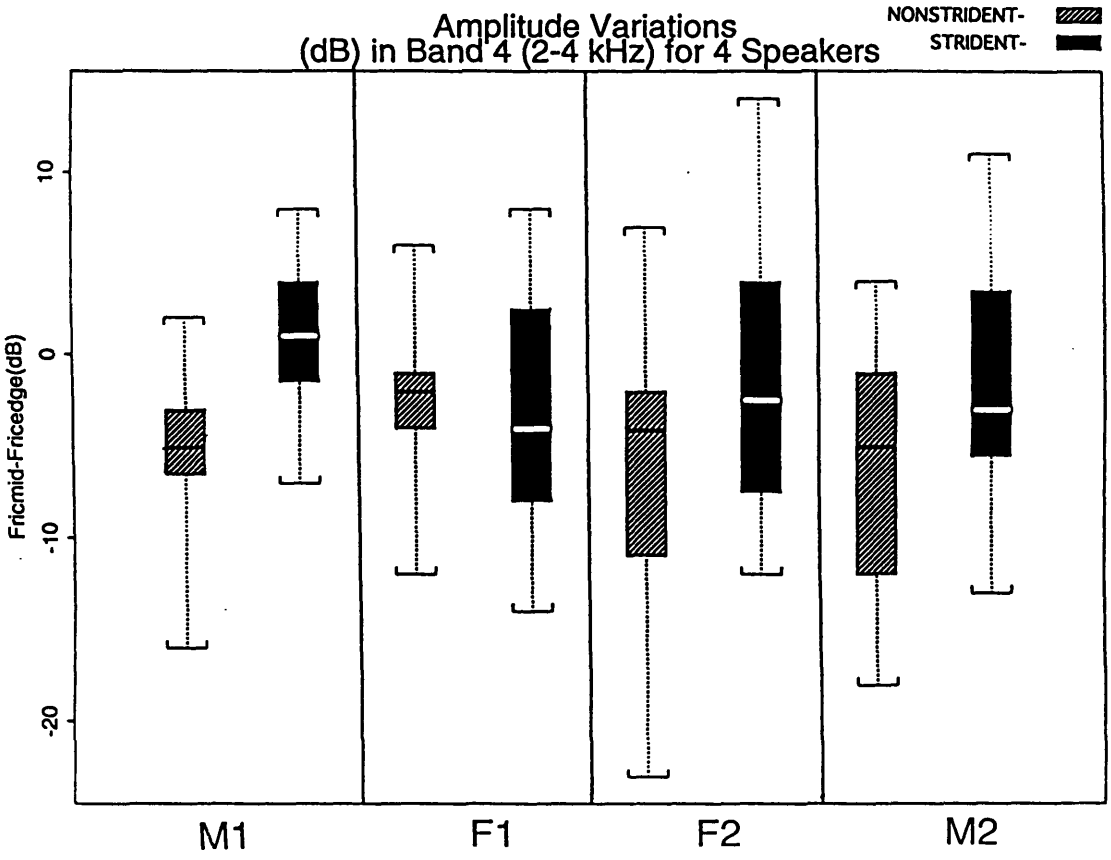

Figure 4.16: The medians (lines) and interquartiles ranges (IQR=box height) illustrate the magnitude of amplitude variations (dB) in Band 4. Each box represents 24 data points (two repetitions of two voiceless fricatives before each of six vowels), calculated by subtracting the amplitude at the right edge ($CV - 20$) from the amplitude in the middle ($CV - duration/2$), contrasted for the nonstrident vs. strident voiceless fricatives and shown separately for each speaker. The whiskers (the dotted lines extending from the top and bottom of the box) extend to the extreme values of the data or a distance 1.5xIQR from the center, whichever is less.

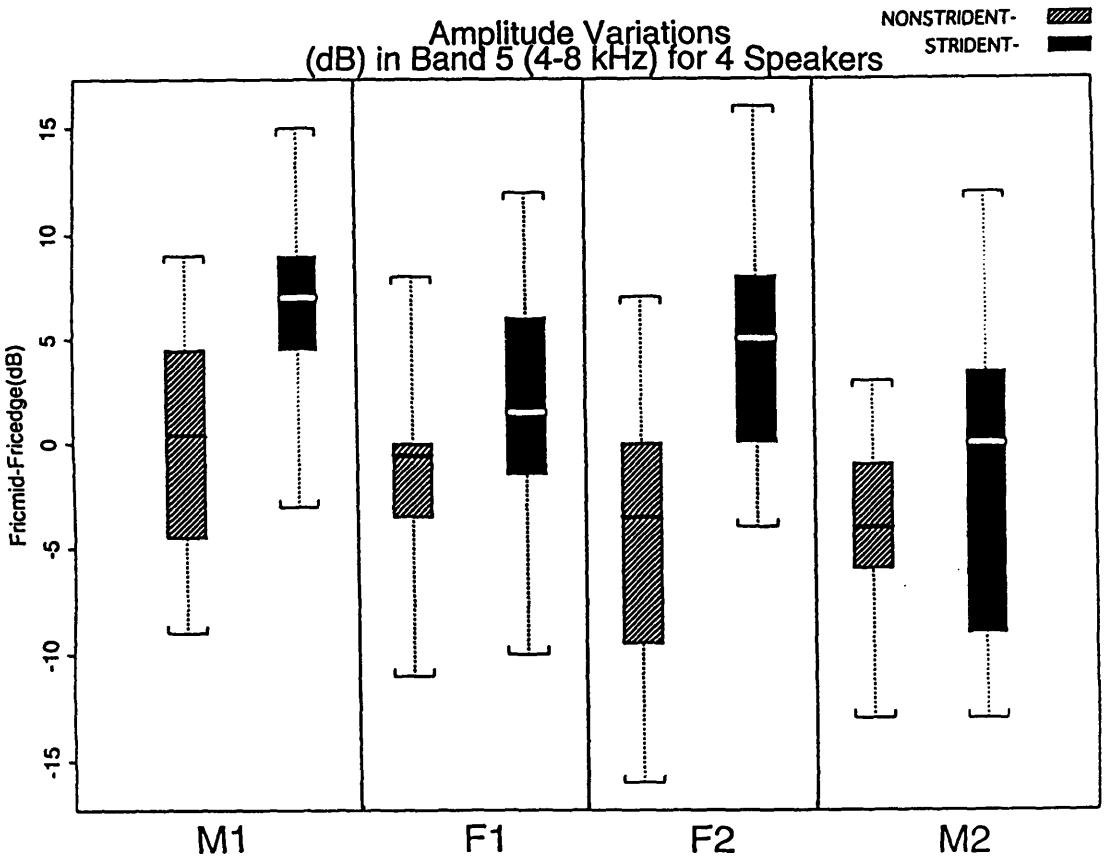

Figure 4.17: The medians (lines) and interquartiles ranges (IQR=box height) illustrate the magnitude of amplitude variations (dB) in Band 5. Each box represents 24 data points (two repetitions of two voiceless fricatives before each of six vowels), calculated by subtracting the amplitude at the right edge ($CV - 20$) from the amplitude in the middle ($CV - duration/2$), contrasted for the nonstrident vs. strident voiceless fricatives and shown separately for each speaker. The whiskers (the dotted lines extending from the top and bottom of the box) extend to the extreme values of the data or a distance 1.5xIQR from the center, whichever is less.

4.4.3 Quantifying Stridency

An approach for quantifying the feature [stridency] was developed by refining the application of the A_1 normalization procedure, which subtracts the amplitude of F1 (A_1), to time-varying measures of the high frequency content of the fricative. Figure 4.18 shows the average values for each speaker, obtained by subtracting the F1 amplitude normalization factor from the maximum amplitude peak above 2000 Hz in the consonant (i.e., maximum (Band4 amplitude, Band5 amplitude). Plots in Figure 4.18(a) represent consonant measures taken at the fricative midpoint ($CV - duration/2$). Results for M1 and F1 can be derived from the normalized amplitudes previously shown in Figures 4.10 through 4.13, neglecting the offset. In Figure 4.18(b), all measures are made in the vicinity of the CV boundary ($CV + 20$ for the vowel and $CV - 20$ for the fricative) in order to compare the amplitudes between the fricative and the vowel at times when the vocal tract is most similar in shape. The more negative the results, the more they signify a relatively weaker consonant. As expected, the weak fricatives are well-separated from the strong fricatives. For example, the mean amplitude differences between /θ/ and /s/, which have the closest relative location of supraglottal constrictions, range from 13.7 to 19.9 dB for measures made at fricative midpoint and from 12 to 21.5 dB for measures made at the right edge of the fricative.

We can also compare these normalized amplitudes averaged separately for the weak voiceless fricatives (/f, θ/) and for the strong voiceless fricatives (/s, š/), which we will now call nonstrident and strident fricatives, respectively. When we compare measures taken near the CV boundary, we can consider the relative contributions of the vocal tract sources and filters. In the vicinity of this landmark, the vocal sources of frication, aspiration and voicing are turning on and off, but the vocal tract shape does not change very much. The average normalized amplitudes for nonstrident and strident fricatives, measured at the edge of the fricative (at $relative\ time = CV - 20$) and normalized with respect to A_1 amplitude (at relative time = $CV + 20$) are shown in Table 4.1. The difference between the grand average means for nonstrident and strident fricatives, computed as the average of the means of individual subjects, is 17 dB.

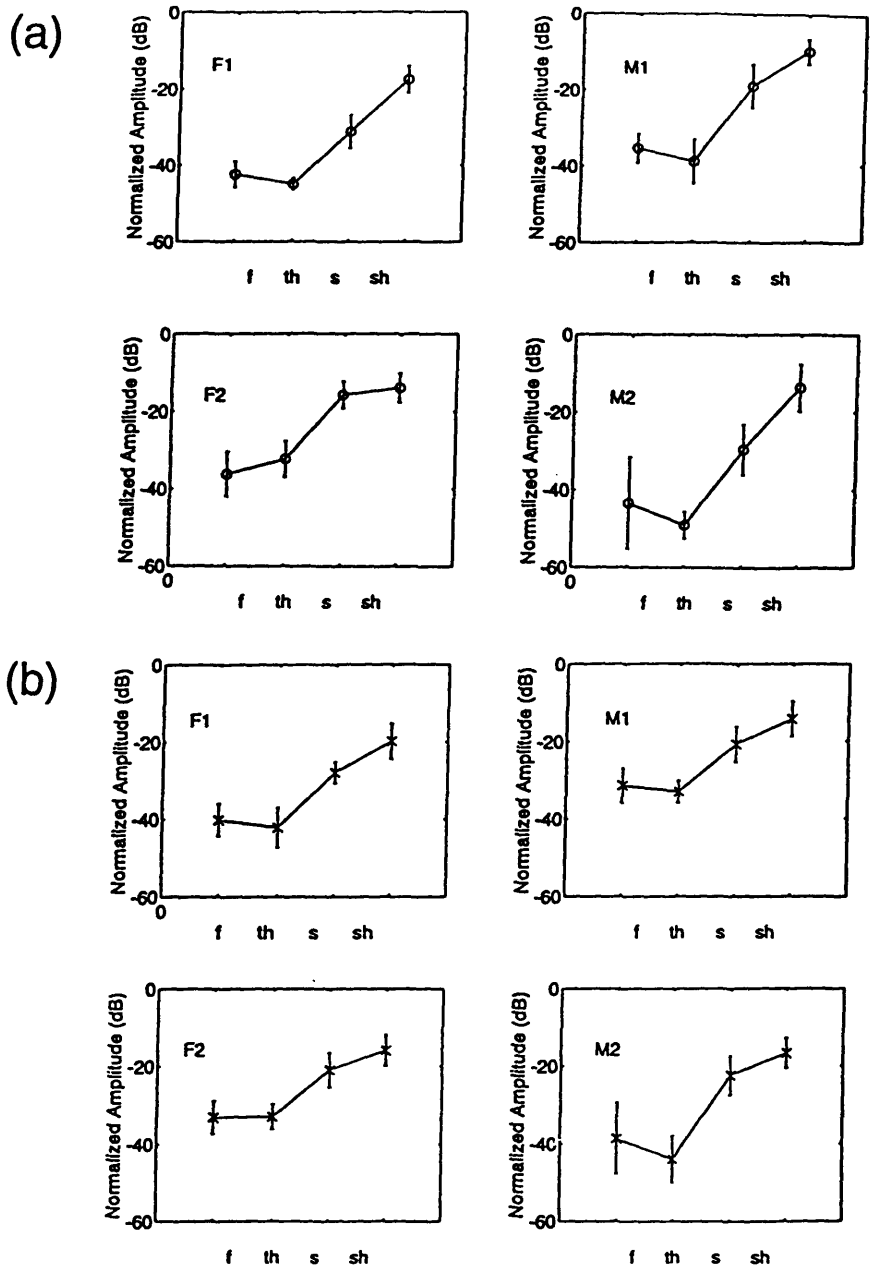

Figure 4.18: Means and standard deviations of normalized amplitudes contrasted between nonstrident (labiodental /f/ and dental /th/) and strident (alveolar /s/ and palatoalveolar /sh/) fricatives in English. All values are normalized by subtracting the amplitude of F1 in the vowel (*relative time* = $CV + 20$) from the maximum amplitude peak above 2000 Hz ((a) circle: measured at fricative midpoint; (b) x: measured at fricative edge (*relative time* = $CV - 20$)).

Table 4.1: Means and standard deviations of normalized amplitudes (dB) for the nonstrident and strident fricatives for all four subjects. The normalized amplitudes were found by subtracting the amplitude of the first formant (at *relative time* = $CV + 20$) from the maximum peak above 2000 Hz (at *relative time* = $CV - 20$).

Normalized Amplitudes (dB) for Strident vs. Nonstrident Fricatives			
Speaker		Means	Standard Deviations
F1	Nonstrident	-41	4.74
	Strident	-23	5.58
F2	Nonstrident	-31	3.88
	Strident	-16	5.46
M1	Nonstrident	-33	3.71
	Strident	-18	4.93
M2	Nonstrident	-41	8.11
	Strident	-19	5.31
Grand Average	Nonstrident	-36	7.07
	Strident	-19	5.90

4.5 Summary and Interpretation of Results

In describing and quantifying the time variations in English fricatives, a goal was to reduce the dimensionality of the data. We chose to study the peak amplitudes in octave frequency bands of time-averaged speech. For our purposes, which included applying our results to the synthesis of more natural vowel-fricative-vowel utterances, we always considered relative amplitudes. We chose normalization with respect to F1 amplitude in the following vowel, because it maintained gross spectral characteristics consistent with the place of the supraglottal constriction.

This series of studies did not highlight differences with respect to the following vowel. Our observations of a tendency for Band 5 (4-8 kHz) intensity to increase for alveolars before rounded vowels are consistent with Klatt's (in preparation) observations of lip-rounding effects on high-frequency amplitude increases for /s/. We would not expect to directly detect within-band frequency changes, such as the slight lowering in third formant peak frequencies for palato-alveolars, which Klatt observed. Finally, we would expect that coarticulation effects, such as anticipatory rounding, would be greater in running speech, where durations are shorter.

In quantifying the observed time-variations, we focused on finding answers to the following questions: How big a change and in which frequency regions. These questions were complicated by the nature of the random noise source in fricative generation. Fant (in press) provided a way of estimating the error in noise spectra from empirical data. According to the parameters we used for time averaging, the random amplitude variations would be on the order of 1-2 dB. We often observed considerably more amplitude change between the fricative midpoint to the fricative-vowel boundary, in the range of -12 to +7 dB. Presumably, these greater variations also reflect movements of the major articulators in forming and releasing the supraglottal constriction.

Our calculations of the amplitude variations in selected frequency bands for voiceless fricatives yield several interpretations for the considerable variability. The findings obtained in the 1-2 and 2-4 kHz ranges are consistent with the presence of

aspiration in the vicinity of the fricative-vowel boundary. It should be noted that back cavity excitation can also reflect incomplete pole-zero cancellation which can occur when there is coupling between the front and back cavities. Often, there is a short (less than 20 ms) gap, where neither the frication nor aspiration noise is very strong. Presumably this reflects the mistiming between turning off the noise source for the fricative and turning on the voicing source for the following vowel. That is, as the supraglottal constriction size increases, but the glottal opening remains relatively large, there is a period of time with no strong excitation. In addition, trends for the voicing source in the preceding vowel to turn off gradually as the closure is made for a voiceless fricative and turn on abruptly at the release into the following vowel have been observed before (Klatt, in preparation) but may not have been incorporated into strategies for synthesizing fricatives.

Significant spectrum amplitude differences were observed at high frequencies between the nonstrident and strident fricatives. For the strident fricatives, the highest frequencies are strongest in the middle of the consonant, when the area of the supraglottal constriction may reach its minimum. The nonstrident fricatives in English show greater overall variability in amplitude than the stridents. Results of Utman and Blumstein (1994) for English and Ewe, in which the feature [strident] serves to contrast the labiodental and bilabial fricative, suggest that the realization of an acoustic property is influenced by the linguistic role its associated feature plays in a particular language's sound inventory. They found that the most optimal acoustic measure corresponding to the feature [strident] was the relative amplitude of the fricative noise in the 6-10 kHz (vs. 1-10 kHz) region, measured at 51.2 ms (vs. 25.6 ms) before vowel onset. The ranges of their normalized amplitudes, calculated as the amplitude difference between the frication noise and vowel segment, were as follows: greater than 21 dB for the bilabial fricatives of Ewe, from 3-21 dB for labiodental fricatives and below 3 dB for the alveolar fricatives. Their finding was that the range of values corresponding to the feature [strident] remained stable across the two languages, but the phonetic contrast between bilabial and labiodental fricatives in Ewe was enhanced , i.e., a skewed distribution toward the higher range for [f, v] was found

in Ewe as compared to English. In general, their interpretations of findings is consistent with what we expect: amplitude differences show up in the "higher" frequencies, and the smaller the front cavity, the higher the resonant frequencies.

It is non-trivial to map our current acoustic findings into production models, as discussed in Chapter 2. The average of the means between nonstrident and strident fricatives, shown for the four subjects in Table 4.1, was 17 dB. In accordance with the acoustic theory of speech production (Fant, 1960), we can attempt to interpret our results in terms of the sources and the filtering of the sources. In our present acoustic findings, the high-frequency amplitude differences between nonstrident /θ/, which has essentially no front cavity, and the strident /s/ was found to be approximately 12-18 dB. In Chapter 2, the front cavity was calculated to contribute approximately 15-21 dB to the output spectrum for the strident fricative /s/. Incorporating the effect of the zeros, which are affected by the position of the source, yielded a range of 12-18 dB in the amplitude of the transfer function for /s/. Therefore, what's the contribution of the efficiency of the noise sources between /θ/ and /s/? The findings of Stevens (1971) suggested a source effect of approximately 5-10 dB. Our results leave a smaller margin for a source effect.

The spectral differences between nonstrident and strident fricatives suggest that models of filtering of the noise source by the cavity in front of the constriction could be improved if the losses in the vocal tract were better represented, and if better estimates could be made of the source location. It is currently difficult to specify how much of the increased high-frequency noise output for stridents is the result of a more effective obstacle in the air stream and how much is the result of high-frequency emphasis due to filtering effects of the vocal tract. Source-tract interactions are inherent in the generation of turbulence noise for fricative consonants; source properties are a direct result of this interaction between the shape of the vocal tract and the airflow through it (Stevens, 1987). Turbulence noise used to signal a phonetic distinction in speech primarily involves turbulence generated at an obstacle or surface (Stevens, in preparation). Shadle (1990) suggests that there is a continuum between an obstacle and no-obstacle case. The expected magnitudes of effects discussed in Chapter 2 lead

us to conclude that the filtering effects can dominate.

In this chapter, we have shown a way to measure movement over time in the spectral characteristics of fricatives. However, it is left to be determined whether these variations have perceptual significance. In the next chapter, we present how we used our acoustic results in setting the time-varying controls in a speech synthesizer. Listening tests with utterances containing synthetic fricatives were used to help determine which acoustic aspects play a role in our perception of naturalness.

Chapter 5

Applying Acoustic Findings to Speech Synthesis and Perceptual Evaluation

We assessed the effect of the observed time-variations in noise over the duration of a fricative on the perceived naturalness of synthetic fricatives. The main objectives were to determine which aspects of the inherently noisy speech signal contribute to fricatives that are preferred by listeners, and which aspects can be ignored.

5.1 Introduction

Estimation of the saliency of cues requires an appropriate stimulus inventory. Successful imitation of a natural utterance has been shown to be possible through matching observed short-term spectra. Klatt notes:

> The speech copying techniques described earlier succeed, in part, because they reproduce essentially all of the potential cues present in the waveform or spectrum, even though we may not know which cues are most important to the human listener. A synthesis-by-rule program, on the other hand, constitutes a set of rules for generating what are often highly stylized and simplified approximations to natural speech. As such, the rules are an

> embodiment of a theory as to exactly which cues are important for each phonetic contrast (Klatt, 1987, p. 752).

We will attempt to model the speech waveform in the acoustic domain, using a formant synthesizer. This work does not purport to be a study in articulatory synthesis, i.e., it does not attempt to model the mechanical motions of the articulators or the shapes of the vocal tract. However, the temporal considerations in the model being developed by Scully et al. (1992) are relevant. In addition, consideration of physiologic measures, such as those described in Chapter 2, provide additional data for relating acoustic features to articulatory events. For perception, explicit knowledge of the production mechanism may not matter (Fant, 1960).

5.2 Baseline Perceptual Testing Using Natural Utterances

Identification tests were run to establish a baseline measure of the intelligibility of natural fricative-vowel tokens, according to a protocol established by Klatt (unpublished manuscript). The listening tests were designed to evaluate the intelligibility of the eight English fricatives before each of the 12 English vowels. Each stimulus consisted of the first C and 2/3 of the first V of a CVCVCVC in the Klatt database for one male speaker (10 kHz sampling rate). The initial vowel was edited in order to remove the evidence of transitions to the following fricative, but maintain natural variations in vowel duration. Separate identification tests for voiced and voiceless fricatives were prepared for each set of the 48 CV waveforms. Each taped test consisted of randomized presentations of three repetitions of each individual token. Each test was played by earphones in a sound-treated room to five subjects who were experienced in speech testing and transcription.

As expected, this was a very easy listening task for the subjects. All of the strident alveolar and palatal fricatives /s, š/ and /z, ž/ were identified correctly (0 errors out of a total of 360 responses on each of the voiceless and voiced tests). The

only confusions were between the labiodental and dental tokens. There were more errors for the voiceless fricatives (38 errors out of 360 responses) than for the voiced (4 errors out of 360 responses) in the separate tests. The possibility of confusion across the voicing feature (e.g., /v/ vs. /f/) was not assessed in the forced choice format.

The pattern of errors according to vowel context for the voiceless fricatives illustrated the highest percentage of errors for the low and mid back-vowel context (i.e., 55.5% of these CV's containing /f/ were misidentified as /θ/). These results for the misidentified stimuli can be partially explained by the observation of limited formant transition information, which has been shown to be important for discriminating among fricatives when strong spectral cues are absent (Harris, 1958). Additional testing with natural speech, for example using /f, v, θ, ð/ to probe for confusion across the voicing feature, may be warranted.

5.3 Synthesis Methods

Earlier attempts at fricative synthesis in this project involved a hybrid approach. First time-varying parameters, such as formant specifications, were schematized and implemented. Then an iterative spectral match was made until the noise spectra matched the appropriate grand-averaged spectra (Klatt, unpublished manuscript) for the desired place of articulation. The current study reversed the process by starting with speech-copying in order to avoid adopting generalizations that could lead to nonconvergence of test and target stimuli.

The target signals were VCV utterances spoken by one male speaker (M1), that were excised from the 'CVCV'CVC nonsense in the recorded database described in the previous analysis chapters. A variety of speech-copying techniques were used to synthesize /əfɛ/, /əfɑ/, /əsɛ/ and /əsɑ/.

The non-reduced vowel contexts for synthesis were chosen with consideration of results from the baseline perceptual identification test. The vowel /ɑ/ was selected in view of the relatively high percentage of errors observed for the weak fricatives in low and mid back-vowel context. A front vowel context was included for comparison.

A relatively open front-vowel, such as /ɛ/, was used to avoid the complication of additional noise that may be generated during the more constricted vocal tract configuration for the high front vowel /i/. Both /ɑ/ and /ɛ/ are unconstricted enough to minimize noise generation due to airflow through the vowel configuration.

We will summarize the general speech synthesis strategy and illustrate it with an example. All synthesis was accomplished using KLSYN93 (Klatt, unpublished manuscript), an experimental version of the formant synthesizer developed by Dennis Klatt. Readers are referred to Klatt and Klatt (1990) for further details on the Klatt synthesizer.

5.3.1 General Strategy for Formant Synthesis

Since we were synthesizing a male voice, and we wished to provide a reasonable approximation to the fricative spectra, we decided to use six formants, the maximum number available on KLSYN93. The average spacing between formants for a male speaker is 1 kHz. Therefore, we redigitized the natural speech to a 12048 Hz sampling rate. The constraints on the bandwidth of the signal meant that the higher formants were aliased, and this aliasing was used for the higher pole correction.

Parameters for synthesis were specified in terms of the amplitudes of the sources: 1) amplitude of voicing (AV), 2) amplitude of frication (AF) and 3) amplitude of aspiration (AH). The filtering of the fricative noise source was controlled by specifying the amplitudes (A2F-A6F) and bandwidths (B2F-B6F) of digital filters in the parallel branch, which generally correspond to front cavity resonances.

The time-varying parameters of the synthetic best match to the natural /əfɛ/ is shown in Figure 5.1. The synthesis parameter (.doc) file for this utterance and all others synthesized for this study are included in Appendix C. The following general strategy (indicated in bold) was used for copying the natural speech VCV's. A detailed overview of the strategy used in synthesizing the utterance /əfɛ/ is included.

1. **Define the total duration of the utterance.**

   The total duration of the utterance was taken to be 500 ms.

2. **Establish the pitch contour for the entire utterance by assigning values for the fundamental frequency.**

   Sampled and smoothed values output from the pitch tracker of the Klattools (using lspecto -syn), shown in Figure 5.1(a) was used to copy F0.

3. **Draw in the first three vowel formants (F1, F2, F3) leaving an interval of unspecified values for the consonant portion.**

   Sampled and smoothed values from the formant tracker of the Klattools (using lspecto -syn) were used to copy F1, F2 and F3. Formant parameter tracks are shown in Figure 5.1(b). The values of the formant onset frequencies obtained from the acoustic analysis of formant patterns for speaker M1 are included in Appendix B, and were used to anchor the formant onset values for synthesis.

4. **Establish formant values for the consonant portion to set the cascade formant parameters.**

   The values of the formants during the consonantal interval were estimated from the spectrogram and spectral slices of the natural speech. These could be further adjusted during the subsequent spectral match. Extra poles that appeared to be due to tracheal resonances were ignored in this phase.

   One reason for not achieving an ideal match was due to a limitation encountered in setting AH in KLSYN93: The spectral tilt (TL) parameter did not modify AH. Therefore it was impossible to obtain the energy needed at 1500 Hz without changing tilt of the entire spectrum. This meant that F4, F5, and F6 amplitudes would be compromised. In view of this limitation, the experimental synthesis program was later changed in order to introduce a parameter that changes AH spectrum.

   It was found that a peak near 1400 Hz could be replicated by making leaving AB constant and varying A2F over time, as shown in Figure 5.1(f).

5. **Set the time-varying controls for amplitudes of voicing (AV), amplitude of frication (AF) and amplitude of aspiration (AH).**

The relative timing and amplitudes of the periodic and aperiodic sources, as well as overall intensity, are illustrated in Figure 5.1(c), (d) and (e), respectively.

6. **Use trial and error to match the DFT spectra of the synthetic and natural speech tokens.**

   The the bandwidths of F4 and F5 were changed: B4 from 200 to 300 Hz and B5 from 200 to 400 Hz. It is not unreasonable for the bandwidth of F5 to double due to the effect of the radiation loss on the high frequencies. F5 was changed in order to match the correct spectral tilt, as previously discussed.

The following section describes how time variations observed in Chapter 4 were incorporated into the synthesis process.

5.3.2 Capturing Time-Variations in Noise Using Copy Synthesis

Two strategies for synthesizing the consonant portion of the fricatives were used as the basis for two main conditions in the subsequent perceptual tests. In the time-varying noise method, the averaged spectra at 20 ms before the CV boundary and at the temporal midpoint were used as templates for spectral matching. Unaveraged spectra taken every 10 ms were also used to help interpolate the values of the noise parameters for matching the changing spectra of the fricative. For the steady noise condition, the parameters used for the spectral match for the averaged spectra at the fricative midpoint were replicated throughout the duration of the fricative consonant. Spectrograms of the two versions of the synthetic utterance /əfɛ/, the time-varying (left) and steady (right), are shown in Figure 5.2.

Eight parameter (.doc) files, one time-varying (tv) and one steady noise (steady) for each of the four natural utterances copied, are included in Appendix C.

Since the primary focus of this study was on the fricative portion of the stimuli, in relation to the surrounding vowel context, some compromises were made in synthesizing the vowel portions. That is, vowel parameters were reused when possible.

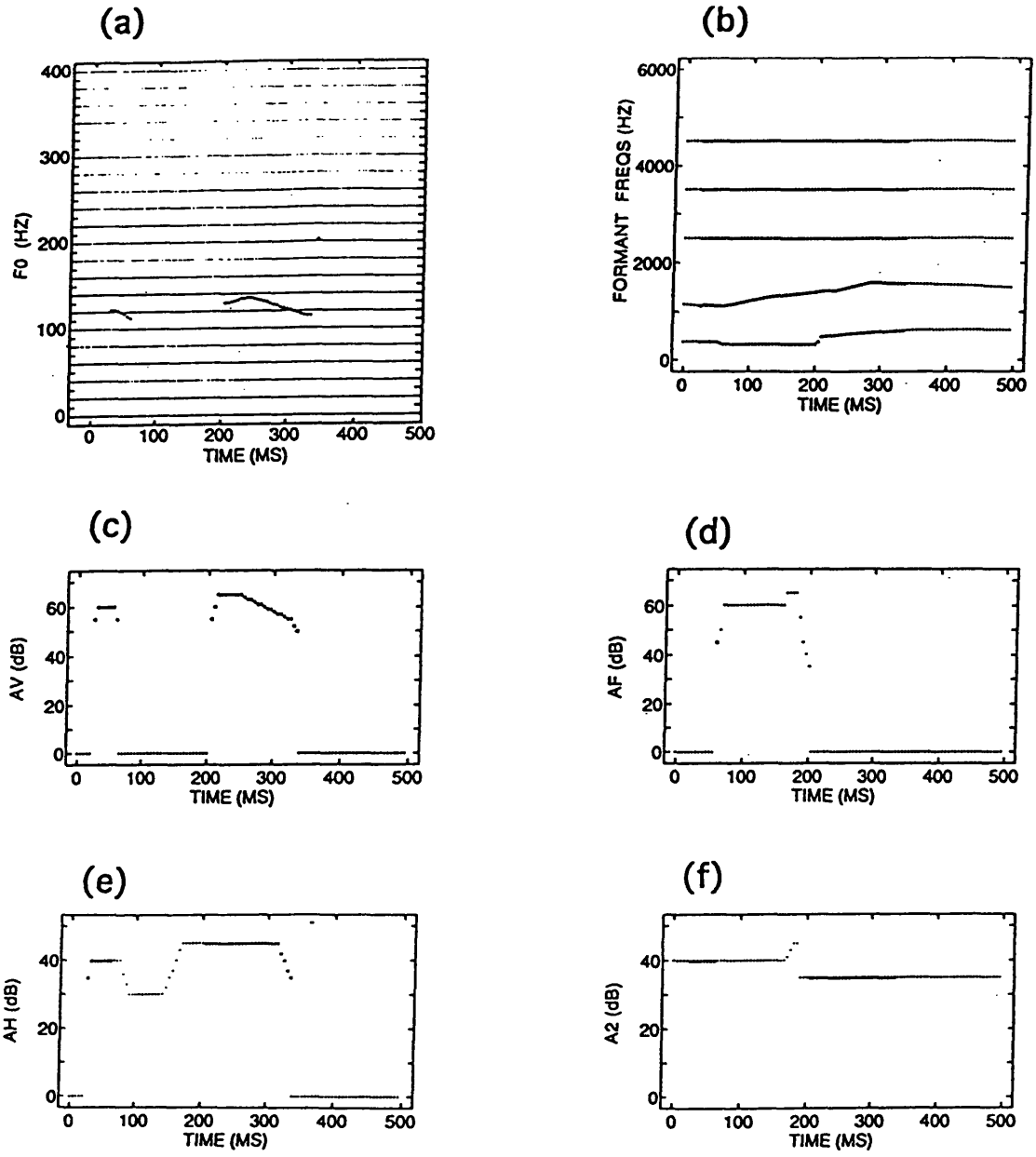

Figure 5.1: The trajectories show the settings for the parameters for synthesizing the utterance /əfɛ/ that varied over time. The bold tracks indicate when the voicing source AV is on, except in the formant tracks which are bold when any glottal source is on.

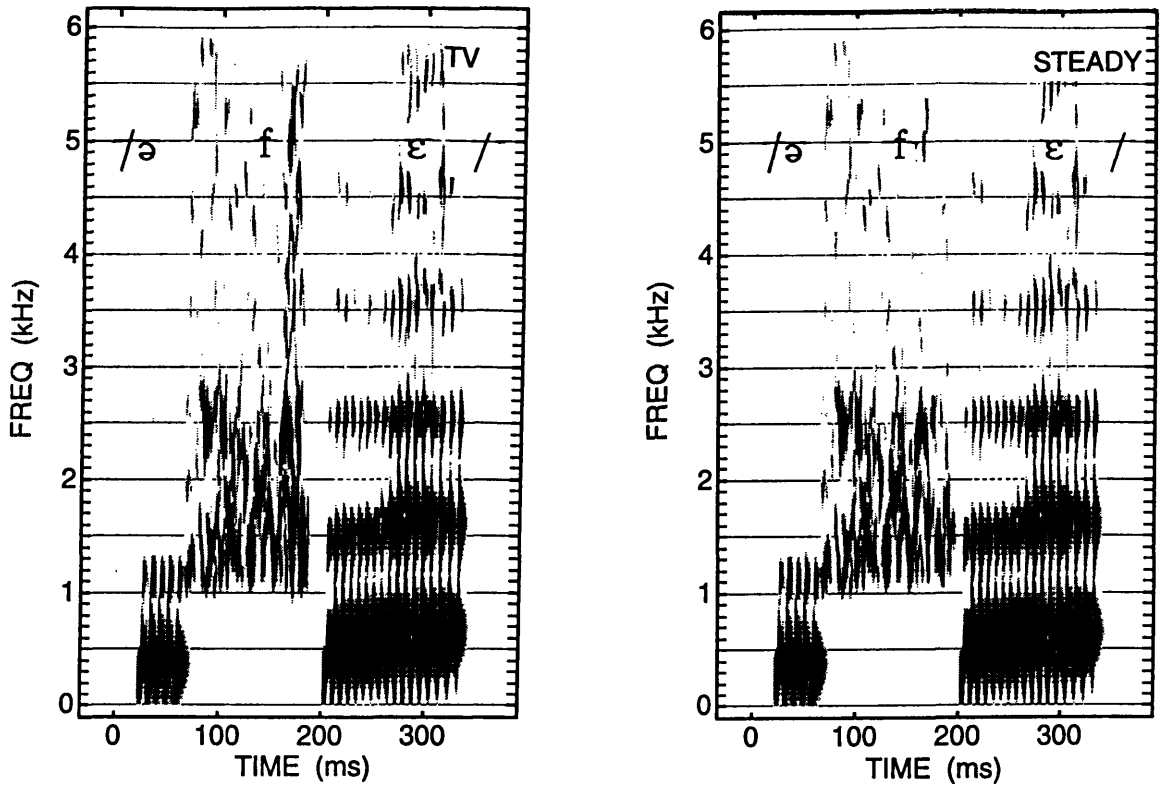

Figure 5.2: Spectrograms of synthetic /əfɛ/ are the result of the time-varying (top) and steady (bottom) synthesis methods.

The final vowels in the VCV utterances were edited to meet the following criteria: the vowels were long enough to reach maximum amplitude and short enough to avoid the transitions into the final fricative of the 'CVCV'CVC utterance. It was later determined that the /ɑ/ in the natural /əsɑ/ had been inadvertently shortened by an additional 50 ms when excised. Still, all vowel durations met the desired criteria and the VCV pairs used in the final listening tests always contained identical vowels.

We added auditory matching to the iterative process of spectral matching to provide criteria for ending the iteration. Listening tests used to evaluate the match between synthetic utterances and the natural targets were designed in order to minimize the influence of differences in the quality of the vowels. We asked several sophisticated listeners whether they could tell the difference between the natural and synthetic fricatives. In these comparisons, the natural fricative in the target utterance was excised and replaced by the synthetic copy. The programs RECORD and CONCAT (Klatt, 1984) were used to edit these utterances. These listeners reported that they could not tell which stimuli contained the natural vs. synthetic fricatives. When forced to choose, their choices were randomly split between calling the utterence with the natural or synthetic fricative the natural target. However, when presented with pairs of the entire natural and synthetic VCV's, most listeners could readily identify the natural target. These informal findings supported two main conclusions: (1) the time-varying fricative consonants were reasonably reproduced (2) natural tests that included comparison of both the vocalic and consonantal synthetic stimuli to the natural stimuli could be unfairly biased based on the quality of the match between the vowels. It was, therefore, decided to use only synthetic stimuli in the subsequent perceptual evaluations.

5.4 Perceptual Tests

The primary objective of the perceptual evaluation was to determine if faithfully replicating the time variations observed during fricative consonants would affect listeners' preferences for synthetic fricatives. The time-varying copy method is more costly in

time and effort than schematizing fricatives with current rule-based synthesis. Is this additional cost well spent?

5.4.1 Stimulus Preparation

The listening tests were designed to investigate the primary question: Can listeners perceive the effects of temporal detail in synthetic fricatives patterned after natural stimuli and do they prefer this method of synthesis? Secondly, how do amplitude changes in the noise affect these decisions? That is, how are judgments of naturalness influenced when the fricative-vowel ratio is changed by a set amount.

The reference stimuli were the results of the copy synthesis "best match", as described in the previous section. The preceding synthetic vowel was /ə/ and the following vowel was either /ɑ/ or /ɛ/. The vowel context was always identical when stimulus pairs were prepared. Only the fricative portion was modified.

Two possible methods were initially proposed for evaluating the perceptual effects of incorporating the observed time variations of natural fricatives in synthetic copies.

1. Time-varying Noise (tv): KLSYN93 noise source parameters AF and AH and the parallel noise parameters A2F-A6F for filtering the noise were varied, as described in the previous section, in order to match the time-varying noise spectra of frication and/or aspiration.

2. Steady Noise without Aspiration (steady): AF and A2F-A6F filter values were taken to be the values matched at the center of the fricative. These parameters remained unchanged during the consonantal portion. AH was replicated in a stereotypical pattern: 0 dB during the fricative and 20 dB down from AV during the vowel.

The amplitude modification in Test 1 was designed to investigate the effect of manipulating the consonant-vowel ratio. This was accomplished by generating a continuum of test stimuli which varied only in the value of the parameter GF, the overall gain scale factor for AF (amplitude of frication in dB). The reference stimuli

Table 5.1: The stimulus conditions for Test 2 to determine if listeners prefer spectral matches with time-varying (tv) or steady fricative parameters The +8 notation, as explained in the text, represents an additional frication noise gain of 8 dB.

Comparisons	Assesses
tv vs. tv+8	Amplitude changes
tv vs. steady	Method changes
steady vs. steady+8	Amplitude changes
tv+8 vs. steady+8	Method change at higher amplitude

were the synthetic VCV's designated the "best match" to the relative amplitudes between the natural consonant and the vowel in the iterative spectral comparisons. That is, for the reference stimuli, 0 dB was added to the gain factor GF. The range of GF modifications (dB) for Test 1 stimuli was as follows: -12, -8, -4, 0 ,+4, +8, +12. Continua for Test 1 were systematically modified versions of the best match to the synthetic /əfɛ/ and /əsɛ/ tokens generated by the steady method.

The Test 2 stimuli were constructed from the time-varying (tv) and steady best matches. The comparisons among conditions are shown in Table 5.1, The +8 notation indicates that the stimuli were synthesized with the frication gain parameter, GF, increased by 8 dB over the setting used for the best match. The increment value of 8 dB was chosen in view of the results of pilot perceptual tests and the results of the acoustic analysis, which showed that 8 dB was approximately one standard deviation of amplitude spread for the normalized peak amplitude measured for /s/. Therefore, this increment could be expected to allow natural variation, but still separate between strong and weak fricatives.

5.4.2 Stimuli Presentation

All tests used a two-alternative forced choice task (2AFC) task where the listener was asked to choose which token from a pair of stimuli had the better sounding fricative. This design was used in order to structure the task for the subjects. The judgments as such were made to be as subjective as possible, given the nature of the task, and the response mode avoided the need for a rating scale, as used on previous pilot tests

(Wilde and Huang, 1991).

Two separate test batteries were constructed for /f/ and /s/. For each amplitude test stimulus in Test 1, steady noise stimuli with parameter GF ranging from -12 to +12 dB in 4 dB increments, there were four repetitions of each comparison with the reference stimulus (GF=0 dB). In order to familiarize the listeners with the task of comparing synthesis methods, which involved some finer distinctions, the corresponding Test 1 was always presented before Test 2. The order of presentation of the test sets for /f/ and /s/ were counter-balanced across subjects.

Each test item of Test 2 presented one of fricatives, /f/ or /s/, in two following vowel contexts. The four comparisons listed in Table 5.1, including all possible combinations of order of presentation, were repeated 10 times, yielding a total of 40 comparisons per subject. The resulting waveforms were assembled into a listening test.

The tape source was the Nakamichi (Model LX-5) which was amplified by a Crown power amplifier (Model D-75) feeding one or more Sennheiser HD 430 headphones. Stimuli were presented binaurally at a comfortable listening level.

Five adults, all native speakers of American English with no known hearing impairments, served as subjects. In view of the fine auditory distinctions presented in these tests, testing with a portable audiometer was performed to further confirm that hearing thresholds were within normal limits.

In summary, Test 1 presented a continua of stimuli generated with steady noise in which only the amplitude of the frication was varied, in 4 dB increments from -12 to +12 dB where 0 dB was the reference. Test 2 presented the stimulus comparisons listed in Table 5.1, in which stimuli were generated with either steady or tv synthesis strategies, and some variation in amplitude.

5.5 Results from Perceptual Evaluation of Synthetic Stimuli

The general result was that the time variations presented did not affect the naturalness judgements, but the consonant-vowel ratio did. The results of Test 1 are shown in Figure 5.3. The most preferred stimuli were the reference stimuli (also indicated as the 0 dB condition), which represented the best match to the consonant-vowel ratio of the natural speech. Preference profiles showed that listeners were more willing to accept some variation in level for /f/, but were less tolerant for /s/. Indeed, by subject report, the /s/ in the stimuli with relatively less noise amplitude were noted to approach /θ/ and were rejected. One listener noticed that the weakest /f/ stimulus sounded more stop-like, and he preferred those. This preference showed up rather prominently in the inflated group scores for the -12 dB condition for /f/.

The results of the four VCV conditions of Test 2 are shown in Table 5.2. The results for the amplitude comparisons in Test 2 revealed that the subjects always preferred the time-varying stimuli at the 0 dB level to the +8 dB level. This preference for the spectral match to the natural utterance can be seen in the group totals as well as individual results (for all listeners except L3, who preferred the noisier stimuli in all cases except /əf ɛ/). Preferences for the steady stimuli at the 0 dB level over the +8 dB were observed for both /f/ and /s/ preceding /ɛ/. These results, i.e., clear preferences (approximately 2:1) for the natural consonant-vowel ratio over a relatively stronger fricative, are consistent with the Test 1 findings. For steady /əsɑ/, the +8 dB stimulus was preferred over the stimulus at the 0 dB level.

When stimuli comparing synthesis strategies at the higher level are paired, the group preferences indicate that the steady+8 stimulus was chosen over the tv+8 stimulus, with strong preferences shown for /əfɑ/ and /əs ɑ/. However, there was no difference between the tv and steady stimuli at the best match reference level in three out of four cases. The only clear preference was for the tv version of /əsɑ/.

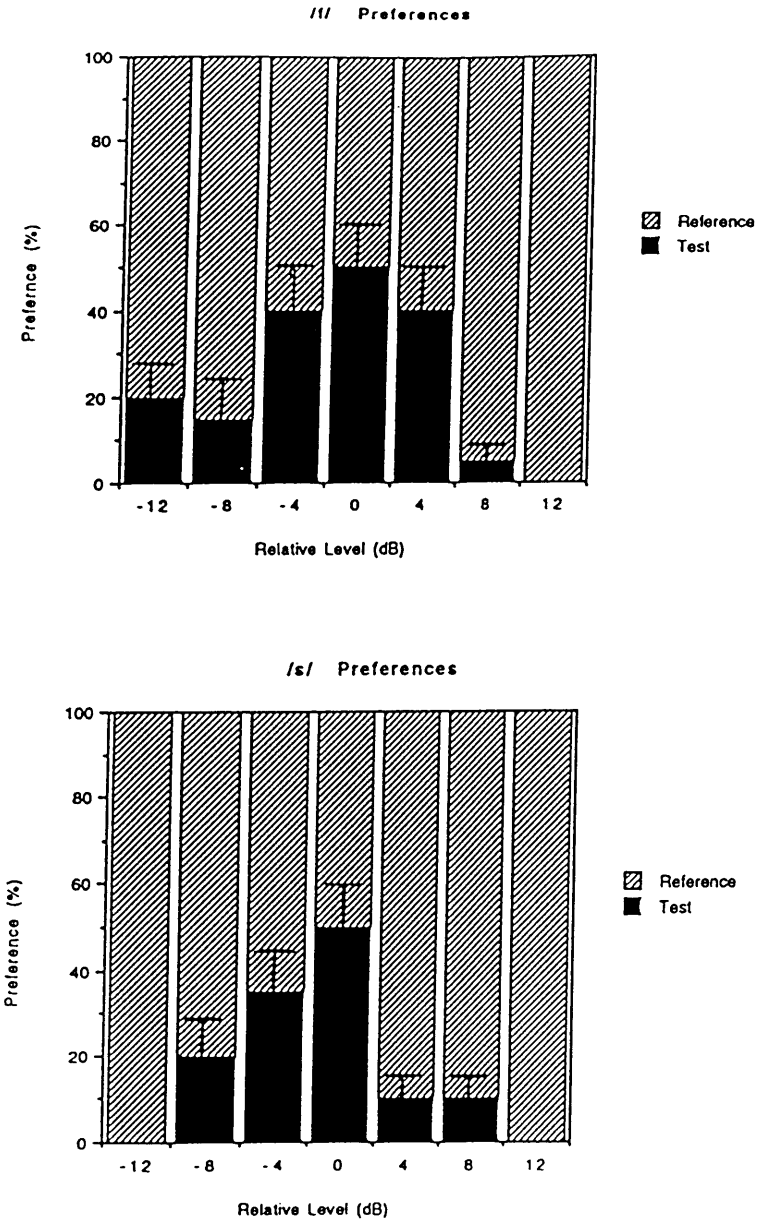

Figure 5.3: Listener preferences (in percent) are shown for two tests of synthetic stimuli in which the amplitude of the noise was varied in relation to the level of following vowel. The results for /f/ are shown in the top bargraph and the results for /s/ in the bottom bargraph.

Table 5.2: Number of times, out of 10 comparisons, that five listeners (L1-L5) preferred one synthetic stimuli over another in Listening Test 2.

Condition for /əfɛ/				
	tv vs. steady	tv vs. tv+8	tv+8 vs. steady+8	steady vs. steady+8
L1	5:5	10:0	4:6	9:1
L2	1:9	10:0	3:7	9:1
L3	5:5	10:0	3:7	7:3
L4	7:3	9:1	5:5	10:0
L5	3:7	9:1	3:7	8:2
Total	21:29	48:2	18:32	43:7
Condition for /əfɑ/				
	tv vs. steady	tv vs. tv+8	tv+8 vs. steady+8	steady vs. steady+8
L1	5:5	5:5	4:6	3:7
L2	5:5	9:1	4:6	7:3
L3	6:4	1:9	1:9	1:9
L4	5:5	8:2	3:7	4:6
L5	8:2	8:2	1:9	8:2
Total	29:21	31:19	13:37	23:27

Condition for /əsɛ/				
	tv vs. steady	tv vs. tv+8	tv+8 vs. steady+8	steady vs. steady+8
L1	4:6	9:1	4:6	10:0
L2	8:2	9:1	4:6	8:2
L3	7:3	3:7	6:4	0:10
L4	2:8	6:4	4:6	5:5
L5	4:6	9:1	2:8	10:0
Total	25:25	36:14	20:30	33:17
Condition for /əsɑ/				
	tv vs. steady	tv vs. tv+8	tv+8 vs. steady+8	steady vs. steady+8
L1	7:3	8:2	4:6	1:9
L2	6:4	9:1	2:8	4:6
L3	6:4	0:10	4:6	0:10
L4	8:2	7:3	2:8	6:4
L5	5:5	10:0	0:10	7:3
Total	32:18	34:16	12:38	18:32

5.6 Comparing Perceptual Results to Acoustic Findings

In view of the perceptual results indicating that judgements of the naturalness of a fricative in these tests were consistently influenced by the consonant-vowel ratio, further analysis of the synthetic stimuli, and their natural targets, was performed. Listeners always preferred the time-varying synthetic stimuli with the best match to the natural token over corresponding stimuli with an increased consonant-vowel ratio. Findings favoring the time-varying best match over the corresponding +8 dB condition, and the steady best match over the corresponding +8 dB condition observed for the /əCɛ/ contexts, confirm this expected result.

In general, there was not a strong preference expressed between the time-varying or the steady stimuli synthesis methods at expected consonant-vowel ratios. However, when the frication noise was increased, listeners were less tolerant of the time-varying stimuli, which already had some amplitude increments at the edges representing the emergence of the back cavity resonances.

Furthermore, for responses to /əsɑ/, there was a slight preference for the time-varying synthetic stimulus. The findings for the /ɑ/ context could be interpreted as supporting evidence for a vowel dependence or could be considered to be a token-specific effect. The noise amplitudes taken at the fricative midpoint might be an underestimate and might, therefore, explain why subjects prefer the steady + 8 dB stimuli, as well. We ruled out the possible confound from the shortened /ɑ/. A shorter vowel would lead to a preference for a softer fricative; for the steady case, we actually found the opposite effect.

For speaker M1, the natural /s/ tokens before /ɑ/ were relatively weak, compared to typical values for /s/. We re-examined the averaged data for all tokens in the original database and found that this speaker's alveolar fricatives before /ɑ/ were consistently at the lower limits for the strident fricatives. Amplitude differences for two instances of prestressed fricatives, the first CV and the third VCV from the 'CVCV'CVC, are shown in Table 5.3. The amplitude differences correspond to the

Table 5.3: The amplitude differences (dB) for CV and VCV tokens in the natural database for speaker M1 were found by subtracting F1 amplitude from the the maximum spectral peak above 2000 Hz

Token 1 (CV)	Amplitude Difference (dB)	Token 3 (VCV)	Amplitude Difference (dB)
/fɛ/	-42	/əfɛ/	-38
/fɑ/	-39	/əfɑ/	-34
/sɛ/	-16	/əsɛ/	-13
/sɑ/	-30	/əsɑ/	-25

noise amplitudes normalized by subtracting F1 amplitude (A_1) as were presented in Chapter 4. The VCV indicated as Token 3 was the natural target for the synthesis match. Note that while the amplitude differences for /f/ stimuli are quite similar in the two instances of each vowel context, the values for /s/ before /ɑ/ are much weaker (greater differential) than /s/ before /ɛ/. That is, for this speaker and this pair of tokens, noise for /s/ preceding /ɛ/ is much stronger in relation to the vowel than the noise for /s/ preceding /ɑ/. The results would seem to suggest that the listeners preferred the time-varying stimulus for the synthetic /əsɑ/ because it compensated by being stronger at the fricative edges.

Additional acoustic analysis was used to investigate the finding that the subjects tended to prefer an increase in frication for /s/ in the /əsɑ/ context. The F1 amplitude of the following vowel and the amplitude of the highest peak in the fricative were compared for the natural and synthetic stimuli. It was observed that for the /s/ stimuli, the amplitude difference between the F1 peak and the fricative peak for the vowel /ɑ/ is approximately 8 dB higher than difference for the /s/ and the following /ɛ/. This is considerably higher than the approximately 3 dB difference between the vowels alone, as predicted from the literature (Peterson and Barney, 1952). The spectral peak frequency is a little higher for the /s/ preceding /ɛ/ than the /s/ preceding /ɑ/.

As noted in Tables 5.3 and 5.4, the clearly strident /s/ preceding /ɛ/ is quite strong in relation to the following vowel, while the /s/ preceding /ɑ/ is relatively weak. There is no difference between the /f/ stimuli in the different vowel contexts. We then re-

Table 5.4: The difference between the amplitude of F1 in the vowel and the amplitude of the highest spectral peak in the synthetic fricatives modelled on the natural speech of Speaker M1.

Segement	Spectral peak amplitude (dB)	F1 amplitude (dB)	Amplitude Difference
/f ɛ/	18	56	-38
/f ɑ/	18	56	-38
/s ɛ/	38	52	-14
/s ɑ/	33	60	- 27

examined the time variations that occur for the natural /s/ in /əsɑ/ according to the octave-band description in Chapter 4. We confirmed that the highest energy for the /s/ was in the 4-8 kHz range (Band 5) in the middle of the fricative. However, when the the noise increases in the F3 region, exciting the back-cavity resonance, the energy in the 2-4 kHz range (Band 4) dominates.

5.6.1 Summary of Main Findings

Our perceptual measures were designed to investigate the importance of observed acoustic events, associated with the movement between a fricative and a vowel. Our tasks involved listener judgments of synthetic fricatives, as one way to study cues for naturalness. The reader is further referred to recent studies of the multiplicity of cues in the perception of English fricatives (e.g., Behrens and Blumstein, 1988b; Jongman, 1989; Whalen, 1991). The following is a brief summary of our perceptual results, interpreted with respect to our acoustic findings for fricative consonants:

- Details of the time variation in noise did not affect naturalness judgements, as tested by comparing results of two fricative synthesis strategies:

  1. best spectral match of time-varying spectra (tv) vs.

  2. stationary noise from fricative midpoint (steady).

- Consonant-to-vowel amplitude ratio significantly affected naturalness judgements.

    1. The original consonant-vowel ratio was generally preferred.

        - The best match of the time-varying spectral amplitude was vigorously chosen over stronger (and weaker) noise amplitudes.

        - The steady noise stimuli were generally preferred over stronger (and weaker) noise amplitudes, except when the fricative midpoint of the natural target utterance was observably weaker than the rest of the consonant.

- Vowel context effects were observed to interact with the above results.

    1. The original consonant-vowel ratio was preferred for /ɛ/.

    2. The steady noise stimuli based on the original fricative midpoint amplitude were preferred over more intense fricatives preceding /ɛ/; however, this trend was not seen in the /ɑ/ context.

        - There was no preference between noise amplitude levels tested for /əfɑ/.

        - There was a clear preference for boosting the noise amplitude for the /əsɑ/ stimulus, which was observed to have a particularly weak consonant-vowel ratio as measured at the fricatve midpoint.

As we discussed in Chapter 2, speech synthesis by computers is one way to model the generation of human speech. Klatt (1983, p. 95) indicates that "each rule system attempts to make appropriate generalizations and simplifications concerning the form and content of rules for consonant-vowel synthesis." The results of the current perceptual evaluation of synthetic stimuli showed that replicating detailed time-variations, as represented in individual tokens, appears to matter less than choosing the appropriate amplitude in relation to the following vowel. Of course, the spectral shape also needs to be representative of the place of articulation. The present work can further contribute to specification of rules for mapping to higher-level parameters, which greatly simplifies the synthesis process (Stevens and Bickley, 1991).

Chapter 6

Conclusions

This thesis focused on studying the inventory of acoustic events associated with the movement between a fricative consonant and a vowel. Knowledge about fricative production and perception was incorporated into strategies for speech analysis and synthesis.

6.1 Summary and Interpretation of Results

Chapter 2 provided theoretical background in the articulatory, aerodynamic and acoustic aspects of the production of intervocalic fricatives. Acoustic results, in Chapters 3 and 4, were applied to modify the theory, where appropriate, and to set the time-varying control parameters in a speech synthesizer in Chapter 5.

Results of the acoustic analysis in Chapters 3 highlighted the coarticulation resistance between fricatives and vowels by examining the formant transitions near the fricative-vowel boundary. It was shown that variability in the acoustic signal at the release of a fricative decreases when production of the fricative places greater constraints on the position of the tongue blade and body. That is, the constraints are greatest when the tongue body is involved in articulation of the fricative and least when the major articulators are the lips. Chapter 3 results also contributed to rules for synthesizing fricatives by providing tables of carefully timed and measured formant onset values of second and third formant frequencies for synthesizing prevocalic

fricatives spoken by male and female speakers.

We also sought to better model the relative timing of the vocal sources and detailed spectral characteristics of the noise. Time-varying noise results in Chapter 4 raised major questions about model predictions and suggested how the source-filter models might be modified. First, the relative timing in the area trajectories of glottal and supraglottal constrictions could be inferred from acoustic events at the fricative-vowel boundary. For example, a short gap in energy at the boundary between a voiceless fricative and the following vowel could be interpreted as reflecting that the supraglottal constriction is released before the glottis is closed. The acoustic result in that interval is aspiration noise, but the magnitude of that noise is extremely weak. In addition, the spectral differences between strident and nonstrident fricatives suggested that models of the filtering of the noise source by the front cavity might be improved if the losses in the vocal tract were better represented and if better estimates could be made of the source location.

A significant difference between the production constraints between nonstrident and strident fricatives was compatible with perceptual results using synthetic stimuli. Listeners were more tolerant of the manipulations in the consonant-vowel ratio for /f/ than /s/. We failed to demonstrate that listeners are sensitive to time variations in fricative noise when judging naturalness. However, we have shown that listeners reject stimuli with energy at unexpected frequencies or times.

6.2 Implications

In our acoustic studies of noise characteristics, greater amplitude variations over the duration of the consonant were observed for nonstridents than for stridents. Also, labiodentals were found to have the most variation in formant patterns. The greater variability that we found in each speaker's acoustic realization of nonstridents is consistent with the greater interspeaker variability of nonstridents found in a recent MRI study of sustained fricatives (Narayanan et al., 1994).

Examination of the acoustic characteristics of fricatives with respect to adjacent

vowels led to several conclusions. It has been noted that the acoustic correlates of place for fricatives are not strictly invariant. For example, /f/ is usually observed to have the lowest F2 onset of any English fricative; however, when preceding high front vowels, labiodentals have higher F2 onsets than dentals. This can be explained in terms of the reduced articulatory constraints we discussed in the previous section. In Chapter 3 we noticed that the acoustic variability increased when the articulatory constraints decreased. Despite the acoustic variability of the nonstridents, human listeners seem to have no difficulty normalizing the acoustic cues for a particular fricative given a particular vowel context.

What glues together vowels and fricatives in the speech stream? They have different energy concentrations in the low and high frequencies and vastly different temporal characteristics. The source for vowels is periodic and the source for fricatives is aperiodic. However, we have seen that continuous motion of the vocal tract, when moving from a fricative into a vowel, is reflected in the acoustic signal. Formant transitions in the vowel have been shown to be continuous with peaks in the fricative spectra (Xu and Wilde, 1994). This continuity may help the listener to recognize that the fricative and the vowel are being produced by the same speaker (if not the same source).

6.3 Future Work

The practical limitations in our analysis and synthesis tools, such as the 7.8 kHz anti-aliasing filter used for analysis and the limited bandwidth of the formant synthesizer, leave much room for improvement in speech processing techniques. Also, we need to consider more sensitive perceptual measures when assessing how listeners judge the naturalness, as well as intelligibility, of natural and synthetic speech.

The current analysis system was designed to be modular and modifiable. It would be useful to extend the noise analysis to allow more direct comparisions with other measures of fricative acoustic characteristics, such as spectral moment measures. The work of Xu and Wilde (1994) on combining time-averaging with ensemble-averaging

in analyzing fricatives demonstrates one possible extension to the system that has already started to produce encouraging results.

One necessary change to the analysis system is to allow more active control over the frequency regions. It would be desirable to be able to track the formant peaks in the vowel and use this information to choose appropriate frequency bands for noise analysis in the fricative. We need better frequency resolution than octave bands in order to relate the analysis results to production and perception. Incorporating improved models of the auditory system is a long-range goal.

Studying noisy speech sounds yields inherently noisy findings. We observe noise variations over time, variations from one token to another and inter-speaker variability. It is critical that we further examine the sources of this variability.

In our preliminary analysis of the unaveraged 'CVCV'CVC utterances in our database, we observed substantial intra-speaker variability in the acoustic realization of fricatives. There was a pattern that co-occurred with the known prosodic constraints. Fricatives tended to appear weaker and shorter before the reduced vowels in this controlled database, at times appearing stop-like in manner. For example, intra-speaker variability of time-varying fricatives is especially visible in the utterance /ðɑðəðɑð/ shown in Figure 6.1. The /ð/ in the post-stressed position beginning at approximately 600 ms appears stop-like before the /ə/. In our experience, the phenomenon of fricatives that appear to be produced in a stop-like manner is generally restricted to the weak, typically voiced, fricatives and can be considered to be an extreme case of time-varying noise. The /ð/ in pre-stressed position beginning at approximately 800 ms appears sonorant-like. The initial and final fricatives show an increase in high-frequency energy with time.

We want to determine if time variations, such as a buildup of energy in nonstrident fricatives before non-reduced vowels, could help describe the articulatory kinematics involved in moving from a fricative to a vowel. Analysis of data on fricatives preceding reduced vowels is already in process. We will continue work in the area of searching for acoustic cues in the vicinity of landmarks in the speech signal.

The noise measurement system presented in this thesis has already been configured

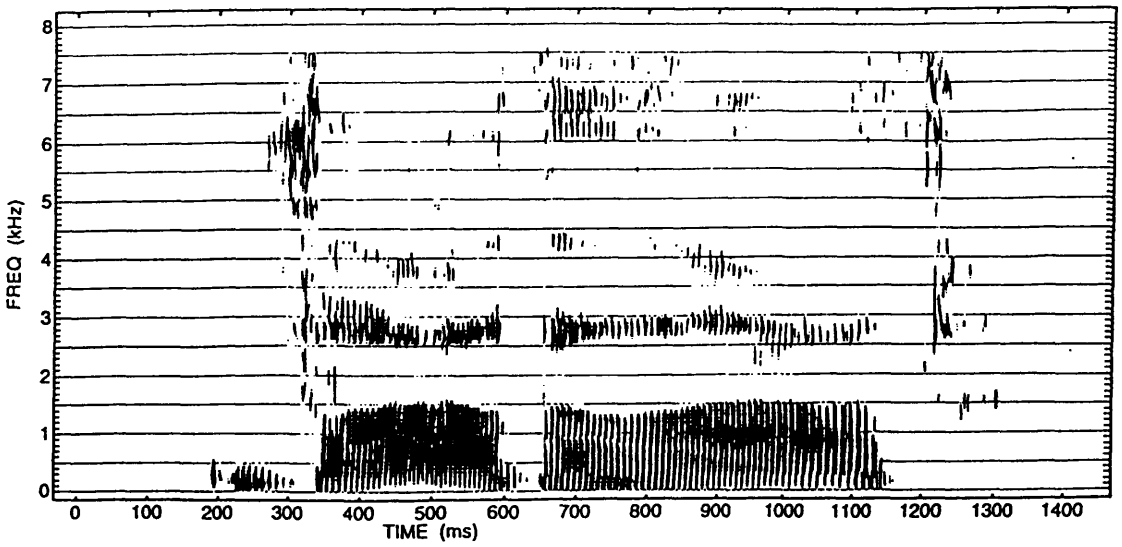

Figure 6.1: A spectrogram of the utterance /ðɑðəðɑð/ spoken by a male speaker (M1). The fricatives start at approximately 300, 600, 800 and 1200 ms.

for application to the TIMIT database. We look forward to testing the findings we obtained with use of the present controlled database to more meaningful speech with a broader base of speakers and phonetic contexts.

Bibliography

1. ESPS (1992). 4.0 Functions, File Types, and Tech Memos. Entropic Research Laboratory, Inc.

2. Badin, P. (1991). "Fricative consonants: Acoustic and x-ray measurements," J. Phonetics 19, 397-408.

3. Badin, P., Shadle, C. H., Pham Thi Ngoc Y., Carter, J. N., Chiu, W. S. C., Sculley, C., and Stromberg, K. (1994). "Frication and aspiration noise sources: Contribution of experimental data to articulatory synthesis." Paper presented at the International Conference on Spoken Language Processing, Yokohama, 163-166.

4. Baer, T., Gore, J. C., Gracco, L. C., and Nye, W. (1991). "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels," J. Acoust. Soc. Am. 90(2), 799-828.

5. Behrens, S. J. and Blumstein, S. E. (1988a). "Acoustic characteristics of English voiceless fricatives: A descriptive analysis," J. Phonetics 16, 295-298.

6. Behrens, S. J. and Blumstein, S. E. (1988b). "On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants," J. Acoust. Soc. Am. 84(3), 861-867.

7. Beranek, L. (1954). *Acoustics.* McGraw-Hill: New York.

8. Bitar, N. (1993). "Strident feature extraction in English fricatives." Paper presented at the 125th meeting of the Acoustical Society of America, Ottawa, Canada.

9. Bothorel, G. J., Simon, P., Wioland, F. and Zerling, J-P. (1986). *Cineradiographie: Des Voyelles et Consonnes du Francais.* (Institut de Phonetique, Strasbourg).

10. Catford, J. C. (1977). *Fundamental Problems in Phonetics.* (Indiana University Press, Bloomington).

11. Chan, C. and Ng, K. W. (1985). "Separation of fricatives from aspirated plosives by means of temporal spectral variation," IEEE Transactions, Vol. ASSP-33, No. 4, 1130-1137.

12. Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English.* (Harper and Row, New York).

13. Denes, P. B. and Pinson, E. N. (1973). *The Speech Chain.* (Anchor Press/Doubleday, Garden City, New York).

14. Deng, L. and Sun, D. (1993, unpublished manuscript). "A statistical approach to automatic speech recognition using the atomic speech units constructed from overlapping articulatory features," Submitted to J. Acoust. Soc. Am.

15. Fant, G. (1960). *Acoustic Theory of Speech Production.* (Mouton, The Hague).

16. Fant, G. (in press). "Acoustical analysis of speech," in M. Crocker (ed.) *Handbook of Acoustics*, (John Wiley and Sons, Inc., New York).

17. Fowler, C. (1994). "Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation," Perception and Psychophysics 55(6), 597-610.

18. Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. (1988). "Statistical analysis of word-initial voiceless obstruents: Preliminary data," J. Acoust. Soc. Am. 84(1), 115-123.

19. Glass, J. R. (1988). Finding Acoustic Regularities in Speech: Applications to phonetic recognition. Ph.D. Dissertation. MIT.

20. Green, D. (1976). *An Introduction to Hearing.* (Lea Associates, New York).

21. Haggard, M. (1978). "The devoicing of voiced fricatives," J. Phonetics 6, 95-102.

22. Halle, M. and Stevens, K. N. (1991). " Knowledge of language and the sounds of speech," in *Music, Language, Speech and Brain*, edited by J. Sundberg, L. Nord, and R. Carlson (MacMillan, Basingstoke, Hampshire, England), pp. 1-19.

23. Harris, K. (1958). "Cues for discrimination of American English fricatives in spoken syllables," Language and Speech 1, 1-17.

24. Heinz, J. M. and Stevens, K. N. (1961). "On the properties of voiceless fricative consonants," J. Acoust. Soc. Am. 33, 589-596.

25. Hughes, G. and Halle, M. (1956). "Spectral properties of fricative consonants," J. Acoust. Soc. Am. 28 (2), 303-310.

26. Isshiki, N. (1964). "Regulating mechanisms of vocal intensity variation," J. Speech Hear. Res. 7, 17-29.

27. Jakobson, R., Fant, C. G. M., and Halle, M. (1965). *Preliminaries to Speech Analysis.* (The MIT Press, Cambridge, MA).

28. Jassem, W. (1965). "The formants of fricative consonants," Language and Speech 8, 1-16.

29. Jongman, A. (1989). "Duration of frication noise required for identification of English fricatives," J. Acoust. Soc. Am. 85 (4), 1718-1725.

30. Kewley-Port, D. (1982). "Measurement of formant transitions in naturally produced stop consonant/vowel syllables," J. Acoust. Soc. Am. 72, 379-389.

31. Klatt, D. H. (1975). "Voice onset time, frication, and aspiration in word-initial consonant clusters," J. Speech Hear. Res. 18., no. 4, 686-706.

32. Klatt, D. H. (1980). "Software for a cascade/parallel formant synthesizer," J. Acoust Soc. Am. 67 (3), 971-995.

33. Klatt, D. H. (1983). Synthesis by rule of consonant-vowel syllables. Speech Communication Group Working Papers III, Research Laboratory of Electronics, MIT, Cambridge, MA, 93-102.

34. Klatt, D. H. (1984). The new MIT Speechvax computer facility. Speech Communication Group Working Papers IV, Research Laboratory of Electronics, MIT, Cambridge, MA, 93-102.

35. Klatt, D. H. (1987). "Review of text-to-speech conversion for English," J. Acoust. Soc. Am. 82, 737-793.

36. Klatt, D. H. (Chapter 6, unpublished manuscript) "Fricative Consonants."

37. Klatt, D. H. and Klatt, L. C. (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. 87, 820-857.

38. Klatt, D. H., Stevens, K. N., and Mead, J. (1968). "Studies of articulatory activity and airflow during speech." Annals of the New York Academy of Sciences 155, 42-55.

39. Ladefoged, P. (1993). *A Course in Phonetics*, Third Edition, (Harcourt Brace Jovanovich, Inc., New York).

40. Lamel, L. F., Kassel, R. H., and Seneff, S. (1986). "Design and analysis of thes acostic-phonetic corpus," in *Proceedings of the DARPA Speech Recognition Workshop*, (Palo Alto, CA), 100-109.

41. Leon-Garcia, A. (1994). *Probability and Random Processes for Electrical Engineers*, Second Edition, (Addison-Wesley Publishing Company, Reading, MA).

42. McCasland, G.P. (1979). "Noise Intensity and Spectrum Cues for Spoken Fricatives," J. Acoust. Soc. Am. 65,S78 (A). in *Speech Communication Papers*, edited by J. J. Wolf and D. H. Klatt (Acoustical Society of America, New York), pp. 303-306.

43. Narayanan, S., Alwan, A. and Haker, K. (1994). "An MRI study of fricative consonants," In *Proceedings of the International Conference on Spoken Language Processing*, Yokohama, Japan, pp. 627-630.

44. Ohde, R. N. and Sharf, D. J. (1992). *Phonetic Analysis of Normal and Abnormal Speech*, (MacMillan Publishing Company, New York).

45. Perkell, J. S. (1969). *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study.* The Research Monograph No. 53, (MIT Press, Cambridge MA).

46. Peterson, G. E., and Barney, H. L. (1952). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. 24(2), 175-184.

47. Rothenberg, M. (1968). "The breath stream dynamics of simple-released-plosive production," Bibliotheca Phonetica, No. 6, S. Karger, Basel.

48. Scully, C., Castelli, E., Brearley, E., and Shirt, M. (1992). "Analysis and simulation of a speaker's aerodynamic and acoustic patterns for fricatives," J. Phonetics (20), 39-51.

49. Scully, C., Georges, E., and Castelli, E. (1992). "Articulatory paths for some fricatives in connected speech," Speech Communication 11, 411-416.

50. Shadle, C. H. (1985). "The acoustics of fricative consonants," *Research Laboratory of Electronics*, Technical Report 506, (MIT, Cambridge, MA)

51. Shadle, C. H. (1990). "Articulatory-acoustic relationships in fricative consonants," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Dordrecht, Netherlands), pp. 187-209.

52. Shadle, C. H., Moulinier, A., Dobelke, C. U., and Scully, C. (1992). "Ensemble averaging applied to the analysis of fricative consonants," in *Proceedings of the International Conference on Spoken Language Processing*, Vol. 1, (Banff, Alberta, Canada), pp. 53-56.

53. Soli, S. (1981). "Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation," J. Acoust. Soc. Am. 70, 976-984.

54. Sondhi, M. and Schroeter, J. (1992). "Articulatory speech analysis/synthesis," ICASSP 1992 Tutorial. March 23-26, San Francisco, CA.

55. Stevens, P. (1960). "Spectra of fricative noise in human speech," Language and Speech (3), 32-49.

56. Stevens, K. N. (1971). "Airflow and turbulence for fricative and stop consonants: Static considerations," J. Acoust. Soc. Am. 50, 1180-1192.

57. Stevens, K. N. (1985). "Evidence for the role of acoustic boundaries in the perception of speech sounds," in Fromkin (ed.) *Phonetic Linguistics*, (Academic Press, New York), pp. 243-255.

58. Stevens, K. N. (1987). "Interaction between acoustic sources and vocal-tract configurations for consonants," in *Proceedings of the Eleventh International Conferences of Phonetic Sciences*, Vol. 3, (Tallin, Estonia, USSR) pp. 385-389.

59. Stevens, K. N. and Keyser, S. J. (1989). "Primary features and their enhancement in consonants," Language 65, (1), 81-106.

60. Stevens, K. N. (1991). "Speech Perception Based on Acoustic Landmarks: Implications for Speech Production," Perilus XIV, Paper from the symposium, *Current Phonetic Research Paradigms: Implications for Speech Motor Control*, (Stockholm, Sweden), pp. 83-87.

61. Stevens, K. N. and Bickley, C. A. (1991). " Constraints among parameters simplify control of Klatt formant synthesizers," J. Phonetics 19, 161-174.

62. Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M. and Kurowski, K. (1992). " Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters," J. Acoust. Soc. Am. 91, 2979-3000.

63. Stevens, K. N., Manuel, S. Y., Shattuck-Hugnagel, S. and Liu, S. (1992). "Implementation of a model for lexical access based on features," in *Proceedings of International Conference on Spoken Language Processing*, Banff, Alberta, Canada, pp. 499-502.

64. Stevens, K. N. (1993). "Modelling affricate consonants," Speech Communication 13, 33-43.

65. Stevens, K. N., Bickley, C. A., and Williams, D. (in preparation). *Notes on HLSYN.*

66. Sussman, H. M. , McCaffrey, H. A. and Matthews, S. A. (1991). "An investigation of locus equations as a source of relational invariance for stop place categorization," J. Acoust. Soc. Am. 90(3), 1309-1325.

67. Tomiak, G. (1990). "An Acoustic and Perceptual Analysis of the Spectral Moments Invariant with Voiceless Fricative Obstruents," Ph.D. Dissertation, SUNY Buffalo, (January, 1990).

68. Titze, I. R. (1992). "Phonation threshold pressure: A missing link in glottal aerodynamics," J. Acoust. Soc. Am. 91(5), 2926-2935.

69. Utman, J. A. and Blumstein, S. E. (1994). "The influence of language on the acoustic properties of phonetic features: A study of the Feature [strident] in Ewe and English." Phonetica 51(4): 221-238.

70. Whalen, D. H. (1991). "Perception of the English /s/-/ʒ/ distinction relies on fricative noises and transitions, not on brief spectral slices," J. Acoust. Soc. Am. 90(4), 1776-1785.

71. Wilde, L. F. and Huang, C. B. (1991). "Acoustic properties at fricative-vowel boundaries in American English," in *Proceedings of the XII International Congress of Phonetic Sciences*, Vol. 5, (Aix-en-Provence, France), pp. 398-401.

72. Wilde, L. (1993). "Inferring articulatory movements from acoustic properties at fricative-vowel boundaries," Speech Communication Group Working Papers, IX, 18-27.

73. Xu, Yi (1989, unpublished manuscript). "Distribution of vowel information in prevocalic fricative noise: Perceptual and acoustic evidence."

74. Yeni-Komshian, G. H. and Soli, S. D. (1981). "Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation," J. Acoust. Soc. Am. 70(4), 966-975.

75. Zue, V. W. (1985). " The use of speech knowledge in automatic speech recognition," Invited paper in *Proceedings of IEEE Special Issue on Human-Machine Communication by Speech*, Vol. 73(11), pp. 1602-1614.

76. Zue, V. Seneff, S., and Glass, J. (1990). "Speech database development at MIT: TIMIT and beyond," Speech Communication 9, 351-356.

Appendix A

Formant Frequencies for Two Female Speakers

Shown in the Tables A.1 through A.4 are the values of the formant onset frequencies (F1onset, F2onset, F3onset) and formant values at vowel center (F1vowel, F2vowel, F3vowel) in Hz for two female speakers (F1 and F2).

token	F1onset	F2onset	F3onset	F1vowel	F2vowel	F3vowel
fiy	250	2500	3087	252	2993	3465
feh	305	1974	2678	630	2111	2898
faa	330	1300	2646	693	1323	2646
fah	630	1386	2678	693	1481	2709
fow	250	1363	2500	693	1449	2646
fuw	270	1449	2615	252	1166	2835
thiy	267	2064	3102	252	2898	3339
theh	504	1928	2921	662	2172	2993
thaa	733	1464	2812	788	1418	2867
thah	525	1495	2678	693	1449	2709
thow	277	1628	2800	693	1481	2678
thuw	270	1905	2520	252	1827	2615
siy	273	2346	3125	252	2741	3119
seh	252	1966	2820	693	2142	2835
saa	567	1796	2835	788	1418	2835
sah	315	1992	2835	725	1481	2709
sow	305	1875	2867	725	1418	2520
suw	227	1890	2898	252	1796	2741
shiy	252	2400	3024	252	2867	3339
sheh	288	2268	3150	630	2268	2867
shaa	598	1891	2772	914	1418	2489
shah	567	1890	2709	725	1355	2961
show	567	1938	2787	693	1512	2457
shuw	283	2243	2835	252	2205	2583
viy	226	2457	2898	252	2835	3213
veh	441	1796	2835	630	2048	2993
vaa	567	1418	2583	788	1418	2804
vah	441	1512	2583	693	1638	2583
vow	441	1323	2520	567	1323	2709
vuw	182	1229	2583	252	1197	2772
dhiy	207	2346	2898	189	2898	3339
dheh	410	1890	2961	567	2268	3087
dhaa	441	1638	2835	788	1638	2835
dhah	441	1701	2867	630	1575	2621
dhow	441	1764	2961	567	1449	2804
dhuw	284	1796	2741	221	1733	2804
ziy	168	2352	3173	221	2678	3182
zeh	230	2142	2898	567	2205	2961
zaa	441	1890	2898	819	1764	2993
zah	387	1827	2867	599	1733	2993
zow	252	1827	2709	567	1764	2709
zuw	196	2111	2807	252	1922	2835
zhiy	226	2369	2910	252	2867	3371
zheh	189	2310	3024	504	2205	3056
zhaa	441	2016	2976	788	1670	2993
zhah	441	1985	2804	693	1764	2993
zhow	441	1890	2961	536	1733	2993
zhuw	252	2299	2923	252	2142	2678

Table A.1: Formant Frequencies: Speaker F1 (Token 1)

token	F1onset	F2onset	F3onset	F1vowel	F2vowel	F3vowel
fiy3	289	2583	2898	252	2772	3213
feh3	325	1906	2804	693	2048	2885
faa3	409	1197	2678	788	1448	2678
fah3	620	1333	2583	756	1449	2646
fow3	567	1355	2583	662	1355	2678
fuw3	315	1260	2741	252	1166	2772
thiy3	283	2583	2867	252	2993	3339
theh3	523	2004	3000	630	2111	2993
thaa3	733	1464	2812	788	1418	2898
thah3	567	1607	2741	756	1449	2835
thow3	340	1575	2615	725	1386	2772
thuw3	294	1701	2709	347	1638	2646
siy3	284	2331	2772	221	2898	3308
seh3	315	2111	2856	662	2111	2930
saa3	654	1764	2700	788	1386	2835
sah3	567	1796	2678	756	1701	2835
sow3	315	1638	2835	662	1481	2583
suw3	259	1741	2835	252	1890	2709
shiy3	227	2346	2982	410	2741	3024
sheh3	418	2100	2919	630	2111	2993
shaa3	630	1859	2772	882	1481	2741
shah3	567	1890	2678	788	1449	2993
show3	504	1764	2772	473	1418	2363
shuw3	252	1985	2583	284	1796	2552
viy3	221	2496	2930	252	2835	3276
veh3	275	1764	2835	662	2079	2961
vaa3	567	1386	2646	788	1481	2678
vah3	378	1512	2583	693	1544	2709
vow3	378	1323	2520	599	1386	2835
vuw3	189	1292	2646	252	1323	2646
dhiy3	315	2426	2993	252	2961	3339
dheh3	441	1764	2898	662	2111	3056
dhaa3	567	1607	2961	788	1544	2819
dhah3	441	1638	2898	756	1575	3024
dhow3	429	1588	2862	567	1386	2867
dhuw3	315	1796	2835	252	1575	2898
ziy3	211	2268	3024	252	2678	3087
zeh3	378	1922	2961	630	2111	2961
zaa3	378	1764	2961	788	1481	2835
zah3	315	1701	2772	662	1733	2930
zow3	378	1796	2646	599	1481	2678
zuw3	189	2268	3056	252	1890	2741
zhiy3	189	2426	3087	221	2898	3276
zheh3	182	2268	3087	504	2174	3056
zhaa3	378	1796	2961	788	1607	2835
zhah3	315	1922	2961	693	1733	2933
zhow3	315	1890	2961	630	1575	2426
zhuw3	252	2268	2961	252	2237	2646

Table A.2: Formant Frequencies: Speaker F1 (Token 3)

token	F1onset	F2onset	F3onset	F1vowel	F2vowel	F3vowel
fiy	252	2268	2741	378	2772	3245
feh	315	1859	2867	662	1827	2898
faa	725	1197	2678	914	1197	2772
fah	599	1229	2646	725	1323	2583
fow	536	1197	2520	567	1328	2583
fuw	347	1701	2520	347	1638	2520
thiy	284	2394	2867	252	2583	3308
theh	441	1890	2646	693	1923	2772
thaa	725	1292	2867	788	1229	2867
thah	473	1512	2835	756	1229	2804
thow	504	1764	2520	567	1512	2583
thuw	315	1890	2741	284	1922	2426
siy	252	2237	2993	252	2646	3119
seh	473	1890	2993	630	1922	2993
saa	504	1386	2741	788	1292	2615
sah	441	1512	2804	693	1386	2741
sow	410	1953	2835	536	1670	2583
suw	315	1890	2867	252	1953	2520
shiy	252	2363	2993	252	2583	3056
sheh	315	2142	2993	662	1922	2835
shaa	441	1922	2898	756	1260	2522
shah	410	2205	2993	693	1733	2583
show	378	1953	2646	504	1701	2526
shuw	252	2111	2646	252	2142	2646
viy	252	2394	2930	315	2741	3371
veh	252	1827	2583	725	1890	2993
vaa	662	1229	2552	756	1166	2993
vah	252	1418	2489	630	1166	2583
vow	252	1260	2363	567	1260	2741
vuw	221	1134	2205	315	1197	2489
dhiy	189	2142	2898	252	2520	3371
dheh	252	1859	2678	662	1859	2993
dhaa	284	1544	2898	756	1166	2993
dhah	284	1638	2741	630	1607	2898
dhow	252	1607	2930	599	1638	2898
dhuw	221	1733	2583	284	1733	2489
ziy	221	2174	2993	284	2583	3213
zeh	284	1733	2867	662	1953	2993
zaa	315	1638	2741	945	1260	2898
zah	284	1701	2804	630	1638	2867
zow	284	1764	2835	567	1638	2583
zuw	221	1796	2898	284	1859	2520
zhiy	252	2268	3087	252	2520	3056
zheh	284	2111	3087	567	2048	2898
zhaa	315	1985	3024	756	1260	2583
zhah	284	1953	2993	630	1733	2583
zhow	284	1733	2678	567	1638	2205
zhuw	221	2174	2741	284	2048	2583

Table A.3: Formant Frequencies: Speaker F2 (Token 1)

token	F1onset	F2onset	F3onset	F1vowel	F2vowel	F3vowel
fiy3	378	2426	2930	252	2678	3182
feh3	315	1896	2646	725	1859	2741
faa3	725	1197	2646	945	1260	2678
fah3	662	1292	2552	662	1323	2678
fow3	567	1197	2552	599	1323	2583
fuw3	378	1418	2489	473	1607	2583
thiy3	189	2174	2993	252	2552	3245
theh3	536	1859	2678	662	1890	2867
thaa3	441	977	2835	851	1166	2867
thah3	630	1323	2835	756	1323	2867
thow3	473	1575	2489	630	1575	2646
thuw3	315	1764	2520	315	1890	2457
siy3	315	2268	2615	252	2678	3119
seh3	504	1859	2961	630	1922	2993
saa3	662	1197	2646	788	1260	2646
sah3	567	1607	2709	630	1292	2772
sow3	410	1701	2646	504	1638	2583
suw3	252	1953	3119	252	1953	2489
shiy3	315	2142	2898	284	2552	3056
sheh3	441	2174	2993	662	1953	2804
shaa3	378	1985	2615	851	1260	2583
shah3	347	1827	2678	725	1733	2583
show3	315	1890	2615	536	1733	3623
shuw3	252	2111	2709	252	2174	2678
viy3	221	2457	2993	284	2646	3087
veh3	441	1796	2583	693	1796	2678
vaa3	693	1229	2520	662	1260	2552
vah3	315	1197	2363	662	1229	2426
vow3	252	1197	2300	599	1229	2426
vuw3	252	1197	2300	410	1323	2394
dhiy3	221	2174	2961	284	2520	3056
dheh3	252	1796	2678	693	1859	2741
dhaa3	284	1670	2678	725	1040	2741
dhah3	315	1544	2489	693	1575	2741
dhow3	252	1575	2772	662	1323	2583
dhuw3	284	1764	2520	284	1670	2237
ziy3	252	2426	3056	284	2520	3087
zeh3	252	1764	2646	630	1859	2835
zaa3	252	1575	2741	775	1323	2772
zah3	252	1575	2898	693	1418	2741
zow3	284	1575	2741	630	1575	2678
zuw3	221	1922	2678	315	1922	2426
zhiy3	252	2426	2993	315	2583	3024
zheh3	315	1922	2898	693	1890	2646
zhaa3	315	1890	2678	725	1229	2552
zhah3	284	1827	2835	630	1638	2520
zhow3	284	1922	2457	662	1701	2237
zhuw3	284	1827	2205	315	1827	2268

Table A.4: Formant Frequencies: Speaker F2 (Token 3)

Appendix B

Formant Frequencies for Two Male Speakers

Shown in the Tables B.1 through B.4 are the values of the formant onset frequencies (F1onset, F2onset, F3onset) and formant values at vowel center (F1vowel, F2vowel, F3vowel) in Hz for two male speakers (M1 and M2).

token	F1onset	F2onset	F3onset	F1vowel	F2vowel	F3vowel
fiy	287	1940	2478	284	2174	2898
feh	355	1497	2316	536	1638	2489
faa	682	961	2623	630	1040	2709
fah	467	882	2583	630	1103	2615
fow	441	850	2576	473	788	2583
fuw	214	1050	2700	347	1071	2048
thiy	292	1927	2573	252	2237	2961
theh	368	1544	2440	500	1575	2520
thaa	479	968	2678	567	1008	2709
thah	418	977	2752	630	1103	2709
thow	399	1000	2520	410	819	2520
thuw	225	1400	2372	284	1134	1953
siy	214	2053	2558	252	2174	2898
seh	315	1481	2583	504	1575	2583
saa	473	1103	2772	599	1040	2709
sah	378	1071	2709	567	1103	2835
sow	252	1071	2772	473	914	2457
suw	207	1473	2016	315	1260	2993
shiy	214	1913	2646	252	2205	2741
sheh	315	1733	2394	441	1638	2489
shaa	378	1386	2253	567	1071	2426
shah	410	1386	2331	567	1134	2426
show	378	1448	2331	473	945	2426
shuw	207	1575	2016	315	1323	1985
viy	277	1913	2495	252	1922	2898
veh	340	1423	2384	536	1575	2457
vaa	462	947	2541	599	1040	2709
vah	409	945	2583	599	1040	2678
vow	315	819	2646	473	819	2583
vuw	267	1071	2111	284	945	2930
dhiy	287	1741	2551	284	2205	2835
dheh	357	1489	2542	504	1670	2583
dhaa	414	1103	2671	567	1008	2804
dhah	378	1174	2672	567	1071	2678
dhow	315	1071	2619	473	945	2520
dhuw	274	1401	2558	315	1323	1953
ziy	250	1850	2631	284	2237	2993
zeh	252	1481	2583	441	1607	2615
zaa	346	1250	2858	567	1103	2835
zah	315	1355	2646	567	1103	2741
zow	315	1197	2583	473	1040	2363
zuw	242	1500	2562	284	1418	2520
zhiy	252	1953	2528	252	2142	2678
zheh	315	1857	2520	441	1796	2520
zhaa	352	1623	2331	567	1166	2520
zhah	315	1670	2268	567	1166	2520
zhow	315	1575	2394	473	1103	2300
zhuw	252	1670	2016	284	1418	1985

Table B.1: Formant Frequencies: Speaker M1 (Token 1)

token	F1onset	F2onset	F3onset	F1vowel	F2vowel	F3vowel
fiy3	201	1771	2669	267	2169	2850
feh3	497	1481	2432	536	1575	2583
faa3	525	957	2520	655	1078	2694
fah3	273	977	2538	630	1071	2615
fow3	463	892	2573	536	882	2426
fuw3	315	1000	2772	315	1050	2400
thiy3	273	1700	2418	270	2195	2744
theh3	441	1481	2678	473	1701	2615
thaa3	494	998	2814	661	1074	2400
thah3	409	1000	2741	630	1197	2678
thow3	426	1015	2631	504	977	2520
thuw3	275	1250	2497	315	1400	2400
siy3	237	1938	2568	270	2184	2735
seh3	378	1481	2646	567	1638	2552
saa3	451	1100	2749	615	1008	2667
sah3	441	1134	2709	599	1166	2489
sow3	378	1134	2646	473	977	1953
suw3	360	1450	2951	289	1150	2350
shiy3	273	2099	2600	273	2169	2606
sheh3	378	1796	2331	536	1638	2457
shaa3	388	1270	2184	637	1104	2500
shah3	378	1386	2268	347	1638	2363
show3	337	1449	2363	504	1008	2268
shuw3	214	1659	2551	303	1400	2006
viy3	315	1522	2321	315	2048	2709
veh3	401	1323	2432	473	1512	2426
vaa3	378	901	2520	599	1008	2709
vah3	466	924	2530	567	977	2646
vow3	378	819	2426	473	882	2457
vuw3	305	967	2300	315	1008	2331
dhiy3	315	1741	2505	284	2205	2804
dheh3	396	1345	2631	504	1575	2583
dhaa3	423	998	2772	567	1040	2867
dhah3	357	1024	2724	536	1197	2678
dhow3	340	1048	2667	504	1040	2426
dhuw3	300	1292	2556	347	1323	2363
ziy3	252	1764	2646	252	2111	2898
zeh3	315	1481	2583	504	1638	2678
zaa3	378	1229	2709	567	1103	2678
zah3	315	1292	2583	567	1229	2678
zow3	315	1260	2583	473	1040	2300
zuw3	252	1418	2520	315	1481	2205
zhiy3	252	1985	2583	315	2174	2678
zheh3	315	1890	2520	410	1827	2520
zhaa3	378	1512	2363	594	1166	2363
zhah3	315	1670	2394	536	1292	2363
zhow3	378	1449	2426	441	1040	2048
zhuw3	315	1607	2174	315	1355	2174

Table B.2: Formant Frequencies: Speaker M1 (Token 3)

token	F1onset	F2onset	F3onset	F1vowel	F2vowel	F3vowel
fiy	371	1777	2344	313	1992	2871
feh	443	1484	2344	547	1563	2383
faa	527	1094	2510	762	1172	2656
fah	488	1074	2520	566	1133	2656
fow	352	918	2480	488	879	2559
fuw	384	1055	2246	293	918	2266
thiy	361	1425	2422	254	1953	2871
theh	419	1269	2656	566	1543	2500
thaa	430	1202	2793	742	1211	2773
thah	430	1211	2656	586	1211	2637
thow	293	1113	2734	508	1055	2363
thuw	382	1220	2509	352	1094	2207
siy	273	1582	2500	293	2051	2715
seh	391	1348	2637	625	1563	2441
saa	469	1230	2754	723	1211	2656
sah	352	1211	2734	586	1172	2578
sow	430	1211	2695	527	1074	2324
suw	254	1328	2500	293	1133	2070
shiy	273	1816	2773	273	1953	2813
sheh	332	1699	2598	527	1543	2305
shaa	371	1641	2539	684	1230	2383
shah	313	1602	2344	625	1328	2402
show	352	1563	2207	488	1133	2031
shuw	254	1797	2285	254	1270	2031
viy	323	1582	2198	234	1914	2832
veh	362	1434	2285	508	1543	2500
vaa	323	1163	2383	781	1172	2734
vah	391	1074	2461	586	1152	2676
vow	430	977	2383	469	820	2422
vuw	352	1094	2119	313	1035	2207
dhiy	306	1386	2530	293	1797	2539
dheh	404	1317	2686	508	1504	2578
dhaa	421	1260	2647	664	1211	2734
dhah	371	1270	2578	566	1211	2676
dhow	391	1270	2715	527	1191	2402
dhuw	365	1321	2558	332	1289	2207
ziy	234	1563	2559	293	1953	2715
zeh	332	1367	2695	508	1582	2441
zaa	332	1328	2773	645	1211	2422
zah	293	1348	2754	586	1270	2617
zow	293	1309	2773	488	1230	2559
zuw	234	1406	2596	313	1328	2266
zhiy	234	1777	2598	254	1895	2539
zheh	313	1797	3203	469	1738	2461
zhaa	332	1758	3281	547	1309	2285
zhah	254	1602	2480	449	1230	2285
zhow	254	1641	2520	410	1211	1973
zhuw	254	1797	2441	293	1484	1895

Table B.3: Formant Frequencies: Speaker M2 (Token 1)

token	F1onset	F2onset	F3onset	F1vowel	F2vowel	F3vowel
fiy3	293	1699	2305	293	1934	2754
feh3	449	1406	2324	547	1465	2344
faa3	508	1016	2480	723	1133	2687
fah3	332	1113	2402	625	1113	2637
fow3	332	820	2500	488	859	2578
fuw3	313	898	2246	449	996	2207
thiy3	313	1309	2480	332	1855	2656
theh3	293	1230	2461	527	1465	2441
thaa3	469	1191	2637	645	1172	2637
thah3	391	1172	2637	586	1172	2617
thow3	313	1113	2695	527	1113	2539
thuw3	391	1211	2480	410	1211	2188
siy3	293	1445	2578	254	1973	2676
seh3	371	1376	2617	586	1406	2480
saa3	371	1211	2637	684	1250	2480
sah3	410	1230	2656	625	1191	2578
sow3	449	1172	2637	527	1133	2383
suw3	273	1309	2441	352	1191	2148
shiy3	234	1816	2813	273	1934	2773
sheh3	449	1602	2559	527	1543	2383
shaa3	235	1387	2480	664	1211	2480
shah3	254	1621	2383	566	1250	2344
show3	234	1504	2285	488	1172	2129
shuw3	293	1758	2148	352	1406	1973
viy3	293	1777	2246	273	2012	2773
veh3	313	1309	2285	527	1465	2422
vaa3	488	1074	2441	684	1152	2734
vah3	391	1016	2480	547	1113	2637
vow3	488	859	2422	469	840	2441
vuw3	293	879	2285	313	977	2266
dhiy3	254	1465	2344	293	1895	2637
dheh3	332	1250	2520	488	1484	2480
dhaa3	469	1172	2539	605	1211	2539
dhah3	391	1191	2656	547	1191	2676
dhow3	352	1133	2676	508	1172	2656
dhuw3	352	1211	2305	352	1309	2246
ziy3	254	1465	2617	293	1895	2500
zeh3	352	1309	2637	449	1504	2383
zaa3	332	1211	2734	645	1211	2461
zah3	371	1230	2695	508	1211	2637
zow3	293	1250	2676	488	1191	2441
zuw3	313	1445	2500	352	1270	2246
zhiy3	254	1855	2637	273	1895	2578
zheh3	293	1660	2422	391	1563	2363
zhaa3	293	1582	2344	488	1270	2363
zhah3	293	1602	2285	449	1270	2324
zhow3	313	1563	2227	332	1211	2031
zhuw3	254	1758	2148	293	1523	1934

Table B.4: Formant Frequencies: Speaker M2 (Token 3)

Appendix C

Synthesis Parameter (.doc) Files

C.1 Synthetic /əfɛ/

C.1.1 Time-varying Noise

```
Synthesis specification for file: 'fehtv.wav'  Fri Jun  3 12:21:58 1994

KLSYN93   Version 2.0  April 2,1993 M.M.(original program by D.H. Klatt)

Max output signal (overload if greater than 0.0 dB) is  -7.2 dB
Total number of waveform samples = 6000


  CURRENT CONFIGURATION:
    63 parameters

  SYM V/C  MIN  VAL   MAX  DESCRIPTION
-----------------------------------------------
  DU   C   30   500  5000  Duration of the utterance, in msec
  UI   C    1     5    20  Update interval for parameter reset, in msec
  SR   C 5000 12000 20000  Output sampling rate, in samples/sec
  NF   C    1     6     6  Number of formants in cascade branch
  SS   C    1     2     4  Source Switch 1:Impulse 2:Natural 3:Anantha 4:LF
  RS   C    1     8  8191  Random seed (initial value of random # generator)
  SB   C    0     1     1  Same noise burst, reset RS if AF=AH=0, 0=no,1=yes
  CP   C    0     0     1  0=Cascade, 1=Parallel tract excitation by AV
  OS   C    0     0    20  Output selector (0=normal,1=voicing source,...)
  GV   C    0    55    80  Overall gain scale factor for AV, in dB
  GH   C    0    60    80  Overall gain scale factor for AH, in dB
  GF   C    0    70    80  Overall gain scale factor for AF, in dB
  GI   C    0    60    80  Overall gain scale factor for AI, in dB
  F0   V    0  1000  5000  Fundamental frequency, in tenths of a Hz
  AV   V    0    60    80  Amplitude of voicing, in dB
  OQ   v   10    50    99  Open quotient (voicing open-time/period), in %
  SQ   v  100   200   500  Speed quotient (rise/fall time, LF model), in %
  TL   V    0     0    41  Extra tilt of voicing spectrum, dB down @ 3 kHz
  FL   v    0     0   100  Flutter (random fluct in f0), in % of maximum
  DI   v    0     0   100  Diplophonia (alt periods closer), in % of max
  AH   V    0     0    80  Amplitude of aspiration, in dB
  AF   V    0     0    80  Amplitude of frication, in dB
  F1   V  180   500  1300  Frequency of 1st formant, in Hz
  B1   v   30    60  1000  Bandwidth of 1st formant, in Hz
  DF1  v    0     0   100  Change in F1 during open portion of period, in Hz
  DB1  v    0     0   400  Change in B1 during open portion of period, in Hz
  F2   V  550  1500  3000  Frequency of 2nd formant, in Hz
  B2   v   40    90  1000  Bandwidth of 2nd formant, in Hz
  F3   V 1200  2500  4800  Frequency of 3rd formant, in Hz
  B3   v   60   150  1000  Bandwidth of 3rd formant, in Hz
  F4   V 2400  3442  4990  Frequency of 4th formant, in Hz
```

```
  B4   v  100   300  1000  Bandwidth of 4th formant, in Hz
  F5   v 3000  4500  4990  Frequency of 5th formant, in Hz
  B5   v  100   400  1500  Bandwidth of 5th formant, in Hz
  F6   v 3000  5500  5500  Frequency of 6th formant, in Hz (applies if NF=6)
  B6   v  100   500  4000  Bandwidth of 6th formant, in Hz (applies if NF=6)
  FNP  v  180   280   500  Frequency of nasal pole, in Hz
  BNP  v   40    90  1000  Bandwidth of nasal pole, in Hz
  FNZ  v  180   280   800  Frequency of nasal zero, in Hz
  BNZ  v   40    90  1000  Bandwidth of nasal zero, in Hz
  FTP  v  300  2150  3000  Frequency of tracheal pole, in Hz
  BTP  v   40   180  1000  Bandwidth of tracheal pole, in Hz
  FTZ  v  300  2150  3000  Frequency of tracheal zero, in Hz
  BTZ  v   40   180  2000  Bandwidth of tracheal zero, in Hz
  A2F  V    0    30    80  Amp of fric-excited parallel 2nd formant, in dB
  A3F  v    0    27    80  Amp of fric-excited parallel 3rd formant, in dB
  A4F  v    0    10    80  Amp of fric-excited parallel 4th formant, in dB
  A5F  v    0    10    80  Amp of fric-excited parallel 5th formant, in dB
  A6F  v    0    15    80  Amp of fric-excited parallel 6th formant, in dB
  AB   V    0    45    80  Amp of fric-excited parallel bypass path, in dB
  B2F  v   40   250  1000  Bw of fric-excited parallel 2nd formant, in Hz
  B3F  v   60   300  1000  Bw of fric-excited parallel 3rd formant, in Hz
  B4F  v  100   320  1000  Bw of fric-excited parallel 4th formant, in Hz
  B5F  v  100   360  1500  Bw of fric-excited parallel 5th formant, in Hz
  B6F  v  100  1500  4000  Bw of fric-excited parallel 6th formant, in Hz
  ANV  v    0     0    80  Amp of voice-excited parallel nasal form., in dB
  A1V  v    0    60    80  Amp of voice-excited parallel 1st formant, in dB
  A2V  v    0    60    80  Amp of voice-excited parallel 2nd formant, in dB
  A3V  v    0    60    80  Amp of voice-excited parallel 3rd formant, in dB
  A4V  v    0    60    80  Amp of voice-excited parallel 4th formant, in dB
  ATV  v    0     0    80  Amp of voice-excited par tracheal formant, in dB
  AI   v    0     0    80  Amp of impulse, in dB
  FSF  v    0     0     1  Formant Spacing Filter (1=on, 0=off)
```

```
Varied Parameters:
time   F0   AV   TL   AH   AF   F1   F2   F3   F4   A2F   AB
   0    0    0    5    0    0  380 1145 2500 3500   40   38
   5    0    0    5    0    0  380 1139 2500 3500   40   38
  10    0    0    5    0    0  380 1133 2500 3500   40   38
  15    0    0    5    0    0  380 1127 2500 3500   40   38
  20    0    0    5    0    0  380 1121 2500 3500   40   38
  25 1220   55    5   35    0  380 1115 2500 3500   40   38
  30 1220   60    5   40    0  380 1138 2500 3500   40   38
  35 1200   60    5   40    0  380 1133 2500 3500   40   38
  40 1176   60    5   40    0  380 1129 2500 3500   40   38
  45 1153   60    5   40    0  380 1125 2500 3500   40   38
  50 1130   60    5   40    0  380 1120 2500 3500   40   38
  55 1105   60    5   40    0  355 1116 2500 3500   40   38
  60 1080   55    5   40   45  330 1112 2500 3500   40   38
  65    0    0    5   40   50  330 1125 2500 3500   40   38
  70    0    0    5   40   60  330 1139 2500 3500   40   38
  75    0    0    5   40   60  330 1152 2500 3500   40   38
  80    0    0    5   36   60  330 1166 2500 3500   40   38
  85    0    0    5   33   60  330 1179 2500 3500   40   38
  90    0    0    5   30   60  330 1193 2500 3500   40   38
  95    0    0    5   30   60  330 1206 2500 3500   40   38
 100    0    0    5   30   60  330 1220 2500 3500   40   38
 105    0    0    5   30   60  330 1234 2500 3500   40   38
 110    0    0    5   30   60  330 1247 2500 3500   40   38
 115    0    0    5   30   60  330 1261 2500 3500   40   38
 120    0    0    5   30   60  330 1274 2500 3500   40   38
 125    0    0    5   30   60  330 1288 2500 3500   40   38
 130    0    0    5   30   60  330 1301 2500 3500   40   38
 135    0    0    5   30   60  330 1308 2500 3500   40   38
 140    0    0    5   30   60  330 1315 2500 3500   40   38
 145    0    0    5   32   60  330 1322 2500 3500   40   38
 150    0    0    5   35   60  330 1329 2500 3500   40   38
```

155	0	0	5	37	60	330	1336	2500	3500	40	38
160	0	0	5	40	60	330	1343	2500	3500	40	38
165	0	0	5	43	65	330	1350	2500	3500	40	38
170	0	0	5	45	65	330	1358	2500	3500	41	38
175	0	0	5	45	65	330	1365	2500	3500	43	38
180	0	0	5	45	65	330	1372	2500	3500	45	38
185	0	0	5	45	55	330	1379	2500	3500	45	38
190	0	0	5	45	45	330	1386	2500	3500	35	38
195	0	0	5	45	40	330	1393	2500	3500	35	38
200	0	0	5	45	35	330	1400	2500	3500	35	38
205	1260	55	5	45	0	384	1411	2500	3500	35	38
210	1280	60	5	45	0	488	1423	2500	3500	35	38
215	1300	65	5	45	0	494	1424	2500	3500	35	38
220	1320	65	5	45	0	501	1426	2500	3500	35	38
225	1340	65	5	45	0	508	1428	2500	3500	35	38
230	1340	65	5	45	0	515	1430	2500	3500	35	38
235	1340	65	5	45	0	522	1445	2500	3500	35	38
240	1335	65	5	45	0	529	1461	2500	3500	35	38
245	1330	65	5	45	0	536	1476	2500	3500	35	38
250	1320	65	5	45	0	543	1492	2500	3500	35	38
255	1310	64	5	45	0	550	1507	2500	3500	35	38
260	1300	63	5	45	0	553	1523	2500	3500	35	38
265	1286	63	5	45	0	557	1538	2500	3500	35	38
270	1273	62	5	45	0	561	1554	2500	3500	35	38
275	1260	61	5	45	0	564	1569	2500	3500	35	38
280	1246	61	5	45	0	568	1585	2500	3500	35	38
285	1233	60	5	45	0	572	1600	2500	3500	35	38
290	1220	59	5	45	0	575	1597	2500	3500	35	38
295	1206	59	5	45	0	579	1594	2500	3500	35	38
300	1192	58	5	45	0	583	1592	2500	3500	35	38
305	1178	57	5	45	0	586	1589	2500	3500	35	38
310	1165	57	5	45	0	590	1587	2500	3500	35	38
315	1151	56	5	45	0	594	1584	2500	3500	35	38
320	1137	55	5	42	0	597	1582	2500	3500	35	38
325	1123	55	5	40	0	601	1579	2500	3500	35	38
330	1110	52	5	37	0	605	1577	2500	3500	35	38
335	1094	50	5	35	0	608	1574	2500	3500	35	38
340	1078	0	5	0	0	612	1571	2500	3500	35	38
345	1062	0	5	0	0	616	1569	2500	3500	35	38
350	1046	0	5	0	0	619	1566	2500	3500	35	38
355	1030	0	5	0	0	619	1564	2500	3500	35	38
360	1002	0	5	0	0	619	1561	2500	3500	35	38
365	975	0	5	0	0	619	1559	2500	3500	35	38
370	0	0	5	0	0	619	1556	2500	3500	35	38
375	0	0	5	0	0	619	1554	2500	3500	35	38
380	0	0	5	0	0	619	1551	2500	3500	35	38
385	0	0	5	0	0	619	1549	2500	3500	35	38
390	0	0	5	0	0	619	1546	2500	3500	35	38
395	0	0	5	0	0	619	1543	2500	3500	35	38
400	0	0	5	0	0	619	1541	2500	3500	35	38
405	0	0	5	0	0	619	1538	2500	3500	35	38
410	0	0	5	0	0	619	1536	2500	3500	35	38
415	0	0	5	0	0	619	1533	2500	3500	35	38
420	0	0	5	0	0	619	1531	2500	3500	35	38
425	0	0	5	0	0	619	1528	2500	3500	35	38
430	0	0	5	0	0	619	1526	2500	3500	35	38
435	0	0	5	0	0	619	1523	2500	3500	35	38
440	0	0	5	0	0	619	1521	2500	3500	35	38
445	0	0	5	0	0	619	1518	2500	3500	35	38
450	0	0	5	0	0	619	1515	2500	3500	35	38
455	0	0	5	0	0	619	1513	2500	3500	35	38
460	0	0	5	0	0	619	1510	2500	3500	35	38
465	0	0	5	0	0	619	1508	2500	3500	35	38
470	0	0	5	0	0	619	1505	2500	3500	35	38
475	0	0	5	0	0	619	1503	2500	3500	35	38
480	0	0	5	0	0	619	1500	2500	3500	35	38
485	0	0	5	0	0	619	1498	2500	3500	35	38
490	0	0	5	0	0	619	1495	2500	3500	35	38
495	0	0	5	0	0	619	1493	2500	3500	35	38

C.1.2 Steady Noise

```
Synthesis specification for file: 'fehsteady.wav'  Tue May 24 14:32:23 1994

KLSYN93   Version 2.0  April 2,1993 M.M.(original program by D.H. Klatt)

Max output signal (overload if greater than 0.0 dB) is  -7.2 dB
Total number of waveform samples = 6000


  CURRENT CONFIGURATION:
    63 parameters

  SYM V/C  MIN  VAL   MAX  DESCRIPTION
-------------------------------------------
  DU   C   30   500  5000  Duration of the utterance, in msec
  UI   C    1     5    20  Update interval for parameter reset, in msec
  SR   C 5000 12000 20000  Output sampling rate, in samples/sec
  NF   C    1     6     6  Number of formants in cascade branch
  SS   C    1     2     4  Source Switch 1:Impulse 2:Natural 3:Anantha 4:LF
  RS   C    1     8  8191  Random seed (initial value of random # generator)
  SB   C    0     1     1  Same noise burst, reset RS if AF=AH=0, 0=no,1=yes
  CP   C    0     0     1  0=Cascade, 1=Parallel tract excitation by AV
  OS   C    0     0    20  Output selector (0=normal,1=voicing source,...)
  GV   C    0    55    80  Overall gain scale factor for AV, in dB
  GH   C    0    60    80  Overall gain scale factor for AH, in dB
  GF   C    0    70    80  Overall gain scale factor for AF, in dB
  GI   C    0    60    80  Overall gain scale factor for AI, in dB
  F0   V    0  1000  5000  Fundamental frequency, in tenths of a Hz
  AV   V    0    60    80  Amplitude of voicing, in dB
  OQ   v   10    50    99  Open quotient (voicing open-time/period), in %
  SQ   v  100   200   500  Speed quotient (rise/fall time, LF model), in %
  TL   V    0     0    41  Extra tilt of voicing spectrum, dB down @ 3 kHz
  FL   v    0     0   100  Flutter (random fluct in f0), in % of maximum
  DI   v    0     0   100  Diplophonia (alt periods closer), in % of max
  AH   V    0     0    80  Amplitude of aspiration, in dB
  AF   V    0     0    80  Amplitude of frication, in dB
  F1   V  180   500  1300  Frequency of 1st formant, in Hz
  B1   v   30    60  1000  Bandwidth of 1st formant, in Hz
  DF1  v    0     0   100  Change in F1 during open portion of period, in Hz
  DB1  v    0     0   400  Change in B1 during open portion of period, in Hz
  F2   V  550  1500  3000  Frequency of 2nd formant, in Hz
  B2   v   40    90  1000  Bandwidth of 2nd formant, in Hz
  F3   V 1200  2500  4800  Frequency of 3rd formant, in Hz
  B3   v   60   150  1000  Bandwidth of 3rd formant, in Hz
  F4   V 2400  3442  4990  Frequency of 4th formant, in Hz
  B4   v  100   300  1000  Bandwidth of 4th formant, in Hz
  F5   v 3000  4500  4990  Frequency of 5th formant, in Hz
  B5   v  100   400  1500  Bandwidth of 5th formant, in Hz
  F6   v 3000  5500  5500  Frequency of 6th formant, in Hz (applies if NF=6)
  B6   v  100   500  4000  Bandwidth of 6th formant, in Hz (applies if NF=6)
  FNP  v  180   280   500  Frequency of nasal pole, in Hz
  BNP  v   40    90  1000  Bandwidth of nasal pole, in Hz
  FNZ  v  180   280   800  Frequency of nasal zero, in Hz
  BNZ  v   40    90  1000  Bandwidth of nasal zero, in Hz
  FTP  v  300  2150  3000  Frequency of tracheal pole, in Hz
  BTP  v   40   180  1000  Bandwidth of tracheal pole, in Hz
  FTZ  v  300  2150  3000  Frequency of tracheal zero, in Hz
  BTZ  v   40   180  2000  Bandwidth of tracheal zero, in Hz
  A2F  V    0    30    80  Amp of fric-excited parallel 2nd formant, in dB
  A3F  v    0    27    80  Amp of fric-excited parallel 3rd formant, in dB
  A4F  v    0    10    80  Amp of fric-excited parallel 4th formant, in dB
```

```
  A5F  v    0    10    80  Amp of fric-excited parallel 5th formant, in dB
  A6F  v    0    15    80  Amp of fric-excited parallel 6th formant, in dB
  AB   V    0    45    80  Amp of fric-excited parallel bypass path, in dB
  B2F  v   40   250  1000  Bw of fric-excited parallel 2nd formant, in Hz
  B3F  v   60   300  1000  Bw of fric-excited parallel 3rd formant, in Hz
  B4F  v  100   320  1000  Bw of fric-excited parallel 4th formant, in Hz
  B5F  v  100   360  1500  Bw of fric-excited parallel 5th formant, in Hz
  B6F  v  100  1500  4000  Bw of fric-excited parallel 6th formant, in Hz
  ANV  v    0     0    80  Amp of voice-excited parallel nasal form., in dB
  A1V  v    0    60    80  Amp of voice-excited parallel 1st formant, in dB
  A2V  v    0    60    80  Amp of voice-excited parallel 2nd formant, in dB
  A3V  v    0    60    80  Amp of voice-excited parallel 3rd formant, in dB
  A4V  v    0    60    80  Amp of voice-excited parallel 4th formant, in dB
  ATV  v    0     0    80  Amp of voice-excited par tracheal formant, in dB
  AI   v    0     0    80  Amp of impulse, in dB
  FSF  v    0     0     1  Formant Spacing Filter (1=on, 0=off)
```

```
Varied Parameters:
time   F0   AV   TL   AH   AF   F1   F2   F3   F4   A2F   AB
   0    0    0    5    0    0  380 1145 2500 3500   40   38
   5    0    0    5    0    0  380 1139 2500 3500   40   38
  10    0    0    5    0    0  380 1133 2500 3500   40   38
  15    0    0    5    0    0  380 1127 2500 3500   40   38
  20    0    0    5    0    0  380 1121 2500 3500   40   38
  25 1220   55    5   35    0  380 1115 2500 3500   40   38
  30 1220   60    5   40    0  380 1138 2500 3500   40   38
  35 1200   60    5   40    0  380 1133 2500 3500   40   38
  40 1176   60    5   40    0  380 1129 2500 3500   40   38
  45 1153   60    5   40    0  380 1125 2500 3500   40   38
  50 1130   60    5   40    0  380 1120 2500 3500   40   38
  55 1105   60    5   40    0  355 1116 2500 3500   40   38
  60 1080   55    5   35   45  330 1112 2500 3500   40   38
  65    0    0    5    0   50  330 1125 2500 3500   40   38
  70    0    0    5    0   60  330 1139 2500 3500   40   38
  75    0    0    5    0   60  330 1152 2500 3500   40   38
  80    0    0    5    0   60  330 1166 2500 3500   40   38
  85    0    0    5    0   60  330 1179 2500 3500   40   38
  90    0    0    5    0   60  330 1193 2500 3500   40   38
  95    0    0    5    0   60  330 1206 2500 3500   40   38
 100    0    0    5    0   60  330 1220 2500 3500   40   38
 105    0    0    5    0   60  330 1234 2500 3500   40   38
 110    0    0    5    0   60  330 1247 2500 3500   40   38
 115    0    0    5    0   60  330 1261 2500 3500   40   38
 120    0    0    5    0   60  330 1274 2500 3500   40   38
 125    0    0    5    0   60  330 1288 2500 3500   40   38
 130    0    0    5    0   60  330 1301 2500 3500   40   38
 135    0    0    5    0   60  330 1308 2500 3500   40   38
 140    0    0    5    0   60  330 1315 2500 3500   40   38
 145    0    0    5    0   60  330 1322 2500 3500   40   38
 150    0    0    5    0   60  330 1329 2500 3500   40   38
 155    0    0    5    0   60  330 1336 2500 3500   40   38
 160    0    0    5    0   60  330 1343 2500 3500   40   38
 165    0    0    5    0   60  330 1350 2500 3500   40   38
 170    0    0    5    0   60  330 1358 2500 3500   40   38
 175    0    0    5    0   60  330 1365 2500 3500   40   38
 180    0    0    5    0   60  330 1372 2500 3500   40   38
 185    0    0    5    0   60  330 1379 2500 3500   40   38
 190    0    0    5    0   60  330 1386 2500 3500   40   38
 195    0    0    5    0   50  330 1393 2500 3500   40   38
 200    0    0    5    0   45  330 1400 2500 3500   40   38
 205 1260   55    5   35    0  384 1411 2500 3500   40   38
 210 1280   60    5   40    0  488 1423 2500 3500   40   38
 215 1300   65    5   45    0  494 1424 2500 3500   40   38
 220 1320   65    5   45    0  501 1426 2500 3500   40   38
 225 1340   65    5   45    0  508 1428 2500 3500   40   38
 230 1340   65    5   45    0  515 1430 2500 3500   40   38
```

```
235 1340   65    5   45    0  522 1445 2500 3500   40   38
240 1335   65    5   45    0  529 1461 2500 3500   40   38
245 1330   65    5   45    0  536 1476 2500 3500   40   38
250 1320   65    5   45    0  543 1492 2500 3500   40   38
255 1310   64    5   45    0  550 1507 2500 3500   40   38
260 1300   63    5   45    0  553 1523 2500 3500   40   38
265 1286   63    5   45    0  557 1538 2500 3500   40   38
270 1273   62    5   45    0  561 1554 2500 3500   40   38
275 1260   61    5   45    0  564 1569 2500 3500   40   38
280 1246   61    5   45    0  568 1585 2500 3500   40   38
285 1233   60    5   45    0  572 1600 2500 3500   40   38
290 1220   59    5   45    0  575 1597 2500 3500   40   38
295 1206   59    5   45    0  579 1594 2500 3500   40   38
300 1192   58    5   45    0  583 1592 2500 3500   40   38
305 1178   57    5   45    0  586 1589 2500 3500   40   38
310 1165   57    5   45    0  590 1587 2500 3500   40   38
315 1151   56    5   45    0  594 1584 2500 3500   40   38
320 1137   55    5   42    0  597 1582 2500 3500   40   38
325 1123   55    5   40    0  601 1579 2500 3500   40   38
330 1110   52    5   37    0  605 1577 2500 3500   40   38
335 1094   50    5   35    0  608 1574 2500 3500   40   38
340 1078    0    5    0    0  612 1571 2500 3500   40   38
345 1062    0    5    0    0  616 1569 2500 3500   40   38
350 1046    0    5    0    0  619 1566 2500 3500   40   38
355 1030    0    5    0    0  619 1564 2500 3500   40   38
360 1002    0    5    0    0  619 1561 2500 3500   40   38
365  975    0    5    0    0  619 1559 2500 3500   40   38
370    0    0    5    0    0  619 1556 2500 3500   40   38
375    0    0    5    0    0  619 1554 2500 3500   40   38
380    0    0    5    0    0  619 1551 2500 3500   40   38
385    0    0    5    0    0  619 1549 2500 3500   40   38
390    0    0    5    0    0  619 1546 2500 3500   40   38
395    0    0    5    0    0  619 1543 2500 3500   40   38
400    0    0    5    0    0  619 1541 2500 3500   40   38
405    0    0    5    0    0  619 1538 2500 3500   40   38
410    0    0    5    0    0  619 1536 2500 3500   40   38
415    0    0    5    0    0  619 1533 2500 3500   40   38
420    0    0    5    0    0  619 1531 2500 3500   40   38
425    0    0    5    0    0  619 1528 2500 3500   40   38
430    0    0    5    0    0  619 1526 2500 3500   40   38
435    0    0    5    0    0  619 1523 2500 3500   40   38
440    0    0    5    0    0  619 1521 2500 3500   40   38
445    0    0    5    0    0  619 1518 2500 3500   40   38
450    0    0    5    0    0  619 1515 2500 3500   40   38
455    0    0    5    0    0  619 1513 2500 3500   40   38
460    0    0    5    0    0  619 1510 2500 3500   40   38
465    0    0    5    0    0  619 1508 2500 3500   40   38
470    0    0    5    0    0  619 1505 2500 3500   40   38
475    0    0    5    0    0  619 1503 2500 3500   40   38
480    0    0    5    0    0  619 1500 2500 3500   40   38
485    0    0    5    0    0  619 1498 2500 3500   40   38
490    0    0    5    0    0  619 1495 2500 3500   40   38
495    0    0    5    0    0  619 1493 2500 3500   40   38
```

C.2 Synthetic /əfɑ/

C.2.1 Time-varying Noise

```
Synthesis specification for file: 'faatv.wav'  Fri Jun  3 11:43:42 1994

KLSYN93   Version 2.0  April 2,1993 M.M.(original program by D.H. Klatt)

Max output signal (overload if greater than 0.0 dB) is  -6.3 dB
Total number of waveform samples = 6000
```

```
CURRENT CONFIGURATION:
  63 parameters

SYM V/C  MIN  VAL   MAX  DESCRIPTION
-----------------------------------------------
 DU   C   30   500  5000  Duration of the utterance, in msec
 UI   C    1     5    20  Update interval for parameter reset, in msec
 SR   C 5000 12000 20000  Output sampling rate, in samples/sec
 NF   C    1     6     6  Number of formants in cascade branch
 SS   C    1     2     4  Source Switch 1:Impulse 2:Natural 3:Anantha 4:LF
 RS   C    1     8  8191  Random seed (initial value of random # generator)
 SB   C    0     1     1  Same noise burst, reset RS if AF=AH=0, 0=no,1=yes
 CP   C    0     0     1  0=Cascade, 1=Parallel tract excitation by AV
 OS   C    0     0    20  Output selector (0=normal,1=voicing source,...)
 GV   C    0    55    80  Overall gain scale factor for AV, in dB
 GH   C    0    60    80  Overall gain scale factor for AH, in dB
 GF   C    0    60    80  Overall gain scale factor for AF, in dB
 GI   C    0    60    80  Overall gain scale factor for AI, in dB
 F0   V    0  1000  5000  Fundamental frequency, in tenths of a Hz
 AV   V    0    60    80  Amplitude of voicing, in dB
 OQ   v   10    50    99  Open quotient (voicing open-time/period), in %
 SQ   v  100   200   500  Speed quotient (rise/fall time, LF model), in %
 TL   V    0     0    41  Extra tilt of voicing spectrum, dB down @ 3 kHz
 FL   v    0     0   100  Flutter (random fluct in f0), in % of maximum
 DI   v    0     0   100  Diplophonia (alt periods closer), in % of max
 AH   V    0     0    80  Amplitude of aspiration, in dB
 AF   V    0     0    80  Amplitude of frication, in dB
 F1   V  180   500  1300  Frequency of 1st formant, in Hz
 B1   v   30    60  1000  Bandwidth of 1st formant, in Hz
 DF1  v    0     0   100  Change in F1 during open portion of period, in Hz
 DB1  v    0     0   400  Change in B1 during open portion of period, in Hz
 F2   V  550  1500  3000  Frequency of 2nd formant, in Hz
 B2   v   40    90  1000  Bandwidth of 2nd formant, in Hz
 F3   V 1200  2500  4800  Frequency of 3rd formant, in Hz
 B3   v   60   150  1000  Bandwidth of 3rd formant, in Hz
 F4   V 2400  3442  4990  Frequency of 4th formant, in Hz
 B4   v  100   300  1000  Bandwidth of 4th formant, in Hz
 F5   v 3000  4500  4990  Frequency of 5th formant, in Hz
 B5   v  100   400  1500  Bandwidth of 5th formant, in Hz
 F6   v 3000  5500  5500  Frequency of 6th formant, in Hz (applies if NF=6)
 B6   v  100   500  4000  Bandwidth of 6th formant, in Hz (applies if NF=6)
 FNP  v  180   280   500  Frequency of nasal pole, in Hz
 BNP  v   40    90  1000  Bandwidth of nasal pole, in Hz
 FNZ  v  180   280   800  Frequency of nasal zero, in Hz
 BNZ  v   40    90  1000  Bandwidth of nasal zero, in Hz
 FTP  v  300  2150  3000  Frequency of tracheal pole, in Hz
 BTP  v   40   180  1000  Bandwidth of tracheal pole, in Hz
 FTZ  v  300  2150  3000  Frequency of tracheal zero, in Hz
 BTZ  v   40   180  2000  Bandwidth of tracheal zero, in Hz
 A2F  V    0    30    80  Amp of fric-excited parallel 2nd formant, in dB
 A3F  v    0    23    80  Amp of fric-excited parallel 3rd formant, in dB
 A4F  v    0    10    80  Amp of fric-excited parallel 4th formant, in dB
 A5F  v    0    10    80  Amp of fric-excited parallel 5th formant, in dB
 A6F  v    0    15    80  Amp of fric-excited parallel 6th formant, in dB
 AB   V    0    45    80  Amp of fric-excited parallel bypass path, in dB
 B2F  v   40   250  1000  Bw of fric-excited parallel 2nd formant, in Hz
 B3F  v   60   300  1000  Bw of fric-excited parallel 3rd formant, in Hz
 B4F  v  100   320  1000  Bw of fric-excited parallel 4th formant, in Hz
 B5F  v  100   360  1500  Bw of fric-excited parallel 5th formant, in Hz
 B6F  v  100  1500  4000  Bw of fric-excited parallel 6th formant, in Hz
 ANV  v    0     0    80  Amp of voice-excited parallel nasal form., in dB
 A1V  v    0    60    80  Amp of voice-excited parallel 1st formant, in dB
 A2V  v    0    60    80  Amp of voice-excited parallel 2nd formant, in dB
 A3V  v    0    60    80  Amp of voice-excited parallel 3rd formant, in dB
 A4V  v    0    60    80  Amp of voice-excited parallel 4th formant, in dB
 ATV  v    0     0    80  Amp of voice-excited par tracheal formant, in dB
 AI   v    0     0    80  Amp of impulse, in dB
```

```
FSF  v    0     0     1  Formant Spacing Filter (1=on, 0=off)
```

Varied Parameters:

time	F0	AV	TL	AH	AF	F1	F2	F3	F4	A2F	AB
0	0	0	5	0	0	432	1011	2549	3032	40	45
5	0	0	5	0	0	432	1011	2546	3032	40	45
10	0	0	5	0	0	432	1011	2544	3032	40	45
15	0	0	5	0	0	431	1011	2542	3032	40	45
20	1200	0	5	0	0	431	1010	2540	3032	40	45
25	1165	55	5	35	0	431	1010	2538	3032	40	45
30	1130	60	5	40	0	430	1010	2536	3032	40	45
35	1117	60	5	40	0	430	1010	2534	3032	40	45
40	1105	60	5	40	0	430	1010	2532	3032	40	45
45	1092	60	5	40	0	430	1010	2530	3032	40	45
50	1080	60	5	40	0	416	1010	2521	3032	40	45
55	1067	60	5	40	0	403	1010	2512	3032	40	45
60	1055	55	5	40	45	348	1010	2504	3032	40	45
65	1042	0	5	40	50	294	1010	2506	3032	40	45
70	1020	0	5	40	60	302	906	2508	3032	40	45
75	0	0	5	40	60	311	906	2510	3032	40	45
80	0	0	5	36	60	319	906	2512	3032	40	45
85	0	0	5	33	60	328	906	2514	3032	40	45
90	0	0	5	30	60	337	906	2516	3032	40	45
95	0	0	5	30	60	345	906	2518	3032	40	45
100	0	0	5	30	60	354	906	2520	3032	40	45
105	0	0	5	30	60	363	906	2522	3032	40	45
110	0	0	5	30	60	371	906	2524	3032	40	45
115	0	0	5	30	60	380	906	2526	3032	40	45
120	0	0	5	30	60	389	906	2529	3032	40	45
125	0	0	5	30	60	397	906	2531	3032	40	45
130	0	0	5	30	60	406	906	2533	3032	40	45
135	0	0	5	30	60	414	906	2535	3032	40	45
140	0	0	5	30	60	423	906	2537	3032	40	45
145	0	0	5	32	60	432	906	2539	3032	40	45
150	0	0	5	35	60	440	906	2541	3032	40	45
155	0	0	5	37	60	449	906	2543	3032	40	45
160	0	0	5	40	60	458	906	2545	3032	40	45
165	0	0	5	43	65	466	906	2547	3032	40	45
170	0	0	5	45	65	475	906	2549	3032	41	45
175	0	0	5	45	65	484	906	2552	3032	43	45
180	0	0	5	45	65	492	906	2554	3032	45	45
185	0	0	5	45	55	501	906	2556	3032	45	45
190	0	0	5	45	45	509	906	2558	3032	35	45
195	0	0	5	45	40	518	906	2560	3032	35	45
200	0	0	5	45	35	527	906	2562	3032	35	45
205	1210	55	5	45	0	535	906	2564	3032	35	45
210	1210	60	5	45	0	530	906	2566	3032	35	45
215	1210	65	5	45	0	525	906	2568	3032	35	45
220	1210	65	5	45	0	529	906	2570	3032	35	45
225	1210	65	5	45	0	534	906	2569	3032	35	45
230	1210	65	5	45	0	538	915	2569	3032	35	45
235	1210	65	5	45	0	543	924	2568	3032	35	45
240	1210	65	5	45	0	549	934	2568	3032	35	45
245	1210	65	5	45	0	556	943	2567	3032	35	45
250	1210	65	5	45	0	562	952	2567	3032	35	45
255	1210	64	5	45	0	569	971	2567	3032	35	45
260	1202	63	5	45	0	576	990	2566	3032	35	45
265	1194	63	5	45	0	582	1009	2566	3032	35	45
270	1186	62	5	45	0	589	1028	2565	3032	35	45
275	1179	61	5	45	0	595	1029	2565	3032	35	45
280	1171	61	5	45	0	602	1031	2565	3032	35	45
285	1163	60	5	45	0	609	1033	2575	3032	35	45
290	1156	59	5	45	0	615	1035	2585	3032	35	45
295	1148	59	5	45	0	622	1036	2595	3032	35	45
300	1140	58	5	45	0	628	1038	2605	3032	35	45
305	1133	57	5	45	0	632	1040	2616	3032	35	45

```
310 1125   57    5   45    0  636 1042 2626 3032   35   45
315 1117   56    5   45    0  640 1043 2636 3032   35   45
320 1110   55    5   42    0  644 1045 2646 3032   35   45
325 1101   55    5   40    0  648 1047 2646 3032   35   45
330 1093   52    5   37    0  652 1049 2646 3032   35   45
335 1085   50    5   35    0  656 1051 2646 3032   35   45
340 1076    0    5    0    0  660 1053 2646 3032   35   45
345 1068    0    5    0    0  664 1056 2646 3032   35   45
350    0    0    5    0    0  668 1058 2646 3032   35   45
355    0    0    5    0    0  668 1060 2646 3032   35   45
360    0    0    5    0    0  668 1062 2646 3032   35   45
365    0    0    5    0    0  668 1064 2646 3032   35   45
370    0    0    5    0    0  668 1066 2646 3032   35   45
375    0    0    5    0    0  668 1066 2646 3032   35   45
380    0    0    5    0    0  668 1066 2646 3032   35   45
385    0    0    5    0    0  668 1066 2646 3032   35   45
390    0    0    5    0    0  668 1066 2646 3032   35   45
395    0    0    5    0    0  668 1066 2646 3032   35   45
400    0    0    5    0    0  668 1066 2646 3032   35   45
405    0    0    5    0    0  668 1066 2646 3032   35   45
410    0    0    5    0    0  668 1066 2646 3032   35   45
415    0    0    5    0    0  668 1066 2646 3032   35   45
420    0    0    5    0    0  668 1066 2646 3032   35   45
425    0    0    5    0    0  668 1066 2646 3032   35   45
430    0    0    5    0    0  668 1066 2646 3032   35   45
435    0    0    5    0    0  668 1066 2646 3032   35   45
440    0    0    5    0    0  668 1066 2646 3032   35   45
445    0    0    5    0    0  668 1066 2646 3032   35   45
450    0    0    5    0    0  668 1066 2646 3032   35   45
455    0    0    5    0    0  668 1066 2646 3032   35   45
460    0    0    5    0    0  668 1066 2646 3032   35   45
465    0    0    5    0    0  668 1066 2646 3032   35   45
470    0    0    5    0    0  668 1066 2646 3032   35   45
475    0    0    5    0    0  668 1066 2646 3032   35   45
480    0    0    5    0    0  668 1066 2646 3032   35   45
485    0    0    5    0    0  668 1066 2646 3032   35   45
490    0    0    5    0    0  668 1066 2646 3032   35   45
495    0    0    5    0    0  668 1066 2646 3032   35   45
```

C.2.2 Steady Noise

```
Synthesis specification for file: 'faasteady.wav'  Mon Jun 13 14:15:29 1994

KLSYN93   Version 2.0  April 2,1993 D.M.(original program by D.H. Klatt)

Max output signal (overload if greater than 0.0 dB) is  -6.3 dB
Total number of waveform samples = 6000


  CURRENT CONFIGURATION:
    63 parameters

  SYM V/C  MIN  VAL   MAX  DESCRIPTION
-------------------------------------------
  DU   C   30   500  5000  Duration of the utterance, in msec
  UI   C    1     5    20  Update interval for parameter reset, in msec
  SR   C 5000 12000 20000  Output sampling rate, in samples/sec
  NF   C    1     6     6  Number of formants in cascade branch
  SS   C    1     2     4  Source Switch 1:Impulse 2:Natural 3:Anantha 4:LF
  RS   C    1     8  8191  Random seed (initial value of random # generator)
  SB   C    0     1     1  Same noise burst, reset RS if AF=AH=0, 0=no,1=yes
  CP   C    0     0     1  0=Cascade, 1=Parallel tract excitation by AV
  OS   C    0     0    20  Output selector (0=normal,1=voicing source,...)
  GV   C    0    55    80  Overall gain scale factor for AV, in dB
  GH   C    0    60    80  Overall gain scale factor for AH, in dB
  GF   C    0    60    80  Overall gain scale factor for AF, in dB
  GI   C    0    60    80  Overall gain scale factor for AI, in dB
```

F0	V	0	1000	5000	Fundamental frequency, in tenths of a Hz
AV	V	0	60	80	Amplitude of voicing, in dB
OQ	v	10	50	99	Open quotient (voicing open-time/period), in %
SQ	v	100	200	500	Speed quotient (rise/fall time, LF model), in %
TL	V	0	0	41	Extra tilt of voicing spectrum, dB down @ 3 kHz
FL	v	0	0	100	Flutter (random fluct in f0), in % of maximum
DI	v	0	0	100	Diplophonia (alt periods closer), in % of max
AH	V	0	0	80	Amplitude of aspiration, in dB
AF	V	0	0	80	Amplitude of frication, in dB
F1	V	180	500	1300	Frequency of 1st formant, in Hz
B1	v	30	60	1000	Bandwidth of 1st formant, in Hz
DF1	v	0	0	100	Change in F1 during open portion of period, in Hz
DB1	v	0	0	400	Change in B1 during open portion of period, in Hz
F2	V	550	1500	3000	Frequency of 2nd formant, in Hz
B2	v	40	90	1000	Bandwidth of 2nd formant, in Hz
F3	V	1200	2500	4800	Frequency of 3rd formant, in Hz
B3	v	60	150	1000	Bandwidth of 3rd formant, in Hz
F4	V	2400	3442	4990	Frequency of 4th formant, in Hz
B4	v	100	300	1000	Bandwidth of 4th formant, in Hz
F5	v	3000	4500	4990	Frequency of 5th formant, in Hz
B5	v	100	400	1500	Bandwidth of 5th formant, in Hz
F6	v	3000	5500	5500	Frequency of 6th formant, in Hz (applies if NF=6)
B6	v	100	500	4000	Bandwidth of 6th formant, in Hz (applies if NF=6)
FNP	v	180	280	500	Frequency of nasal pole, in Hz
BNP	v	40	90	1000	Bandwidth of nasal pole, in Hz
FNZ	v	180	280	800	Frequency of nasal zero, in Hz
BNZ	v	40	90	1000	Bandwidth of nasal zero, in Hz
FTP	v	300	2150	3000	Frequency of tracheal pole, in Hz
BTP	v	40	180	1000	Bandwidth of tracheal pole, in Hz
FTZ	v	300	2150	3000	Frequency of tracheal zero, in Hz
BTZ	v	40	180	2000	Bandwidth of tracheal zero, in Hz
A2F	V	0	30	80	Amp of fric-excited parallel 2nd formant, in dB
A3F	v	0	23	80	Amp of fric-excited parallel 3rd formant, in dB
A4F	v	0	10	80	Amp of fric-excited parallel 4th formant, in dB
A5F	v	0	10	80	Amp of fric-excited parallel 5th formant, in dB
A6F	v	0	15	80	Amp of fric-excited parallel 6th formant, in dB
AB	V	0	45	80	Amp of fric-excited parallel bypass path, in dB
B2F	v	40	250	1000	Bw of fric-excited parallel 2nd formant, in Hz
B3F	v	60	300	1000	Bw of fric-excited parallel 3rd formant, in Hz
B4F	v	100	320	1000	Bw of fric-excited parallel 4th formant, in Hz
B5F	v	100	360	1500	Bw of fric-excited parallel 5th formant, in Hz
B6F	v	100	1500	4000	Bw of fric-excited parallel 6th formant, in Hz
ANV	v	0	0	80	Amp of voice-excited parallel nasal form., in dB
A1V	v	0	60	80	Amp of voice-excited parallel 1st formant, in dB
A2V	v	0	60	80	Amp of voice-excited parallel 2nd formant, in dB
A3V	v	0	60	80	Amp of voice-excited parallel 3rd formant, in dB
A4V	v	0	60	80	Amp of voice-excited parallel 4th formant, in dB
ATV	v	0	0	80	Amp of voice-excited par tracheal formant, in dB
AI	v	0	0	80	Amp of impulse, in dB
FSF	v	0	0	1	Formant Spacing Filter (1=on, 0=off)

Varied Parameters:

time	F0	AV	TL	AH	AF	F1	F2	F3	F4	A2F	AB
0	0	0	5	0	0	432	1011	2549	3032	35	45
5	0	0	5	0	0	432	1011	2546	3032	35	45
10	0	0	5	0	0	432	1011	2544	3032	35	45
15	0	0	5	0	0	431	1011	2542	3032	35	45
20	1200	0	5	0	0	431	1010	2540	3032	35	45
25	1165	55	5	35	0	431	1010	2538	3032	35	45
30	1130	60	5	40	0	430	1010	2536	3032	35	45
35	1117	60	5	40	0	430	1010	2534	3032	35	45
40	1105	60	5	40	0	430	1010	2532	3032	35	45
45	1092	60	5	40	0	430	1010	2530	3032	35	45
50	1080	60	5	40	0	416	1010	2521	3032	35	45
55	1067	60	5	40	0	403	1010	2512	3032	35	45
60	1055	55	5	35	45	348	1010	2504	3032	35	45

65	1042	0	5	0	50	294	1010	2506	3032	35	45
70	1020	0	5	0	60	302	906	2508	3032	35	45
75	0	0	5	0	60	311	906	2510	3032	35	45
80	0	0	5	0	60	319	906	2512	3032	35	45
85	0	0	5	0	60	328	906	2514	3032	35	45
90	0	0	5	0	60	337	906	2516	3032	35	45
95	0	0	5	0	60	345	906	2518	3032	35	45
100	0	0	5	0	60	354	906	2520	3032	35	45
105	0	0	5	0	60	363	906	2522	3032	35	45
110	0	0	5	0	60	371	906	2524	3032	35	45
115	0	0	5	0	60	380	906	2526	3032	35	45
120	0	0	5	0	60	389	906	2529	3032	35	45
125	0	0	5	0	60	397	906	2531	3032	35	45
130	0	0	5	0	60	406	906	2533	3032	35	45
135	0	0	5	0	60	414	906	2535	3032	35	45
140	0	0	5	0	60	423	906	2537	3032	35	45
145	0	0	5	0	60	432	906	2539	3032	35	45
150	0	0	5	0	60	440	906	2541	3032	35	45
155	0	0	5	0	60	449	906	2543	3032	35	45
160	0	0	5	0	60	458	906	2545	3032	35	45
165	0	0	5	0	60	466	906	2547	3032	35	45
170	0	0	5	0	60	475	906	2549	3032	35	45
175	0	0	5	0	60	484	906	2552	3032	35	45
180	0	0	5	0	60	492	906	2554	3032	35	45
185	0	0	5	0	55	501	906	2556	3032	35	45
190	0	0	5	0	45	509	906	2558	3032	35	45
195	0	0	5	0	40	518	906	2560	3032	35	45
200	0	0	5	0	35	527	906	2562	3032	35	45
205	1210	55	5	35	0	535	906	2564	3032	35	45
210	1210	60	5	40	0	530	906	2566	3032	35	45
215	1210	65	5	45	0	525	906	2568	3032	35	45
220	1210	65	5	45	0	529	906	2570	3032	35	45
225	1210	65	5	45	0	534	906	2569	3032	35	45
230	1210	65	5	45	0	538	915	2569	3032	35	45
235	1210	65	5	45	0	543	924	2568	3032	35	45
240	1210	65	5	45	0	549	934	2568	3032	35	45
245	1210	65	5	45	0	556	943	2567	3032	35	45
250	1210	65	5	45	0	562	952	2567	3032	35	45
255	1210	64	5	44	0	569	971	2567	3032	35	45
260	1202	63	5	43	0	576	990	2566	3032	35	45
265	1194	63	5	42	0	582	1009	2566	3032	35	45
270	1186	62	5	42	0	589	1028	2565	3032	35	45
275	1179	61	5	41	0	595	1029	2565	3032	35	45
280	1171	61	5	40	0	602	1031	2565	3032	35	45
285	1163	60	5	40	0	609	1033	2575	3032	35	45
290	1156	59	5	39	0	615	1035	2585	3032	35	45
295	1148	59	5	38	0	622	1036	2595	3032	35	45
300	1140	58	5	37	0	628	1038	2605	3032	35	45
305	1133	57	5	37	0	632	1040	2616	3032	35	45
310	1125	57	5	36	0	636	1042	2626	3032	35	45
315	1117	56	5	35	0	640	1043	2636	3032	35	45
320	1110	55	5	35	0	644	1045	2646	3032	35	45
325	1101	55	5	33	0	648	1047	2646	3032	35	45
330	1093	52	5	31	0	652	1049	2646	3032	35	45
335	1085	50	5	30	0	656	1051	2646	3032	35	45
340	1076	0	5	0	0	660	1053	2646	3032	35	45
345	1068	0	5	0	0	664	1056	2646	3032	35	45
350	0	0	5	0	0	668	1058	2646	3032	35	45
355	0	0	5	0	0	668	1060	2646	3032	35	45
360	0	0	5	0	0	668	1062	2646	3032	35	45
365	0	0	5	0	0	668	1064	2646	3032	35	45
370	0	0	5	0	0	668	1066	2646	3032	35	45
375	0	0	5	0	0	668	1066	2646	3032	35	45
380	0	0	5	0	0	668	1066	2646	3032	35	45
385	0	0	5	0	0	668	1066	2646	3032	35	45
390	0	0	5	0	0	668	1066	2646	3032	35	45
395	0	0	5	0	0	668	1066	2646	3032	35	45
400	0	0	5	0	0	668	1066	2646	3032	35	45
405	0	0	5	0	0	668	1066	2646	3032	35	45

```
410    0    0    5    0    0  668 1066 2646 3032   35   45
415    0    0    5    0    0  668 1066 2646 3032   35   45
420    0    0    5    0    0  668 1066 2646 3032   35   45
425    0    0    5    0    0  668 1066 2646 3032   35   45
430    0    0    5    0    0  668 1066 2646 3032   35   45
435    0    0    5    0    0  668 1066 2646 3032   35   45
440    0    0    5    0    0  668 1066 2646 3032   35   45
445    0    0    5    0    0  668 1066 2646 3032   35   45
450    0    0    5    0    0  668 1066 2646 3032   35   45
455    0    0    5    0    0  668 1066 2646 3032   35   45
460    0    0    5    0    0  668 1066 2646 3032   35   45
465    0    0    5    0    0  668 1066 2646 3032   35   45
470    0    0    5    0    0  668 1066 2646 3032   35   45
475    0    0    5    0    0  668 1066 2646 3032   35   45
480    0    0    5    0    0  668 1066 2646 3032   35   45
485    0    0    5    0    0  668 1066 2646 3032   35   45
490    0    0    5    0    0  668 1066 2646 3032   35   45
495    0    0    5    0    0  668 1066 2646 3032   35   45
```

C.3 Synthetic /əsɛ/

C.3.1 Time-varying Noise

```
Synthesis specification for file: 'sehtv.wav'  Tue May 31 18:56:15 1994

KLSYN93   Version 2.0  April 2,1993 M.M.(original program by D.H. Klatt)

Max output signal (overload if greater than 0.0 dB) is  -7.4 dB
Total number of waveform samples = 6000


  CURRENT CONFIGURATION:
    63 parameters

  SYM V/C  MIN  VAL   MAX  DESCRIPTION
-----------------------------------------------
  DU   C   30   500  5000  Duration of the utterance, in msec
  UI   C    1     5    20  Update interval for parameter reset, in msec
  SR   C 5000 12000 20000  Output sampling rate, in samples/sec
  NF   C    1     6     6  Number of formants in cascade branch
  SS   C    1     2     4  Source Switch 1:Impulse 2:Natural 3:Anantha 4:LF
  RS   C    1     8  8191  Random seed (initial value of random # generator)
  SB   C    0     1     1  Same noise burst, reset RS if AF=AH=0, 0=no,1=yes
  CP   C    0     0     1  0=Cascade, 1=Parallel tract excitation by AV
  OS   C    0     0    20  Output selector (0=normal,1=voicing source,...)
  GV   C    0    55    80  Overall gain scale factor for AV, in dB
  GH   C    0    60    80  Overall gain scale factor for AH, in dB
  GF   C    0    70    80  Overall gain scale factor for AF, in dB
  GI   C    0    60    80  Overall gain scale factor for AI, in dB
  F0   V    0  1000  5000  Fundamental frequency, in tenths of a Hz
  AV   V    0    60    80  Amplitude of voicing, in dB
  OQ   v   10    50    99  Open quotient (voicing open-time/period), in %
  SQ   v  100   200   500  Speed quotient (rise/fall time, LF model), in %
  TL   V    0     0    41  Extra tilt of voicing spectrum, dB down @ 3 kHz
  FL   v    0     0   100  Flutter (random fluct in f0), in % of maximum
  DI   v    0     0   100  Diplophonia (alt periods closer), in % of max
  AH   V    0     0    80  Amplitude of aspiration, in dB
  AF   V    0     0    80  Amplitude of frication, in dB
  F1   V  180   500  1300  Frequency of 1st formant, in Hz
  B1   v   30    60  1000  Bandwidth of 1st formant, in Hz
  DF1  v    0     0   100  Change in F1 during open portion of period, in Hz
  DB1  v    0     0   400  Change in B1 during open portion of period, in Hz
  F2   V  550  1500  3000  Frequency of 2nd formant, in Hz
```

```
  B2   v   40    90  1000  Bandwidth of 2nd formant, in Hz
  F3   V 1200  2500  4800  Frequency of 3rd formant, in Hz
  B3   v   60   150  1000  Bandwidth of 3rd formant, in Hz
  F4   V 2400  3442  4990  Frequency of 4th formant, in Hz
  B4   v  100   300  1000  Bandwidth of 4th formant, in Hz
  F5   V 3000  4500  4990  Frequency of 5th formant, in Hz
  B5   v  100   400  1500  Bandwidth of 5th formant, in Hz
  F6   v 3000  5500  5500  Frequency of 6th formant, in Hz (applies if NF=6)
  B6   v  100   500  4000  Bandwidth of 6th formant, in Hz (applies if NF=6)
  FNP  v  180   280   500  Frequency of nasal pole, in Hz
  BNP  v   40    90  1000  Bandwidth of nasal pole, in Hz
  FNZ  v  180   280   800  Frequency of nasal zero, in Hz
  BNZ  v   40    90  1000  Bandwidth of nasal zero, in Hz
  FTP  v  300  2150  3000  Frequency of tracheal pole, in Hz
  BTP  v   40   180  1000  Bandwidth of tracheal pole, in Hz
  FTZ  v  300  2150  3000  Frequency of tracheal zero, in Hz
  BTZ  v   40   180  2000  Bandwidth of tracheal zero, in Hz
  A2F  V    0    30    80  Amp of fric-excited parallel 2nd formant, in dB
  A3F  V    0    23    80  Amp of fric-excited parallel 3rd formant, in dB
  A4F  V    0    10    80  Amp of fric-excited parallel 4th formant, in dB
  A5F  V    0    10    80  Amp of fric-excited parallel 5th formant, in dB
  A6F  V    0    15    80  Amp of fric-excited parallel 6th formant, in dB
  AB   V    0    45    80  Amp of fric-excited parallel bypass path, in dB
  B2F  v   40   250  1000  Bw of fric-excited parallel 2nd formant, in Hz
  B3F  v   60   300  1000  Bw of fric-excited parallel 3rd formant, in Hz
  B4F  v  100   320  1000  Bw of fric-excited parallel 4th formant, in Hz
  B5F  v  100   360  1500  Bw of fric-excited parallel 5th formant, in Hz
  B6F  v  100  1500  4000  Bw of fric-excited parallel 6th formant, in Hz
  ANV  v    0     0    80  Amp of voice-excited parallel nasal form., in dB
  A1V  v    0    60    80  Amp of voice-excited parallel 1st formant, in dB
  A2V  v    0    60    80  Amp of voice-excited parallel 2nd formant, in dB
  A3V  v    0    60    80  Amp of voice-excited parallel 3rd formant, in dB
  A4V  v    0    60    80  Amp of voice-excited parallel 4th formant, in dB
  ATV  v    0     0    80  Amp of voice-excited par tracheal formant, in dB
  AI   v    0     0    80  Amp of impulse, in dB
  FSF  v    0     0     1  Formant Spacing Filter (1=on, 0=off)
```

```
Varied Parameters:
time   F0   AV   TL   AH   AF   F1   F2   F3   F4   F5   A2F   A3F   A4F   A5F   A6F   AB
   0    0    0    5    0    0  350 1361 2720 3500 4500    0    33    43    50    29    45
   5    0    0    5    0    0  350 1361 2715 3500 4505    0    33    43    50    29    45
  10    0    0    5    0    0  350 1361 2710 3500 4510    0    33    43    50    29    45
  15    0    0    5    0    0  350 1361 2705 3500 4515    0    33    43    50    29    45
  20    0    0    5    0    0  350 1362 2700 3500 4520    0    33    43    50    29    45
  25    0   55    5    0    0  350 1362 2695 3500 4525    0    33    43    50    29    45
  30 1441   57    5   40    0  350 1362 2690 3500 4530    0    33    43    50    29    45
  35 1429   60    5   40    0  366 1370 2685 3500 4535    0    33    43    50    29    45
  40 1417   60    5   40    0  383 1379 2680 3500 4540    0    33    43    50    29    45
  45 1405   60    5   40    0  400 1388 2675 3500 4545    0    33    43    50    29    45
  50 1394   60    5   40    0  390 1396 2670 3500 4550    0    33    43    50    29    45
  55 1369   60    5   40    0  381 1405 2665 3500 4555    0    33    43    50    29    45
  60 1344   60    5   40    0  371 1400 2660 3500 4560    0    33    43    50    29    45
  65 1319   60    5   40    0  362 1429 2655 3500 4565    0    33    43    50    29    45
  70 1295   60    5   40    0  353 1459 2650 3500 4570    0    33    43    50    29    45
  75 1238   57    5   40    0  320 1489 2615 3500 4575    0    33    43    50    29    45
  80 1182   55    5   40   45  288 1518 2580 3500 4580    0    33    43    50    29    45
  85 1126   50    5   40   50  256 1541 2545 3500 4585    0    33    43    50    29    45
  90    0   45    5   37   52  256 1564 2510 3500 4590    0    33    43    50    29    45
  95    0    0    5   35   54  256 1587 2513 3500 4595    0    33    43    50    29    45
 100    0    0    5   32   57  256 1611 2517 3500 4600    0    33    43    50    29    45
 105    0    0    5   30   60  256 1634 2521 3500 4605    0    33    43    50    29    45
 110    0    0    5   30   60  256 1657 2525 3500 4610    0    33    43    50    29    45
 115    0    0    5   30   60  256 1681 2529 3500 4615    0    33    43    50    29    45
 120    0    0    5   30   60  256 1704 2533 3500 4620    0    33    43    50    29    45
 125    0    0    5   30   60  256 1727 2537 3500 4625    0    33    43    50    29    45
 130    0    0    5   30   60  256 1750 2541 3500 4630    0    33    43    50    29    45
```

135	0	0	5	30	60	256	1750	2545	3500	4635	0	33	43	50	29	45
140	0	0	5	30	60	256	1750	2549	3500	4640	0	33	43	50	29	45
145	0	0	5	30	60	256	1750	2553	3500	4645	0	33	43	50	29	45
150	0	0	5	30	60	256	1750	2557	3500	4650	0	33	43	50	29	45
155	0	0	5	30	60	256	1750	2561	3500	4655	0	33	43	50	29	45
160	0	0	5	30	60	256	1750	2565	3500	4660	0	33	43	50	29	45
165	0	0	5	30	60	256	1750	2569	3500	4665	0	33	43	50	29	45
170	0	0	5	30	60	256	1750	2573	3500	4670	0	33	43	50	29	45
175	0	0	5	30	60	256	1750	2577	3500	4675	0	33	43	50	29	45
180	0	0	5	35	60	256	1750	2581	3500	4680	0	38	43	50	29	45
185	0	0	5	37	60	256	1718	2612	3500	4685	0	43	43	50	29	45
190	0	0	5	40	60	256	1687	2643	3500	4690	0	43	43	50	29	45
195	0	0	5	41	60	256	1656	2674	3500	4695	0	43	43	50	29	45
200	0	0	5	45	60	256	1625	2706	3500	4700	0	43	43	50	29	45
205	0	0	5	45	57	256	1593	2720	3500	4697	0	43	43	50	29	45
210	0	0	5	45	55	256	1562	2720	3500	4694	0	43	43	50	29	45
215	0	0	5	45	52	256	1531	2720	3500	4690	0	43	43	50	29	45
220	0	60	5	45	50	355	1500	2720	3500	4687	0	38	43	50	29	45
225	1440	65	5	45	0	455	1511	2720	3496	4683	0	33	43	50	29	45
230	1432	65	5	45	0	465	1523	2720	3492	4680	0	33	43	50	29	45
235	1424	65	5	45	0	476	1534	2720	3489	4677	0	33	43	50	29	45
240	1420	65	5	45	0	487	1546	2720	3485	4673	0	33	43	50	29	45
245	1416	65	5	45	0	497	1557	2720	3481	4670	0	33	43	50	29	45
250	1412	65	5	45	0	512	1569	2720	3478	4666	0	33	43	50	29	45
255	1408	65	5	45	0	528	1581	2720	3474	4663	0	33	43	50	29	45
260	1404	65	5	45	0	544	1592	2720	3470	4660	0	33	43	50	29	45
265	1397	65	5	45	0	546	1604	2720	3467	4656	0	33	43	50	29	45
270	1391	65	5	45	0	548	1615	2720	3463	4653	0	33	43	50	29	45
275	1385	65	5	45	0	551	1627	2720	3460	4649	0	33	43	50	29	45
280	1379	65	5	45	0	553	1638	2720	3456	4646	0	33	43	50	29	45
285	1373	65	5	45	0	556	1639	2720	3452	4643	0	33	43	50	29	45
290	1367	65	5	45	0	558	1640	2720	3449	4639	0	33	43	50	29	45
295	1361	65	5	45	0	558	1645	2720	3445	4636	0	33	43	50	29	45
300	1355	65	5	45	0	558	1650	2720	3441	4632	0	33	43	50	29	45
305	1336	65	5	45	0	558	1656	2720	3438	4629	0	33	43	50	29	45
310	1318	65	5	45	0	558	1661	2720	3434	4625	0	33	43	50	29	45
315	1299	65	5	45	0	559	1666	2720	3430	4622	0	33	43	50	29	45
320	1281	65	5	42	0	559	1652	2720	3427	4619	0	33	43	50	29	45
325	1263	65	5	40	0	559	1638	2720	3423	4615	0	33	43	50	29	45
330	1244	65	5	37	0	559	1624	2720	3420	4612	0	33	43	50	29	45
335	1226	65	5	35	0	560	1610	2720	3416	4608	0	33	43	50	29	45
340	1208	65	5	0	0	560	1596	2720	3412	4605	0	33	43	50	29	45
345	1188	65	5	0	0	560	1582	2720	3409	4602	0	33	43	50	29	45
350	1169	65	5	0	0	560	1568	2720	3405	4598	0	33	43	50	29	45
355	1150	65	5	0	0	561	1554	2720	3401	4595	0	33	43	50	29	45
360	1131	65	5	0	0	561	1540	2720	3398	4591	0	33	43	50	29	45
365	1118	65	5	0	0	561	1526	2720	3394	4588	0	33	43	50	29	45
370	1104	65	5	0	0	561	1512	2720	3390	4585	0	33	43	50	29	45
375	1090	62	5	0	0	561	1512	2720	3387	4581	0	33	43	50	29	45
380	1077	60	5	0	0	561	1512	2720	3383	4578	0	33	43	50	29	45
385	1063	57	5	0	0	561	1512	2720	3380	4574	0	33	43	50	29	45
390	1050	55	5	0	0	561	1512	2720	3376	4571	0	33	43	50	29	45
395	0	0	5	0	0	561	1512	2720	3372	4568	0	33	43	50	29	45
400	0	0	5	0	0	561	1512	2720	3369	4564	0	33	43	50	29	45
405	0	0	5	0	0	561	1512	2720	3365	4561	0	33	43	50	29	45
410	0	0	5	0	0	561	1512	2720	3361	4557	0	33	43	50	29	45
415	0	0	5	0	0	561	1512	2720	3358	4554	0	33	43	50	29	45
420	0	0	5	0	0	561	1512	2720	3354	4551	0	33	43	50	29	45
425	0	0	5	0	0	561	1512	2720	3350	4547	0	33	43	50	29	45
430	0	0	5	0	0	561	1512	2720	3347	4544	0	33	43	50	29	45
435	0	0	5	0	0	561	1512	2720	3343	4540	0	33	43	50	29	45
440	0	0	5	0	0	561	1512	2720	3340	4537	0	33	43	50	29	45
445	0	0	5	0	0	561	1512	2720	3336	4534	0	33	43	50	29	45
450	0	0	5	0	0	561	1512	2720	3332	4530	0	33	43	50	29	45
455	0	0	5	0	0	561	1512	2720	3329	4527	0	33	43	50	29	45
460	0	0	5	0	0	561	1512	2720	3325	4523	0	33	43	50	29	45
465	0	0	5	0	0	561	1512	2720	3321	4520	0	33	43	50	29	45
470	0	0	5	0	0	561	1512	2720	3318	4517	0	33	43	50	29	45
475	0	0	5	0	0	561	1512	2720	3314	4513	0	33	43	50	29	45

```
480    0    0    5    0    0  561 1512 2720 3310 4510    0   33   43   50   29   45
485    0    0    5    0    0  561 1512 2720 3307 4506    0   33   43   50   29   45
490    0    0    5    0    0  561 1512 2720 3303 4503    0   33   43   50   29   45
495    0    0    5    0    0  561 1512 2720 3300 4500    0   33   43   50   29   45
```

C.3.2 Steady Noise

```
Synthesis specification for file: 'sehsteady.wav'  Tue May 31 19:34:50 1994

KLSYN93   Version 2.0  April 2,1993 M.M.(original program by D.H. Klatt)

Max output signal (overload if greater than 0.0 dB) is  -7.4 dB
Total number of waveform samples = 6000


  CURRENT CONFIGURATION:
    63 parameters

  SYM V/C  MIN  VAL   MAX  DESCRIPTION
-----------------------------------------
  DU   C   30   500  5000  Duration of the utterance, in msec
  UI   C    1     5    20  Update interval for parameter reset, in msec
  SR   C 5000 12000 20000  Output sampling rate, in samples/sec
  NF   C    1     6     6  Number of formants in cascade branch
  SS   C    1     2     4  Source Switch 1:Impulse 2:Natural 3:Anantha 4:LF
  RS   C    1     8  8191  Random seed (initial value of random # generator)
  SB   C    0     1     1  Same noise burst, reset RS if AF=AH=0, 0=no,1=yes
  CP   C    0     0     1  0=Cascade, 1=Parallel tract excitation by AV
  OS   C    0     0    20  Output selector (0=normal,1=voicing source,...)
  GV   C    0    55    80  Overall gain scale factor for AV, in dB
  GH   C    0    60    80  Overall gain scale factor for AH, in dB
  GF   C    0    70    80  Overall gain scale factor for AF, in dB
  GI   C    0    60    80  Overall gain scale factor for AI, in dB
  F0   V    0  1000  5000  Fundamental frequency, in tenths of a Hz
  AV   V    0    60    80  Amplitude of voicing, in dB
  OQ   v   10    50    99  Open quotient (voicing open-time/period), in %
  SQ   v  100   200   500  Speed quotient (rise/fall time, LF model), in %
  TL   V    0     0    41  Extra tilt of voicing spectrum, dB down @ 3 kHz
  FL   v    0     0   100  Flutter (random fluct in f0), in % of maximum
  DI   v    0     0   100  Diplophonia (alt periods closer), in % of max
  AH   V    0     0    80  Amplitude of aspiration, in dB
  AF   V    0     0    80  Amplitude of frication, in dB
  F1   V  180   500  1300  Frequency of 1st formant, in Hz
  B1   v   30    60  1000  Bandwidth of 1st formant, in Hz
  DF1  v    0     0   100  Change in F1 during open portion of period, in Hz
  DB1  v    0     0   400  Change in B1 during open portion of period, in Hz
  F2   V  550  1500  3000  Frequency of 2nd formant, in Hz
  B2   v   40    90  1000  Bandwidth of 2nd formant, in Hz
  F3   V 1200  2500  4800  Frequency of 3rd formant, in Hz
  B3   v   60   150  1000  Bandwidth of 3rd formant, in Hz
  F4   V 2400  3442  4990  Frequency of 4th formant, in Hz
  B4   v  100   300  1000  Bandwidth of 4th formant, in Hz
  F5   V 3000  4500  4990  Frequency of 5th formant, in Hz
  B5   v  100   400  1500  Bandwidth of 5th formant, in Hz
  F6   v 3000  5500  5500  Frequency of 6th formant, in Hz (applies if NF=6)
  B6   v  100   500  4000  Bandwidth of 6th formant, in Hz (applies if NF=6)
  FNP  v  180   280   500  Frequency of nasal pole, in Hz
  BNP  v   40    90  1000  Bandwidth of nasal pole, in Hz
  FNZ  v  180   280   800  Frequency of nasal zero, in Hz
  BNZ  v   40    90  1000  Bandwidth of nasal zero, in Hz
  FTP  v  300  2150  3000  Frequency of tracheal pole, in Hz
  BTP  v   40   180  1000  Bandwidth of tracheal pole, in Hz
  FTZ  v  300  2150  3000  Frequency of tracheal zero, in Hz
  BTZ  v   40   180  2000  Bandwidth of tracheal zero, in Hz
  A2F  V    0    30    80  Amp of fric-excited parallel 2nd formant, in dB
  A3F  V    0    23    80  Amp of fric-excited parallel 3rd formant, in dB
```

A4F	V	0	10	80	Amp of fric-excited parallel 4th formant, in dB
A5F	V	0	10	80	Amp of fric-excited parallel 5th formant, in dB
A6F	V	0	15	80	Amp of fric-excited parallel 6th formant, in dB
AB	V	0	45	80	Amp of fric-excited parallel bypass path, in dB
B2F	v	40	250	1000	Bw of fric-excited parallel 2nd formant, in Hz
B3F	v	60	300	1000	Bw of fric-excited parallel 3rd formant, in Hz
B4F	v	100	320	1000	Bw of fric-excited parallel 4th formant, in Hz
B5F	v	100	360	1500	Bw of fric-excited parallel 5th formant, in Hz
B6F	v	100	1500	4000	Bw of fric-excited parallel 6th formant, in Hz
ANV	v	0	0	80	Amp of voice-excited parallel nasal form., in dB
A1V	v	0	60	80	Amp of voice-excited parallel 1st formant, in dB
A2V	v	0	60	80	Amp of voice-excited parallel 2nd formant, in dB
A3V	v	0	60	80	Amp of voice-excited parallel 3rd formant, in dB
A4V	v	0	60	80	Amp of voice-excited parallel 4th formant, in dB
ATV	v	0	0	80	Amp of voice-excited par tracheal formant, in dB
AI	v	0	0	80	Amp of impulse, in dB
FSF	v	0	0	1	Formant Spacing Filter (1=on, 0=off)

Varied Parameters:

time	F0	AV	TL	AH	AF	F1	F2	F3	F4	F5	A2F	A3F	A4F	A5F	A6F	AB
0	0	0	5	0	0	350	1361	2720	3500	4500	0	33	43	50	29	45
5	0	0	5	0	0	350	1361	2715	3500	4505	0	33	43	50	29	45
10	0	0	5	0	0	350	1361	2710	3500	4510	0	33	43	50	29	45
15	0	0	5	0	0	350	1361	2705	3500	4515	0	33	43	50	29	45
20	0	0	5	0	0	350	1362	2700	3500	4520	0	33	43	50	29	45
25	0	55	5	0	0	350	1362	2695	3500	4525	0	33	43	50	29	45
30	1441	57	5	37	0	350	1362	2690	3500	4530	0	33	43	50	29	45
35	1429	60	5	37	0	366	1370	2685	3500	4535	0	33	43	50	29	45
40	1417	60	5	37	0	383	1379	2680	3500	4540	0	33	43	50	29	45
45	1405	60	5	37	0	400	1388	2675	3500	4545	0	33	43	50	29	45
50	1394	60	5	37	0	390	1396	2670	3500	4550	0	33	43	50	29	45
55	1369	60	5	37	0	381	1405	2665	3500	4555	0	33	43	50	29	45
60	1344	60	5	37	0	371	1400	2660	3500	4560	0	33	43	50	29	45
65	1319	60	5	37	0	362	1429	2655	3500	4565	0	33	43	50	29	45
70	1295	60	5	37	0	353	1459	2650	3500	4570	0	33	43	50	29	45
75	1238	57	5	37	0	320	1489	2615	3500	4575	0	33	43	50	29	45
80	1182	55	5	35	45	288	1518	2580	3500	4580	0	33	43	50	29	45
85	1126	50	5	30	50	256	1541	2545	3500	4585	0	33	43	50	29	45
90	0	45	5	25	52	256	1564	2510	3500	4590	0	33	43	50	29	45
95	0	0	5	0	54	256	1587	2513	3500	4595	0	33	43	50	29	45
100	0	0	5	0	57	256	1611	2517	3500	4600	0	33	43	50	29	45
105	0	0	5	0	60	256	1634	2521	3500	4605	0	33	43	50	29	45
110	0	0	5	0	60	256	1657	2525	3500	4610	0	33	43	50	29	45
115	0	0	5	0	60	256	1681	2529	3500	4615	0	33	43	50	29	45
120	0	0	5	0	60	256	1704	2533	3500	4620	0	33	43	50	29	45
125	0	0	5	0	60	256	1727	2537	3500	4625	0	33	43	50	29	45
130	0	0	5	0	60	256	1750	2541	3500	4630	0	33	43	50	29	45
135	0	0	5	0	60	256	1750	2545	3500	4635	0	33	43	50	29	45
140	0	0	5	0	60	256	1750	2549	3500	4640	0	33	43	50	29	45
145	0	0	5	0	60	256	1750	2553	3500	4645	0	33	43	50	29	45
150	0	0	5	0	60	256	1750	2557	3500	4650	0	33	43	50	29	45
155	0	0	5	0	60	256	1750	2561	3500	4655	0	33	43	50	29	45
160	0	0	5	0	60	256	1750	2565	3500	4660	0	33	43	50	29	45
165	0	0	5	0	60	256	1750	2569	3500	4665	0	33	43	50	29	45
170	0	0	5	0	60	256	1750	2573	3500	4670	0	33	43	50	29	45
175	0	0	5	0	60	256	1750	2577	3500	4675	0	33	43	50	29	45
180	0	0	5	0	60	256	1750	2581	3500	4680	0	33	43	50	29	45
185	0	0	5	0	60	256	1718	2612	3500	4685	0	33	43	50	29	45
190	0	0	5	0	60	256	1687	2643	3500	4690	0	33	43	50	29	45
195	0	0	5	0	60	256	1656	2674	3500	4695	0	33	43	50	29	45
200	0	0	5	0	60	256	1625	2706	3500	4700	0	33	43	50	29	45
205	0	0	5	0	57	256	1593	2720	3500	4697	0	33	43	50	29	45
210	0	0	5	0	55	256	1562	2720	3500	4694	0	33	43	50	29	45
215	0	0	5	0	52	256	1531	2720	3500	4690	0	33	43	50	29	45
220	0	60	5	40	50	355	1500	2720	3500	4687	0	33	43	50	29	45
225	1440	65	5	45	0	455	1511	2720	3496	4683	0	33	43	50	29	45

```
230 1432   65   5   45   0  465 1523 2720 3492 4680   0   33   43   50   29   45
235 1424   65   5   45   0  476 1534 2720 3489 4677   0   33   43   50   29   45
240 1420   65   5   45   0  487 1546 2720 3485 4673   0   33   43   50   29   45
245 1416   65   5   45   0  497 1557 2720 3481 4670   0   33   43   50   29   45
250 1412   65   5   45   0  512 1569 2720 3478 4666   0   33   43   50   29   45
255 1408   65   5   45   0  528 1581 2720 3474 4663   0   33   43   50   29   45
260 1404   65   5   45   0  544 1592 2720 3470 4660   0   33   43   50   29   45
265 1397   65   5   45   0  546 1604 2720 3467 4656   0   33   43   50   29   45
270 1391   65   5   45   0  548 1615 2720 3463 4653   0   33   43   50   29   45
275 1385   65   5   45   0  551 1627 2720 3460 4649   0   33   43   50   29   45
280 1379   65   5   45   0  553 1638 2720 3456 4646   0   33   43   50   29   45
285 1373   65   5   45   0  556 1639 2720 3452 4643   0   33   43   50   29   45
290 1367   65   5   45   0  558 1640 2720 3449 4639   0   33   43   50   29   45
295 1361   65   5   45   0  558 1645 2720 3445 4636   0   33   43   50   29   45
300 1355   65   5   45   0  558 1650 2720 3441 4632   0   33   43   50   29   45
305 1336   65   5   45   0  558 1656 2720 3438 4629   0   33   43   50   29   45
310 1318   65   5   45   0  558 1661 2720 3434 4625   0   33   43   50   29   45
315 1299   65   5   45   0  559 1666 2720 3430 4622   0   33   43   50   29   45
320 1281   65   5   45   0  559 1652 2720 3427 4619   0   33   43   50   29   45
325 1263   65   5   45   0  559 1638 2720 3423 4615   0   33   43   50   29   45
330 1244   65   5   45   0  559 1624 2720 3420 4612   0   33   43   50   29   45
335 1226   65   5   45   0  560 1610 2720 3416 4608   0   33   43   50   29   45
340 1208   65   5   45   0  560 1596 2720 3412 4605   0   33   43   50   29   45
345 1188   65   5   45   0  560 1582 2720 3409 4602   0   33   43   50   29   45
350 1169   65   5   45   0  560 1568 2720 3405 4598   0   33   43   50   29   45
355 1150   65   5   45   0  561 1554 2720 3401 4595   0   33   43   50   29   45
360 1131   65   5   45   0  561 1540 2720 3398 4591   0   33   43   50   29   45
365 1118   65   5   45   0  561 1526 2720 3394 4588   0   33   43   50   29   45
370 1104   65   5   45   0  561 1512 2720 3390 4585   0   33   43   50   29   45
375 1090   62   5   42   0  561 1512 2720 3387 4581   0   33   43   50   29   45
380 1077   60   5   40   0  561 1512 2720 3383 4578   0   33   43   50   29   45
385 1063   57   5   37   0  561 1512 2720 3380 4574   0   33   43   50   29   45
390 1050   55   5   35   0  561 1512 2720 3376 4571   0   33   43   50   29   45
395    0    0   5    0   0  561 1512 2720 3372 4568   0   33   43   50   29   45
400    0    0   5    0   0  561 1512 2720 3369 4564   0   33   43   50   29   45
405    0    0   5    0   0  561 1512 2720 3365 4561   0   33   43   50   29   45
410    0    0   5    0   0  561 1512 2720 3361 4557   0   33   43   50   29   45
415    0    0   5    0   0  561 1512 2720 3358 4554   0   33   43   50   29   45
420    0    0   5    0   0  561 1512 2720 3354 4551   0   33   43   50   29   45
425    0    0   5    0   0  561 1512 2720 3350 4547   0   33   43   50   29   45
430    0    0   5    0   0  561 1512 2720 3347 4544   0   33   43   50   29   45
435    0    0   5    0   0  561 1512 2720 3343 4540   0   33   43   50   29   45
440    0    0   5    0   0  561 1512 2720 3340 4537   0   33   43   50   29   45
445    0    0   5    0   0  561 1512 2720 3336 4534   0   33   43   50   29   45
450    0    0   5    0   0  561 1512 2720 3332 4530   0   33   43   50   29   45
455    0    0   5    0   0  561 1512 2720 3329 4527   0   33   43   50   29   45
460    0    0   5    0   0  561 1512 2720 3325 4523   0   33   43   50   29   45
465    0    0   5    0   0  561 1512 2720 3321 4520   0   33   43   50   29   45
470    0    0   5    0   0  561 1512 2720 3318 4517   0   33   43   50   29   45
475    0    0   5    0   0  561 1512 2720 3314 4513   0   33   43   50   29   45
480    0    0   5    0   0  561 1512 2720 3310 4510   0   33   43   50   29   45
485    0    0   5    0   0  561 1512 2720 3307 4506   0   33   43   50   29   45
490    0    0   5    0   0  561 1512 2720 3303 4503   0   33   43   50   29   45
495    0    0   5    0   0  561 1512 2720 3300 4500   0   33   43   50   29   45
```

C.4 Synthetic /əsɑ/

C.4.1 Time-varying Noise

```
Synthesis specification for file: 'saatv.wav'  Wed Jun 22 15:58:27 1994

KLSYN93   Version 2.0  April 2,1993 D.H.(original program by D.H. Klatt)

Max output signal (overload if greater than 0.0 dB) is  -0.5 dB
Total number of waveform samples = 6000
```

```
CURRENT CONFIGURATION:
  63 parameters

 SYM V/C  MIN  VAL   MAX  DESCRIPTION
-------------------------------------------
 DU   C   30   500  5000  Duration of the utterance, in msec
 UI   C    1     5    20  Update interval for parameter reset, in msec
 SR   C 5000 12000 20000  Output sampling rate, in samples/sec
 NF   C    1     6     6  Number of formants in cascade branch
 SS   C    1     2     4  Source Switch 1:Impulse 2:Natural 3:Anantha 4:LF
 RS   C    1     8  8191  Random seed (initial value of random # generator)
 SB   C    0     1     1  Same noise burst, reset RS if AF=AH=0, 0=no,1=yes
 CP   C    0     0     1  0=Cascade, 1=Parallel tract excitation by AV
 OS   C    0     0    20  Output selector (0=normal,1=voicing source,...)
 GV   C    0    60    80  Overall gain scale factor for AV, in dB
 GH   C    0    60    80  Overall gain scale factor for AH, in dB
 GF   C    0    65    80  Overall gain scale factor for AF, in dB
 GI   C    0    60    80  Overall gain scale factor for AI, in dB
 F0   V    0  1000  5000  Fundamental frequency, in tenths of a Hz
 AV   V    0    60    80  Amplitude of voicing, in dB
 OQ   v   10    50    99  Open quotient (voicing open-time/period), in %
 SQ   v  100   200   500  Speed quotient (rise/fall time, LF model), in %
 TL   V    0     0    41  Extra tilt of voicing spectrum, dB down @ 3 kHz
 FL   v    0     0   100  Flutter (random fluct in f0), in % of maximum
 DI   v    0     0   100  Diplophonia (alt periods closer), in % of max
 AH   V    0     0    80  Amplitude of aspiration, in dB
 AF   V    0     0    80  Amplitude of frication, in dB
 F1   V  180   500  1300  Frequency of 1st formant, in Hz
 B1   v   30    60  1000  Bandwidth of 1st formant, in Hz
 DF1  v    0     0   100  Change in F1 during open portion of period, in Hz
 DB1  v    0     0   400  Change in B1 during open portion of period, in Hz
 F2   V  550  1500  3000  Frequency of 2nd formant, in Hz
 B2   v   40    90  1000  Bandwidth of 2nd formant, in Hz
 F3   V 1200  2500  4800  Frequency of 3rd formant, in Hz
 B3   v   60   150  1000  Bandwidth of 3rd formant, in Hz
 F4   V 2400  3442  4990  Frequency of 4th formant, in Hz
 B4   v  100   300  1000  Bandwidth of 4th formant, in Hz
 F5   V 3000  4500  4990  Frequency of 5th formant, in Hz
 B5   v  100   400  1500  Bandwidth of 5th formant, in Hz
 F6   v 3000  5500  5500  Frequency of 6th formant, in Hz (applies if NF=6)
 B6   v  100   500  4000  Bandwidth of 6th formant, in Hz (applies if NF=6)
 FNP  v  180   280   500  Frequency of nasal pole, in Hz
 BNP  v   40    90  1000  Bandwidth of nasal pole, in Hz
 FNZ  v  180   280   800  Frequency of nasal zero, in Hz
 BNZ  v   40    90  1000  Bandwidth of nasal zero, in Hz
 FTP  v  300  2150  3000  Frequency of tracheal pole, in Hz
 BTP  v   40   180  1000  Bandwidth of tracheal pole, in Hz
 FTZ  v  300  2150  3000  Frequency of tracheal zero, in Hz
 BTZ  v   40   180  2000  Bandwidth of tracheal zero, in Hz
 A2F  V    0    30    80  Amp of fric-excited parallel 2nd formant, in dB
 A3F  V    0    23    80  Amp of fric-excited parallel 3rd formant, in dB
 A4F  V    0    10    80  Amp of fric-excited parallel 4th formant, in dB
 A5F  V    0    10    80  Amp of fric-excited parallel 5th formant, in dB
 A6F  V    0    15    80  Amp of fric-excited parallel 6th formant, in dB
 AB   V    0    45    80  Amp of fric-excited parallel bypass path, in dB
 B2F  v   40   250  1000  Bw of fric-excited parallel 2nd formant, in Hz
 B3F  v   60   300  1000  Bw of fric-excited parallel 3rd formant, in Hz
 B4F  v  100   320  1000  Bw of fric-excited parallel 4th formant, in Hz
 B5F  v  100   360  1500  Bw of fric-excited parallel 5th formant, in Hz
 B6F  v  100  1500  4000  Bw of fric-excited parallel 6th formant, in Hz
 ANV  v    0     0    80  Amp of voice-excited parallel nasal form., in dB
 A1V  v    0    60    80  Amp of voice-excited parallel 1st formant, in dB
 A2V  v    0    60    80  Amp of voice-excited parallel 2nd formant, in dB
 A3V  v    0    60    80  Amp of voice-excited parallel 3rd formant, in dB
 A4V  v    0    60    80  Amp of voice-excited parallel 4th formant, in dB
 ATV  v    0     0    80  Amp of voice-excited par tracheal formant, in dB
 AI   v    0     0    80  Amp of impulse, in dB
```

```
  FSF  v    0     0     1  Formant Spacing Filter (1=on, 0=off)
```

Varied Parameters:

time	F0	AV	TL	AH	AF	F1	F2	F3	F4	F5	A2F	A3F	A4F	A5F	A6F	AB
0	0	0	3	0	0	432	1011	2549	3700	4500	0	30	53	50	45	45
5	0	0	3	0	0	432	1011	2572	3700	4510	0	29	53	50	45	45
10	0	0	3	0	0	432	1011	2595	3700	4520	0	29	53	50	45	45
15	0	0	3	0	0	431	1011	2619	3700	4531	0	28	53	50	45	45
20	1113	0	3	0	0	431	1010	2642	3700	4541	0	28	53	50	45	45
25	1098	55	3	35	0	431	1010	2665	3700	4552	0	28	53	50	45	45
30	1083	60	3	40	0	430	1010	2689	3700	4562	0	27	53	50	45	45
35	1068	60	3	40	0	430	1010	2712	3700	4573	0	27	53	50	45	45
40	1053	60	3	40	0	430	1010	2735	3700	4583	0	27	53	50	45	45
45	1039	60	3	40	0	430	1010	2759	3700	4594	0	26	53	50	45	45
50	1080	60	3	40	0	416	1010	2759	3700	4604	0	26	53	50	45	45
55	0	60	3	40	0	403	1098	2759	3700	4614	0	26	53	50	45	45
60	0	60	3	40	45	348	1187	2759	3700	4625	0	25	53	50	45	45
65	0	57	3	40	50	294	1275	2759	3700	4635	0	25	53	50	45	45
70	0	55	3	40	60	302	1364	2759	3700	4608	0	25	53	50	45	45
75	0	55	3	40	60	288	1453	2759	3700	4581	0	24	53	50	45	45
80	0	50	3	40	60	256	1541	2759	3700	4554	0	24	53	50	45	45
85	0	45	3	37	60	256	1564	2759	3700	4527	0	24	53	50	45	45
90	0	0	3	35	60	256	1587	2759	3700	4500	0	23	53	50	45	45
95	0	0	3	32	60	256	1611	2759	3700	4474	0	23	53	50	45	45
100	0	0	3	30	60	256	1634	2759	3700	4471	0	23	53	50	45	45
105	0	0	3	30	60	256	1647	2759	3697	4467	0	19	53	50	45	45
110	0	0	3	30	60	256	1661	2759	3693	4464	0	15	53	50	45	45
115	0	0	3	30	60	256	1674	2759	3689	4460	0	15	53	50	45	45
120	0	0	3	30	60	256	1688	2759	3685	4457	0	15	53	50	45	45
125	0	0	3	30	60	256	1701	2759	3682	4453	0	15	53	50	45	45
130	0	0	3	30	60	256	1715	2758	3678	4449	0	15	53	50	45	45
135	0	0	3	30	60	256	1715	2758	3674	4446	0	15	53	50	45	45
140	0	0	3	30	60	256	1715	2758	3670	4442	0	15	53	50	45	45
145	0	0	3	30	60	256	1715	2758	3666	4439	0	23	53	50	45	45
150	0	0	3	30	60	256	1715	2758	3663	4435	0	35	53	50	45	45
155	0	0	3	30	60	256	1715	2758	3659	4432	0	47	53	50	45	45
160	0	0	3	30	60	256	1715	2758	3655	4428	0	47	53	50	45	45
165	0	0	3	30	60	256	1715	2758	3651	4424	0	47	53	50	45	45
170	0	0	3	30	60	256	1715	2758	3648	4421	0	47	53	50	45	45
175	0	0	3	35	60	256	1707	2758	3644	4417	0	47	53	50	45	45
180	0	0	3	37	60	256	1700	2758	3640	4414	0	47	53	50	45	45
185	0	0	3	40	60	256	1592	2758	3636	4410	0	47	53	50	45	45
190	0	0	3	41	60	256	1484	2758	3633	4407	0	47	53	50	45	45
195	0	0	3	45	60	256	1351	2758	3629	4403	0	47	53	50	45	45
200	0	0	3	45	57	256	1217	2758	3625	4400	0	47	53	50	45	45
205	0	0	3	45	55	344	1083	2758	3621	4400	0	47	53	50	45	45
210	0	0	3	45	52	432	950	2758	3618	4400	0	47	53	50	45	45
215	0	55	3	45	50	520	906	2758	3611	4400	0	47	53	50	45	45
220	0	60	3	45	45	525	906	2756	3538	4400	0	47	53	50	45	45
225	0	65	3	45	0	525	906	2754	3465	4400	0	47	53	50	45	45
230	0	65	3	45	0	529	906	2746	3347	4400	0	47	53	50	45	45
235	1213	65	3	45	0	534	906	2739	3229	4400	0	47	53	50	45	45
240	1199	65	3	45	0	538	915	2731	3231	4400	0	47	53	50	45	45
245	1185	65	3	45	0	543	924	2724	3217	4400	0	47	53	50	45	45
250	1171	65	3	45	0	549	934	2716	3203	4400	0	47	53	50	45	45
255	1165	65	3	45	0	556	943	2709	3189	4400	0	47	53	50	45	45
260	1159	65	3	45	0	562	952	2701	3175	4400	0	47	53	50	45	45
265	1153	64	3	45	0	569	971	2694	3160	4400	0	47	53	50	45	45
270	1147	63	3	45	0	576	990	2686	3146	4400	0	47	53	50	45	45
275	1141	63	3	45	0	582	1009	2679	3132	4400	0	47	53	50	45	45
280	1137	62	3	45	0	589	1028	2672	3118	4400	0	47	53	50	45	45
285	1134	61	3	45	0	595	1029	2669	3104	4400	0	47	53	50	45	45
290	1131	61	3	45	0	602	1031	2666	3113	4400	0	47	53	50	45	45
295	1127	60	3	45	0	609	1033	2663	3123	4400	0	47	53	50	45	45
300	1124	59	3	45	0	615	1035	2660	3132	4400	0	47	53	50	45	45
305	1121	59	3	45	0	622	1036	2657	3142	4400	0	47	53	50	45	45

```
310 1117   58    3   45    0  628 1038 2654 3152 4400    0   47   53   50   45   45
315 1114   57    3   45    0  632 1040 2651 3161 4400    0   47   53   50   45   45
320 1111   57    3   45    0  636 1042 2648 3171 4400    0   47   53   50   45   45
325 1107   56    3   45    0  640 1043 2645 3180 4400    0   47   53   50   45   45
330 1104   55    3   42    0  644 1045 2642 3190 4400    0   47   53   50   45   45
335 1101   55    3   40    0  648 1047 2640 3200 4400    0   47   53   50   45   45
340    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
345    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
350    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
355    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
360    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
365    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
370    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
375    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
380    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
385    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
390    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
395    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
400    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
405    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
410    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
415    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
420    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
425    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
430    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
435    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
440    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
445    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
450    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
455    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
460    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
465    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
470    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
475    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
480    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
485    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
490    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
495    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
```

C.4.2 Steady Noise

```
Synthesis specification for file: 'saasteady.wav'  Wed Jun 22 15:57:45 1994

KLSYN93   Version 2.0  April 2,1993 M.M.(original program by D.H. Klatt)

Max output signal (overload if greater than 0.0 dB) is  -0.5 dB
Total number of waveform samples = 6000


  CURRENT CONFIGURATION:
    63 parameters

  SYM V/C  MIN  VAL   MAX  DESCRIPTION
-----------------------------------------------
  DU   C   30   500  5000  Duration of the utterance, in msec
  UI   C    1     5    20  Update interval for parameter reset, in msec
  SR   C 5000 12000 20000  Output sampling rate, in samples/sec
  NF   C    1     6     6  Number of formants in cascade branch
  SS   C    1     2     4  Source Switch 1:Impulse 2:Natural 3:Anantha 4:LF
  RS   C    1     8  8191  Random seed (initial value of random # generator)
  SB   C    0     1     1  Same noise burst, reset RS if AF=AH=0, 0=no,1=yes
  CP   C    0     0     1  0=Cascade, 1=Parallel tract excitation by AV
  OS   C    0     0    20  Output selector (0=normal,1=voicing source,...)
  GV   C    0    60    80  Overall gain scale factor for AV, in dB
  GH   C    0    60    80  Overall gain scale factor for AH, in dB
```

```
  GF   C    0    65    80  Overall gain scale factor for AF, in dB
  GI   C    0    60    80  Overall gain scale factor for AI, in dB
  F0   V    0  1000  5000  Fundamental frequency, in tenths of a Hz
  AV   V    0    60    80  Amplitude of voicing, in dB
  OQ   v   10    50    99  Open quotient (voicing open-time/period), in %
  SQ   v  100   200   500  Speed quotient (rise/fall time, LF model), in %
  TL   V    0     0    41  Extra tilt of voicing spectrum, dB down @ 3 kHz
  FL   v    0     0   100  Flutter (random fluct in f0), in % of maximum
  DI   v    0     0   100  Diplophonia (alt periods closer), in % of max
  AH   V    0     0    80  Amplitude of aspiration, in dB
  AF   V    0     0    80  Amplitude of frication, in dB
  F1   V  180   500  1300  Frequency of 1st formant, in Hz
  B1   v   30    60  1000  Bandwidth of 1st formant, in Hz
  DF1  v    0     0   100  Change in F1 during open portion of period, in Hz
  DB1  v    0     0   400  Change in B1 during open portion of period, in Hz
  F2   V  550  1500  3000  Frequency of 2nd formant, in Hz
  B2   v   40    90  1000  Bandwidth of 2nd formant, in Hz
  F3   V 1200  2500  4800  Frequency of 3rd formant, in Hz
  B3   v   60   150  1000  Bandwidth of 3rd formant, in Hz
  F4   V 2400  3442  4990  Frequency of 4th formant, in Hz
  B4   v  100   300  1000  Bandwidth of 4th formant, in Hz
  F5   V 3000  4500  4990  Frequency of 5th formant, in Hz
  B5   v  100   400  1500  Bandwidth of 5th formant, in Hz
  F6   v 3000  5500  5500  Frequency of 6th formant, in Hz (applies if NF=6)
  B6   v  100   500  4000  Bandwidth of 6th formant, in Hz (applies if NF=6)
  FNP  v  180   280   500  Frequency of nasal pole, in Hz
  BNP  v   40    90  1000  Bandwidth of nasal pole, in Hz
  FNZ  v  180   280   800  Frequency of nasal zero, in Hz
  BNZ  v   40    90  1000  Bandwidth of nasal zero, in Hz
  FTP  v  300  2150  3000  Frequency of tracheal pole, in Hz
  BTP  v   40   180  1000  Bandwidth of tracheal pole, in Hz
  FTZ  v  300  2150  3000  Frequency of tracheal zero, in Hz
  BTZ  v   40   180  2000  Bandwidth of tracheal zero, in Hz
  A2F  V    0    30    80  Amp of fric-excited parallel 2nd formant, in dB
  A3F  V    0    23    80  Amp of fric-excited parallel 3rd formant, in dB
  A4F  V    0    10    80  Amp of fric-excited parallel 4th formant, in dB
  A5F  V    0    10    80  Amp of fric-excited parallel 5th formant, in dB
  A6F  V    0    15    80  Amp of fric-excited parallel 6th formant, in dB
  AB   V    0    45    80  Amp of fric-excited parallel bypass path, in dB
  B2F  v   40   250  1000  Bw of fric-excited parallel 2nd formant, in Hz
  B3F  v   60   300  1000  Bw of fric-excited parallel 3rd formant, in Hz
  B4F  v  100   320  1000  Bw of fric-excited parallel 4th formant, in Hz
  B5F  v  100   360  1500  Bw of fric-excited parallel 5th formant, in Hz
  B6F  v  100  1500  4000  Bw of fric-excited parallel 6th formant, in Hz
  ANV  v    0     0    80  Amp of voice-excited parallel nasal form., in dB
  A1V  v    0    60    80  Amp of voice-excited parallel 1st formant, in dB
  A2V  v    0    60    80  Amp of voice-excited parallel 2nd formant, in dB
  A3V  v    0    60    80  Amp of voice-excited parallel 3rd formant, in dB
  A4V  v    0    60    80  Amp of voice-excited parallel 4th formant, in dB
  ATV  v    0     0    80  Amp of voice-excited par tracheal formant, in dB
  AI   v    0     0    80  Amp of impulse, in dB
  FSF  v    0     0     1  Formant Spacing Filter (1=on, 0=off)
```

Varied Parameters:

time	F0	AV	TL	AH	AF	F1	F2	F3	F4	F5	A2F	A3F	A4F	A5F	A6F	AB
0	0	0	3	0	0	432	1011	2549	3700	4500	0	23	53	50	45	45
5	0	0	3	0	0	432	1011	2572	3700	4510	0	23	53	50	45	45
10	0	0	3	0	0	432	1011	2595	3700	4520	0	23	53	50	45	45
15	0	0	3	0	0	431	1011	2619	3700	4531	0	23	53	50	45	45
20	1113	0	3	0	0	431	1010	2642	3700	4541	0	23	53	50	45	45
25	1098	55	3	35	0	431	1010	2665	3700	4552	0	23	53	50	45	45
30	1083	60	3	40	0	430	1010	2689	3700	4562	0	23	53	50	45	45
35	1068	60	3	40	0	430	1010	2712	3700	4573	0	23	53	50	45	45
40	1053	60	3	40	0	430	1010	2735	3700	4583	0	23	53	50	45	45
45	1039	60	3	40	0	430	1010	2759	3700	4594	0	23	53	50	45	45
50	1080	60	3	40	0	416	1010	2759	3700	4604	0	23	53	50	45	45

55	0	60	3	40	0	403	1098	2759	3700	4614	0	23	53	50	45	45
60	0	60	3	40	45	348	1187	2759	3700	4625	0	23	53	50	45	45
65	0	57	3	37	50	294	1275	2759	3700	4635	0	23	53	50	45	45
70	0	55	3	35	60	302	1364	2759	3700	4608	0	23	53	50	45	45
75	0	55	3	32	60	288	1453	2759	3700	4581	0	23	53	50	45	45
80	0	50	3	30	60	256	1541	2759	3700	4554	0	23	53	50	45	45
85	0	45	3	25	60	256	1564	2759	3700	4527	0	23	53	50	45	45
90	0	0	3	0	60	256	1587	2759	3700	4500	0	23	53	50	45	45
95	0	0	3	0	60	256	1611	2759	3700	4474	0	23	53	50	45	45
100	0	0	3	0	60	256	1634	2759	3700	4471	0	23	53	50	45	45
105	0	0	3	0	60	256	1647	2759	3697	4467	0	23	53	50	45	45
110	0	0	3	0	60	256	1661	2759	3693	4464	0	23	53	50	45	45
115	0	0	3	0	60	256	1674	2759	3689	4460	0	23	53	50	45	45
120	0	0	3	0	60	256	1688	2759	3685	4457	0	23	53	50	45	45
125	0	0	3	0	60	256	1701	2759	3682	4453	0	23	53	50	45	45
130	0	0	3	0	60	256	1715	2758	3678	4449	0	23	53	50	45	45
135	0	0	3	0	60	256	1715	2758	3674	4446	0	23	53	50	45	45
140	0	0	3	0	60	256	1715	2758	3670	4442	0	23	53	50	45	45
145	0	0	3	0	60	256	1715	2758	3666	4439	0	23	53	50	45	45
150	0	0	3	0	60	256	1715	2758	3663	4435	0	23	53	50	45	45
155	0	0	3	0	60	256	1715	2758	3659	4432	0	23	53	50	45	45
160	0	0	3	0	60	256	1715	2758	3655	4428	0	23	53	50	45	45
165	0	0	3	0	60	256	1715	2758	3651	4424	0	23	53	50	45	45
170	0	0	3	0	60	256	1715	2758	3648	4421	0	23	53	50	45	45
175	0	0	3	0	60	256	1707	2758	3644	4417	0	23	53	50	45	45
180	0	0	3	0	60	256	1700	2758	3640	4414	0	23	53	50	45	45
185	0	0	3	0	60	256	1592	2758	3636	4410	0	23	53	50	45	45
190	0	0	3	0	60	256	1484	2758	3633	4407	0	23	53	50	45	45
195	0	0	3	0	60	256	1351	2758	3629	4403	0	23	53	50	45	45
200	0	0	3	0	57	256	1217	2758	3625	4400	0	23	53	50	45	45
205	0	0	3	0	55	344	1083	2758	3621	4400	0	23	53	50	45	45
210	0	0	3	0	52	432	950	2758	3618	4400	0	23	53	50	45	45
215	0	55	3	35	50	520	906	2758	3611	4400	0	23	53	50	45	45
220	0	60	3	40	45	525	906	2756	3538	4400	0	23	53	50	45	45
225	0	65	3	45	0	525	906	2754	3465	4400	0	23	53	50	45	45
230	0	65	3	45	0	529	906	2746	3347	4400	0	23	53	50	45	45
235	1213	65	3	45	0	534	906	2739	3229	4400	0	23	53	50	45	45
240	1199	65	3	45	0	538	915	2731	3231	4400	0	23	53	50	45	45
245	1185	65	3	45	0	543	924	2724	3217	4400	0	23	53	50	45	45
250	1171	65	3	45	0	549	934	2716	3203	4400	0	23	53	50	45	45
255	1165	65	3	45	0	556	943	2709	3189	4400	0	23	53	50	45	45
260	1159	65	3	45	0	562	952	2701	3175	4400	0	23	53	50	45	45
265	1153	64	3	44	0	569	971	2694	3160	4400	0	23	53	50	45	45
270	1147	63	3	43	0	576	990	2686	3146	4400	0	23	53	50	45	45
275	1141	63	3	43	0	582	1009	2679	3132	4400	0	23	53	50	45	45
280	1137	62	3	42	0	589	1028	2672	3118	4400	0	23	53	50	45	45
285	1134	61	3	41	0	595	1029	2669	3104	4400	0	23	53	50	45	45
290	1131	61	3	40	0	602	1031	2666	3113	4400	0	23	53	50	45	45
295	1127	60	3	40	0	609	1033	2663	3123	4400	0	23	53	50	45	45
300	1124	59	3	39	0	615	1035	2660	3132	4400	0	23	53	50	45	45
305	1121	59	3	38	0	622	1036	2657	3142	4400	0	23	53	50	45	45
310	1117	58	3	38	0	628	1038	2654	3152	4400	0	23	53	50	45	45
315	1114	57	3	37	0	632	1040	2651	3161	4400	0	23	53	50	45	45
320	1111	57	3	36	0	636	1042	2648	3171	4400	0	23	53	50	45	45
325	1107	56	3	36	0	640	1043	2645	3180	4400	0	23	53	50	45	45
330	1104	55	3	35	0	644	1045	2642	3190	4400	0	23	53	50	45	45
335	1101	55	3	35	0	648	1047	2640	3200	4400	0	23	53	50	45	45
340	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
345	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
350	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
355	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
360	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
365	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
370	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
375	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
380	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
385	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
390	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
395	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

400	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
405	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
410	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
415	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
420	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
425	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
430	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
435	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
440	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
445	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
450	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
455	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
460	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
465	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
470	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
475	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
480	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
485	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
490	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
495	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0