The Acoustics of Fricative Consonants

Christine Helen Shadle

Technical Report 506

March 1985

Massachusetts Institute of Technology
Research Laboratory of Electronics
Cambridge, Massachusetts 02139

The Acoustics of Fricative Consonants

Christine Helen Shadle

Technical Report 506

March 1985

Massachusetts Institute of Technology
Research Laboratory of Electronics
Cambridge, Massachusetts 02139

This work has been supported in part by the National Institutes of Health Grant 5 R01 NS04332

THE ACOUSTICS OF FRICATIVE CONSONANTS

by Christine Helen Shadle

Submitted to the Department of Electrical Engineering and Computer Science on March 21, 1985, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

ABSTRACT

The acoustic mechanism of fricative consonants was studied in the context of three domains: speech, mechanical models, and theoretical models. All fricative configurations have in common a small turbulence-producing constriction within the vocal tract. Thus, preliminary experiments were conducted using a mechanical model having this basic configuration of a constriction in a tube. Parameters such as constriction area, length, location, and degree of inlet tapering, and presence of an obstacle, were varied. It was found that acoustically the most significant parameters are the presence of an obstacle, the length of the front cavity, and the flowrate. Therefore, configurations in which only these parameters were varied, referred to as the obstacle and no-obstacle cases, were examined more thoroughly and modeled theoretically.

A source function for the obstacle case was derived from the far-field sound pressure measured when the obstacle was located in space, downstream of a constriction in a baffle. The directivity pattern produced by the obstacle in this position was similar to that of a dipole, as expected. A dipole source located inside a duct is equivalent to a pressure source in a transmission-line model, when only the longitudinal modes of a duct are considered. The filter function, corresponding to the effect of the duct on such a pressure source, was derived for the transmission-line representations of two configurations in which the obstacle was located inside a front cavity of nonzero length. The spectra predicted by this source-filter model, when compared to the far-field pressure measurements of the equivalent mechanical models, provided a very close match in both absolute sound pressure level and spectral shape. Thus in this case the presence of the duct does not significantly alter the sound source, and a simple linear source-filter model works well.

For the no-obstacle case, it was not possible to derive a source from a free-field measurement, so the validity of a source-tract model could not be checked directly. Investigation of the pressure versus flow-velocity power laws showed evidence of source-tract interaction for the no-obstacle case, but none for the obstacle case.

Spectral measures were developed that characterized the acoustic differences caused by the different types of sources for the obstacle and no-obstacle cases. Analysis of real speech in terms of these spectral measures revealed that fricatives are more similar to the obstacle than to the no-obstacle case. More complex speech-like mechanical models were developed and the acoustic characteristics of the sounds generated by those models were compared to the characteristics of speech sounds, again in terms of the spectral measures. Very good models of /s/ and /š/ were obtained by using an obstacle at right angles to the flow and varying the constriction location. For the fricatives /ϕ, f, θ/, in which the constriction is located forward in the tract, the shape of the constriction was crucial. Constrictions that allowed the jet to come in contact with a surface produced sounds that most closely resembled the analyzed examples of these fricatives. For /x/ and /ç/, in which the constriction is located 4 to 6 cm back from the mouth, a surface also caused the mechanical model spectra to become the most like the speech spectra. However, in general it appears that the constriction shape affects the far-field spectrum less as the front cavity is lengthened. The model for /s, š/ appears to be that of a series pressure source,

located at the teeth. For the fricatives other than /s,š/, a "distributed obstacle", modeled as a distributed pressure source, may be the dominant acoustic mechanism.

The whistles generated by several configurations were also investigated. Orifice tones occurred for untapered constrictions, at a greater range of constriction sizes than predicted by previous studies. The frequencies were related to the constriction length. Edgetones occurred for configurations including an obstacle, at frequencies related to the flowrate and distance to the obstacle. The occurrence, frequency and amplitude of the whistle tones were affected by the constriction shape and the presence of a tube surrounding the obstacle.

Edgetones were generated by the /s/- and /š/-like mechanical models that were similar acoustically to the whistles produced by a subject with the /s/ and /š/ configurations. An even more striking parallel between the whistles produced by speech-like models and humans was found for the typical bilabial whistles produced by /ϕ/-like configurations. Two constrictions, reproducing the role of tongue and lips, were necessary to model these whistles. The agreement in flow range, frequency, and control parameters of the holetones thus produced and the whistles generated by humans was good enough to suggest that the same acoustic mechanism occurred in both the models and the human vocal tract.

Thesis supervisor: Kenneth N. Stevens
Title: Clarence J. LeBel Professor of Electrical Engineering

To
my father
my mother
and my grandmother

for giving me the desire to learn
and
for making it possible

Acknowledgments

Doing this thesis has been of necessity a solitary task, yet, paradoxically, a time most richly populated with people. It is with delight that I acknowledge those who helped me in such a variety of ways.

First, I would like to thank Ken Stevens for teaching and advising me patiently and thoughtfully. His willingness to let me go off on tangents increased immeasurably the intellectual pleasure I derived from my work; lo and behold, some of the tangents did finally turn into a thesis.

Lou Braida, my graduate counsellor, has also been an indispensable figure of the last six years. Many thanks to him for popping in unexpectedly with interesting and relevant problems and advice, and for concrete assistance in setting up my lab.

The members of my committee, Bill Rabinowitz, Victor Zue, Sheila Widnall, and Uno Ingard, spanned beautifully orthogonal areas of interest and expertise. I thank them for their help and advice, and for reading all those drafts.

I am deeply grateful to Max Mathews, Charles Thompson, and Gunnar Fant, for helping me in a most difficult phase, that of problem definition. Their comments in the beginning stages, as well as later on, greatly influenced my approach.

I wish to thank Rich McGowan, Tom Quatieri, Louis Goldstein, Rich Goldhor, John Wyatt, and Shinji Maeda, who, when asked specific technical questions, took the time to discuss them thoroughly, and taught me a great deal as a result.

Thanks to Patrick Hosein, for his careful work in revising TBFDA; Peter Vitale, for programming the LSI-11; and Dick Lyon, for his help in designing my muffler. Thanks to Joe Perkell, Larry Frishkopf, and Bob Hillman, for their generosity in lending equipment for "a year" that became two or three. And thanks to Keith North, for explaining the intricacies of tape recording and sound level meters, and finding oscillators at a moment's notice. Thanks to Jim Byrne, for teaching me to use a machine shop, and Manny Cabal, for machining all those constrictions.

Special thanks go to the members of the Communication Biophysics Group, for help with an unfamiliar computer system, and for my extensive use of their facilities.

I thank Phyllis Vandermolen and Amy Hendrickson, for their help with figures and text formatting (see, for example, Figure 2.3, and Table 4.2).

Thanks to KS, PP, EM, AS and G for serving as subjects. (Note: upon reading this thesis, you will wonder, who was AS? AS was a subject for a recording session so preliminary that it did not appear in these pages, but for whose time I am nevertheless grateful.)

Many thanks to NIH and the LeBel Foundation, for financial support.

Finally, I want to thank the people who have supported, encouraged and expressed their faith in me so long and well. Among these, I would like especially to thank: Rich Goldhor and Stephanie Seneff, for our thesis-encounter-group lunches that, among other things, helped me to keep a sense of humor; Marie Southwick, Patti Jo Price, and Corine Bickley, for sharing so much with me, and convincing me that there is life during as well as after thesis; Anne Black, John Wyatt, Kent Pitman, Adele Proctor, and Stefanie Shattuck-Hufnagel, for celebratory dinners, a sense of perspective when I needed it, cookies in the lab at 3 A.M., and the friendship that made all of that possible; Janet Koehnke, Lorraine Delhorne, Neil Macmillan, Rosalie Uchanski, Diane Bustamante, Dan Leotta, Tom Lee, Nat Durlach, and the other members of the Communication Biophysics Group at M.I.T., who responded to my use of their facilities

with friendship, encouragement, and philosophical discussions, and in general made me feel like an honorary CBG-er; Rosalind Fine, Lois Eichler, John Rust, Yehuda and Joy, for helping me make the most out of the process of being a graduate student; Max Mathews, Sandra Pruzansky, Moise Goldstein, Marcia Bush, Gary Kopec, Karen Landahl, and Emel Gokcen, for help in spite of being miles away; and my family, Paul, Elinor, Anna and Paula Shadle and Sally and Roger Gottlieb, who found so many ways to tell me to hang in there.

Biographical Note

Christine Shadle was born in Pasadena, California, on February 6, 1954. From age seven onwards she studied classical piano, an interest that became an official avocation when she double-majored in piano performance and electrical engineering at Stanford University. Graduating in 1976 with an A.B. and M.S. respectively, she began working with Dr. Bishnu Atal at Bell Telephone Laboratories, Murray Hill, New Jersey, in the Speech and Communications Research Department. Gradually changing focus from computer music to speech analysis and synthesis, she left Bell Labs in January 1979 in order to learn more about the basics of speech research. In February 1979 she became a graduate student of Ken Stevens in the Speech Communications Group of the Department of Electrical Engineering and Computer Science at M.I.T. There she has worked on three main research projects: the intrinsic fundamental frequency of vowels in sentence context, the acoustics of whistling, and this thesis, which grew out of the work on whistling. In addition, she was a teaching assistant for courses in probability and acoustics.

Over the last six years, the ups and downs of graduate school have been, shall we say, tempered by her habit of playing the preludes and fugues of J.S. Bach's Well-Tempered Clavichord. She currently intends to switch to Bach's Goldberg Variations, to continue research in her thesis area, and to resume her favorite hobbies: chamber music, photography, hiking and sailing.

Contents

List of Figures

List of Tables

List of Symbols

c Velocity of sound, = 34480 cm/sec for dry room-temperature air.

ρ Density of air, $= 1.18 \times 10^{-3}$g/cm^3

ν Kinematic viscosity of air, = 0.15 cm^2/sec

f Frequency (Hz).

ω Angular frequency, $= f/2\pi$ rad/sec.

s Complex frequency (rad/sec).

k Propagation constant, $= \omega/c$.

λ Wavelength of sound, $= c/f$ cm.

t Time (sec).

p Instantaneous sound pressure, either the time waveform $[p(t)]$ or the complex amplitude $[p(s)]$.

P Sound power, usually total sound power (spatial average).

U DC volume velocity, cc/sec.

V Mean velocity of jet, cm/sec.

Δp Static pressure drop between two points, cm H_2O or dynes/cm^2.

R_i Incremental flow resistance, g/sec-cm^4

b Friction factor of a pipe.

Re Reynolds number, $= Vd/\nu$, where V is a representative linear flow velocity and d is a representative length; dimensionless.

M Mach number, $= V/c$, dimensionless.

S Strouhal number, $= fV/d$, f = frequency, d = relevant dimension, usually diameter of jet; dimensionless.

I Complex magnitude of current.

V Complex magnitude of voltage.

Z Complex magnitude of impedance, $= V/I$ or $= p/U$.

Y Complex magnitude of admittance, $= 1/Z$.

Z_0 Characteristic impedance of an infinitely long tube of area A, $= \rho c/A$.

d Diameter of nozzle, cm; also theoretical distance between point sources of dipole.

Ud Dipole strength, where two point sources, a distance d apart, pulse with volume velocity U.

S Dipole strength, $= Ud$; also circumference of duct, cm.

l_x Length of entity x, cm.

A_x Cross-sectional area of entity x, cm^2 (except for A_T and A_0, see below).

f When used as subscript, refers to front cavity of configuration.

b When used as subscript, refers to back cavity of configuration.

c When used as subscript, refers to constriction of configuration.

m When used as subscript, refers to mouth (downtream opening) of configuration.

g When used as subscript, refers to glottis (upstream opening) of configuration.

o When used as subscript, refers to obstacle of configuration.

l_o Distance between constriction and obstacle in configuration.

θ Angle between jet axis and microphone, with origin located at either mouth of tube or obstacle.

r Distance from mouth of tube or from obstacle to microphone.

ϕ Angle between jet axis and microphone, with origin located at image obstacle.

R Distance from image obstacle to microphone.

A_S Spectral measure of a single averaged power spectrum: absolute overall amplitude, measured in dB SPL, found by summing the squares of the sound pressures over the range 500-10200 Hz.

A_T Spectral measure of a single averaged power spectrum: total dynamic range, defined as maximum amplitude minus minimum amplitude, over the range 500-10200 Hz.

A_0 Spectral measure of a single averaged power spectrum: low-frequency dynamic range, defined by maximum amplitude over the range 500-10200 Hz, minus amplitude at 500 Hz.

/ϕ/ The unvoiced bilabial fricative, as in the italicized portion of the word *wh*ew.

/f/ The unvoiced labiodental fricative, as in the italicized portion of the word *f*in.

/θ/ The unvoiced dental fricative, as in the italicized portion of the word *th*in.

/s/ The unvoiced alveolar fricative, as in the italicized portion of the word *s*in.

/š/ The unvoiced palatal-alveolar fricative, as in the italicized portion of the word *sh*in.

/ç/ The unvoiced palatal fricative, as in the italicized portion of the German word i*ch*.

/x/ The unvoiced velar fricative, as in the italicized portion of the German word a*ch*.

/h/ The unvoiced glottal fricative, as in the italicized portion of the word *h*er.

/v/ The voiced labiodental fricative, as in the italicized portion of the word li*v*e.

/δ/ The voiced dental fricative, as in the italicized portion of the word *th*is.

/z/ The voiced alveolar fricative, as in the italicized portion of the word *z*est.

/ž/ The voiced palatal-alveolar fricative, as in the italicized portion of the word a*z*ure.

Chapter 1

Introduction and Literature Review

1.1 Introduction

Speech is produced by passing air through the vocal tract. A speaker is able to vary the volume flow of air through the tract and the position of the articulators, such as tongue and vocal folds, and by this means vary the acoustic output so as to produce the desired sequence of sounds. The different speech sounds can be broadly divided into vowels and consonants. Vowel sounds are periodic and tend to have energy concentrated in the low-frequency regions (from roughly 50 to 5000 Hz). Contrasting with vowels is the class of consonants known as fricatives, which are noisy rather than periodic, and tend to have the energy concentrated at higher frequencies (roughly 3000 to 10000 Hz). The main question addressed in this thesis is, what controls the nature of these fricative sounds? In other words, what is the acoustic mechanism for fricative consonants?

An acoustic theory of speech production, developed primarily by Fant (1960), is concerned with the articulatory-acoustic transformation for all speech sounds. It is in effect based on a set of simplifying assumptions that allow us to model the acoustical behavior of the vocal tract by a distributed linear circuit in which sound sources are independent of the system. The circuit parameters describing the model can then be used to perform synthesis, analysis or recognition of speech. Our understanding of circuit behavior coupled with the physical basis for the circuit model allows us to predict the acoustic effect of articulatory or anatomical changes.

When a vowel is being uttered, the vocal tract is relatively unconstricted and the vocal folds vibrate periodically, causing the volume of air flowing through the glottis to fluctuate periodically as well. A typical model for production of a vowel treats the vocal tract as a tube of nonuniform cross-sectional area, in which only plane-wave sound propagation is considered. Regardless of the shape of the tract at a given cross-section, only the area is incorporated in the model, a simplification that is justifiable for frequencies below about 5000 Hz. The nonuniform tube is modeled as a concatenation of short uniform tubes, each of which is represented by a transmission line. The system is a linear filter for the sound produced by the vocal folds. The waveform of the glottal volume flow becomes the excitation function for this linear filter, and is assumed to be independent of the vocal tract configuration (e.g., Flanagan et al., 1975).

For fricative consonants the acoustic mechanism is not as well understood. A fricative is produced when the vocal tract is constricted somewhere along its length enough to produce a noisy sound when air is forced through the constriction. Such a constriction is indicated by an arrow in Fig. 1.1, which shows a schematized midsagittal cross-section of the vocal tract during production of a typical fricative, /š/ (as in *sh*in). As with vowels, the location of the constriction affects the timbre of the resulting sound, as can be seen by the following sequence of fricatives, in which the constriction moves from the lips towards the glottis: /ϕ, f, θ, s, š, ç, x, h/. (These phonemes are pronounced, respectively, as the italicized portion of the following words: *wh*ew, *f*in, *th*in, *s*in, *sh*in, German i*ch*, German a*ch*, *h*it.) In addition, the vocal folds may vibrate simultaneously, generating a periodic sound at the glottis and modulating the airflow through the constriction. Examples of voicing occur in the minimal pairs /v,f/ and /z,s/ (li*v*e,

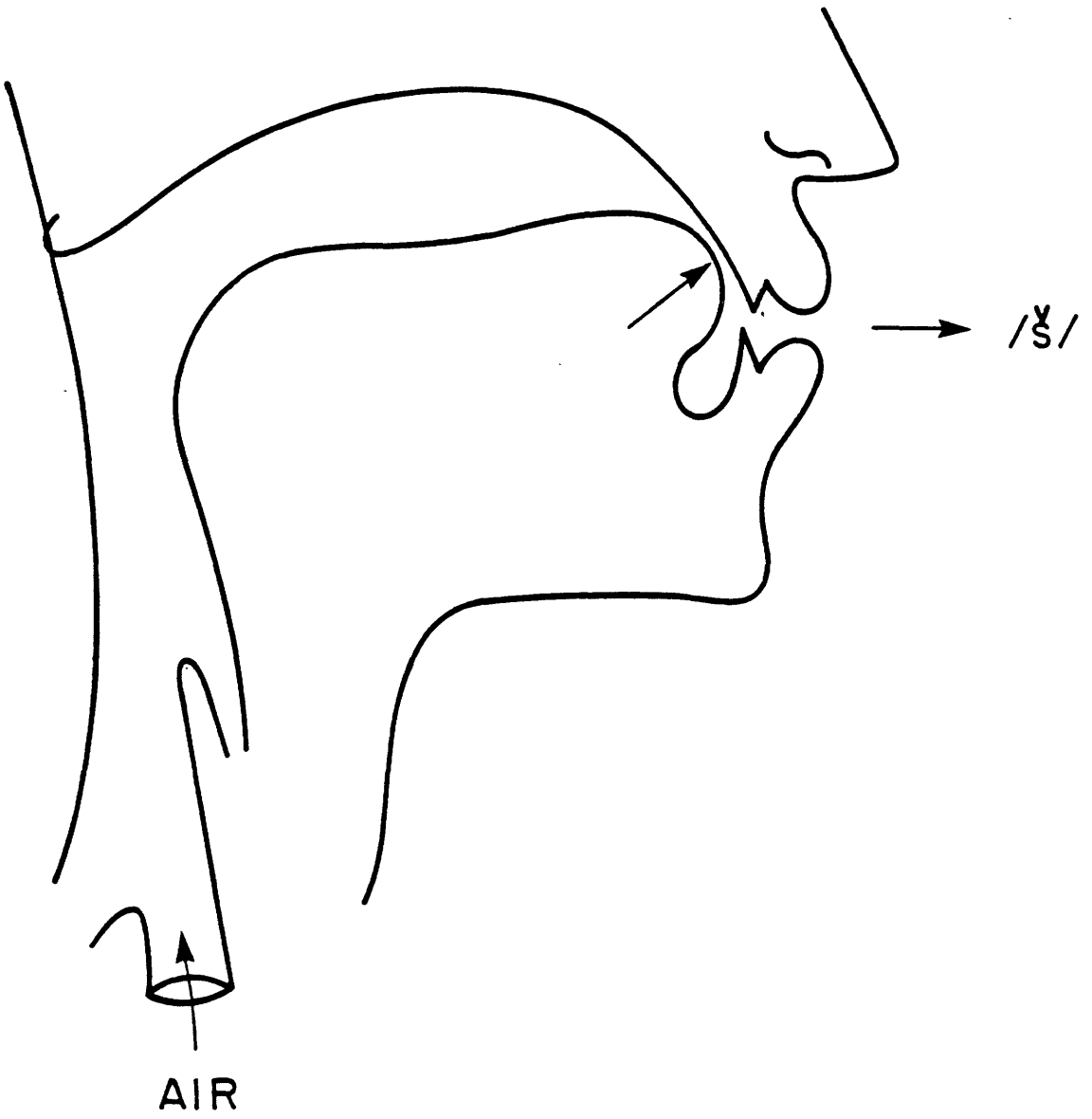


Figure 1.1: Diagram of a midsagittal cross-section of the vocal tract during the production of the fricative /š/. The arrow at the tip of the tongue indicates the point of greatest constriction in the vocal tract.

li*f*e; *z*ip, *s*ip), where the order is voiced, unvoiced.

Due to their profound acoustic effect, these two articulatory parameters, constriction location and presence of voicing, are the primary means of classifying fricatives. An additional feature sometimes used is that of stridency (Jakobson, Fant and Halle, 1963) or sibilancy (Lindblad, 1980). This feature identifies the fricatives in which the airstream is directed towards an obstacle such as the teeth downstream of the constricted region, at which, presumably, additional sound is generated. These fricatives include /z,s, ž,š/, and in some systems /v,f/.

Aerodynamic theories of turbulence are not well worked out. As a result, the effect of each of these articulatory parameters can be predicted analytically only with difficulty, if at all. Therefore, fricative models used to date, such as those of Fant (1960) or Flanagan et al. (1975), are not based on a set of simplifying assumptions applied to a well-understood physical mechanism. Instead they consist of empirically-based elaborations of the vowel models. The vocal tract is still represented as a tube of varying cross-sectional area, and again only plane-wave propagation is considered, but pressure sources are placed at the downstream edges of the noise-producing constrictions, or at the location of the teeth. The number of such noise sources and their spectral characteristics are two parameters that Fant and Flanagan et al. experimented with in efforts to make the model correspond more closely to the physical situation. Results were somewhat inconclusive. For some fricatives, no one source configuration provided a match that was equally good across the entire frequency range, and there was no other criterion with which to judge the physical accuracy of the source representations.

The vowel models based on source-tract independence have been largely successful in applications such as speech synthesis and in the more basic task of predicting the acoustic effect of articulatory changes. There is a direct correspondence between the physical action of the vocal folds and the source function, which makes it possible to modify the source to represent different patterns of vibration of the vocal cords. Researchers are still working on refining the vowel model, particularly with regard to such modifications of the source. In recent years, the tract impedance has been shown to affect the glottal volume flow (Rothenberg, 1981; Fant, 1983), although this is a second-order effect. Our intention is not by any means to suggest that the problem of modeling vowels is solved. However, it does seem to be an easier problem, at least initially, than that of modeling fricatives, because the source is fairly localized, and is located at one end of the tract (by virtue of the high glottal impedance that acoustically separates the vocal tract and the subglottal portions of the anatomy). Also, the model initially devised for the voiced source represents the physical situation more accurately, and because of that it works better and is easier to know how to alter.

Both vowel and fricative models were tested by Fant by comparing the predictions of the models to speech spectra. Due to the inaccessibility of the vocal tract, it is not clear whether discrepancies found by the comparison are due to failures of the model or an inexact measurement of the configuration. Since the filter functions for vowels and fricatives are derived in the same way, and the vowel model is generally more successful than the fricative model, it is clear that problems with the fricative model must be due to an inaccurate source representation. The pressure source used in fricative models to date does not derive from specific knowledge of the sound generation process, and therefore we do not know how to alter it for a different shape of constriction, or a higher flowrate. We do not know if it interacts with the tract.

Before we develop a way in which to address these issues, let us consider previous work in more detail.

1.2 Literature Review

1.2.1 General Aspects of Unstable and Turbulent Jets

Techniques such as flow visualization have established that as air exits from a constriction it forms a jet, which gradually mixes with the surrounding air. The Reynolds number (Re) characterizes the degree of turbulence generated as this mixing takes place. It is defined by

$$\mathrm{Re} = \frac{Vd}{\nu}$$

where V = a representative flow velocity, usually taken to be that in the center of the constriction exit. d = a representative dimension, usually the constriction diameter, and ν = the kinematic viscosity of the fluid, which for air is 0.15 cm^2/sec. As Re increases, an initially laminar flow will pass through an unstable region and finally become fully turbulent. Turbulent flow is distinguished by irregular, high-frequency fluctuations in velocity and pressure at a given point in space (Schlichting, 1979). The critical Reynolds numbers, Re_{crit}, separating these regions vary according to the geometry and degree of prior turbulence of the fluid. For a jet issuing from a circular hole, the unstable region would typically occur for $160 < \mathrm{Re} < 1200$ (Goldstein, 1976).

The dimensions of a fully turbulent jet in the subsonic range depend only on the constriction diameter and shape; thus, visually, all jets can be scaled to look the same. Theoretical work by Lighthill (1954) and others established that the sound generated by jets scales as well, that is, that the spectral characteristics of the sound generated by a jet depends only on the jet velocity and diameter.

Sound is generated by the random pressure fluctuations of the turbulent fluid. A good summary of the theoretical and empirical efforts to describe this sound generation process may be found in Goldstein (1976). For our purposes, the essential facts are as follows. For a jet emerging from a constriction of diameter d at $\mathrm{Re} > \mathrm{Re}_{crit}$, three regions, the mixing, transition and fully developed regions, can be defined, as shown in Fig. 1.2. From both theory and experiment it appears that nearly all of the sound power is generated in the mixing and transition regions, possibly with most of it coming from the mixing region (Goldstein, 1976). If half of the sound power is assumed to be generated in the mixing region, the total power, P, generated by the jet is proportional to V^8 (where V is the flow velocity), which agrees with Lighthill's prediction (1952). The total sound power spectrum has a broad peak at about SV/d Hz, where V is the flow velocity in the center of the jet as it exits the constriction, d is the jet diameter, and S, the Strouhal number, defined by this equation, is equal to 0.15. (The frequency of the spectral peak depends on the type of spectrum chosen. The Strouhal number at the peak is $S = 0.15$ when the noise spectral density, an equal-bandwidth representation, is plotted; $S = 1.0$ for the third-octave spectrum.) The sound pressure measured at a particular point in the far field will have a similar spectrum, with a peak frequency dependent on the angle at which the measurement is made. Measurements within the jet itself show that the high-frequency sound originates closer to the nozzle than does the low-frequency sound (Fletcher and Thwaites, 1983).

Lighthill described three types of sound sources that are present to varying degress in the sound produced by turbulent flow: a monopole source, (which is equivalent to a sphere pulsing in and out), a dipole source (two spheres pulsing in opposite phase), and a quadrupole source. A monopole source obeys a V^4 power law, meaning that the total sound power generated by

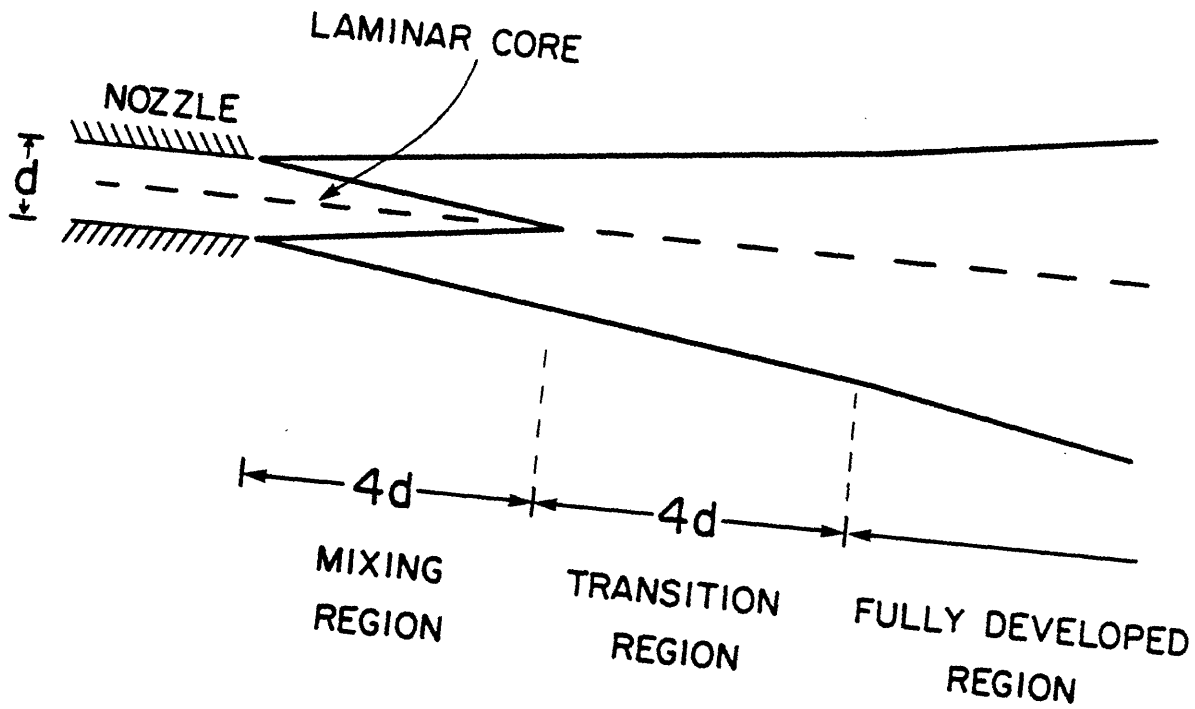


Figure 1.2: Diagram of the mixing, transition and fully developed regions of a fully turbulent jet. After Goldstein (1976)

a flow monopole increases as the fourth power of the flow velocity V; the dipole source obeys a V^6 power law; the quadrupole source, V^8. An equivalent statement is that the efficiency of conversion of the kinetic energy of the flow into sound is proportional to $M = V/c$ for the monopole, M^3 for the dipole, and M^5 for the quadrupole source (Morse and Ingard, 1968). For subsonic velocities ($M < 1.0$) the quadrupole is thus the least efficient source, but its relative contribution to the total sound generated should become progressively more important as M increases. Dipole sources occur along rigid boundaries, which exert an alternating force on the fluid; quadrupole sources exist in free jets. Thus if a jet impinges on an obstacle such as a flat plate, the sound generation can be modeled by a combination of quadrupole and dipole sources. It is not well understood how the energy divides between quadrupole and dipole sources in such a case (Goldstein, 1976).

Finally, air flowing through a constriction can also produce whistles. The acoustic mechanism depends on an instability of the jet resulting in a periodic generation of vortex rings. The sound generated by the vortices acts to increase the oscillations of the jet, thus closing a feedback loop. This type of mechanism is not limited to a jet emerging from a single constriction; it has been shown to operate within the constriction itself (Anderson, 1954; Succi, 1977), when air passes through axial holes in two plates spaced slightly apart (Chanaud and Powell, 1965), or when air passes around a cylinder (Thomas, 1955). Likewise, it can occur when a jet strikes a wedge (Powell 1961, 1962; Holger et al 1977, 1980), the edge of a cavity (Pollack, 1980; Elder et al, 1982), or a ring (Chanaud and Powell, 1965). A given geometry will often possess a range of Reynolds numbers within which different unstable modes occur. For example, according to Goldstein, circular jets are unstable when the Reynolds number is in the range $160 < \mathrm{Re} < 1200$. If a feedback mechanism is present when such an unstable mode occurs, a whistle is likely. Whistles are not confined to the unstable region, however. Goldstein reports the conjecture that a persistent periodic structure may exist in turbulent jets of circular cross-section, for $\mathrm{Re} > 1200$. The conjecture is borne out by Elder (1982), Succi (1977), and Fletcher and Thwaites (1983), who all report whistle-like behavior in fully turbulent jets.

For most of the configurations, the unstable behavior is typified by a sequence of stages, each consisting of a characteristic stable pattern of vortices. The patterns are determined by the flow velocity and the particular dimensions of that configuration: e.g. diameter and length of the orifice for an orifice tone, or diameter of the nozzle, and distance from nozzle to edge, for an edgetone. Within a stage, the frequency of the whistle increases linearly with flow velocity. Between stages, the frequency jumps abruptly, with hysteresis. If a resonator is present, the whistle will couple into the resonances. A similar pattern of transitions from stage to stage will occur, but the frequencies generated within each stage will center on a resonance frequency. Thus, in order to predict the whistle frequency, the ranges within which the jet will be unstable, the factors determining which stage will be present, and the factors determining the frequency within that stage must be established. Such predictions have been made with varying degrees of theoretical justification for each of the configurations mentioned above. Of these, we will mention just the three that are relevant. For the orifice tone, the frequency is near the coincidence of the half-wave resonance frequency due to the orifice itself, and the feedback frequency, corresponding to intervals at which the jet perturbations are reinforced. Both frequencies depend on the orifice length and jet velocity (Succi, 1977). The frequency of the edgetone is the frequency at which the jet flips from one side to the other of the edge. That frequency is determined by the jet velocity and the distance between the nozzle and the edge. The frequency of the holetone depends on the distance between orifices and the jet velocity.

For human whistling it is, of course, much more difficult to measure flow velocity or articulatory dimensions with precision, making frequency predictions difficult to check. It has been shown that the typical bilabial whistle is produced with Reynolds numbers between 1400 and 2000, which in view of the irregular shape of the vocal tract, is almost certainly fully turbulent. The range of frequencies extends from 500 to 4000 Hz, and is mostly controlled by the anterior-posterior position of the constriction formed by the tongue. Since the tongue position also determines the frequencies of the second and third formants, and since frequency jumps between those formants have been observed, it appears that the whistle frequency is always captured by either the second or third formant of the vocal tract (Shadle, 1983). Although these observations leave the exact mechanism of the bilabial whistle unclear, it does exhibit behavior much like that of the mechanical whistles that have been so extensively studied. The capture behavior common to all whistling configurations makes them particularly ill-suited to being modeled by a system in which the source and filter are independent.

1.2.2 Acoustic Models of Fricatives and Fricative-Like Configurations

One of the earliest efforts to consider fricative production in terms of the aerodynamics involved was that of Meyer-Eppler in 1953. He compared sound-pressure vs. flow relationships for plastic tubes with different-sized elliptical openings. He derived effective-width formulas that allowed sound-pressure vs. Re plots for different ellipse sizes to coincide, and then used these derived formulas on data of human subjects uttering fricatives to infer articulatory parameters from pressure measurements.

It is not clear, however, that Re_{crit}, which Meyer-Eppler defined to be the lowest Reynolds number at which measurable sound was generated, should be the same for the plastic tubes and the three fricatives /s, š, f/ that Meyer-Eppler studied. First, Re_{crit} is probably lower for the irregularly shaped vocal tract than for the plastic tubing. Second, sound is most likely being generated for these three fricatives both when the air passes through a constriction (over the tongue for /s, š/ or between teeth and lower lip for /f/) and when the jet of air strikes an obstacle (the teeth for /s, š/ or upper lip for /f/). The intensity of the sound generated may therefore be related more to the distance between the constriction and the obstacle and the physical properties of the obstacle than to the effective area of the constriction. Since the tubes he used had a constriction but no obstacle, application of the effective-width formulas developed for the tubes to the strident fricatives may give misleading results.

Heinz (1958) carried out experiments using a 17 cm tube imbedded in a wooden sphere to approximate the dimensions and radiation impedance of a vocal tract. By placing a cylindrical plug with a 0.2 cm diameter axial hole at the mouth and 4 cm back from the mouth, he approximated the articulatory configurations for fricatives such as /ϕ/ and /ç/, respectively. He obtained far-field directivity patterns and spectra for a variety of frequencies and flow rates. As expected, intensity rose with flow rate, except at the half-wavelength resonance of the constriction. Heinz ascribed the behavior at this resonance to effective movement of the source relative to the constriction. However, the source he used in calculating the system response was a localized pressure source that did not change position with flowrate. The source spectra derived by subtracting the system response from the measured sound spectrum were fairly flat, but with dips at resonance frequencies, which Heinz assumed were a consequence of greater than expected resonance bandwidths due to turbulence losses.

As a first attempt towards including these losses in the model, he calculated the incremental flow resistance, R_i, for one of the configurations using two methods. The first method used

measurements relating the static pressure drop across the constriction to the flow velocity through it. R_i was set equal to the slope of this function at a single, intermediate flow velocity. The second method was based on the conclusion that, at the constriction resonance, the input impedance of the constriction would reduce to R_i. Therefore, the second value of R_i could be based on the values of the frequency and bandwidth of the constriction resonance measured at the same flow velocity. At the one flow rate used, the R_i values computed by these two methods agreed within 8 percent.

Fant (1960) used a distributed model of the vocal tract to investigate the effects of source location, source spectrum, and constriction resistance. Concluding that the theory of turbulence was too undeveloped to be useful, he judged the accuracy of his models by how closely the predicted spectra matched the spectra measured from the speech of a single subject. Models for all fricatives used a series pressure source that generated either white noise for /s, š, š, č/ or integrated white noise (i.e., with a -6dB/octave slope) for /x, f/.

Fant did not attempt to make a physical argument relating the two types of source spectra to the distinguishing features of the fricatives. The location of the source – whether at tongue or teeth – is more clearly linked to the place of articulation. The model for /x/, which used a source located at the tongue, produced the best match of all of the fricatives. For /s/ and /š/, although sources at both tongue and teeth were used, it appeared that neither location by itself would provide a good match at all frequencies. Fant suggested that quite possibly /s/ was produced with sources at both locations, with a low spectral level below 1 kHz, but did not attempt a physical justification for this particular source characteristic. Fant was aware that changes in the source location would alter the frequencies of the zeros it produced in the output, but he wrote that this effect would probably prove to be perceptually unimportant.

Flanagan et al. (Flanagan, Ishizaka, and Shipley, 1975, 1980; Flanagan and Ishizaka, 1976) elaborated on this model by allowing for multiple noise sources, one for each section of the uniform tube model, with the strength of each source (the variance of the white noise pressure source) depending on Re of the section (computed from its area and the glottal volume velocity). A given pressure source was included only if its Reynolds number exceeded the value of Re_{crit} determined from Meyer-Eppler (1953). The model allows for a dependence of both intensity and acoustic resistance on flow rate. The sound sources are now distributed throughout the vocal tract, but the source due to a single section is still localized. Further, the spectral characteristic of the source is unchanged by flow rate (except for its overall amplitude), which contradicts findings of Heinz (1958) and Thwaites and Fletcher (1982), among others. Likewise, Re_{crit} is constant regardless of upstream conditions or the tract configuration. The method of computing Re based on the cross-sectional area is not sensitive to the turbulence generated when a jet of air impinges on an obstacle. This model, like Fant's, assumes linear elements, independence of source and filter functions, and one-dimensional sound propagation.

1.2.3 Analysis of Fricatives

Efforts to model fricatives, such as that of Fant (1960), and work with mechanical models, such as that of Heinz (1958), form two approaches to understanding the acoustic mechanisms involved in fricative production. A third approach is simply through acoustic analysis of spoken fricatives. Hughes and Halle (1956) examined the fricatives /f, s, š/ and their voiced counterparts, /v, z, ž/, as produced by three speakers. A 50-ms portion of each fricative was analyzed by measuring the total amount of energy in each 500 Hz band between 0 and 10 kHz. From these spectra three measurements were derived that could be used to classify the tokens auto-

matically, and perceptual tests established that the automatic procedure generally worked quite well, exhibiting the same pattern of errors as the human listeners. However, it appeared much easier to generalize about the acoustic distinctions between fricatives within a given speaker than across speakers. To quote from the article (p. 305), "The discrepancies among the spectra of a given fricative as spoken by different speakers in different contexts are so great as to make the procedure of plotting these spectra on one set of axes a not very illuminating one. On the other hand, the differences among the three classes of fricatives (labial, dental, and palatal) are quite consistent, *particularly for sounds spoken by a single speaker*" (emphasis added).

Strevens (1960) analyzed a larger set of fricatives, /ϕ, f, θ, s, š, ç, x, χ, h/, but was not much more successful at describing spectral differences that consistently distinguished between fricatives. Distinguishing between the three groups of fricatives, defined by front, mid, and back points of greatest constriction, was possible. Briefly, the front fricatives have the lowest intensity and the smoothest spectra; the mid fricatives have the highest intensity and significant peaks in the middle frequency range; back fricatives have medium intensity and a marked formant-like structure.

Heinz and Stevens (1961) analyzed the fricatives /f, s, š/, matched the spectra with judicious combinations of two conjugate-pole pairs and one pair of conjugate zeros, and conducted perceptual tests of the fricatives synthesized from this model. The pole-zero combinations, which form a simplified transfer function for a localized pressure source in a uniform tube model, were shown to be perceived as any of the fricatives /θ or f, s, š, ç/ depending on the frequencies of the poles. Their perceptual approach yielded generalizations about the acoustic distinctions between fricatives similar to those found by Hughes and Halle.

Finally, Catford (1977) discussed the difference between "channel turbulence" (that accompanying a jet emerging from a constriction) and "wake turbulence" (that produced by a jet impinging on an obstacle) and presented spectra of the fricatives /θ, s, š/ as produced by subjects with and without their upper and lower dentures. Noting that /θ/ changes very little, /š/ somewhat, and /s/ the most between the two conditions, he concluded that the latter two fricatives have significant wake turbulence generated at the teeth, which adds high frequency energy to the spectrum. Perhaps this distinction would have proved useful to Strevens.

1.3 Plan of the Thesis

The literature reviewed above draws on a combination of theoretical and empirical approaches to turbulence, theoretical models of fricatives, and acoustic analysis of fricatives. Each of these domains offers a different type of explanation and insight, and a different set of limitations. First, given the irregular shape of the vocal tract, an exact theoretical analysis is quite difficult and numerical integration techniques would provide little insight, if any. An empirical approach is thus indicated. Since the vocal tract is relatively inaccessible, the use of mechanical models, which proved to be so productive for Heinz (1958), is appealing. However, such an approach requires a separate justification of the applicability of the results to speech. Theoretical models, such as those developed by Fant (1960), provide useful conceptualizations of the acoustic mechanisms. However, as we have seen, developing such models directly from speech data can be difficult since the configuration is not known exactly and cannot be controlled arbitrarily, and because there is no independent check on the success of the model. Finally, attempts to unravel the acoustic mechanism of fricatives by simply analyzing fricatives are limited because the distinctions between fricatives are not clearcut and the articulatory

configuration is difficult both to measure and to vary systematically.

Since many of the limitations of the previous studies are due to the choice of domain, this thesis is conducted in three domains: speech, mechanical models, and theoretical models. We begin by combining knowledge of the basic articulatory configuration common to all fricatives with a rudimentary consideration of the aerodynamics of turbulence. These lead to a simple mechanical model, that of a constriction in a tube with the dimensions of a typical vocal tract. In Ch. 2, the parameters of this model are varied methodically in order to establish which articulatory parameters have the greatest effect acoustically.

Following this sorting-out process, we arrive at two simplified models, in which the parameters that are varied are those that have the greatest acoustic effect and represent simplifications of the articulatory differences observed in fricative configurations. These idealized cases are then theoretically modeled in Ch. 3, and the predicted sound is compared to the actual sound produced by the mechanical models.

We would like a theoretical model for fricative production that is both physically representative and tractable. It seems likely that such a model will still include source and resonator elements corresponding to the sound generated by air flow and the modifications induced on that sound by the vocal tract. In order to reduce the mathematical complexity and thus perhaps increase our physical intuition, we will assume, as others have, that the airflow-to-sound conversion process can be modeled by a linear system, that the source is independent of the resonator, and that the three-dimensional motion of turbulence can be modeled as a one-dimensional transmission line. We are therefore seeking a model that is not physically representative in some fundamental ways. However, these constraints will allow us to establish with some precision the set of source parameters, such as spectral characteristic and location, that are necessary to the model. Further, we will be able to explicitly test some of the assumptions on which the model is based, such as that of no source-tract interaction.

Source-tract interaction in its simplest sense means that the output of the source is influenced by the surrounding tract. Clearly, then, the particular parameters used to define the source and tract affect whether or not such interaction can be said to occur. We will base our assessment of interaction on an initial set of source-controlling parameters derived from the results of Ch. 2 and the previous work with jets reviewed above. By the end of this thesis the process of making that assessment will allow us to refine our definition of source-tract interaction so that it becomes more useful conceptually.

Basing the theoretical models on the mechanical models makes it much easier to check the accuracy of the theoretical models, but also makes it incumbent on us to justify the application of the final results to speech. Therefore in Ch. 4 speech is analyzed, in the same way the mechanical model data were analyzed, and the two types of data are compared. This comparison allows us to establish whether fricatives exhibit the same extremes of acoustic behavior that the articulatorily similar idealized cases do. The comparison is then refined by the development of more complex mechanical models that are designed to mimic the articulatory configuration of each fricative more closely. Again, the speech and mechanical model results are compared.

At this point, we will be in a position to answer two final questions. First, are we justified in using simplified mechanical model results to study real speech, that is, do the mechanical models offer us a useful tool for understanding acoustic mechanisms of fricatives? Finally, depending on that answer, what can we say about theoretical models for each fricative? A summary of all results and a discussion of these questions is presented in Ch. 5, together with an assessment of the logical directions in which to proceed from this point.

Chapter 2

A Constriction in a Tube: Methodology and Preliminary Experiments

All fricative configurations have in common a constriction located within a duct. In this chapter we investigate the acoustic and aerodynamic effects of the various parameters describing such a configuration. The purpose of these preliminary experiments is threefold: to establish a methodology for use in all experiments involving mechanical models: to derive parameters, such as flow resistance, that we will need to develop theoretical models in Chapter 3; and to identify the parameters that affect the sound the most, thereby shortening the theoretical modeling task.

2.1 Method

Figure 2.1 shows the experimental setup used for all mechanical model experiments. It consisted of the sound generating system, in which air was forced through a plastic tube containing the configuration under examination, and the sound analysis system, which analyzed the generated sound and stored the resulting power spectra.

2.1.1 Sound Generating System

The sound generating system was fed by an air tank containing pressurized dry air, and controlled by a regulator which provided a constant pressure of 15 psi to the rest of the system. Dry air at room temperature has a speed of sound of 34480 cm/sec (Beranek, 1949); this value of c was used in all computations. Air passed from the tank through a shutoff valve and through a flowmeter (Flowrator 3/8-25-G-5/36). Following the flowmeter, a T-junction allowed measurement of the static pressure by a water manometer. Following the T-junction, the air passed either directly into the "tract", in the case of the earlier preliminary experiments, or into a muffler and then into the "tract".

The muffler was used in order to attenuate sound generated in the upstream tubing and to provide a reflectionless termination for the input end of the "tract". It was constructed of plywood, and contained two channels for the air to pass around a central body, which was also made of plywood and tapered on both ends. Both sides of each channel were covered with foam rubber, tapered to a maximum depth of 5 cm. The inlet and outlet of the muffler were tapered to allow a gradual transition from a small circular cross-section to two large rectangular channels. The effect of the muffler was measured by positioning the microphone 42 cm from the outlet tube, first with and then without the muffler in place, and recording the sound at the same flowrates. The inlet and outlet tubes were of the same inner diameter, 1.11 cm. Figure 2.2 contrasts two such sound pressure spectra at a flowrate near the maximum of those used, showing that the muffler attenuates the peaks by about 20 dB and smooths out the spectrum considerably. The curve labeled "room noise" in the figure is the spectrum of the ambient noise, measured when no air was flowing through the system, smoothed by performing logarithmic regression.

Figure 2.3 shows the setup in the area downstream of the muffler in greater detail. The "tract" consisted of a tube containing any of a number of constrictions and obstacles. The upstream end, referred to as the "glottis", fit tightly around the outlet tube of the muffler. The downstream end - the "mouth" - was either supported by a ring stand or, more commonly, fitted flush with a plexiglass baffle. A traverse mechanism clamped to the bench allowed obstacles to be positioned precisely either inside or outside the tract. The microphone was taped to a rotating arm, which facilitated measurement of directivity patterns. The microphone position, relative to the "mouth", can be specified by r and θ, as shown. Typically (and, throughout this thesis, unless stated otherwise) $r = 26$ cm and $\theta = 28°$. During acoustic experiments fiberglass was placed over the traverse mechanism, rotating arm, and adjacent areas of the bench in order to reduce reflections from those surfaces.

A detailed diagram of the "tract" is shown in Fig. 2.4. The tube consisted of a section of Tygon clear plastic tubing with an inner diameter of 2.54 cm, an outer diameter of 3.18 cm, and usually of length $l_T = 17$ cm. The tubing had a circular cross-section but was not always perfectly circular since it is slightly stretchy. The stretchiness allowed the tubing to seal airtight around any object with a circular cross-section and a diameter in the vicinity of the tube's inner diameter. This property was exploited to connect the tube to the muffler: the first 0.5 cm of the upstream end of the tube was stuffed with concentric rings of smaller diameter tubing, forming a tight seal around the brass outlet of the muffler (o.d. = 1.27 cm).

The constrictions were cylindrical plugs machined out of aluminum. Those considered in this chapter possessed either a circular or rectangular axial hole. The plug lengths (l_c) ranged from 1.0 to 3.0 cm; the circular hole diameters (d_c) used were 0.16, 0.24, 0.32, 0.64 and 0.95 cm. The rectangular holes measured 1.27 cm on the long side and 0.06 or 0.25 cm on the short side, creating constriction areas (A_c) equal to those of the 0.32 and 0.64 diameter circular constrictions. A tapped blind hole on the downstream face of each plug allowed placement of the constriction in the tube with a long screw. These holes were plugged by inset screws after insertion of the constrictions.

An obstacle was used in some of the experiments. It consisted of a semicircular piece of aluminum of the same radius as the tube, and 2 mm long. When used, it was downstream of the constriction a distance l_o; a brass rod was threaded into a hole on the downstream face, so that it could be positioned in the tube by the traverse mechanism. Other relevant parameters are indicated in Fig. 2.4.

Flowrates between 50 and 690 cc/sec were used, which covered the range used by people producing fricatives (e.g., 300 to 500 cc/sec; Catford, 1977). For a constriction of cross-sectional area 0.079 cm^2, the one most commonly used, this resulted in a flow velocity through the constriction of between Mach 0.02 and 0.25.

2.1.2 Sound Analysis System

The sound analysis system utilized a half-inch condenser microphone (Bruel and Kjaer 4133), which has a flat free-field frequency response (± 1 dB for grazing incidence, from 30 Hz to 40 kHz). At small values of θ a windscreen was used. The windscreen was tested for its sound attenuation properties by observing its effect on the spectra of sine waves played over a loudspeaker. For sine waves in the range 140 to 1990 Hz, it attenuated the peak amplitude by 0.3 dB or less with one exception: at 990 Hz, it attenuated 1.4 dB. Note that all of these values were within the ± 2.0 dB confidence limits of an averaged ($n = 16$) power spectrum.

The microphone output was amplified (Bruel and Kjaer type 2604) and then fed into a

spectrum analyzer (Hewlett-Packard 3582A). Data from the spectrum analyzer could be transferred via an LSI 11 to magnetic tape, and from there to a VAX-11/750, on which further data analysis was performed.

In order to retain absolute sound pressure level information, a reference level was established for the microphone as described in Appendix A. The computed calibration factor was added to all spectra during the data analysis. As a result, all graphs shown here give the absolute sound pressure level, in dB re 0.0002 dyne/cm^2, measured at the microphone.

2.1.3 Signal Analysis

The sounds generated by turbulence are essentially random in nature, and therefore are best characterized by statistical properties such as the mean and standard deviation averaged over time. To observe the spectral characteristics, it would thus seem reasonable to perform Fourier analysis on a long segment of the input signal, and treat the resulting coefficient at each frequency as the mean value of the amplitude of that frequency "component". It turns out that this is an inconsistent estimator of the mean, in the sense that it does not improve as the length of the signal being analyzed increases. A better estimate is found by averaging several finite-duration spectra together, since the expected value of the average is equal to the true mean value, and the variance of the estimate decreases as the number of averages increases. This estimator is known as the average power spectrum (Bendat and Piersol, 1971).

The spectrum analyzer performs such an analysis by sampling the input signal 1024 times, at a rate determined by the selected frequency range, computing a DFT of those points, and combining the DFT coefficients (by squaring, averaging, and taking the square root) with the previous sets of coefficients to form an RMS-averaged spectrum. These steps are repeated until the requested number of averages has been obtained. Let us call the input signal $x(t)$, and the samples forming the mth time record $x(n,m)$, where $n = 0, 1, \ldots N-1$. These samples are spaced a time h apart, giving the entire time record a duration $T = 1024 \cdot h$. The time h is determined by the frequency span chosen for the Fourier analysis; $h = 1/2f_c$, where f_c = the cutoff frequency. In order to minimize processing time and maximize accuracy of the output, the FFT algorithm used by the spectrum analyzer scrambles the 1024 samples (using a special scrambling algorithm, the details of which need not concern us here) so that a 512-point DFT may be performed. The DFT results in 256 unique, non-aliased, complex values in the final spectrum. The cutoff frequency is therefore two times the maximum frequency seen by the user (Pendergrass and Farnbach, 1978). As a result, the width of each of the 256 frequency bins is

$$f_0 = \frac{f_{max}}{256} = \frac{f_c}{2 \cdot 256} = \frac{1}{1024h} = \frac{1}{T}$$

The frequency range most often used extends from 0 to 10240 Hz (thus f_{max} = 10240 Hz), has a frequency spacing f_0 of 40 Hz, a time record length T of 25 msec, and 24 μsec between samples (h).

If we denote the Fourier transform of the mth record of the input signal by $X(k,m)$, the average power spectrum at frequency $f_k = kf_0$, computed by averaging M time records, is then:

$$P_X(k) = \frac{1}{M} \sum_{m=1}^{M} G_{XX}(k,m), \qquad k = 0, 1, \ldots N/4$$

where

$$G_{XX}(k,m) = X(k,m) \cdot X^*(k,m)$$

Since the phase of a random process is meaningless, it is not computed.

Increasing the number of averages should remove the random noise, leaving behind a smoother more representative spectrum. This change is illustrated in Fig. 2.5, which was derived by using the random noise generator built in to the spectrum analyzer as the input signal. The increase in accuracy of the estimate can be expressed in terms of the 90 percent confidence limits corresponding to the number of averages selected. These are given in Table 2.1.

	M = number of averages						
	4	8	16	32	64	128	256
Upper limit dB	+4.7	+3.0	+2.0	+1.4	+1.0	+0.7	+0.5
Lower limit dB	-2.9	-2.2	-1.6	-1.2	-0.8	-0.6	-0.4

Table 2.1: 90% Confidence Limits for Spectral Amplitude as a Function of Number of Averages. From HP Application Note 245-1, 1978.

The increase in accuracy also increases the signal analysis and computation time. The compromise used for most measurements was 16 averages, which resulted in a ± 2 dB confidence limit.

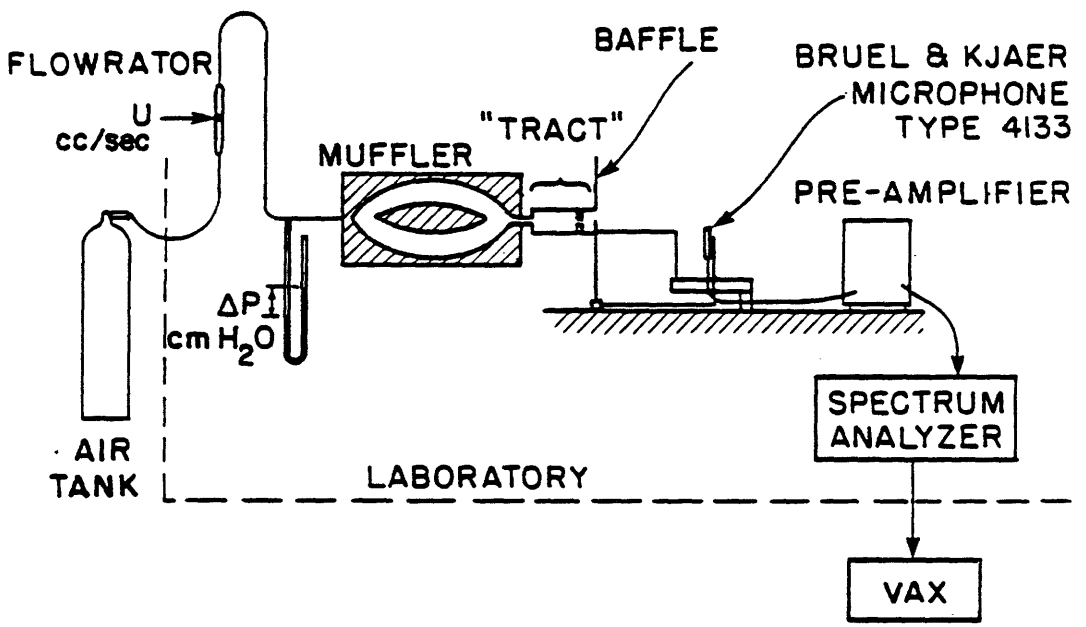


Figure 2.1: Diagram of the experimental setup. The tract, whose outer wall is a cylindrical tube, is shown in cross section. The dimensions of the baffle and other details of the setup downstream of the muffler are given in Fig. 2.3.

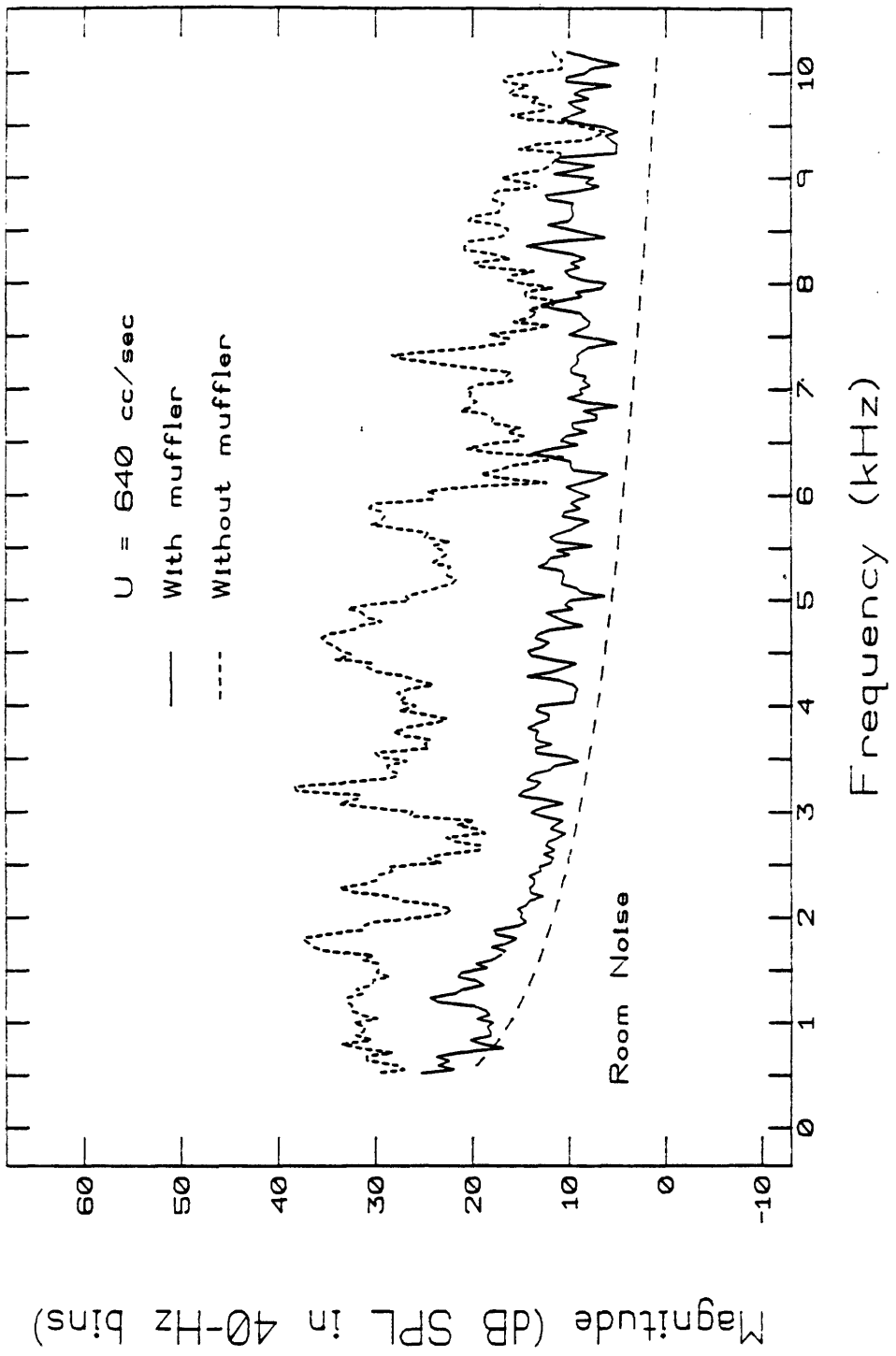


Figure 2.2: Averaged pressure spectra of the sound generated by the system upstream of the "tract" with and without the muffler. The dashed line is the curve fit by logarithmic regression to the averaged pressure spectrum of the room noise. For all three curves, number of averages, n, is 16.

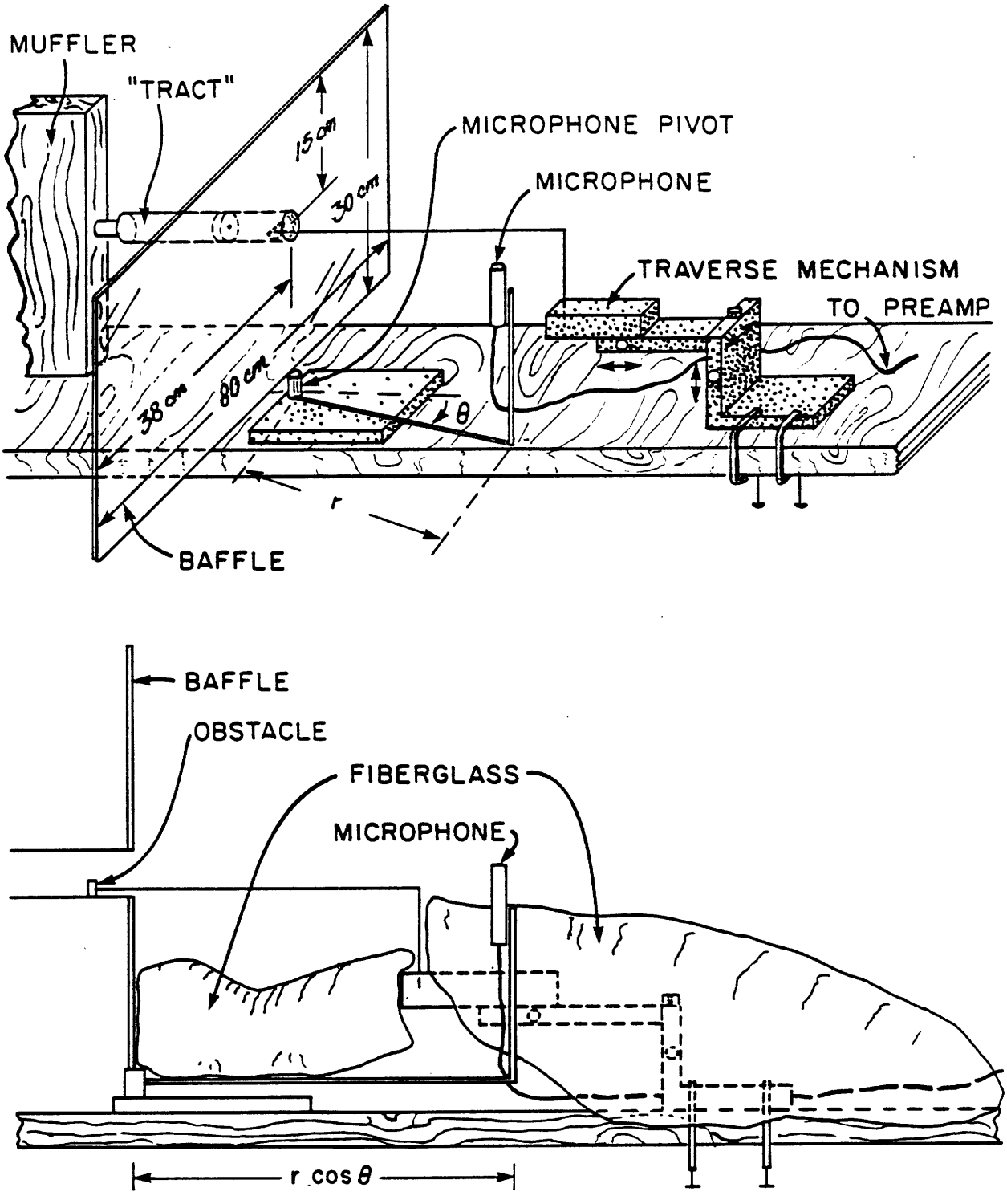


Figure 2.3: Diagram of the experimental setup downstream of the muffler (top). Expanded view showing placement of sound-attenuating fiberglass (bottom). Both sketches are not to scale.

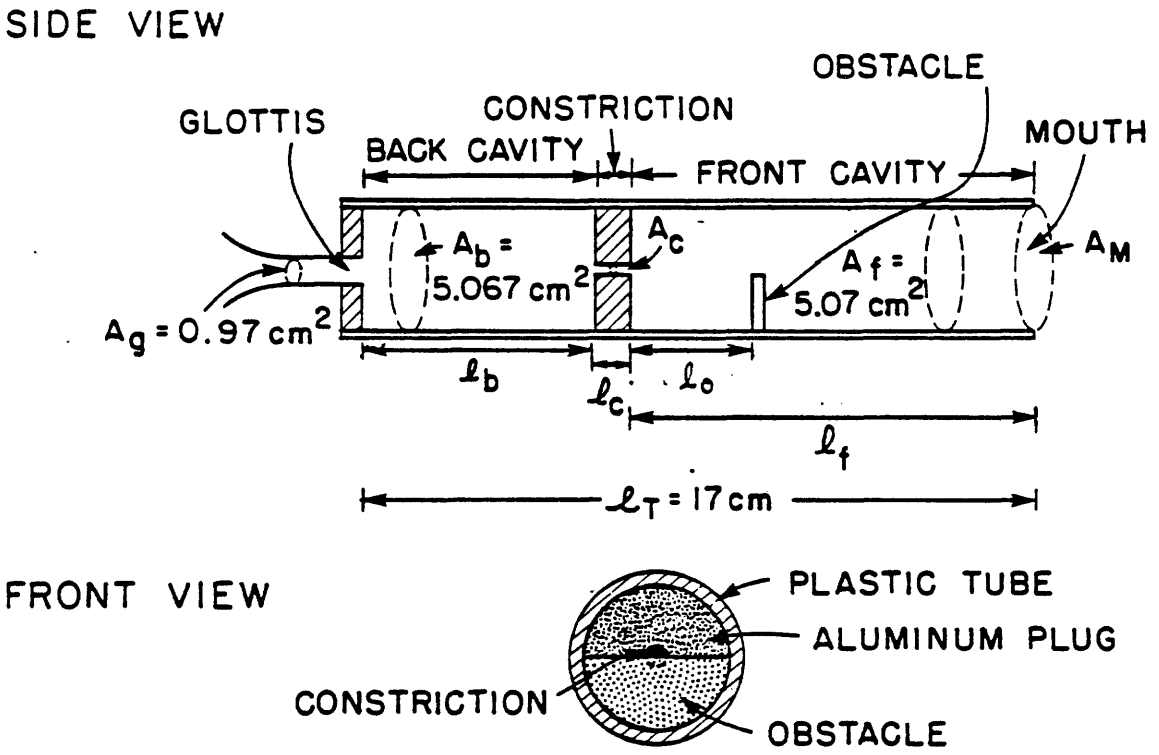


Figure 2.4: Detailed diagram of the "tract", giving relevant dimensions and the terms used to refer to specific regions.

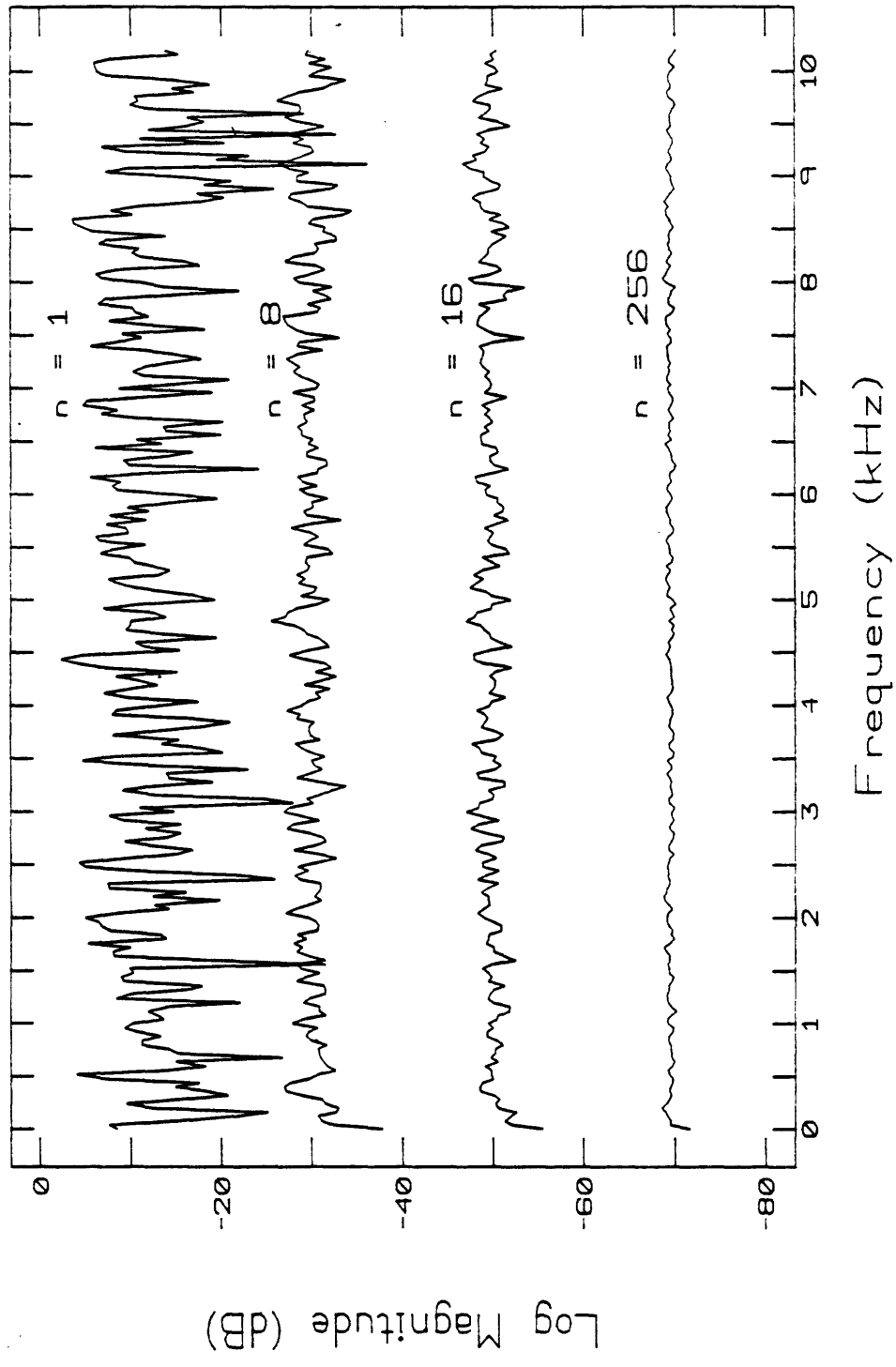


Figure 2.5: Power spectra of white noise, averaged n times. Traces are offset by 20 dB to show the effect of averaging.

2.2 Pressure Drop Across a Constriction

As the flowrate through a constriction is increased, the pressure drop across it increases in a nonlinear fashion. The flow resistance of a constriction is primarily a resistance loss due to this pressure drop Δp at a given volume velocity U. Since Δp is a nonlinear function of U, the slope of that function at a given point is the incremental resistance of the constriction at that U (Beranek, 1954). Since in the experiments to be described the AC fluctuations in pressure and velocity due to a sound wave are small compared to Δp and DC flow U, this incremental resistance is the loss experienced by the acoustic disturbance (superimposed on the DC flow). It is therefore useful in modeling the acoustic behavior of a constriction in a duct.

The flow resistance can be easily derived if one has an analytical relationship for Δp as a function of U. The form of the Δp versus U relationship can be derived theoretically by setting up an energy equation for the movement of fluid from one point to another and computing the resultant head loss, as done by Binder (1962). The total pressure drop across a constriction is proportional to the sum of three separately computed head losses, due to the inlet, outlet, and straight part of the constriction. All three terms are of the form

$$h = K_1 \frac{V_c^2}{2g} \tag{2.1}$$

where

h = head loss (cm)
V_c = velocity through constriction (cm/sec)
g = acceleration due to gravity (cm/sec^2)

For the outlet, the theoretical formulation of K_1 depends only on the duct and constriction areas, and matches data well. However, for the inlet, K_1 is an empirical constant heavily dependent on the abruptness of the inlet, and therefore not easy to specify nor to extrapolate to new shapes. For the straight section, $K_1 = bl_c/d_c^2$, where d_c is the constriction diameter, and b, the friction factor, is a complicated function of velocity and pipe roughness. In different ranges of the Reynolds number (Re), analytic relationships exist for b, but cannot be solved directly for that value. Therefore, a different approach is needed in order to derive the incremental resistance, R_i, from the data.

Rather than consider the above complexities, Heinz (1956) assumed the following form for Δp:

$$\Delta p = K_2 \frac{\rho U^2}{2A^2} + K_3 U \tag{2.2}$$

where

Δp_{tot} = pressure drop across constriction, dyne/cm^2
U = volume velocity through constriction, cc/sec
K_2 = empirically derived constant, dimensionless
K_3 = empirically derived constant, g/sec-cm^4
A_c = cross-sectional area of constriction

K_2 and K_3 are determined by curve-fitting procedures on the data for a particular constriction and range of flowrates. The first term is justifiable on theoretical grounds, since it is equivalent to the form of Eqn. 2.1. The second term is exactly equal to the pressure drop due to the straight part of the constriction when Re < 2000 (laminar flow), roughly equal to that drop when Re > 4000 (for smooth pipes), and completely wrong for $2000 <$ Re < 4000. These differing ranges

are due to the nonlinear relationship between the friction factor b and Re (Binder, 1962). The physical significance of the term therefore depends heavily on the Reynolds number range over which the curve-fitting is performed.

Heinz reported K_2 and K_3 values for five constrictions of similar areas and lengths, for flowrates between 40 and 180 cc/sec. The curves shown in Fig. 2.6a were drawn using the K_2 and K_3 values given by Heinz, extrapolating well beyond the original flowrate range of the data. (The various symbols along each curve represent extrapolated values rather than original data points, but serve to identify each curve.) In Fig. 2.6b the data are normalized with respect to area by computing, for each volume velocity U, the corresponding Mach number

$$M = \frac{U}{A_c c}$$

where A_c = the cross-sectional area of the constriction. This normalization causes the curves for different cross-sectional areas to approximately superimpose. A difference between constrictions of different length is still maintained.

Similar data were obtained for the current setup by inserting constrictions of different lengths, cross-sectional areas, and shapes in the mouth of the tube, and measuring the pressures at flowrates between 50 and 690 cc/sec. For this measurement, the muffler was left out of the system. The pressure values used were those generated by a pressure tap 5 cm upstream of the mouth of the tube.

Figure 2.7a shows a plot of Δp versus the volume velocity U. For the assortment of constriction sizes used, the parameter most affecting the size of the pressure drop was the area of the constriction. At a given flowrate, the pressure drop increases as the area of the constriction decreases. The pressure drop also increases as the length of the constriction increases, and when the constriction shape is changed from circular to rectangular. These last two effects are more noticeable when the area of the constriction is smaller.

In Fig. 2.7b the data are again presented in normalized form. The curves generated from Heinz' data have a slightly smaller slope in general than those measured in this study, but the same trends with area and length are observed. In view of the extrapolation performed beyond the original range of U, the data match quite closely.

The following procedure was used to find the optimal values of K_2 and K_3 for each constriction. Assuming the form in Eq. 2.2, we solved for K_3 at each flowrate used, for incremental values of K_2. For each value of K_2 the mean and standard deviation of K_3 were computed. For a given constriction, the value of K_2 used was that minimizing the standard deviation of K_3. The mean value of K_3 computed across all flowrates for the chosen value of K_2 was then used for K_3. These values of K_2 and K_3 are given together with Heinz' values in Table 2.2.

If the linear term of Eqn. 2.2 does indeed retain the physical significance of the pressure drop due to the constriction length, K_3 should be proportional to constriction length and inversely proportional to the square of the constriction area. Both trends are evident for the data of this study, although an increase in length does not produce as large an increase in K_3 as predicted. The increase in the standard deviation of K_3 for the smaller areas indicates that the fit is less accurate in these cases. Since the range of Reynolds numbers used is higher for the smaller areas, the decrease in accuracy is probably due to the nonlinearities of the friction factor with respect to the Reynolds number that occur for Re > 2000. For Heinz' data the predicted proportionalities are much less accurate, though the trends are still in the predicted direction. Since his reconstituted data match the current data well, his unspecified method of finding

Constriction Shape	A_c	l_c	K_2	K_3	std dev of K_3
Data from this Study					
Circle	.020	1.0	.34	175.7	14.76
Circle	.020	2.5	.43	297.6	21.67
Circle	.079	1.0	1.00	8.7	1.19
Circle	.079	2.5	1.07	15.5	4.57
Circle	.318	1.0	1.03	.88	.25
Rectangle	.079	1.0	1.10	19.0	2.39
Rectangle	.318	1.0	1.01	1.36	.22
Data from Heinz' Study					
Circle	.031	1.0	.59	17.0	
Circle	.031	2.5	.78	22.0	
Circle	.049	5.0	.99	6.5	
Circle	.126	1.0	.66	0.6	
Circle	.126	2.5	.84	0.8	

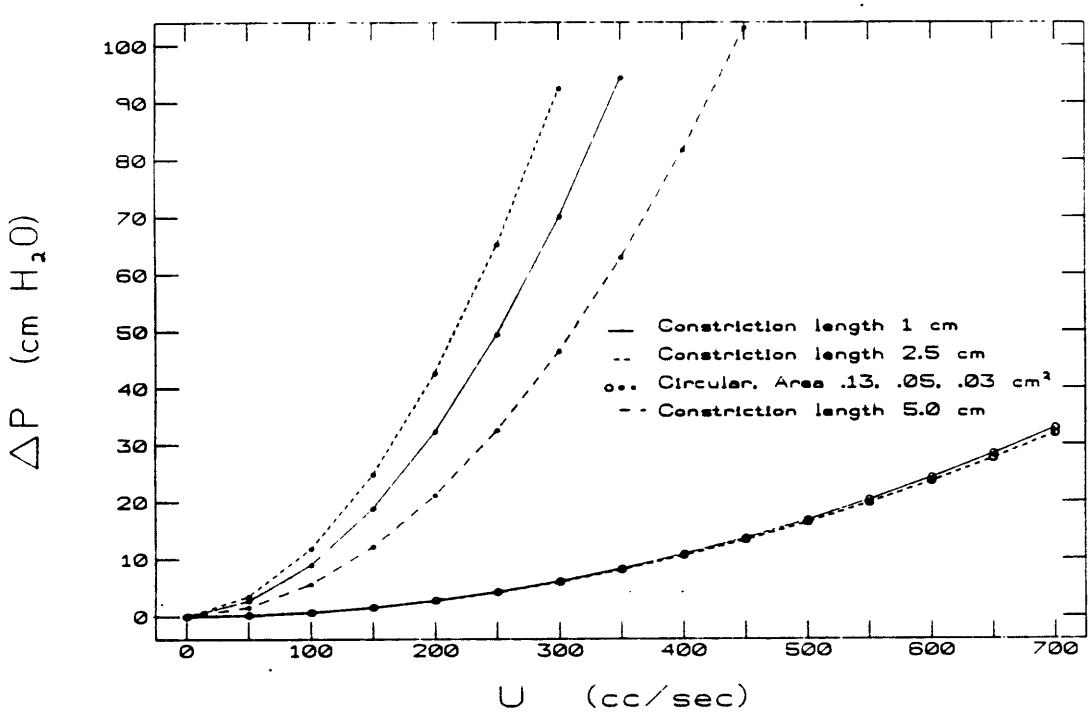
Table 2.2: Pressure-Flow Coefficients K_2, K_3 for Different Constrictions. Heinz' data are from his 1956 paper.

the coefficients appears to differ from the method used here in such a way as to obscure these relationships of K_3 to l_c and A_c.

Once an analytical expression has been fit to the pressure-drop data, the incremental resistance follows via simple differentiation:

$$R_i = \frac{\partial \Delta p}{\partial U} = K_2 \frac{\rho U}{A_c^2} + K_3$$

The coefficients tabulated above were used to compute values of R_i, which are graphed versus U in Fig. 2.8 and are used in the remainder of this thesis. Recognize that extrapolation to other constrictions or ranges of U may be inaccurate.



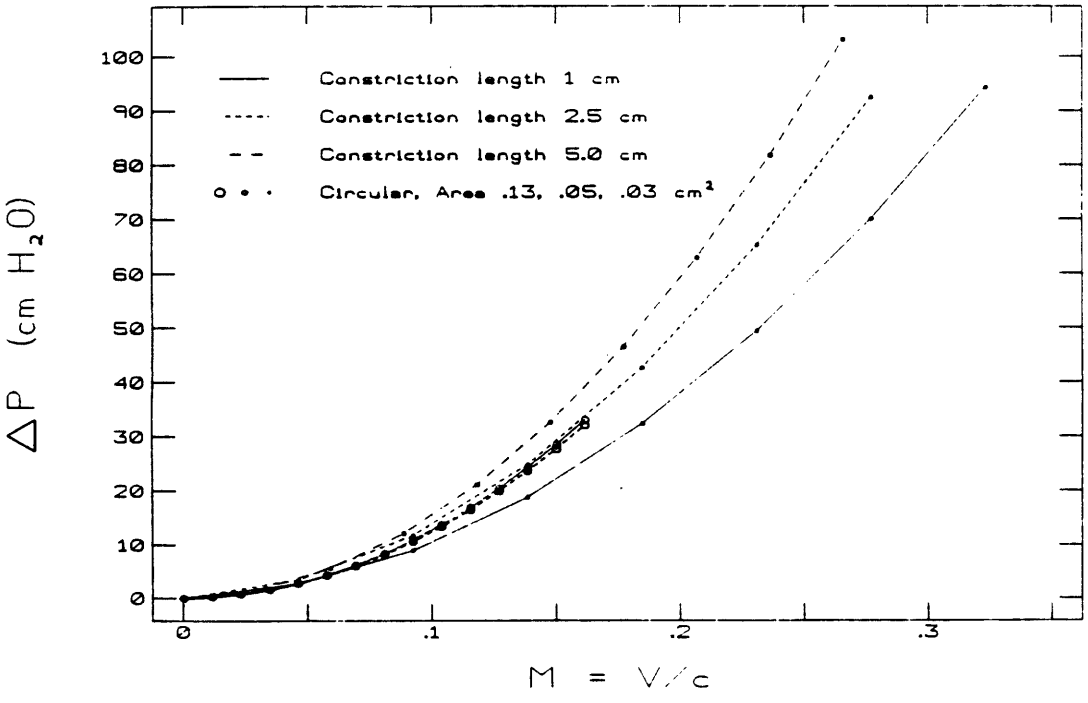


Figure 2.6: a) Curves generated from Heinz' data for the pressure drop Δp across a constriction versus the volume velocity U through it.
b) Same data as in a), plotted against the Mach number $M = V/c$, where V = linear flow velocity = U/A_c.

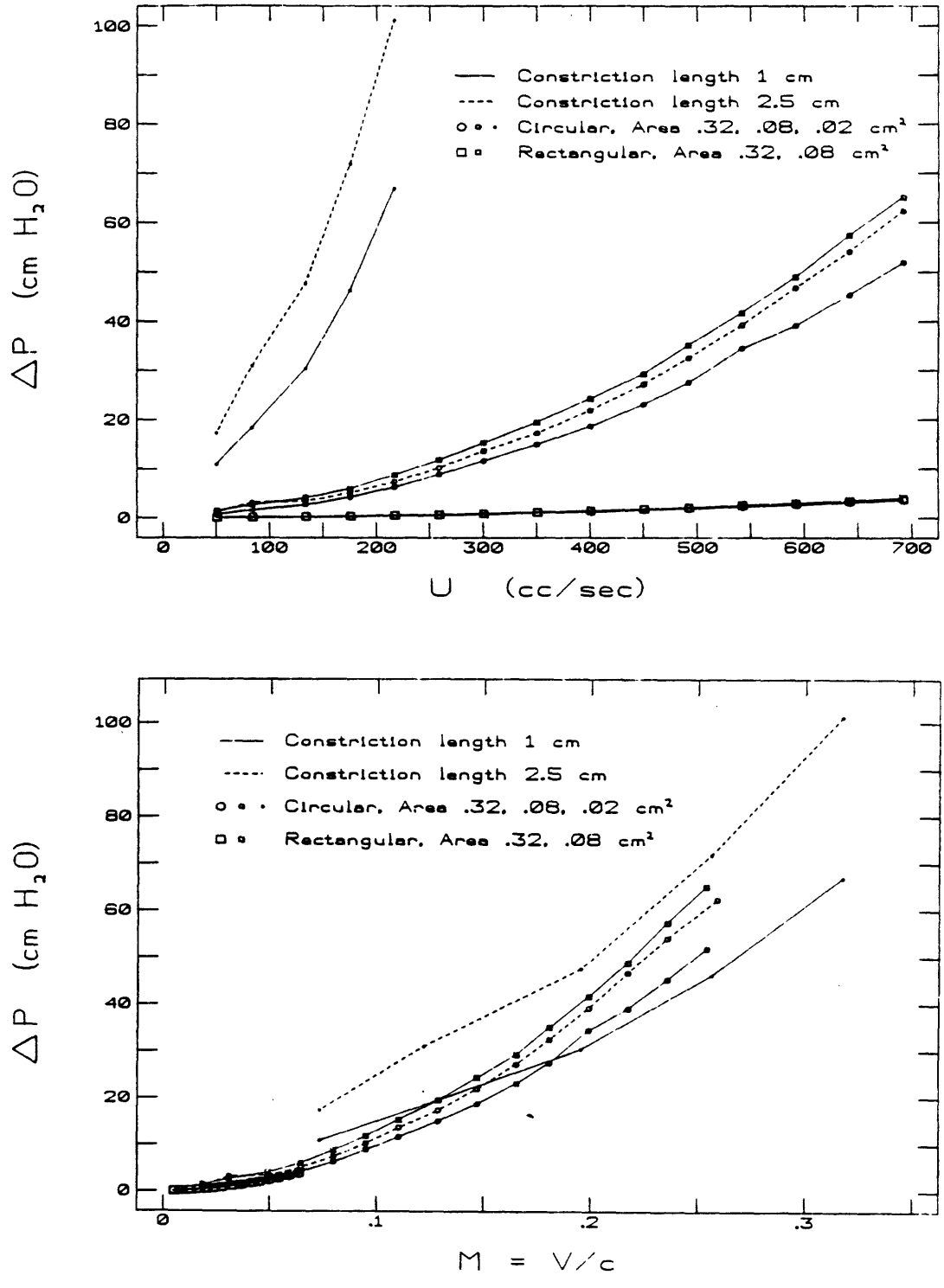


Figure 2.7: a) Experimental data for the pressure drop Δp across a constriction versus the volume velocity U through it.
b) Same data as in a), plotted against the Mach number $M = V/c$, where V = linear flow velocity = U/A_c.

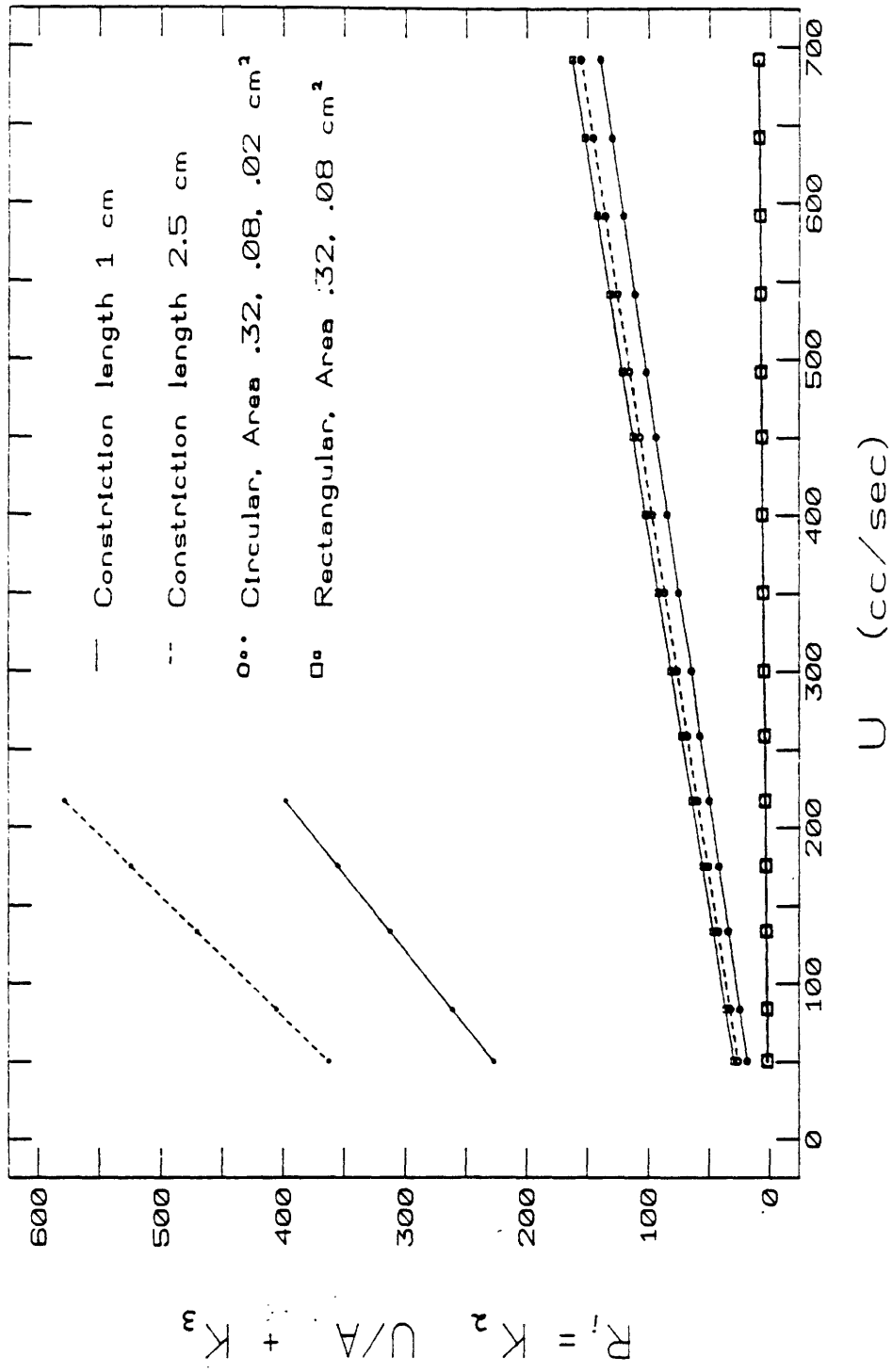


Figure 2.8: Incremental flow resistance of a constriction calculated from differentiation of Eqn. 2.2, fit to the data of Fig. 2.7.

2.3 Acoustic Results

In this section we consider stable configurations, with and without an obstacle present, and unstable configurations. The word "stability" refers not to the characteristics of the sound that is generated, but to characteristics of the flow. Unstable configurations are those that give rise to unstable boundary layers; unstable boundary layers are sensitive to acoustic feedback, and under the right conditions will oscillate, producing whistles.

2.3.1 Stable, No-Obstacle Configurations

Figure 2.9 shows spectra for two constriction diameters, two front-cavity lengths, and a range of volume velocities. The curves are distinguishable primarily by the frequencies at which the peaks occur, the bandwidths of the peaks, and the overall amplitudes. The parameter appearing to control the first two characteristics is the length of the front cavity; the flowrate mainly seems to control the overall amplitude.

The peak frequencies are in a ratio of 2:1 for $l_f = 3 : 6$ cm, and correspond roughly to the quarter-wavelength frequencies of the front cavity, i.e. $f = nc/4l$, $n = 1, 3, \ldots$. The bandwidths of the resonances are smaller for $l_f = 6$ cm, and increase with frequency for both front cavity lengths. The energy for these resonances is primarily stored in the front cavity, and subject to the various sources of loss in that cavity. As we shall see in Ch. 3, chief among these is radiation loss out of the mouth of the tube. Since the size of this loss increases with frequency, it accounts for the increase in bandwidth with frequency. Further, since more energy can be stored in a longer front cavity, the effect of the radiation becomes smaller as the front cavity grows longer, which accounts for the sharpening of the peaks for $l_f = 6$ cm.

The total sound power produced by a free jet is due mainly to that produced in the mixing region, that is, in the first four diameters downstream of the nozzle. According to Goldstein (1976), this total power is:

$$P_T \propto V^8 d^2$$

where d is the diameter of the nozzle, and V is the exit velocity of the jet. The spectrum of the free jet noise has a broad peak at $0.15V/d$. In the vicinity of the peak frequency, the power increases as the seventh power of the flow velocity; below, as the fifth power; above, the ninth power. We cannot compare our data directly to these numbers because the sound produced by the "free-jet" case, when the constriction is at the mouth of the tube, was too quiet to be measurable in the farfield below 5000 Hz. (Either a larger jet or quieter room is necessary.) However, we can make a rough comparison using the data for a non-zero front cavity length by measuring the differences between the amplitudes of the resonance peaks for sound pressure spectra at different flow velocities. The comparison is rough because we are working with spectra measured at a single angle, rather than spatial averages, and because we can only compare amplitudes at the frequencies of the resonances. The use of a single angle is not a serious problem because the directivity pattern is quite uniform, particularly at frequencies below 5000 Hz, for nonzero front cavity lengths. Thus the spectra measured at a single angle can be expected to show the same flow relationships as spectra averaged over all angles. The second problem, that of the limitation in frequencies at which we can make comparisons, means that we have only isolated comparisons and that we cannot pick our points of comparison arbitrarily close to the predicted peak frequency.

The results are shown in Table 2.3. Each constriction was measured for two front-cavity lengths. The extreme flow velocities used are given in the table; for each combination of l_f and

d cm	l_f cm	U cc/s	$.15V/d$ Hz	f_{res} Hz	$20\log(p_2/p_1)$ dB	n_{res}
0.32	3.2	$U_1 = 260$	1540	2000	18	7.6
		$U_2 = 450$	2670			
	6.0	$U_1 = 230$	1360	1200	20	6.9
		$U_2 = 450$	2670	4000	18	6.2
0.16	3.2	$U_1 = 140$	8750	2000	14	6.0
		$U_2 = 240$	15000			
	6.0	$U_1 = 140$	8750	1200	12	6.1
		$U_2 = 220$	13800	4000	10	5.1

Table 2.3: Sound pressure ratios for circular constrictions of diameter d, with front cavity of length l_f. Amplitudes measured at resonance frequencies f_{res}, as indicated. n_{res} is the power exponent at the frequency f_{res}.

d, U_1 is the lowest volume velocity, and U_2 the highest. f_{res} is the frequency of the resonance at which the measurement was made. Since power is proportional to pressure squared, the pressure-velocity relationship is

$$\begin{aligned} p &\propto V^{n/2} d^2 \\ \frac{p_2}{p_1} &= \left(\frac{V_2}{V_1}\right)^{n/2} \\ \log_{10}\left(\frac{p_2}{p_1}\right) &= \frac{n}{2}\log_{10}\left(\frac{V_2}{V_1}\right) \\ n &= \frac{20\log_{10} p_2 - 20\log_{10} p_1}{10\log_{10}(V_2/V_1)} \end{aligned}$$

where p_i is the sound pressure at flow velocity V_i, and n is the power exponent. In the table, the power exponents derived from pressure differences measured at the resonance frequencies are indicated by n_{res}.

For the 0.32 cm diameter constriction, the peak amplitudes follow a seventh power law; for the 0.16 cm diameter constriction, they follow a sixth power law. Recall that the power exponent for a free jet decreases gradually from a value of seven at the peak to five at frequencies below the peak. Since for the smaller constriction the amplitude is measured at a lower frequency relative to the predicted source spectrum peak, these results are consistent with the free jet power law.

In order to see whether the sound spectra scale by diameter as well as by flow velocity, spectra are compared for different constriction areas in Fig. 2.10, with U chosen to give, as

nearly as possible, the same linear flow velocity. Although the identical velocity could not be shown due to limitations of the apparatus, the velocities are roughly equivalent. The curves show large amplitude differences, with the larger area generally producing the higher amplitude. For the velocities shown, the V^6d^2 proportionality actually predicts the reverse, that is, that the sound spectra measured with the larger area and lower velocity should be slightly lower in amplitude in each case. Since, from Fig. 2.9 and Table 2.3, the data for each constriction diameter considered separately exhibited the expected relationship of amplitude to flow velocity, the differences shown in this figure must be due to the effect of the tube section downstream of the constriction.

Let us consider how the tube might have this observed effect of increasing the amplitude of the spectra produced by constrictions with larger areas. The average diameter of a free jet increases linearly as a function of the numbers of diameters downstream of the nozzle (Goldstein, 1976). This affects excitation of the tube resonances in the following way: a larger diameter jet will intersect with the walls of the tube nearer to the constriction outlet, whether measured in absolute terms or in constriction diameters. Thus, if we consider two jets of different diameters but the same flow velocity, additional sound generation at the walls is more likely for the larger jet, since it will strike them fewer diameters downstream and therefore with a higher proportionate velocity.

Figure 2.11 shows the effect of changing the constriction length. For the 2.5 cm constriction, the bandwidth of the lowest peak is smaller, especially for the higher flowrate shown. From Section 2.2 and Fig. 2.8, we know that the increased length raises the flow resistance of the constriction. It also increases the inductance of the constriction. As we shall see in Ch. 3, if both the resistance and inductance of the constriction increase, the net effect is a decrease in bandwidth of the front-cavity resonances. The effect is quite small, however. More noticeable is the effect on the resonance structure. The resonance related to the half-wavelength of the constriction by $f = c/2l_c$ is predicted to be 13 kHz for $l_f = 1.0$ cm, but 6.2 kHz for $l_f = 2.5$ cm. As a result, for the 2.5 cm case, both the front-cavity resonance and the constriction resonance occur at nearly the same frequency, which increases the amplitude around 6000 Hz quite noticeably.

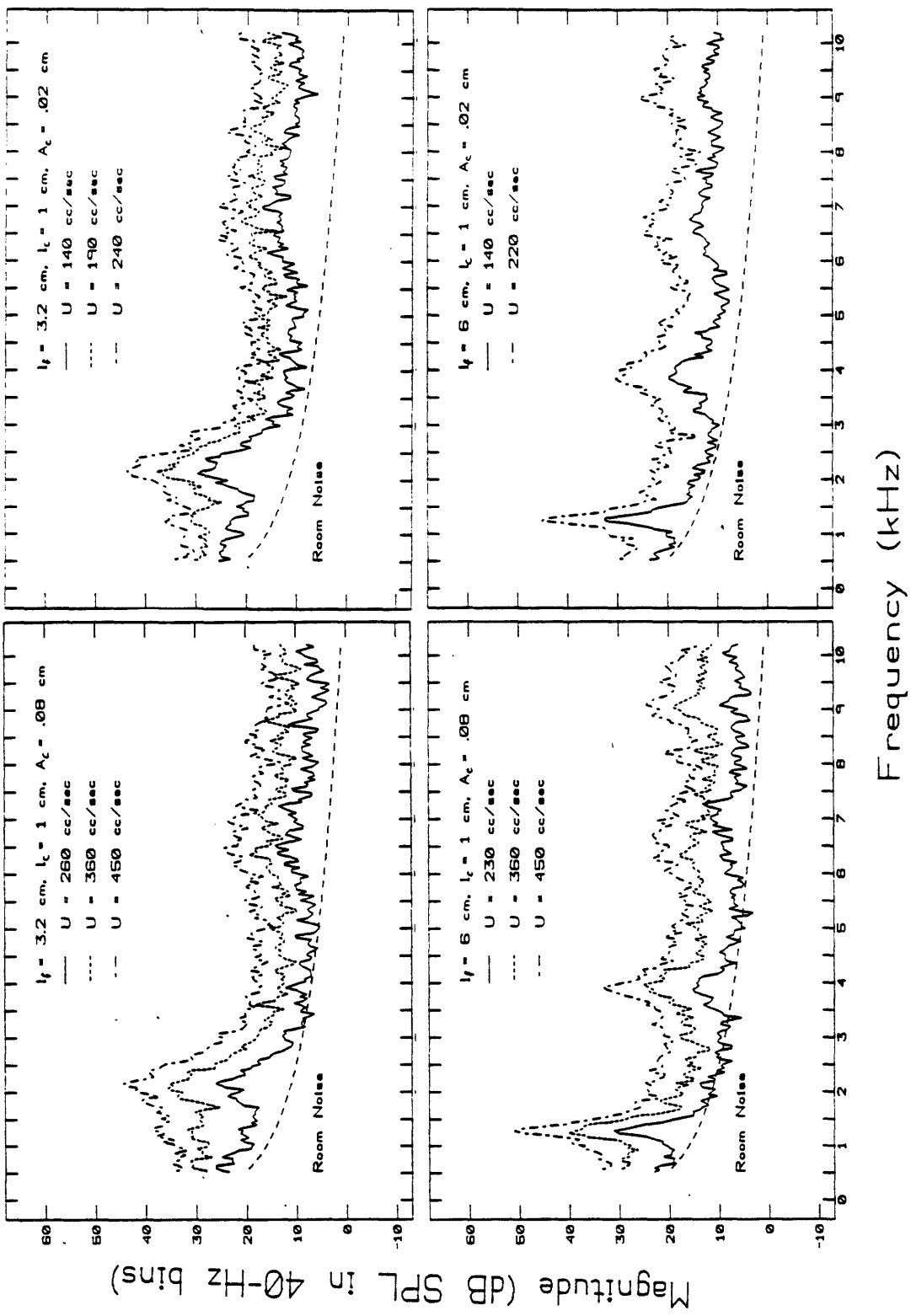


Figure 2.9: Averaged power spectra for front-cavity lengths l_f = 3.2 and 6 cm, and constriction areas A_c = 0.08 and 0.02 cm^2. For definition of other variables, see Figs. 2.3 and 2.4.

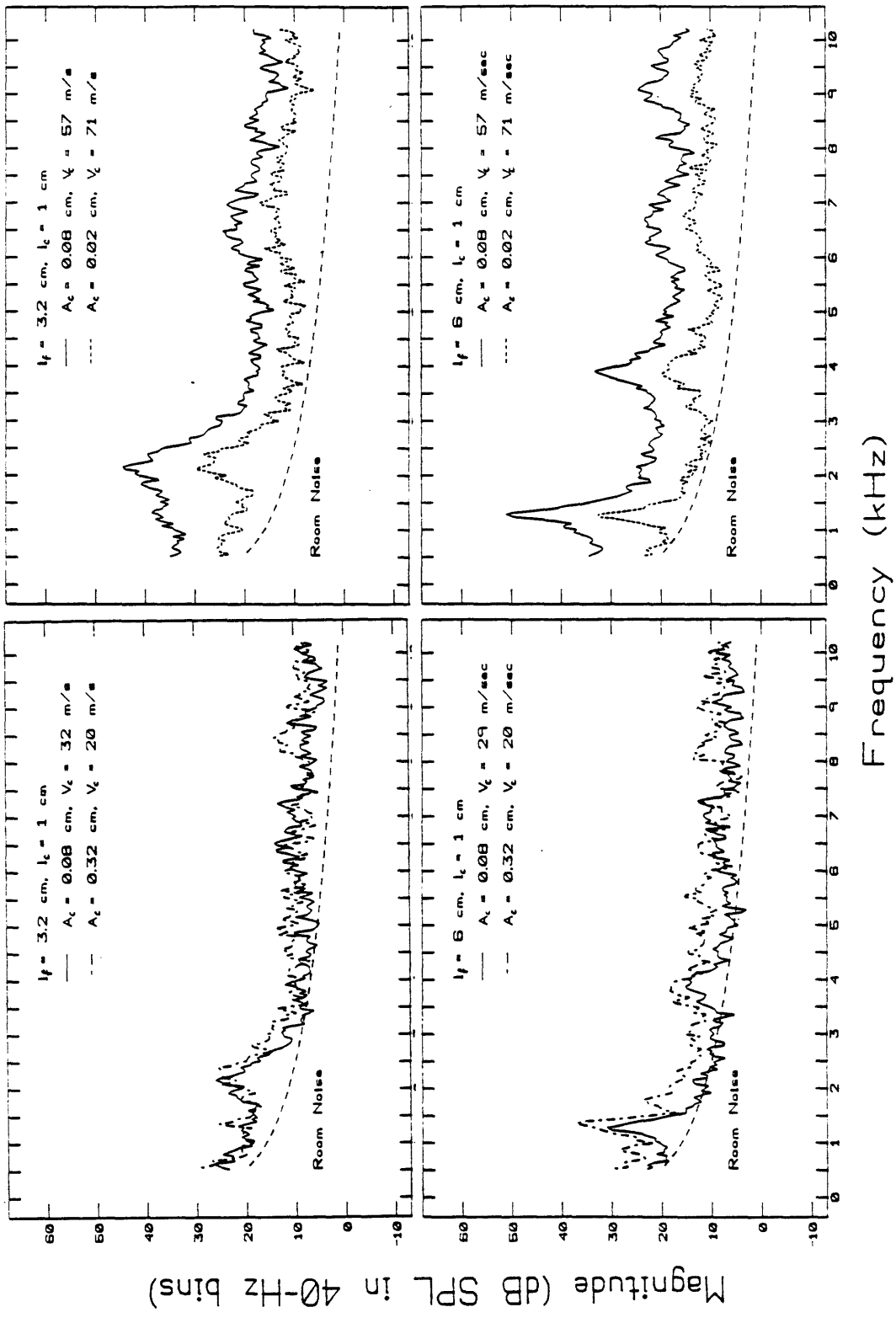


Figure 2.10: Averaged power spectra for l_f = 3.2 and 6 cm and A_c = 0.02, 0.08 and 0.32 cm^2. As nearly as possible, each graph shows the same velocity through the constriction, V_c. For definition of other variables, see Figs. 2.3 and 2.4.

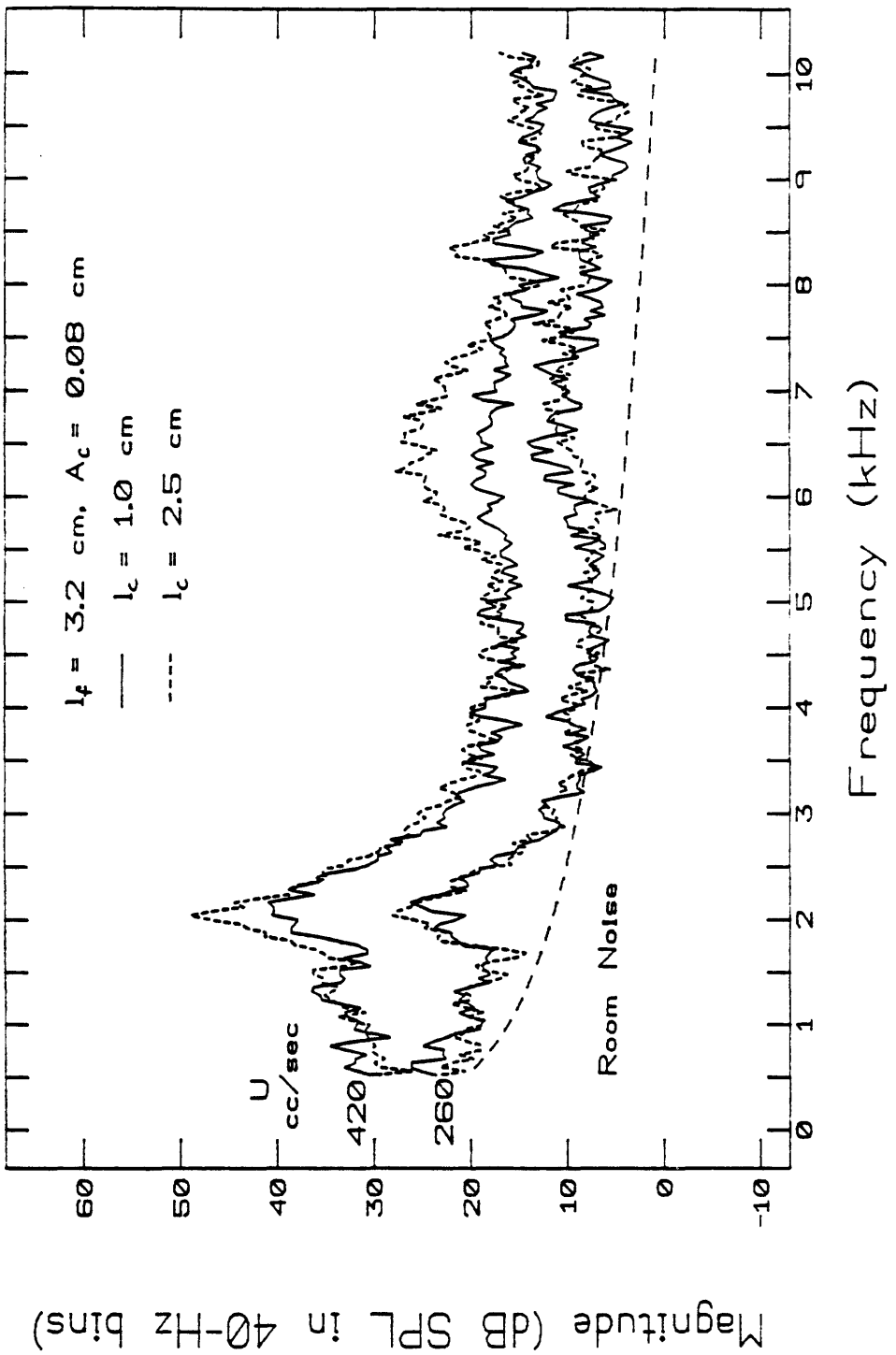


Figure 2.11: Averaged power spectra for l_c = 1.0 and 2.5 cm. For definition of variables, see Figs. 2.3 and 2.4.

2.3.2 Unstable Behavior

Constriction Tapering

The above results utilized constrictions with tapered inlets, and no whistles occurred. (Whistles were identified both by their distinctive sound and the appearance in the spectrum of peaks that were high amplitude, very narrow bandwidth, and very steady in frequency compared to the peaks in an audibly non-whistling spectrum.) When an untapered constriction is used, whistles occur at low flowrates. Two examples are shown in Fig. 2.12, due to constrictions of lengths 1.0 and 2.5 cm positioned at the mouth of the tube. A comparison of the two lengths reveals that the 2.5 cm constriction whistles dramatically; its spectrum exhibits large amplitude narrowband components at 6.2 and 15 kHz. The 1 cm constriction whistles with a fundamental near 12 kHz. In general, the whistle frequency is at the half-wavelength of the effective length of the constriction, that is, $f = c/2l_e$. The effective length is the constriction length, l_c, plus an end effect correction due to radiation of $8d_c/3\pi$ (Morse, 1976). For $l_c = 2.5$ cm, a resonance is predicted at 6.22 kHz, which is close to the observed whistle at 6.2 kHz (particularly given that the measurement resolution bandwidth was 100 Hz). For $l_c = 1.0$ cm, the resonance is predicted at 13.55 kHz, and the whistle occurs at 12.9 kHz.

The frequency at which these configurations whistle, plus the fact that the whistles cease when a tapered constriction is used, make it easy to identify the acoustic mechanism involved. Specifically, they are "orifice tones" (as described by Anderson 1954, Succi 1977, and others), and are caused by unstable boundary layers within the constriction itself, which form at the sharp-edged inlet. Succi identified the length to diameter ratio as a critical parameter that determined whether or not a given constriction would whistle in this manner. He found that, of constrictions with l_c/d between 0.25 and 4.0, those with $l_c/d = .25$ or 4.0 did not whistle, while those with l_c/d between 0.5 and 2.0 did. In the present study, however, untapered constrictions with ratios from 3.1 up to 8.0 all whistled. The discrepancy may be due to differences in the experimental setups, which could, for instance, result in differences in the degree of turbulence of the air as it enters the constriction. Although the value of the l_c/d parameter is not sufficient to predict the presence of whistles, its use still makes sense physically. Viscous forces cause the jet to expand back to the walls of the constriction. If the flow reattaches, the potentially unstable length of the boundary layer is limited, and the feedback loop is broken. Thus, increasing l_c or decreasing d, either of which increases the viscosity, should limit the ability of the constriction to whistle (Succi, 1977).

None of the whistles described above occur if the constriction inlet is tapered. Although humans can produce whistles, abrupt edges in the vocal tract generally do not occur, and thus the human whistle mechanism must be of a different sort. Therefore, to avoid the generation of whistles which are not physiologically realizable, all other experiments reported in this thesis used constrictions with tapered inlets.

Whistles due to Obstacle

The use of an obstacle to investigate the extent of the jet revealed that an obstacle significantly changes the character of the sound produced and also can, under some circumstances, produce whistles. Since the sound produced varies considerably with distance to obstacle, flowrate, and portion of the jet that is occluded, some simplification was achieved by using a semicircular obstacle, since an obstacle of this shape occludes half of the jet regardless of its distance from the constriction. Whistles produced with this obstacle were then systematically

investigated by varying the obstacle-constriction distance and the flowrate. One goal of this investigation was to characterize the whistle behavior of the obstacle. The other goal was to determine stable, non-whistling regions, for experiments to be discussed in the next section and in Ch. 3.

Two constrictions with the same cross-sectional area and having the shapes of a circle and a rectangle were used. The area was 0.08 cm^2; the circle's radius was 0.318 cm, and the rectangle was 1.27 by 0.062 cm. When the rectangle was used, the flat edge of the obstacle was parallel to the long side of the rectangle. For both constrictions, spectra were measured with each of several obstacle positions, for two constriction positions, at a few flowrates. In addition, the entire range of flowrates was scanned at a few obstacle positions in order to identify whistling and non-whistling regions.

Figure 2.13 summarizes the combinations of l_o and U that produce whistles for these two constrictions, with and without a front cavity. It can be seen that the rectangular constriction whistled much more readily than the circular one, and that the presence of a surrounding resonator - the $l_f = 3$ cm case - increased the whistling range for both constrictions. Figure 2.14 contrasts typical spectra for the circular and rectangular constrictions. In general, the rectangle produced louder whistles, and clearer whistles (in the sense of having fewer and more obvious peaks), as well as whistling over a greater range of distance and flowrates. These differences arise, presumably, because a flat jet, as produced by the rectangular slot, impinging on a straight edge will have a more uniform flow velocity along that edge and therefore a more uniform pattern of vortex generation than a round jet impinging on the same edge will have.

When either the circular or rectangular constriction is located at the mouth of the tube ($l_f = 0$ cm), an increase in flowrate causes the whistle frequency to increase steadily and then at some point to jump abruptly upwards. The transition may include situations in which two whistle peaks are present simultaneously, or in which the whistle warbles between two frequencies. This behavior is typical of whistles in general. Likewise, as the obstacle is moved away from the constriction, the frequency of the whistle gradually falls, then abruptly increases and begins to fall again. This behavior is typical of edgetones.

Small changes in obstacle position can produce complex changes in the whistle patterns. In Fig. 2.15, shifting l_0 from 0.41 to 0.38 cm not only increases the main peak frequency (near 5 kHz) and its harmonics by about 9 percent, but introduces new peaks at intermediate frequencies.

When $l_f > 0$ cm, that is, when the constriction and obstacle are within the tube, the general trend of the whistle frequencies as l_0 and U are changed is similar to the case where $l_f = 0$ cm, but the spectra at comparable points will be quite different. Two examples obtained using the rectangular constriction are shown in Fig. 2.16. For both the circular and rectangular constrictions, the presence of the tube stabilized the whistle frequency somewhat and increased the likelihood that a given configuration would whistle. For the circular constriction, the latter change was marked. It may be that the tendency of the tube to emphasize the longitudinal modes serves in this case to reduce the effect of the nonuniformity of the flow velocity along the obstacle. In general, the amplitude of the whistles was increased when the tube was present, although the reverse is sometimes true for particular values of l_o and U.

These effects can be seen more clearly by extracting peak frequencies and amplitudes from the various spectra and presenting them together. Figure 2.17 shows plots of the normalized peak frequency, the Strouhal number $S = fd/V$, versus the normalized distance from the constriction, l_0/d. Only the frequency of the highest amplitude peak was used as the value of

f in computing S. The velocity V was taken to be U/A_c. For the circle, d was taken equal to the radius of the circle; for the rectangle, d was taken to be the shortest dimension. For the rectangle, S gradually falls as l_0/d increases, and then abruptly increases (e.g., at $l_o/d = 6$ or 9) and begins falling again. This behavior keeps S fairly constant. For the circular constriction, S is not nearly as constant. To the extent that the whistles occur, we do see the same pattern of gradual fall and abrupt rise, but this definition of S does not seem to be as powerful a normalizer as for the rectangular constriction.

Whistles tend to be weakest when they are near a transition. The plot of the peak amplitude versus l_0/d clearly shows an alternation of high and low amplitudes as l_0/d increass, and a drop in amplitude when the tube is added, for the rectangular constriction. For the circular constriction, the amplitude is apparently enhanced by the presence of the tube.

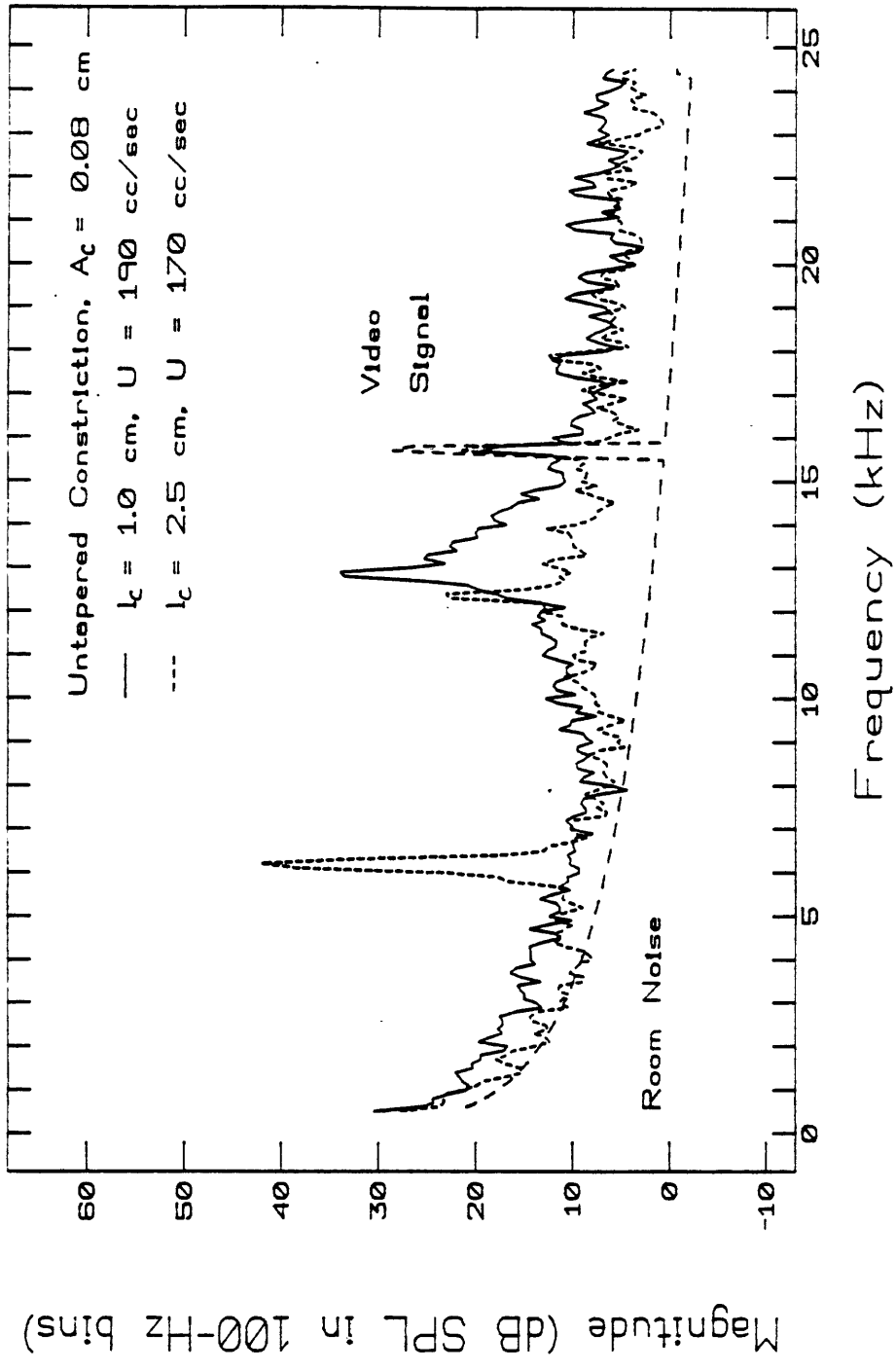


Figure 2.12: Averaged power spectra of untapered constrictions of lengths 1.0 and 2.5 cm, located at mouth of tube ($l_f = 0$ cm).

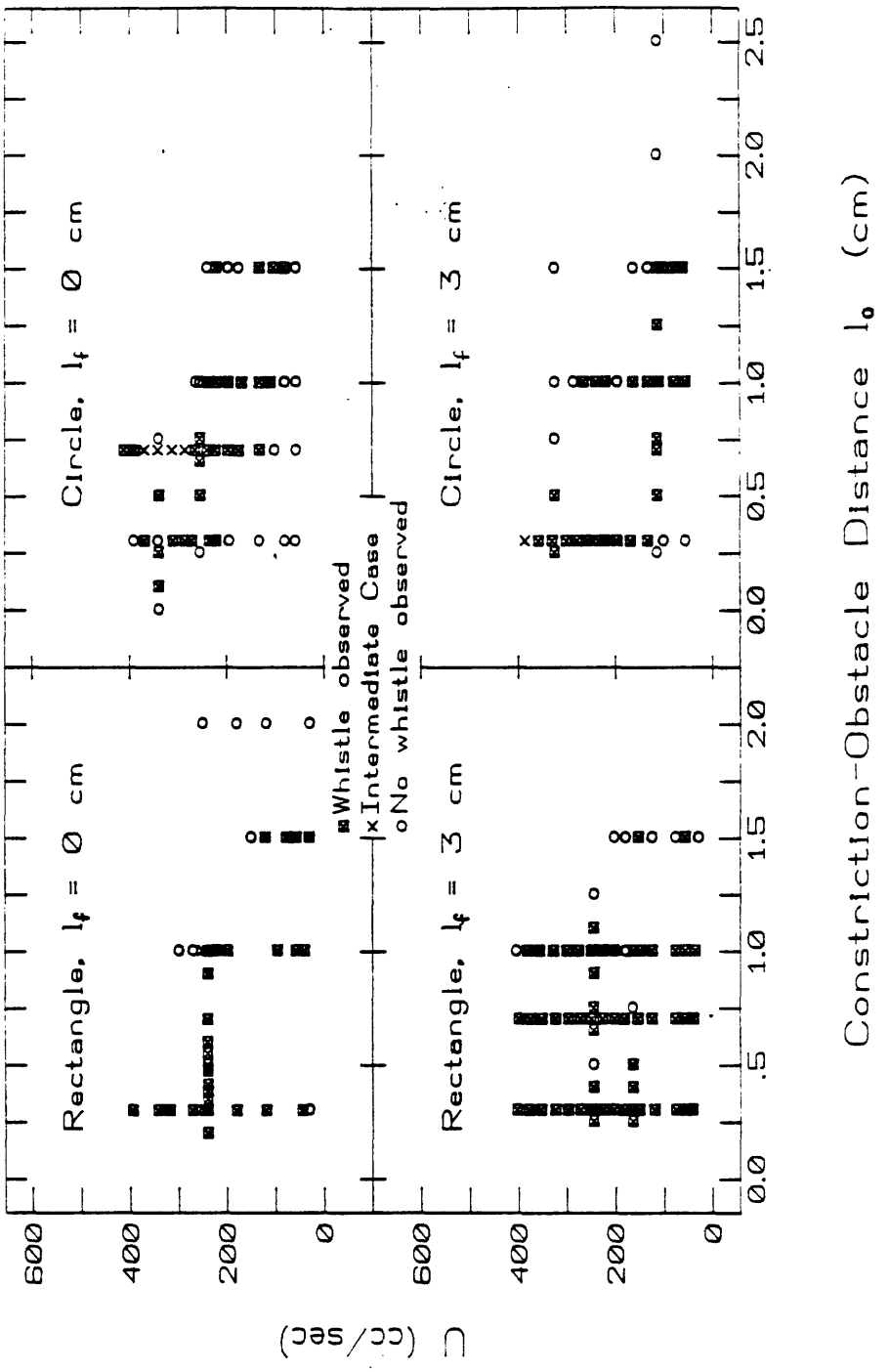


Figure 2.13: Summary of the conditions giving rise to whistles for a constriction-obstacle configuration.

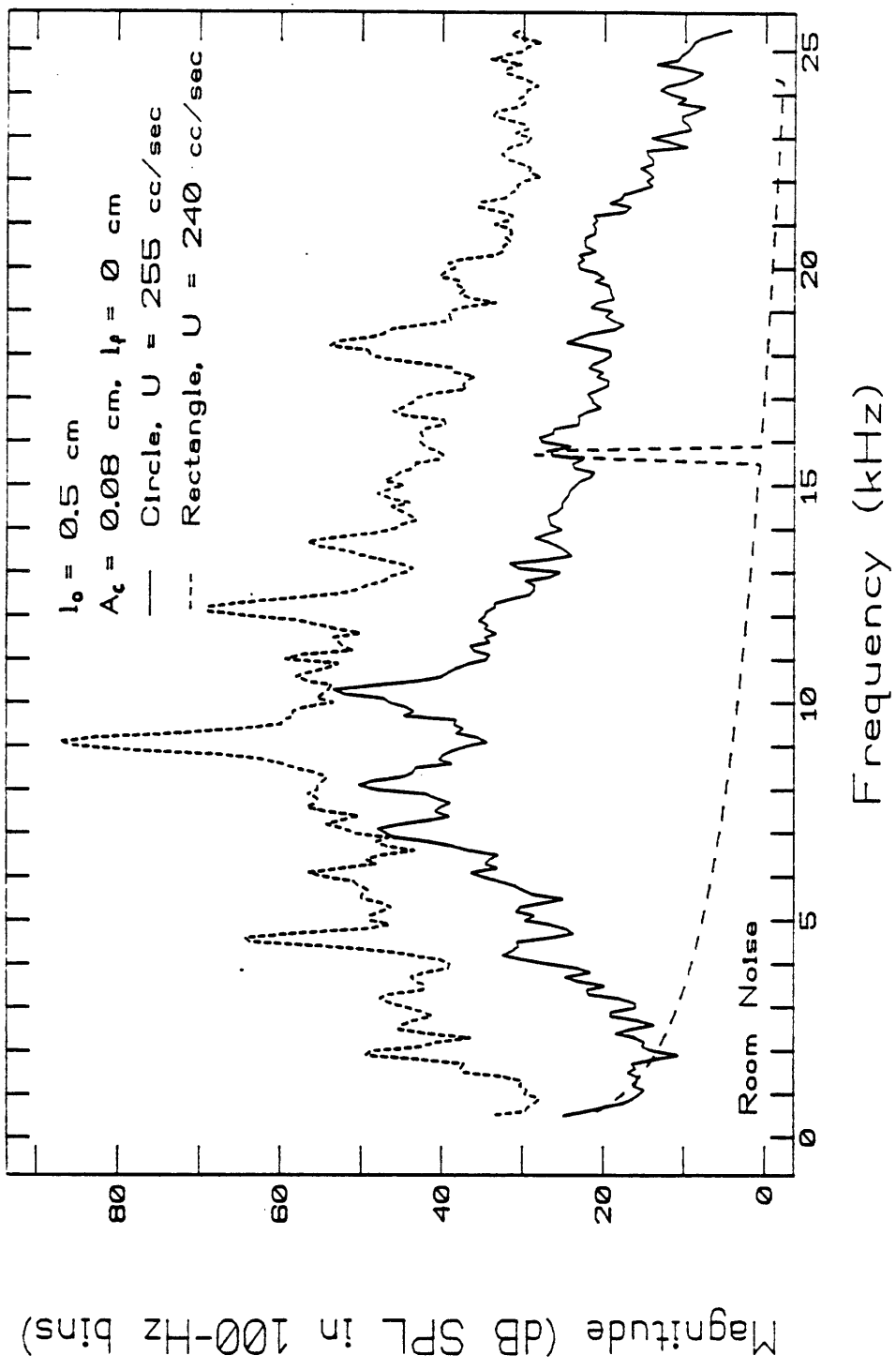


Figure 2.14: Typical spectra for the circular and rectangular constrictions at mouth of tube, with semicircular obstacle a distance l_0 downstream.

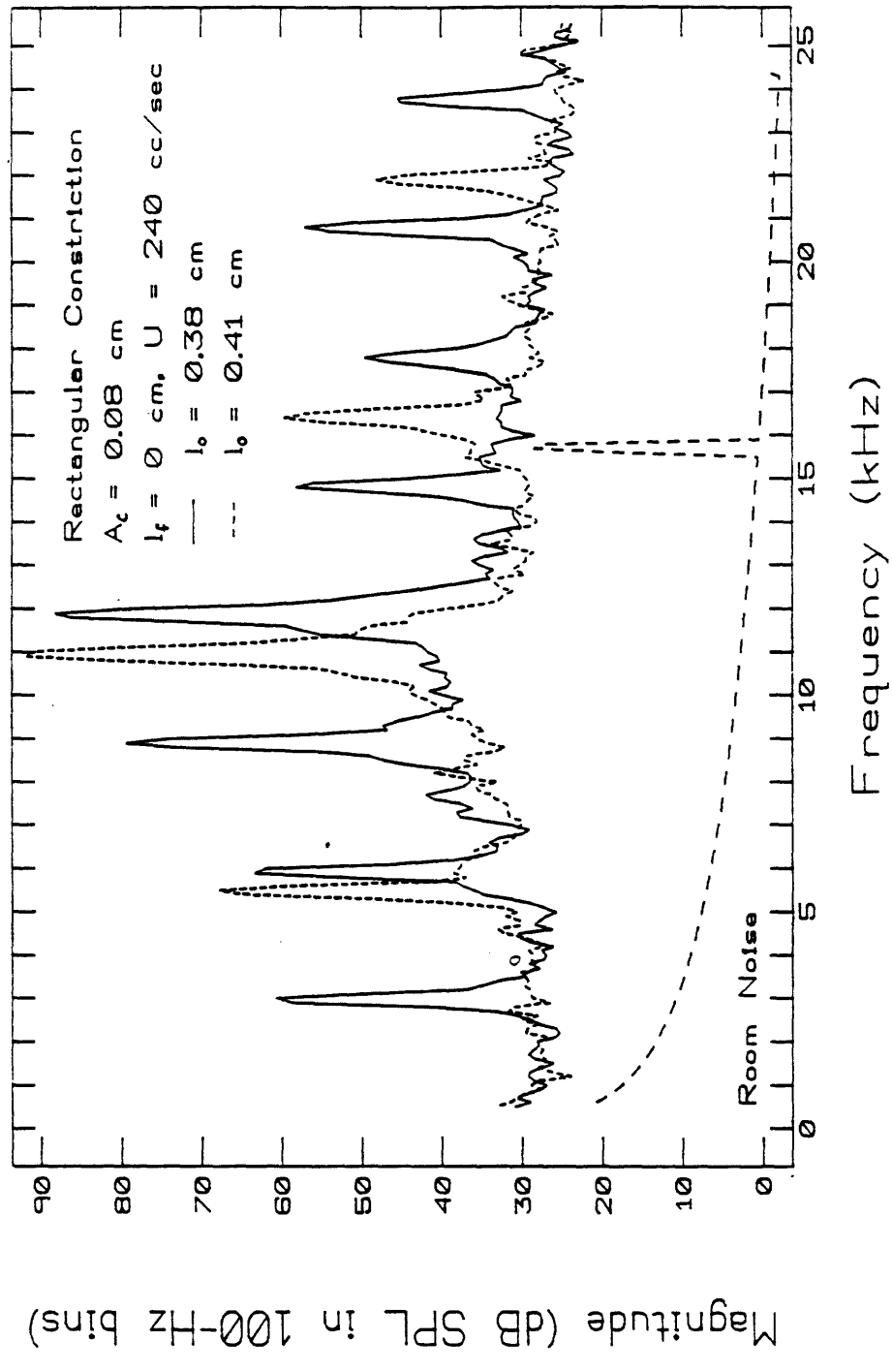


Figure 2.15: Spectra generated by constriction-obstacle configurations differing only in their values of l_0.

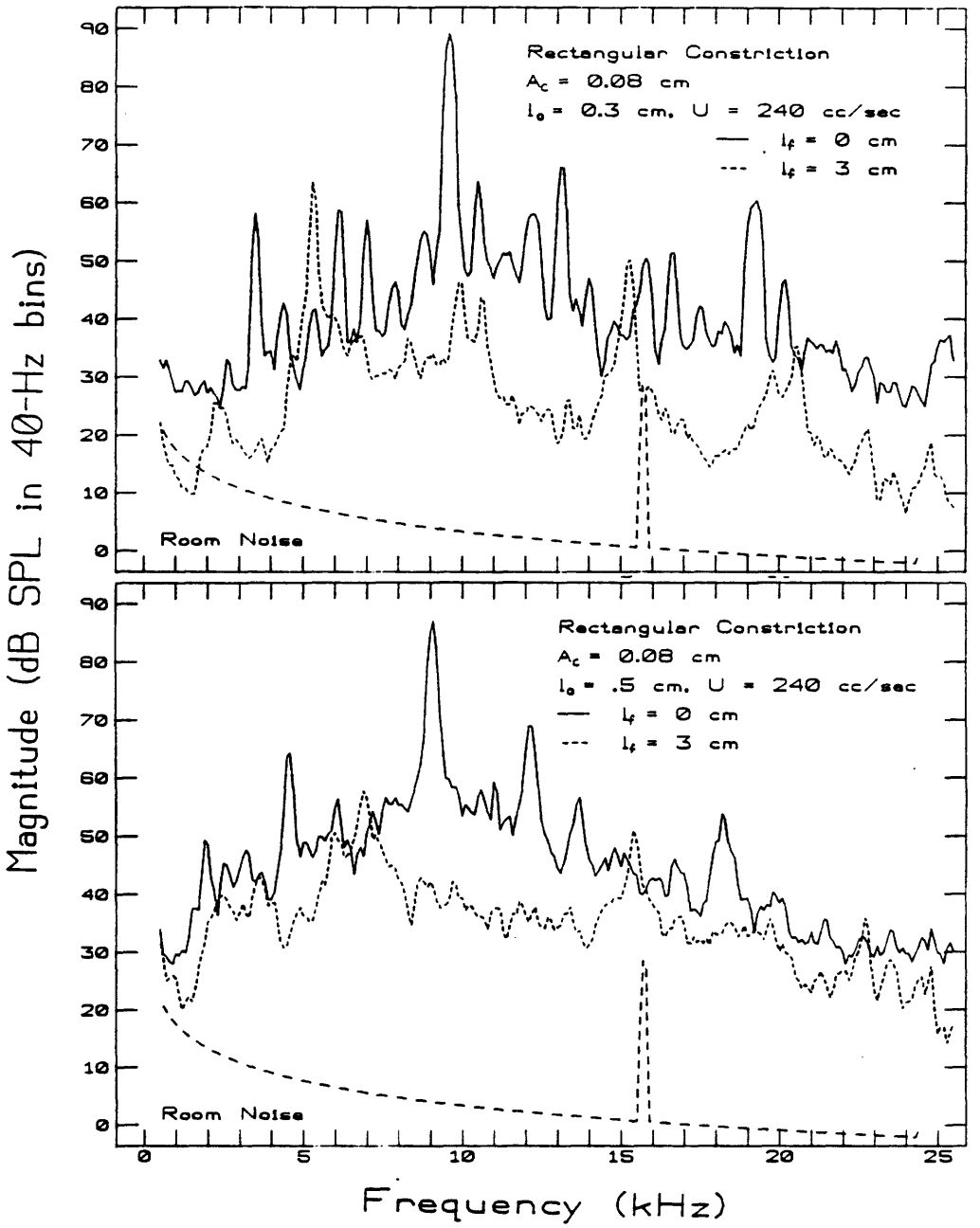


Figure 2.16: Spectra generated by constriction-obstacle configurations, contrasting length of the front cavity at $l_0 = 0.3$ and 0.5 cm.

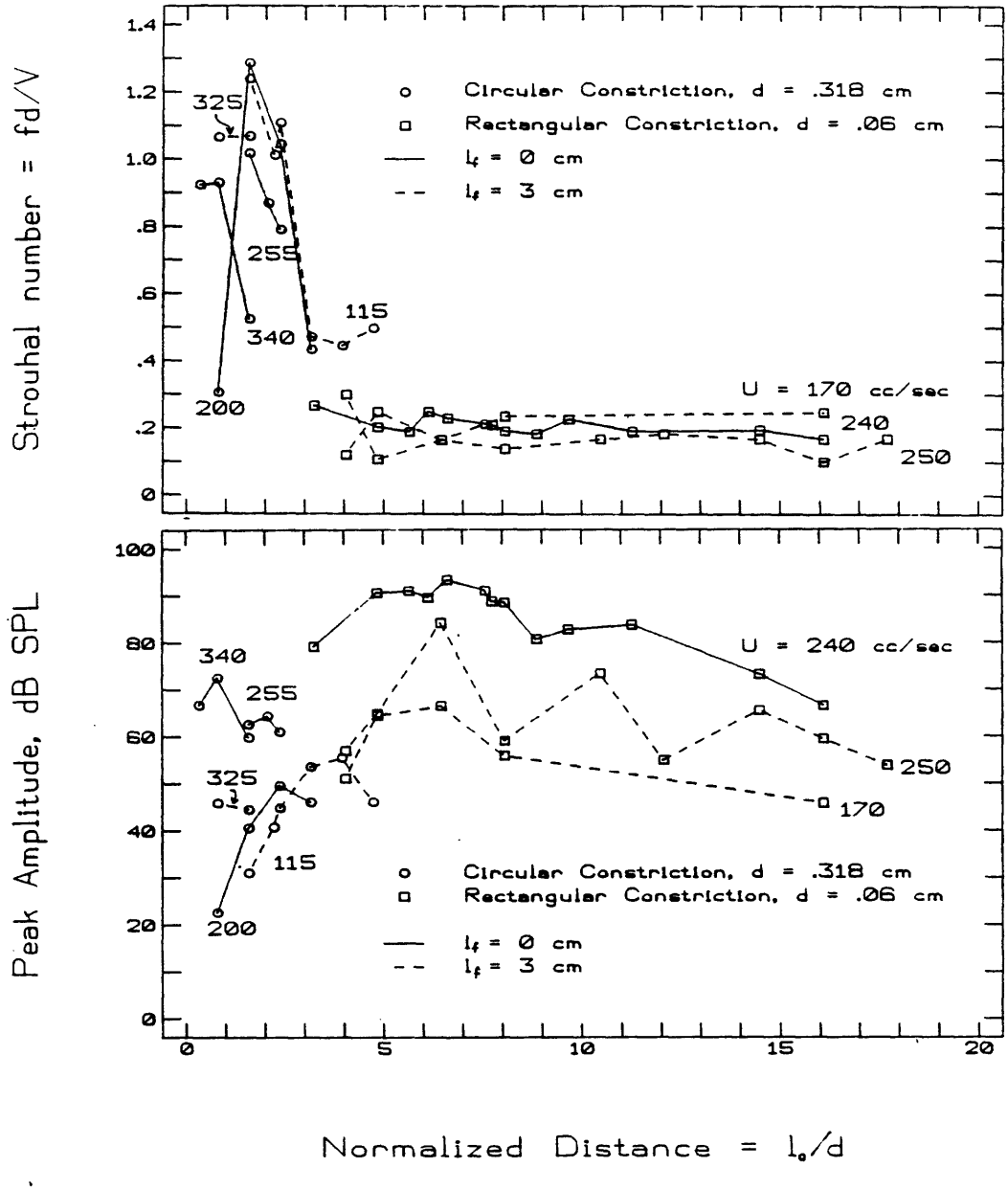


Figure 2.17: Strouhal number and peak amplitude versus normalized distance to obstacle, for whistles generated by the constriction-obstacle configuration. Values shown represent the highest-amplitude peak in the spectrum of each whistle. See text for the definition of d.

2.3.3 Stable Obstacle Configurations

Clearly the whistles are substantially different from sound produced without an obstacle. But outside the whistle range, the sound is different also. Figure 2.18 contrasts spectra generated by a constriction in a tube with and without an obstacle downstream. When the obstacle was present, it was 3 cm downstream, well out of the whistle-generating region. The obstacle increases the amplitude of the spectrum considerably and also increases the dynamic range, that is, the prominence of both poles and zeros, as evidenced by the size of the amplitude excursions across the entire frequency range of a given spectrum.

Figure 2.19 shows spectra for the constriction positioned flush with the baffle ($l_f = 0$ cm), and three different distances to the obstacle. As l_0 increases, the frequencies of the troughs in the spectra decrease; these frequencies occur approximately at $f = nc/2l_0$, where $n = 1, 2, \ldots$, that is, the half-wavelength resonances of the constriction-obstacle distance.

Variations in the angle θ between the jet axis and the microphone also influence the measured spectra. As θ increases, the trough frequencies shift and the spectral amplitude generally decreases (see Fig. 2.20). The amplitude reduction is further demonstrated in terms of the overall sound level as a function of θ (see Fig. 2.21).

A source of sound positioned near a baffle will generate reflections and therefore interference at frequencies related to the distance between source and baffle. This explains the orderly shifting of the minima as a function of both θ and l_0. The amplitude minimum over all frequencies as a function of θ, which in this case occurs around 90°, is related to the characteristics of the source alone. In fact, it is quite reminiscent of the directivity pattern of a dipole. As will be discussed more thoroughly in Ch. 3, there is good reason to expect a dipole-like directivity pattern for sound generated by a jet striking an obstacle.

The smooth curves plotted on Fig. 2.21 constitute the predicted directivity pattern for a dipole positioned 3 cm from a baffle. They were computed using the expression for the far-field sound pressure generated by a dipole (Beranek, 1954) and adding a term to represent the image dipole:

$$p(r,\theta,\omega) = \frac{\omega^2 \rho}{4\pi c} Ud \left[\frac{\cos\theta}{r} e^{-jkr} - \frac{\cos\phi}{R} e^{-jkR} \right]$$

where

Ud = dipole strength
r = distance between dipole and observer (microphone)
θ = angle between downstream jet axis and r vector
R = distance between image dipole and observer (microphone)
ϕ = angle between downstream jet axis and R vector
l_0 = distance between dipole or image dipole and baffle
$\omega = 2\pi f$, f = frequency
ρ = density of air
c = speed of sound in air

Note that R and ϕ are functions of r, θ and l_0.

Since the source strength Ud is unknown, the curves were scaled according to the measured amplitude at 12°. Further, since $p(r,\theta,\omega)/Ud$ increases proportional to frequency squared, an amplitude computed over the entire frequency range would show the effect of the shifting minima, which we are not interested in. Consequently, only values between 2 and 3 kHz were used to compute the theoretical amplitude, since this range excluded the frequency-shifting minima due to the presence of the baffle. The amplitudes for the measured curves were also

computed over this range, which is generally in the vicinity of the maximum amplitude of each spectrum and therefore is the least affected by the ambient noise. Although the measured and theoretical curves exhibit similarly broad minima around 90°, the theoretical curves extend to somewhat lower amplitudes. The discrepancy may be due to the noise generated by the jet itself, which would not be expected to have the directionality of a dipole.

Amplitudes were also computed for the measured data over the range 1500 to 10200 Hz. Summation over a greater number of frequency bins than for the 2-3 kHz range made the directivity patterns somewhat smoother, and also made the minimum near 90° shallower, presumably the effect of the background noise. For either case, the theoretical and measured curves do not match perfectly, but are close enough to suggest quite strongly that the obstacle generates a dipole-like source.

2.4 Discussion

In this chapter the aerodynamic and acoustic effects of various parameters describing the flow and the physical configuration of a constriction in a tube have been investigated. The intent has been threefold: to establish useful methods with which to obtain and analyze mechanical model data, to focus on the parameters that are the most important in terms of their acoustic impact and likelihood of occurrence in fricative production, and to begin to quantify the effects observed.

The qualitative assessment of the various parameters revealed that the presence of an obstacle, the length of the front cavity, and the flowrate have the most effect on the sound produced. Efforts to quantify the effects included the derivation of flow resistance; prediction of the whistle frequencies for the untapered constriction, and of the resonance and anti-resonance frequencies for the stable obstacle and no-obstacle configurations; and comparison of the no-obstacle spectra to Lighthill's V^8 power law. Prediction of resonance frequencies, indeed of the entire spectrum, can be made more accurate by using derived data such as the flow resistance and directivity patterns.

In Chapter 3 we will do just that: derive a more accurate model and compare it to experimental data. In many ways, the data reviewed so far lend themselves to a source-filter model: we speak of the source characteristic of a free jet, and the tract resonances that are excited by that jet. A significant class of exceptions is that of whistling, in which the "source" interacts so strongly with the "filter" as to make it unclear which parts of the configuration and jet contribute to each theoretical entity. Since it is of interest whether other configurations exhibit source-tract interaction, whistling cases were specifically excluded from consideration in Ch. 3.

Further simplification was achieved by varying only the parameters identified as most important (U, l_f, and presence of obstacle), and freezing the other parameters at values that were convenient experimentally and, where possible, typical for speech. Hence, the constriction was tapered (to eliminate the type of whistles that a vocal tract is not likely to produce), of area 0.08 cm^2 (approximately the area used for fricatives), circular (minimum viscosity for its area, no effective-width formulae necessary), and 1 cm long (minimum length for fricatives, but short enough to discourage whistle-like behavior). The obstacle, when used, was semicircular, both because it occluded all parts of the jet regardless of l_0 and because it approximates the shape of teeth, the most likely obstacle in the vocal tract. It was positioned 3 cm away from the constriction in order to avoid regions where whistles could occur. All measurements were made at a single angle of 28°, which was chosen fairly near the jet axis to get maximum amplitude,

but not on it to avoid wind noise.

Two front-cavity lengths, 3.2 and 12 cm, were chosen for two basic reasons. The difference in front cavity lengths affects the resonance frequency and bandwidth noticeably, and places the obstacle in completely different locations relative to the expected standing wave patterns. Yet even with these differences, the fact that the two lengths are nearly integrally related means that they have resonances at the same frequency, and thus have at least one spectral feature in common.

We arrive then at four models total, composed of two constriction positions, each with and without the obstacle. These will be subject to much scrutiny because they are the essence of the acoustic mechanism for fricatives. Thus we will refer to them collectively as the idealized fricative models, or separately as the idealized obstacle and no-obstacle cases.

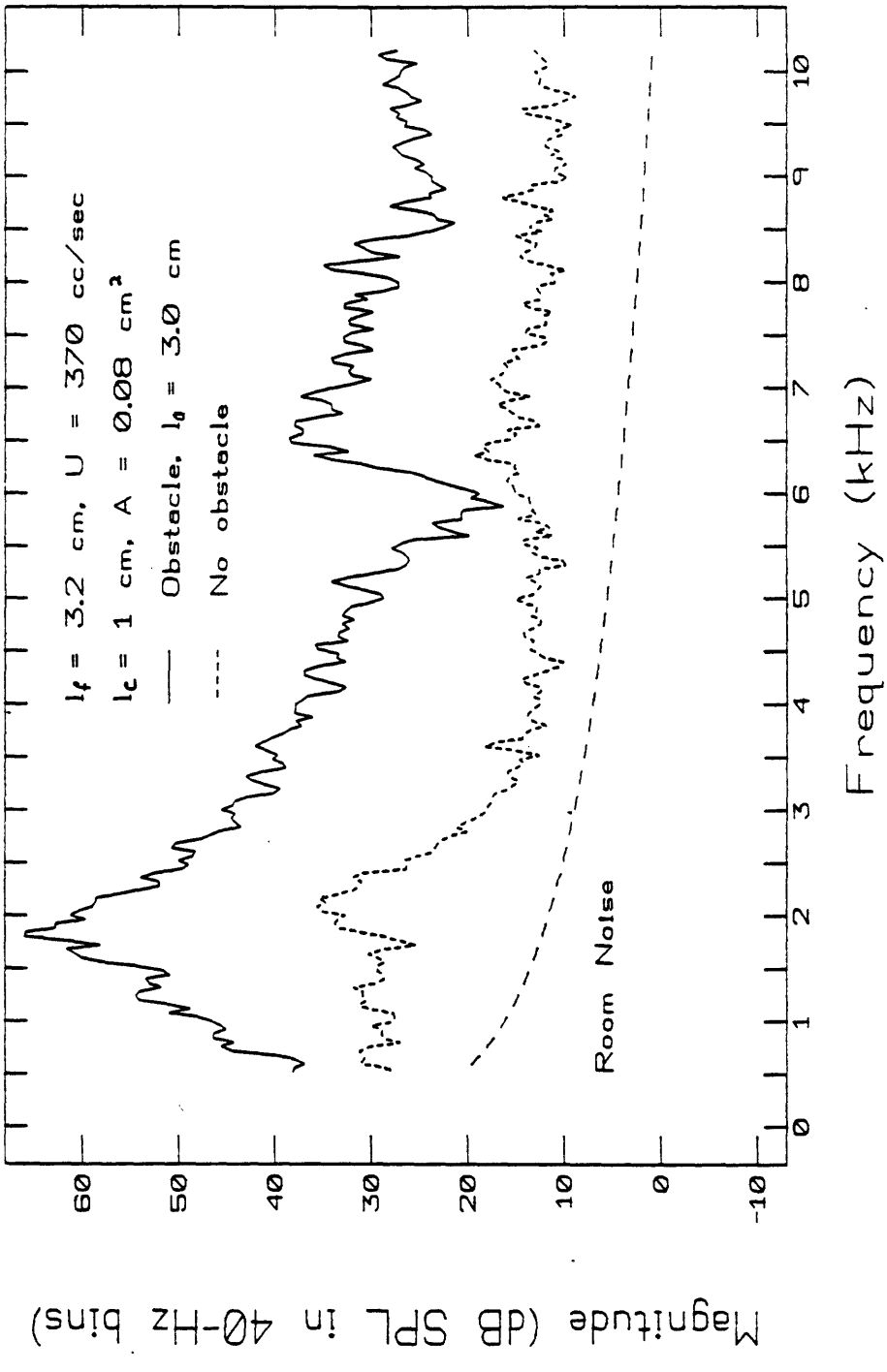


Figure 2.18: Spectra contrasting no-obstacle and non-whistling constriction-obstacle configurations, with all other parameters (U, l_f, A_c) the same.

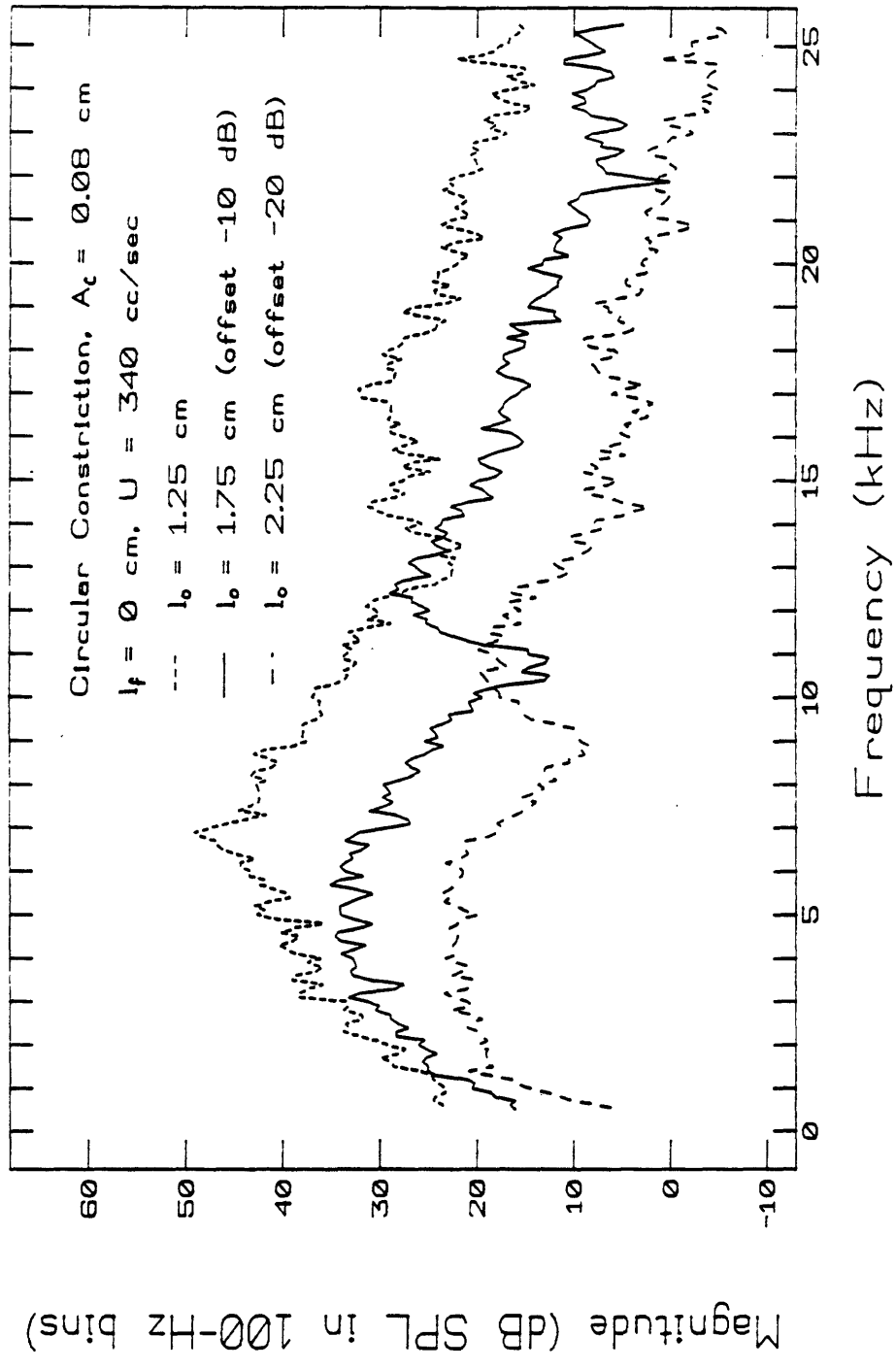


Figure 2.19: Spectra for non-whistling constriction-obstacle configurations, showing the effect of variations in l_0.

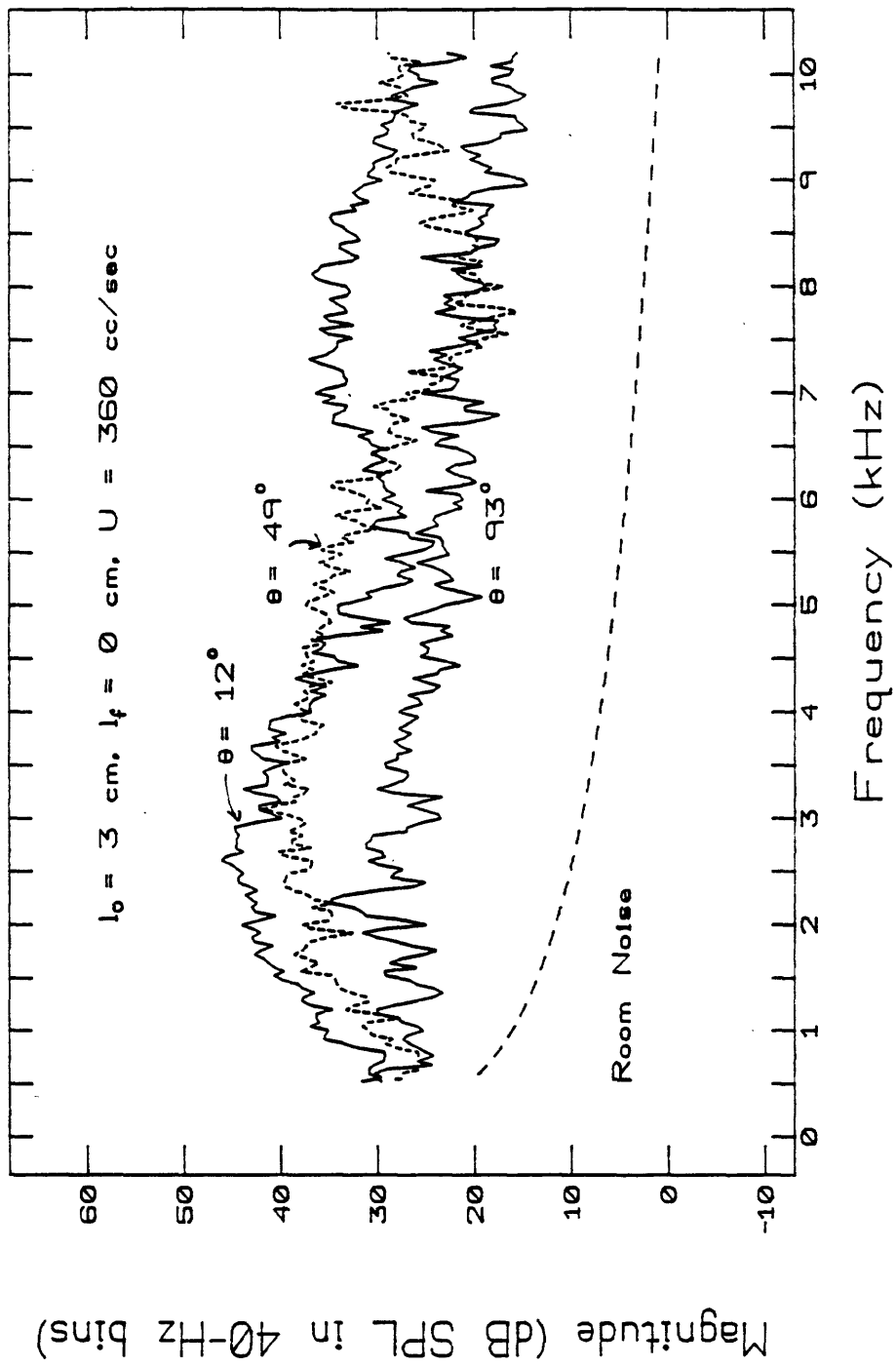


Figure 2.20: Spectra for non-whistling constriction-obstacle configurations, showing the effect of variations in θ.

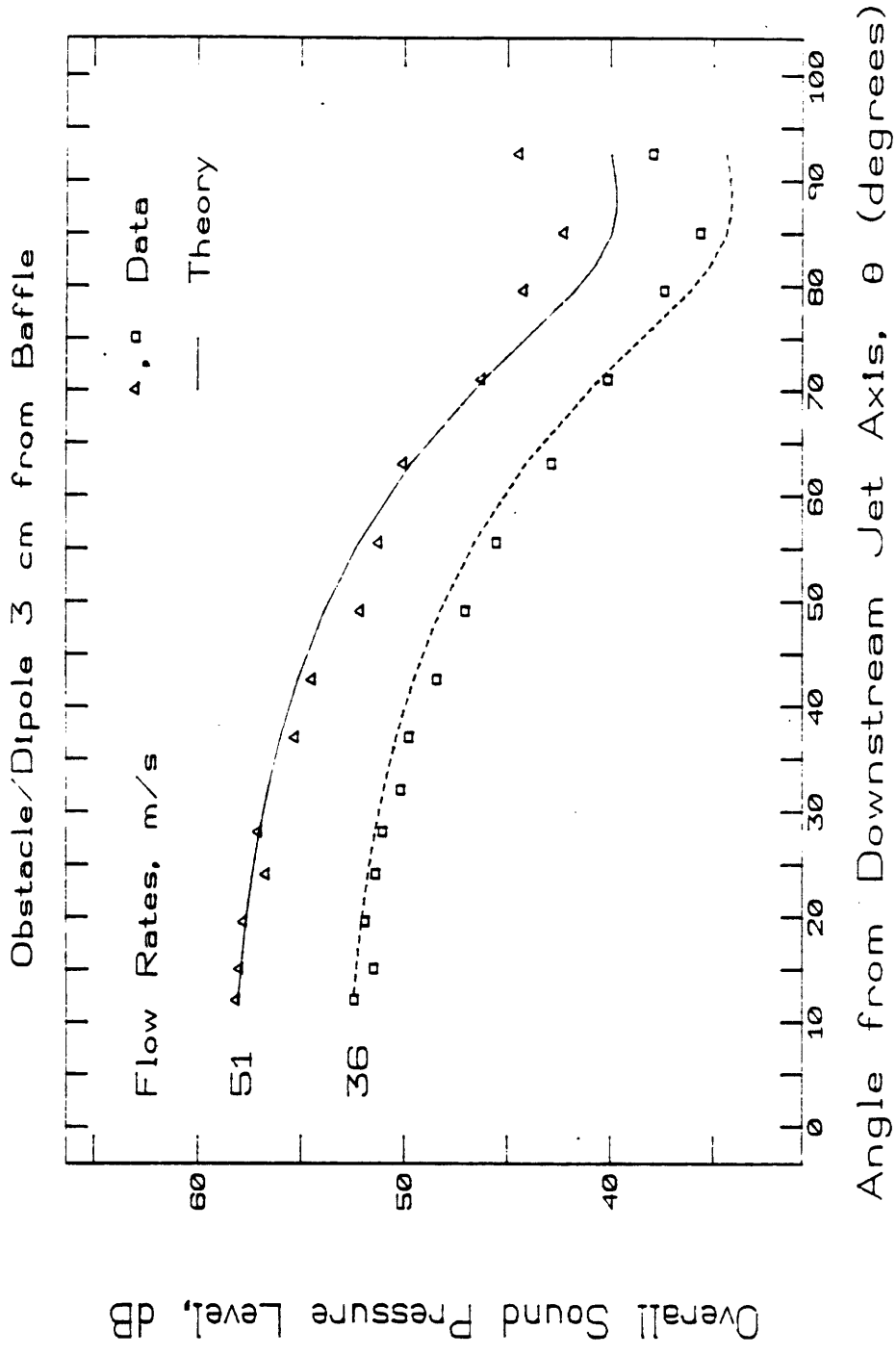


Figure 2.21: Complete directivity patterns for the constriction-obstacle configuration, $l_0 = 3$ cm. $l_f = 0$ cm. The microphone is 26 cm from the center of the obstacle. For both measured and theoretical curves, the sound pressure level is computed over the range 2000 to 3000 Hz. See text for further discussion of theoretical curves.

Chapter 3

The Idealized Obstacle and No-Obstacle Cases

3.1 Theoretical Predictions

In this chapter we attempt to develop theoretical models that account for some of the experimental data in Chapter 2. We restrict our attention to two basic configurations, referred to as the idealized obstacle and no-obstacle cases. We derive models for these cases in Section 3.1, and compare the predictions of the models to experimental data in Section 3.2. We first consider models of the sound-generating mechanism (the sources), and then develop models of the "tract" - the filter - with such sources in it. Such a separation carries an assumption of source-tract independence, which is assessed via the comparisons made in Section 3.2.

3.1.1 Source Model for Obstacle Case

We begin by considering the first idealized case, that of a jet striking a rigid obstacle inside a tube. The details of the system were discussed in Section 2.1, and are diagrammed in Figs. 2.1, 2.3 and 2.4. Throughout this chapter the word "tube" refers to the 17 cm length of plastic tubing, while "tract" refers to the entire assembly of tube, constriction, and obstacle. "Configuration" is used to refer to the set of dimensions describing a particular tract or tract model. Other terms denoting particular regions, such as "glottis", back and front cavities, and "mouth", are defined in Fig. 2.4, with liberal borrowing from vocal tract terminology. In this chapter, "the obstacle" refers to the 2 mm thick semicircle of aluminum, mounted with its plane surfaces perpendicular to the axis of the constriction and the jet, with the center of its edge intersecting the jet axis 3 cm downstream of the constriction. The constriction is a 0.32 cm diameter circular hole of length 1 cm, centered in an aluminum plug. The inlet to the hole is tapered.

Since the lengths of the tube and constriction are constant, the position of the constriction within the tube determines the lengths of the back and front cavities, those regions between glottis and constriction and constriction and mouth. Since the tube diameter and constriction-obstacle distance are constant (2.54 and 3 cm, respectively), different configurations can be identified by the length of the front cavity alone. In this chapter three values of l_f were used: 0, 3.2 and 12 cm, creating corresponding back cavity lengths of 16, 12.8 and 4 cm. The obstacle is therefore suspended in free space (via the rod threaded into its downstream face) for $l_f = 0$ cm; flush with the mouth of the tube (and therefore reducing the area of the mouth) for $l_f = 3.2$ cm; and 9 cm back from the mouth for $l_f = 12$ cm.

When $l_f = 0$ cm, the sound-producing region is not enclosed by the tube, and thus represents the closest possible approximation to a pure source. We therefore proceeded as diagrammed in Fig. 3.1: first, the data measured for the 0 cm configuration were used to derive a one-dimensional source; second, transmission-line theory was used to predict the output when the derived source was located inside the 3.2 and 12 cm configurations; and third, the predictions were compared to data measured for the actual 3.2 and 12 cm configurations.

We begin by considering the sound source. As shown in Chapter 2, the combination of the tapered inlet and the 3 cm constriction-obstacle distance ensure that the configurations will not whistle at any of the flowrates used. However, the presence of the obstacle does substantially increase the amplitude of the sound produced, heightens the peaks, and introduces sharp troughs in the spectrum. As a first approximation, then, we neglect sound produced by the jet, and consider only that produced at the obstacle. Based on the work of Curle (1955) and others, we assume that there are flow dipoles at the obstacle due to the net force exerted by the obstacle on the turbulent air. The orientation of the flow dipoles can be assumed to be longitudinal since the plane surface of the obstacle can only support forces in the longitudinal direction.

A flow dipole is a dipole source generated by fluid flow. A dipole source can be modeled as two out-of-phase simple sources a distance d apart, each generating a volume-velocity U. However, since both U and d are theoretical, unmeasurable entities, we lump them together and characterize the flow dipole by a strength $S = Ud$ that depends on flowrate. The combined parameter S can be determined indirectly, a necessary step towards translating the flow dipole model into a source that can serve as an excitation function for a transmission-line model. First, p_0, the far-field sound pressure generated when $l_f = 0$ cm, was measured. (A word on notation: p_x means the far-field sound pressure measured when $l_f = x$. $\hat{y}$ means the predicted rather than measured value of the variable y.) Although suspending the obstacle in free space did eliminate the sound-altering effects of a surrounding tube, the sound recorded was not only that emanating directly from the obstacle, since the constriction did reflect and scatter sound back towards the microphone. The baffle mounted flush with the mouth of the tube (as shown in Fig. 2.3) controlled the reflections so that the measured sound was that generated by the jet-plus-obstacle and its reflection in the baffle. If we represent the sound source by a single dipole, we have the situation shown in Fig. 3.2a. The quantities r, R, θ, ϕ have been defined in this somewhat inelegant manner in order to correspond to the way in which the measurements are made. At a distance r that is large compared to d but not compared to l_0, the predicted far-field sound pressure $\hat{p}_0$ is found by combining the far-field pressure expressions for a dipole and its image dipole:

$$\hat{p}_0(r,\theta,\omega) = \frac{\omega^2\rho}{4\pi c} S \left[\frac{\cos\theta}{r} e^{-jkr} - \frac{\cos\phi}{R} e^{-jkR}\right] \tag{3.1}$$

where

S = dipole strength
r = distance between dipole and observer (microphone)
θ = angle between downstream jet axis and r vector
R = distance between image dipole and observer (microphone)
ϕ = angle between downstream jet axis and R vector
l_0 = distance between dipole or image dipole and baffle
$\omega = 2\pi f$, f = frequency
ρ = density of air
c = speed of sound in air

Note that R and ϕ are functions of r, θ and l_0.

The terms within square brackets nearly cancel at frequencies where $R - r$ is equal to odd multiples of a half wavelength. The frequencies of the minima caused by such cancelation are very sensitive to changes in θ, especially for large θ, or l_0. The depths of the minima decrease

(i.e., the cancellation becomes less pronounced) as R increases relative to r. As pictured in Fig. 3.2b, a distribution of such dipole pairs along the y axis, modeling the effect of a finite-width obstacle, can be combined incoherently by adding together the $|\hat{p}_0|^2$ quantities generated for each source-image pair. The small frequency shifts in the zeros due to the variations in y for each pair have the effect of increasing the bandwidths of the minima, particularly at large values of θ. A graph of $\hat{p}_0/S$ versus ω for typical values of the parameters is shown in Fig. 3.3.

The effect of the baffle on the source, for a given microphone position, can be thought of as a filter, with a theoretical transfer function of $\hat{p}_0(r,\theta,\omega)/S(\omega)$. We can therefore derive an estimate of the source function, $\widehat{S}$, by inverse filtering the measured pressure $p_0(r,\theta,\omega)$:

$$\widehat{S}(\omega) = \frac{p_0(r,\theta,\omega)}{\hat{p}_0(r,\theta,\omega)/S(\omega)} \tag{3.2}$$

The theoretical transfer function in the denominator was derived by using y equal to the width of the obstacle, 2.54 cm, and three dipoles evenly spaced along the width of the obstacle, that is, $n_y = 3$.

We now wish to find an equivalent source p_s in terms of S. p_s will be used to model sound generation inside the tube only, where only plane waves propagate (for frequencies less than the first cutoff frequency, which as shown in the next section is 8000 Hz). If we make the assumption that all of the sound radiated by the flow dipoles excites longitudinal modes of the tube, we can approximate the spherical sources comprising the dipole by plane-wave sources of the same strength U. The distance d separating them becomes a transmission-line section of (small) length d, which is represented by an acoustic mass of value $\rho d/A$, where A is the cross-sectional area of the duct. We thus obtain the circuit shown in Fig. 3.4. It is then straightforward to derive the pressure drop across the circuit, which is then the strength of the equivalent pressure source:

$$\begin{aligned} \hat{p}_s &= j\left(\frac{\omega\rho}{A}\right)(\widehat{Ud}) \\ &= j\left(\frac{\omega\rho}{A}\right)(\widehat{S}(\omega)) \end{aligned} \tag{3.3}$$

We now have an expression for the sound source generated by the obstacle in a tube that depends only on the sound measured when the obstacle is in free space. In Sections 3.1.3 and 3.1.4 we will derive the rest of the transmission-line model in order to compute the transfer function U_{out}/p_s, where U_{out} is the volume velocity at the mouth of the tube. By combining this transfer function with Eqn. 3.3 we can test the validity of the various assumptions on which the model is based, including that of no source-tract interaction.

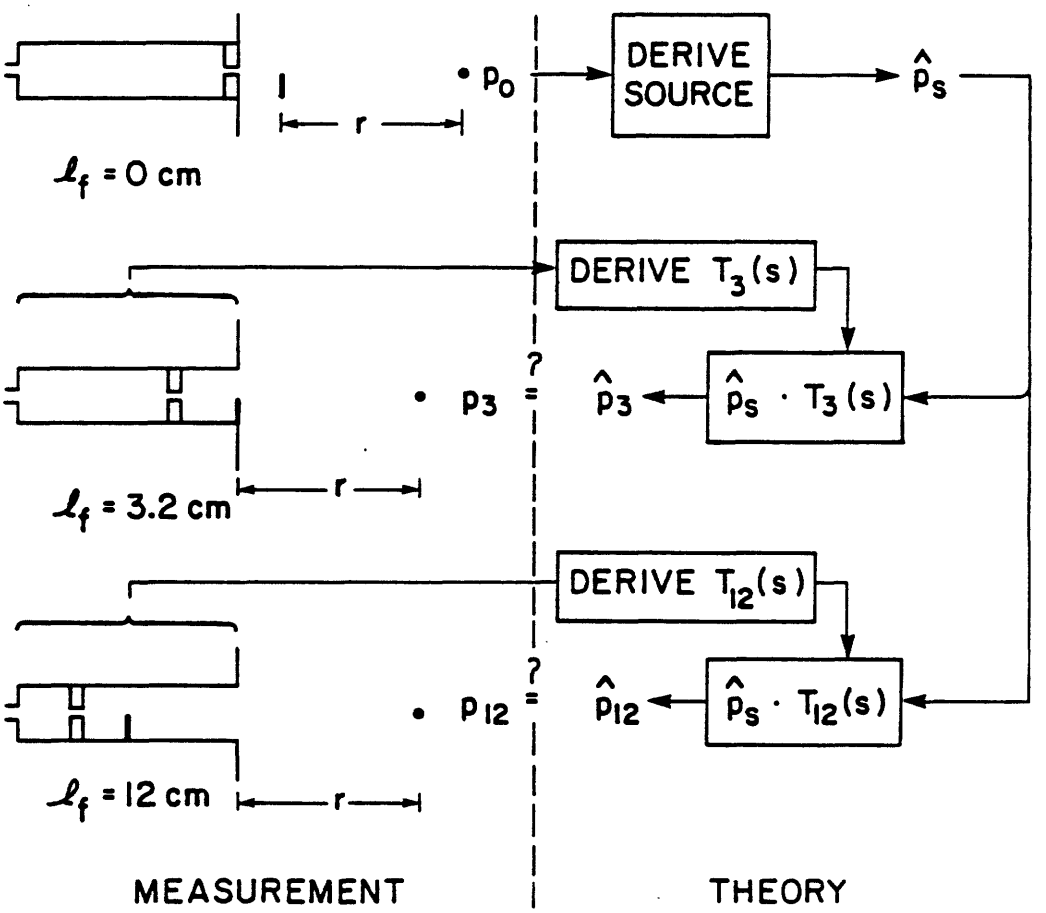


Figure 3.1: Diagram of the predictions and comparisons to be made for the obstacle case.

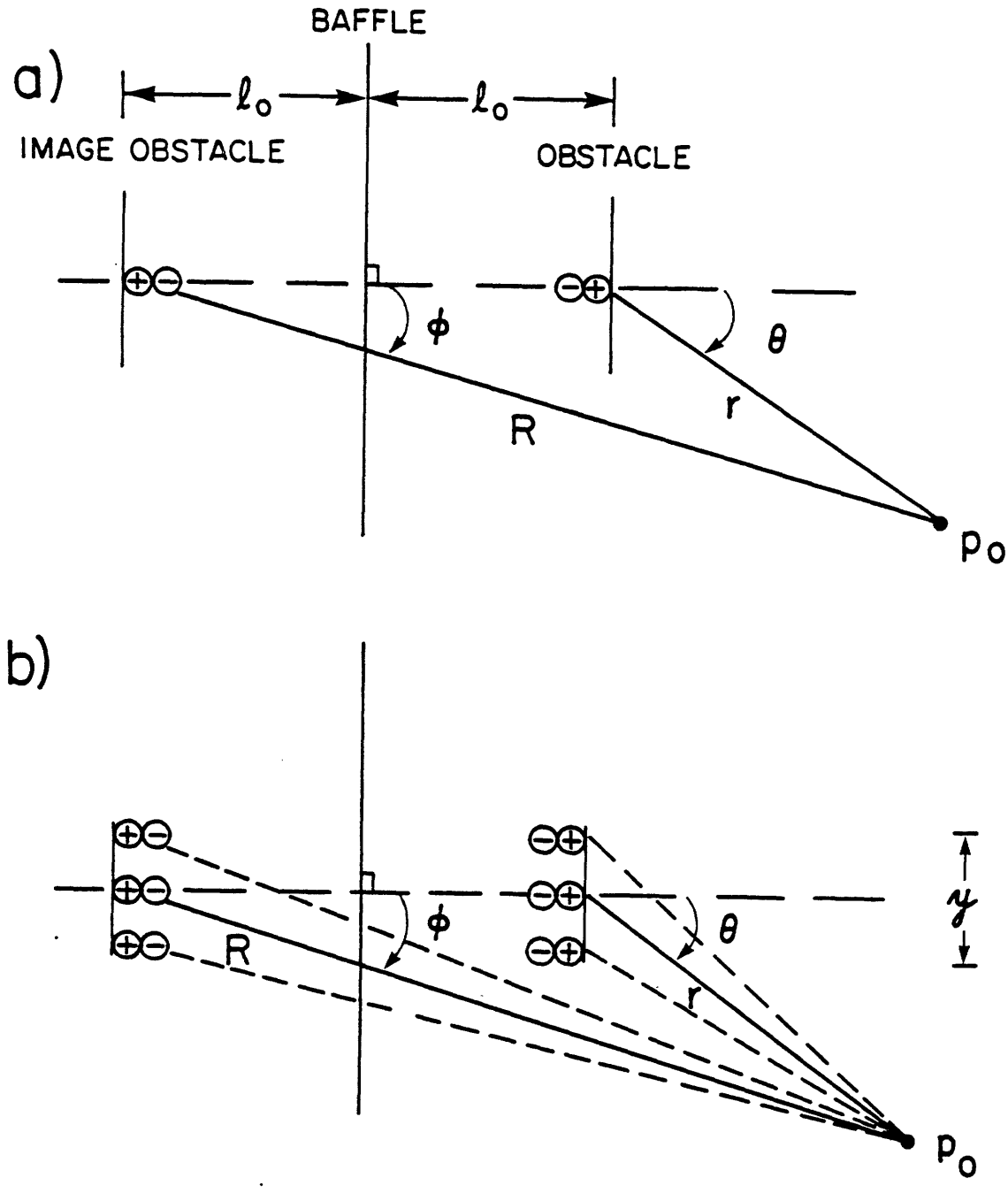


Figure 3.2: a) Schematic of dipole reflected in baffle, defining R, r, θ, ϕ, and l_o.
b) Schematic of distribution of dipoles and image dipoles, defining y and n_y.

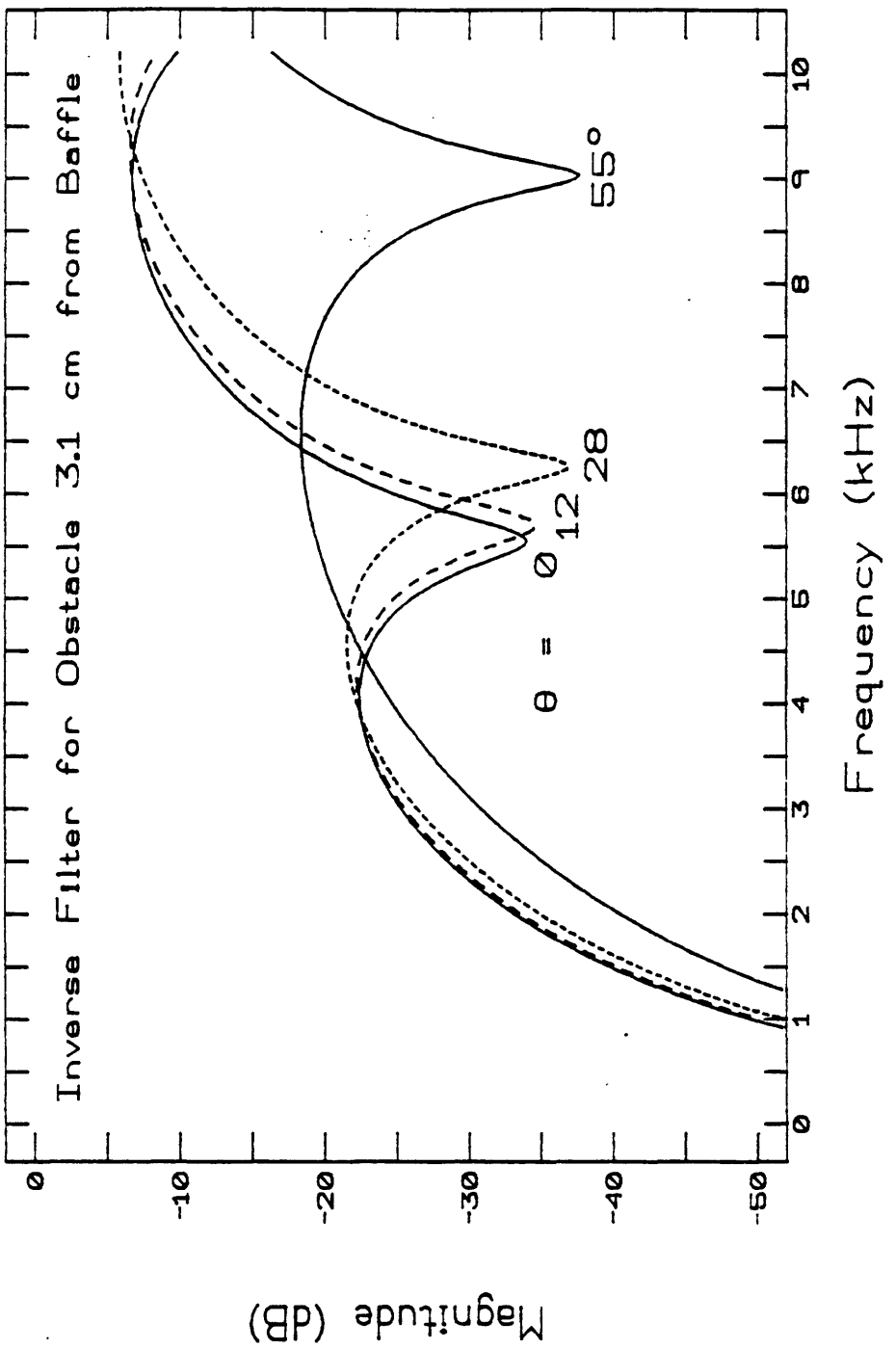


Figure 3.3: $\hat{p}_0/S$ vs. f: predicted far-field pressure due to three dipoles reflected in baffle, versus frequency, at three angles. Dipoles are evenly distributed over 2.5 cm length, all 3.1 cm from baffle, and are assumed to be of the same strength and coherent. S, the dipole strength, normalizes $\hat{p}_0$. $r = 26$ cm, $l_o = 3.1$ cm; θ is shown on the figure.

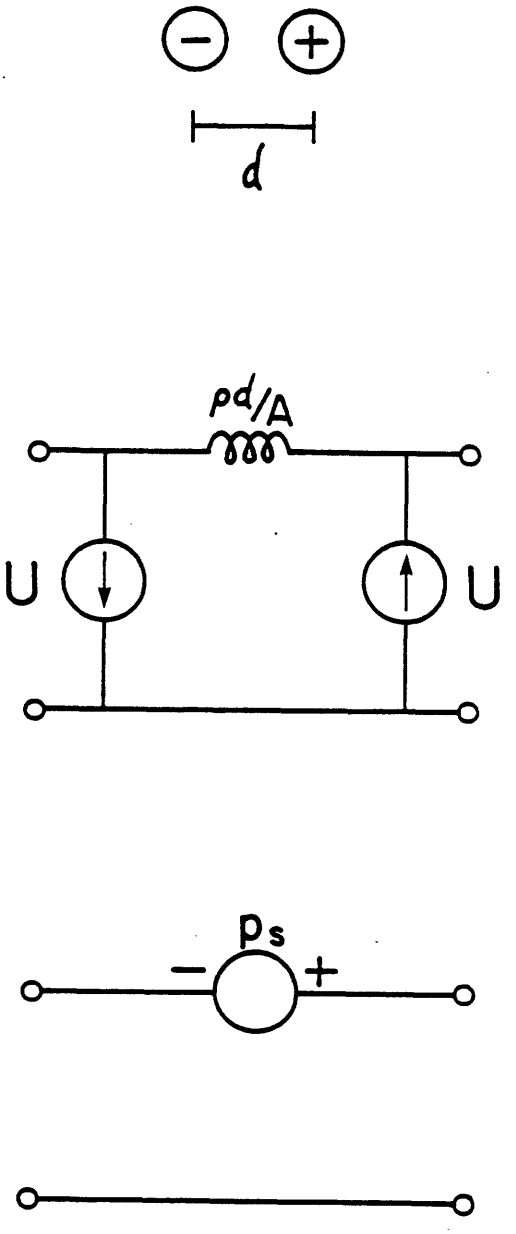


Figure 3.4: Circuits showing relation of dipole of strength Ud, in tube of area A, to pressure source.

3.1.2 Tract Models, General Method

The acoustic behavior of a duct enclosing sound sources can be modeled by distributed or lumped networks, which are composed of circuit elements representing the essential properties of the air and various sources of loss. The distributed network, otherwise known as a transmission line, is a one-dimensional model; its range of validity depends on the length of the individual segments in the transmission line, the frequency of the first cross-mode, and the approximation of the hyperbolic impedance elements (Flanagan, 1972). The lumped networks used here function as low-frequency approximations of the distributed network. They are valid for sounds whose half-wavelength is greater than the length of the tract, which corresponds to frequencies less than 1000 Hz.

In the analysis performed here, the first cross-mode generally is the most stringent limit for the validity of the distributed network model. For a rigid circular tube of radius a, higher modes correspond to solutions of the equation

$$\begin{aligned} & J'_m(\pi q) = 0 \\ \text{or} \quad & \pi q J_{m-1}(\pi q) = m J_m(\pi q) \end{aligned}$$

where J_i denotes the Bessel function of order i. The solutions $q = \alpha_{mn}$ are related to the cutoff frequencies by

$$f_{mn} = \frac{\alpha_{mn}\, c}{2a}$$

The first propagating cross mode occurs for $\alpha_{01} = .5861$; for the tube radius of 1.27 cm used in the experiments, the predicted frequency of this mode is 7950 Hz (Morse and Ingard, 1968).

Whether a lumped or distributed network is being considered, the location of the excitation relative to the output determines the form of the transfer function. If the input is at one end of the circuit and the output at the other, and there are no side branches, the transfer function will be all-pole: as shown in Appendix B, the numerator will be a constant and the denominator will be a function of s, as here:

$$\frac{U_{out}}{U_{in}} = \frac{K}{s^m + b_{m-1}s^{m-1} + \cdots + b_0}$$

A similar result obtains for a pressure source located at one end of the circuit. When s takes on a value equal to a root of the denominator (a pole), the function as a whole goes to infinity. If the source is located elsewhere in the circuit, the transfer function between it and the output takes the form

$$\frac{U_{out}}{U_{source}} = K\, \frac{s^n + a_{n-1}s^{n-1} + \cdots + a_0}{s^m + b_{m-1}s^{m-1} + \cdots + b_0}$$

Since the numerator is now a function of s also, its roots make the function go to zero, and are the zeros of the transfer function. The poles are the natural frequencies of the entire circuit, and are excited by a source positioned anywhere in the circuit. The zeros change as the source location and type change: for a series pressure source, they are the natural frequencies of the circuit behind the source with the source open-circuited. The behavior near zero frequency, where the transfer function equals Ka_0/b_0, and at high frequency, where it is Ks^{n-m}, is affected by the source location also.

There are many ways of deriving the transfer function for a given circuit. Aside from the well-documented, formal methods of solving sets of simultaneous equations for the desired

function, it may be easier in some cases to solve separately for the numerator and denominator. The roots of the denominator are the natural frequencies of the entire network with internal sources set to zero. Any transfer function that is the ratio of an output variable (current or voltage) to an input variable will have the same natural frequencies, so the most convenient set of variables can be chosen. For the numerator, we consider the original circuit with source intact. Suppose that we want the transfer function of the output current over the source voltage, as shown in Fig. 3.5. The zeros occur when the part of the circuit on the side of the source away from the output (labeled "Back Network" in the figure) acts to reduce the output current to zero. For a series voltage source (part a of the figure), this occurs when the impedance looking back from the source is infinite, since all of the source voltage is dropped across Z_{back} without generating any current. (For a parallel current source, as in Fig. 3.5b, the impedance looking back must be zero for analogous reasons.) Setting $Z_{back} = \infty$ is the same thing as finding the poles of the Z_{back} function. Therefore, the numerator of the desired transfer function is the denominator of the impedance looking back at the location of a series voltage source (or the numerator of the impedance, for a parallel current source).

Since $s = \sigma + j\omega$.is complex, the poles and zeros are in general complex. Throughout this chapter, the coordinates of singularities are referred to by the more descriptive parameters of resonance frequency f and bandwidth Δf, where $f = \omega/2\pi$ and $\Delta f = \sigma/\pi$. (If a singularity is pure real, no bandwidth is given, and $f = \sigma/2\pi$.) As shown in Fig. 3.6 for the underdamped case, these parameters define not only the complex s coordinates, but also the center frequency and width between 3 dB-down points of the peak (or trough) in the transfer function $T(j\omega)$. As singularities move away from the $j\omega$ axis, as would happen in, for example, a series R-L-C circuit if R steadily increased, this terminology becomes somewhat misleading. Bandwidth is no longer a visually accurate concept when a resonance has no discernible peak. However, due to the usefulness of the concept in the lightly damped cases and the awkwardness that would be created by constantly changing notation, bandwidth and frequency, as defined above in terms of σ and ω, will be used throughout to specify singularities.

A section of duct of essentially the same cross-sectional area throughout can be modeled by a segment of transmission line with the lumped elements shown in Fig. 3.7. (The model is invalid for half-wavelengths shorter than the length of the section, and may be restricted further, as discussed earlier.) The element R represents the loss due to viscosity at the walls of a section of unit length, G the loss due to heat conductance, while L represents the mass of air in a unit-length section and C its compliance (see Flanagan, 1972). The section can be represented by the T-network in c) of the same figure. The argument of each of the hyperbolic impedances depends on $\gamma = \sqrt{zy} = \alpha + j\beta$, where $z = (R + j\omega L)$ and $y = (G + j\omega C)$. For small losses, $\beta \simeq \omega/c$ and

$$\alpha \simeq \frac{R}{2}\sqrt{\frac{L}{C}} + \frac{G}{2}\sqrt{\frac{C}{L}} \tag{3.4}$$

where

$$\begin{aligned} L &= \frac{\rho}{A} = \text{per-unit-length inductance} \\ C &= \frac{A}{\rho c^2} = \text{per-unit-length compliance} \\ R &= \frac{S}{A^2}\sqrt{\frac{\omega\rho\mu}{2}} = \text{per-unit-length resistance} \end{aligned} \tag{3.5}$$

$G = S\frac{\eta - 1}{\rho c^2}\sqrt{\frac{\lambda\omega}{2c_p\rho}}$ = per-unit-length conductance

A = tube area, cm^2

S = tube circumference, cm

ρ = air density = 1.18×10^{-3} gm/cm^3

c = sound velocity = 34480 cm/s

μ = viscosity coefficient = 1.86×10^{-4} dyne-s/cm^2

λ = coefficient of heat conduction = 5.5×10^{-5} cal/cm-s-deg C

η = adiabatic constant of air = 1.4

c_p = specific heat of air at constant pressure = .24 cal/gm-deg C

If there is no loss, $R = G = 0$, $\alpha = 0$, and the hyperbolic impedances reduce to the corresponding circular functions.

There are two other major impedances not included in the elements described above. The first one, the radiation load of a piston in an infinite baffle, is modeled by the circuit shown in Fig. 3.8, and is valid over the entire range of frequencies that the transmission line is. The impedance is:

$$Z_R = \frac{s^2(M_A R_1 R_2 C_A) + s(M_A R_1 + M_A R_2)}{s^2 M_A R_1 C_A + s(M_A + R_1 R_2 C_A) + R_1 + R_2} \tag{3.6}$$

where

$$\begin{aligned} M_A &= \frac{8}{3}\frac{\rho}{\pi^2 a} \\ R_1 &= 0.1404\frac{\rho c}{a^2} \\ R_2 &= \frac{\rho c}{\pi a^2} \\ C_A &= 5.94\frac{a^3}{\rho c^2} \\ a &= \text{radius of the mouth opening} \end{aligned}$$

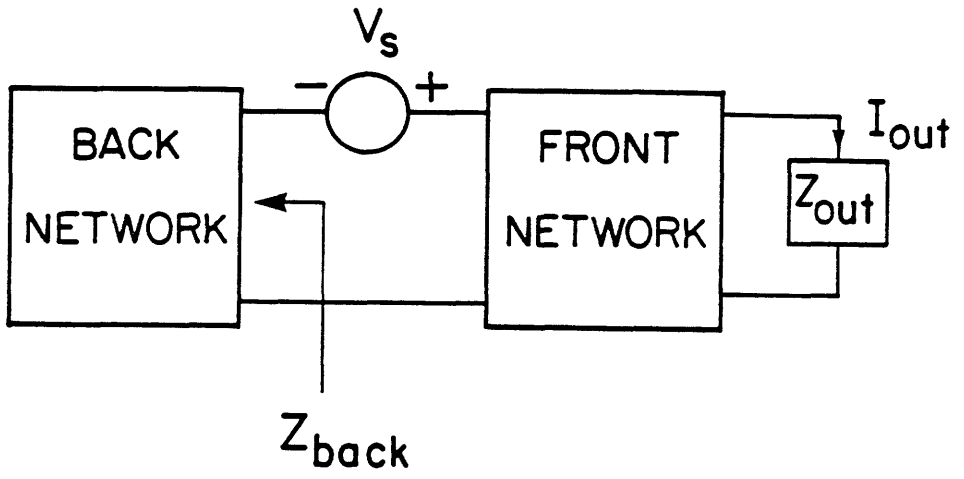
At low frequencies ($ka < 0.5$) the model reduces to a resistance ($R_1 + R_2$) and inductance in parallel: at high frequencies ($ka > 5$) it reduces to a resistance (R_2) with the value of the characteristic impedance of air divided by πa^2, the area of the mouth opening (Beranek, 1954).

Not included in the description of the transmission line elements is the flow resistance of a constriction, which was discussed in Chapter 2. Since in the present experiment the AC fluctuations in pressure and velocity due to a sound wave are small compared to the static pressure drop across and volume velocity through the constriction, the incremental resistance R_i is the loss experienced by the acoustic disturbance (superimposed on the DC flow). The value of R_i at a given flowrate U is found from

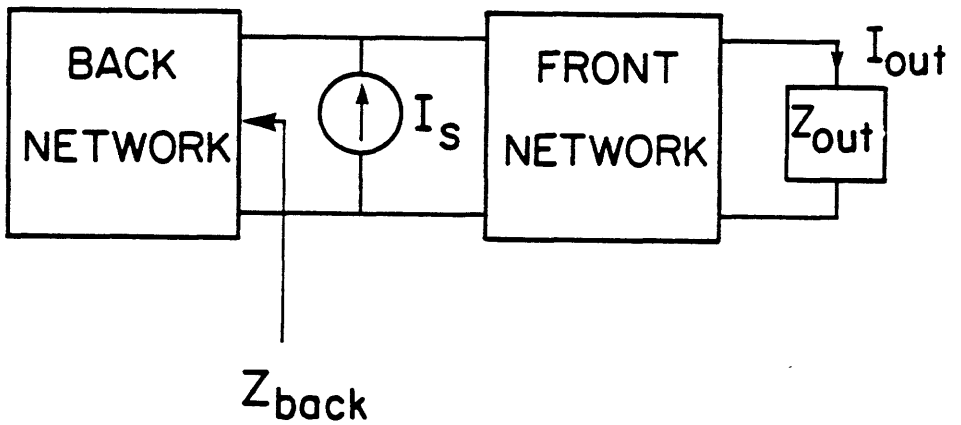
$$R_i = K_2\frac{\rho U}{A_c^2} + K_3 \tag{3.7}$$

using the values of $K_2 = 1.0$ and $K_3 = 8.7$ found in Chapter 2 for the constriction of area $A_c = .079$ cm^2.

Three models of increasing realism and complexity were analyzed. They are shown in Fig. 3.9. Although only the results of Model III are used for the comparisons made in Section 3.2, the simpler models increase our understanding of the effects of the different components of the configurations. For Model I, we assume that the constriction impedance is essentially infinite, leading to a front-cavity-only configuration. In Model II, the constriction is modeled with a finite impedance and a reflectionless termination, and the front cavity is the same as in Model I. In Model III, the back cavity and "glottis" are added to Model II. For each model we derive the low-frequency behavior, using lumped element analysis, and then we extend the frequency range by using distributed transmission line models. Since analysis of transmission-line networks becomes computationally more difficult as the number of sections increases, some of the distributed models were analyzed with the help of a program (TBFDA) that generates the transfer function for a given area function by solving the plane-wave propagation equations iteratively (Henke, 1966; Hosein, 1983). Since TBFDA models the losses in a way more appropriate for a human vocal tract than for plastic tubing, only pole and zero frequencies generated by the program were used, and more accurate bandwidths were calculated separately by considering the effects of heat conduction, viscosity, radiation, and flow resistance. Further details of the calculations will be described in Section 3.1.4.



$$\frac{I_{out}}{V_s} = 0 \quad \text{WHEN} \quad Z_{back} = \infty$$



$$\frac{I_{out}}{I_s} = 0 \quad \text{WHEN} \quad Z_{back} = 0$$

Figure 3.5: Diagram showing relation of network structure to transfer function zeros.

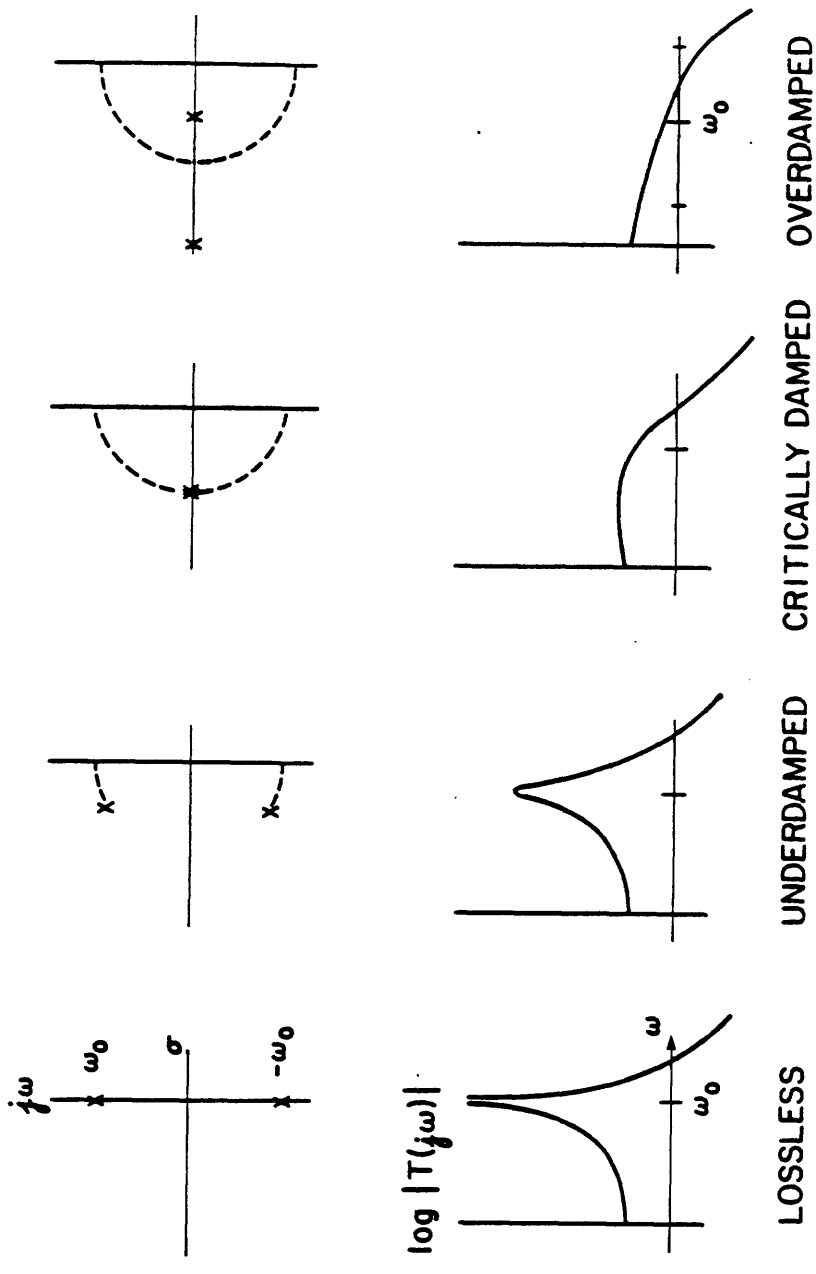
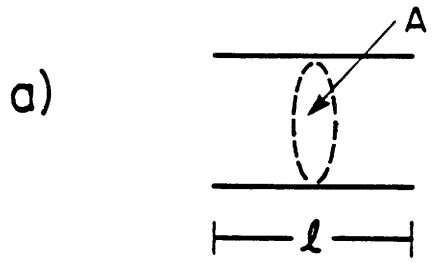
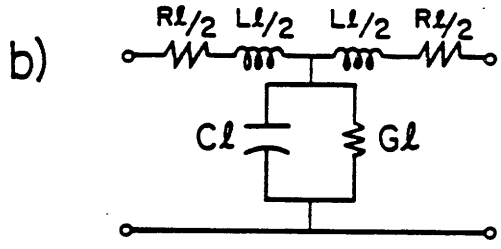
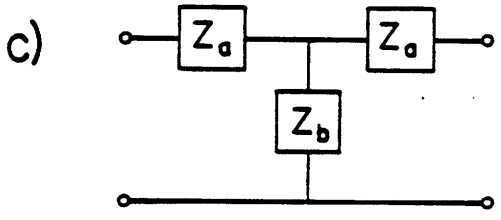


Figure 3.6: Sequence of pole-zero diagrams and plots of transfer function magnitude (log scale) vs. log ω showing the effect of increasing damping.







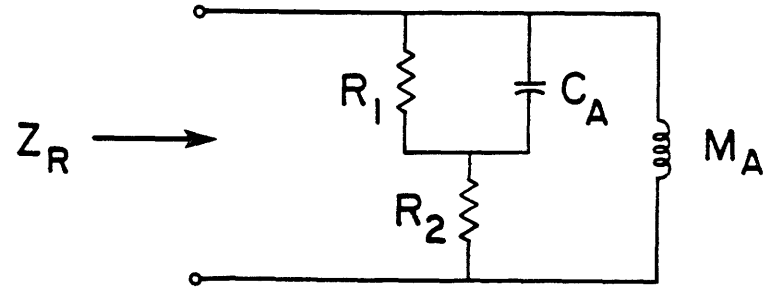
$$Z_a = Z_o \tanh \frac{\gamma \ell}{2}$$

$$Z_b = Z_o \operatorname{csch} \gamma \ell$$

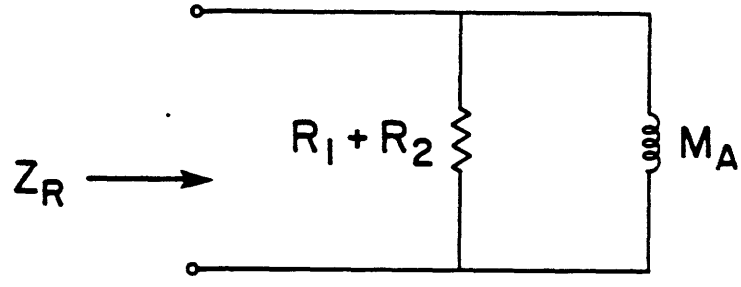
$$Z_o = \frac{\rho c}{A} \qquad \gamma = \sqrt{zy} = \sqrt{(R + j\omega L)(G + j\omega C)}$$

Figure 3.7: a) Section of lossy cylindrical duct.
b) One-dimensional lumped-element equivalent of duct section.
c) Equivalent T-network for a length l of a uniform transmission line.

MODEL OF RADIATION IMPEDANCE



LOW-FREQUENCY APPROXIMATION



HIGH-FREQUENCY APPROXIMATION

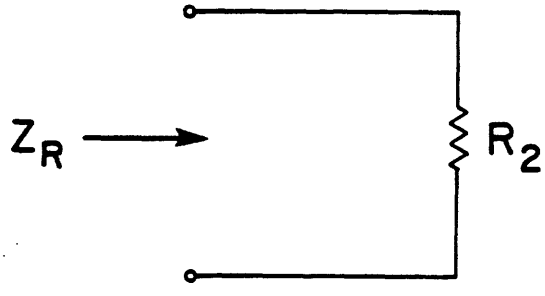


Figure 3.8: Lumped-circuit representation of radiation impedance, and the low- and high-frequency approximations.

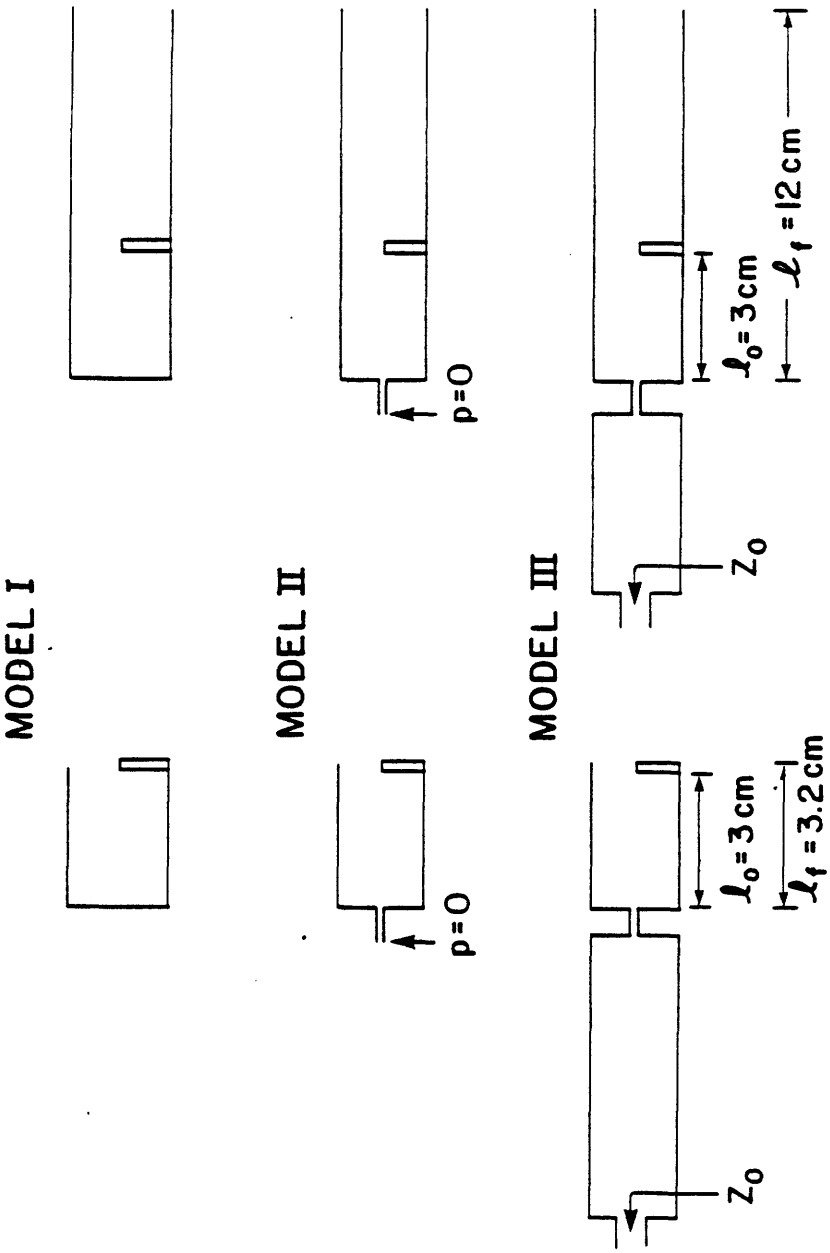


Figure 3.9: Diagrams of Models I, II and III, for front-cavity lengths 3.2 and 12 cm.

3.1.3 Low-Frequency Models

Low-Frequency Tract Model I

Model I is derived by approximating the constriction by an infinite impedance, leaving only the front cavity to be modeled. We begin with a distributed model of the front cavity and reduce it in a series of approximations to the low frequency model. First, only two T-sections are needed, one on either side of the pressure source, since at low frequencies the wavelengths are much longer than the entire front cavity. Near zero frequency, the associated viscosity and heat conduction losses, both proportional to $\sqrt{f}$, approach zero. Likewise, the impedances of the inductances representing the mass of air in the front cavity are proportional to f and thus approach zero. Therefore, the two T-sections reduce at low frequencies to one element each: the compliance of the air in that section of the tube. Only one of these appears in the final circuit, since at low frequencies the radiation impedance, dominated by the inductance, short circuits both the radiation resistance and the compliance of the part of the front cavity forward of the pressure source. We are thus left with the pressure source, the short circuit at the output, and C_f, the compliance of the cavity between the rigid wall of the constriction and the pressure source, as shown in Fig. 3.10a. The transfer function

$$\frac{U_{out}}{p_s} = sC_f$$

has a single zero at 0 Hz, as shown in Fig. 3.10b and c. (The absolute magnitude, using the actual value of C_f, is shown.)

Low Frequency Tract Model II

When the constriction is modeled by a finite impedance terminated on the upstream end by a short circuit, the circuit shown in Fig. 3.11 results. The same low-frequency approximations made for Model I, with the addition of the constriction impedance, reduce the distributed network for this section of the tract to a series inductance-resistance combination in parallel with C_f. R_c models the flow resistance, and L_c models the mass of air in the constriction (since the area of the constriction is relatively small, C_c, the compliance of the constriction, is negligible). Without loss ($R_c = 0$), the transfer function is

$$\frac{U_{out}}{p_s} = C_f \frac{s^2 + \frac{1}{L_c C_f}}{s}$$

which has a pole at 0 Hz, and zeros, corresponding to the Helmholtz resonance of the constriction and the front-cavity compliance, at $\pm(j/2\pi)\sqrt{1/L_c C_f}$. With loss, the transfer function becomes

$$\frac{U_{out}}{p_s} = C_f \frac{s^2 + s\frac{R_c}{L_c} + \frac{1}{L_c C_f}}{s + R_c/L_c}$$

with the singularities:

$$\begin{aligned} \text{pole: } s &= -\frac{R_c}{L_c} \\ \text{zeros: } s &= -\frac{R_c}{2L_c} \pm \frac{1}{2L_c}\sqrt{\frac{R_c}{L_c}^2 - \frac{4}{L_c C_f}} \end{aligned} \tag{3.8}$$

At U = 250 cc/sec, R_c = 55.97 g/sec-cm^4, and for a constriction-obstacle distance of 3 cm, the pole occurs at 610 Hz: the zeros have a frequency of 250 Hz and a "bandwidth" of 610 Hz. At zero frequency, the transfer function equals $1/R_c$. As the frequency increases, it tends towards sC_f, reflecting the fact that there is one more zero than pole in both Models I and II. The pole-zero diagrams are shown in Fig. 3.11b. The transfer function of the case with loss is shown as a solid line in Fig. 3.11c; on the same graph the transfer function for Model I is shown dotted. Note that above about 1000 Hz, where the two curves are within 1 dB of each other, the details of the low-frequency behavior no longer matter.

Low Frequency Tract Model III

At low frequencies the circuit model shown in Fig. 3.12 applies, which adds another resistor and capacitor to model the "glottis" and back cavity. Without loss ($R_c = 0$, $R_g = \infty$), the transfer function is

$$\frac{U_{out}}{p_s} = C_f \frac{s^3 + s\left(\frac{C_b + C_f}{C_f C_b L_c}\right)}{s^2 + \left(\frac{1}{C_b L_c}\right)}$$

which has a zero at 0 Hz, a pole pair due to the Helmholtz resonance of the back cavity and constriction, and a zero pair due to the combination of all three elements. With loss, the transfer function becomes:

$$\frac{U_{out}}{p_s} = C_f \frac{s^3 + s^2\frac{R_c}{L_c} + s\left(\frac{C_b R_g + C_f(R_c + R_g)}{C_f C_b L_c R_g}\right) + \frac{1}{C_f C_b L_c R_g}}{s^2 + s\left(\frac{L_c + C_b R_c R_g}{C_b L_c R_g}\right) + \left(\frac{R_c + R_g}{C_b L_c R_g}\right)}$$

where

$$\begin{aligned}
C_b &= \frac{V_b}{\rho c^2} = \text{ back cavity compliance} \\
R_g &= \frac{\rho c}{A_g} = \text{ "glottal" resistance} \\
C_f &= \frac{V_f}{\rho c^2} = \text{ front cavity compliance} \\
R_c &= K_2\frac{\rho U}{A_c^2} + K_3 = \text{ constriction resistance} \\
L_c &= \frac{\rho l_c}{A_c} = \text{constriction inductance}
\end{aligned}$$

At zero frequency, the transfer function equals $1/(R_c + R_g)$. As the frequency increases, it tends towards sC_f; again there is one more zero than pole, resulting in the same asymptotic behavior.

When the front cavity is 3.2 cm long, R_c = 12.7 g/cm^4-sec, U = 250 cc/sec, and R_g = 41.9 g/cm^4-sec, frequencies and bandwidths of the poles and zeros are as follows:

$$\begin{aligned}
\text{real poles: } f &= 150, 530 \text{ Hz} \\
\text{real zero: } f &= 50 \text{ Hz} \\
\text{zero pair: } f &= 320 \text{ Hz} \\
\Delta f &= 620 \text{ Hz}
\end{aligned}$$

When the front cavity is 12 cm long, for the same U, R_c and R_g, the poles and zeros are:

$$\begin{aligned} \text{pole pair: } f &= 280 \text{ Hz} \\ \Delta f &= 820 \text{ Hz} \\ \text{real zero: } f &= 100 \text{ Hz} \\ \text{zero pair: } f &= 720 \text{ Hz} \\ \Delta f &= 450 \text{ Hz} \end{aligned}$$

These are shown in a pole-zero plot in Fig. 3.12b. The transfer function is shown with sC_f dotted as before in Fig. 3.12c. With this model, the difference in the front and therefore back cavity lengths affects the frequencies of the poles and zeros. When $l_f = 12$ cm, C_f and C_b are about equal; when $L_f = 3$ cm, C_b is roughly four times bigger than C_f. The larger C_b is, the smaller its impedance, and the more current it draws relative to R_g. Since loss in a resistance is proportional to the square of the current through it, this accounts for the smaller amount of damping of the real zero when the back cavity is larger ($l_f = 3$ cm). The Helmholtz resonance between the constriction and the back cavity increases in frequency when the back cavity is shortened; simultaneously the effective damping decreases, so that the poles are underdamped and therefore complex when $l_f = 12$ cm, and are overdamped for $l_f = 3$ cm.

Summing up the three models, we find that it is not too inaccurate to think about the low-frequency behavior in terms of Model I, especially above about 500 Hz. We will, however, use the more realistic Low-frequency Model III results when we assemble the final transfer functions. We turn now to the higher-frequency behavior, using distributed circuits for the same three models.

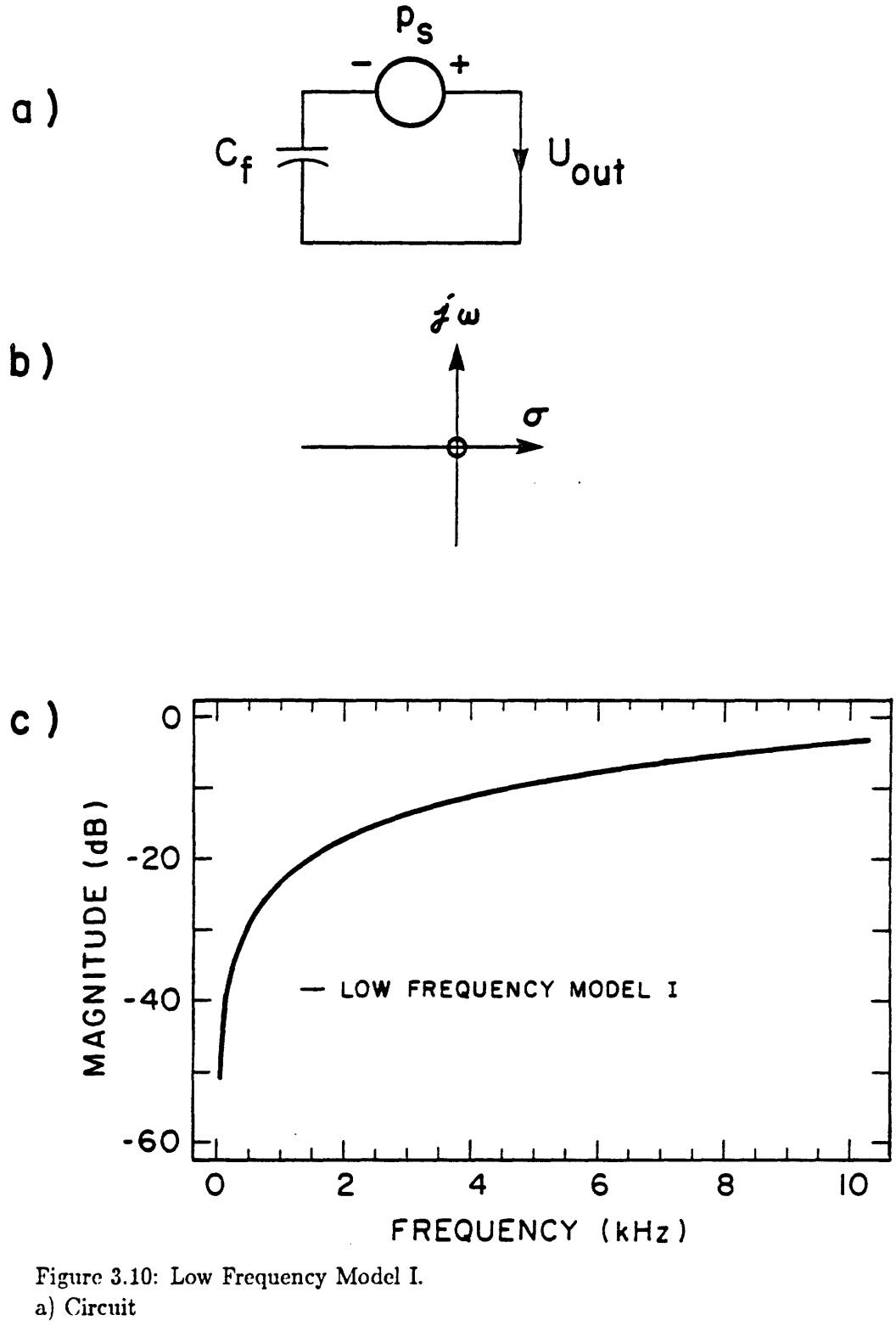


Figure 3.10: Low Frequency Model I.
a) Circuit
b) Pole-zero plot
c) Transfer function magnitude.

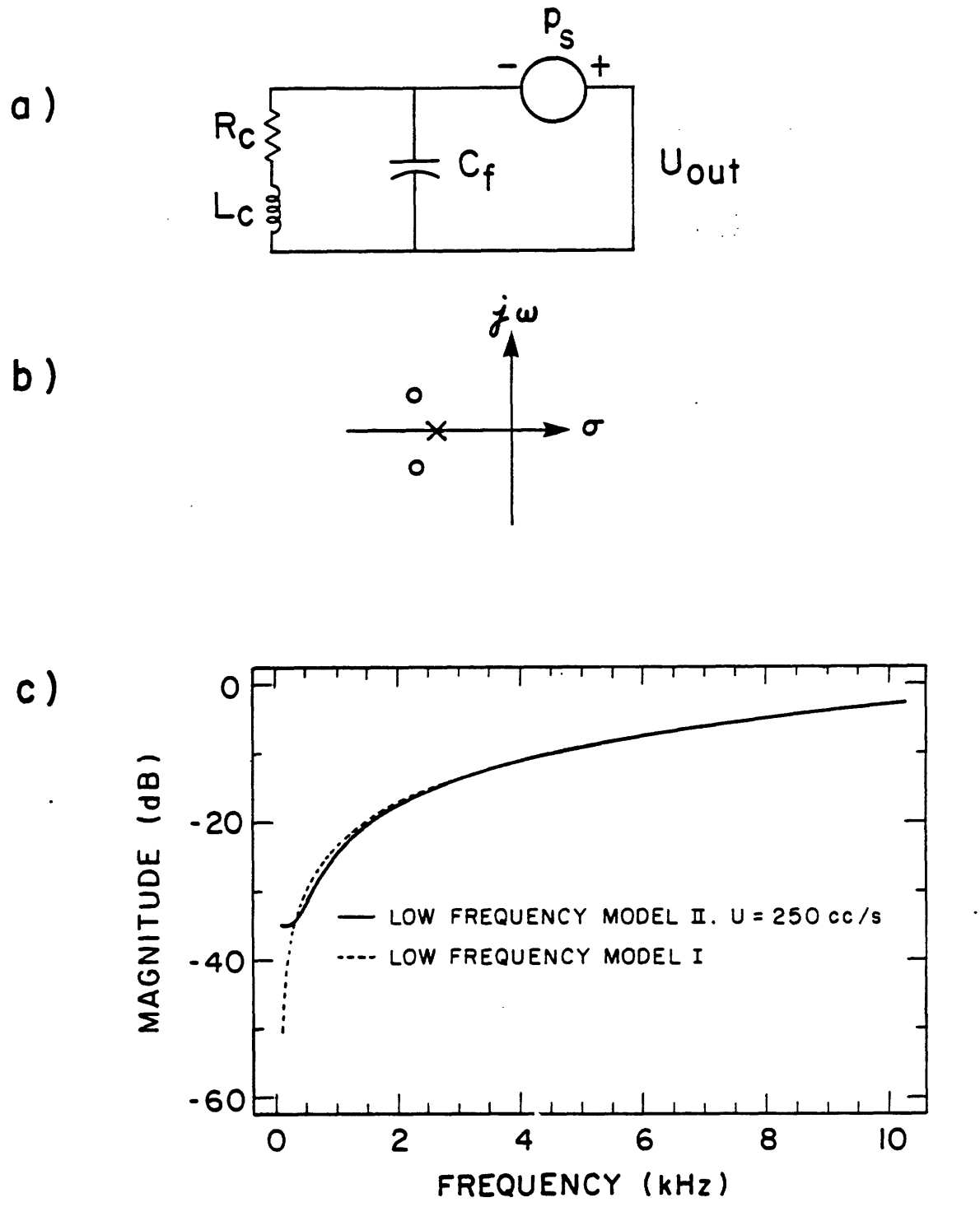


Figure 3.11: Low Frequency Model II.
a) Circuit
b) Pole-zero plot
c) Transfer function magnitude.

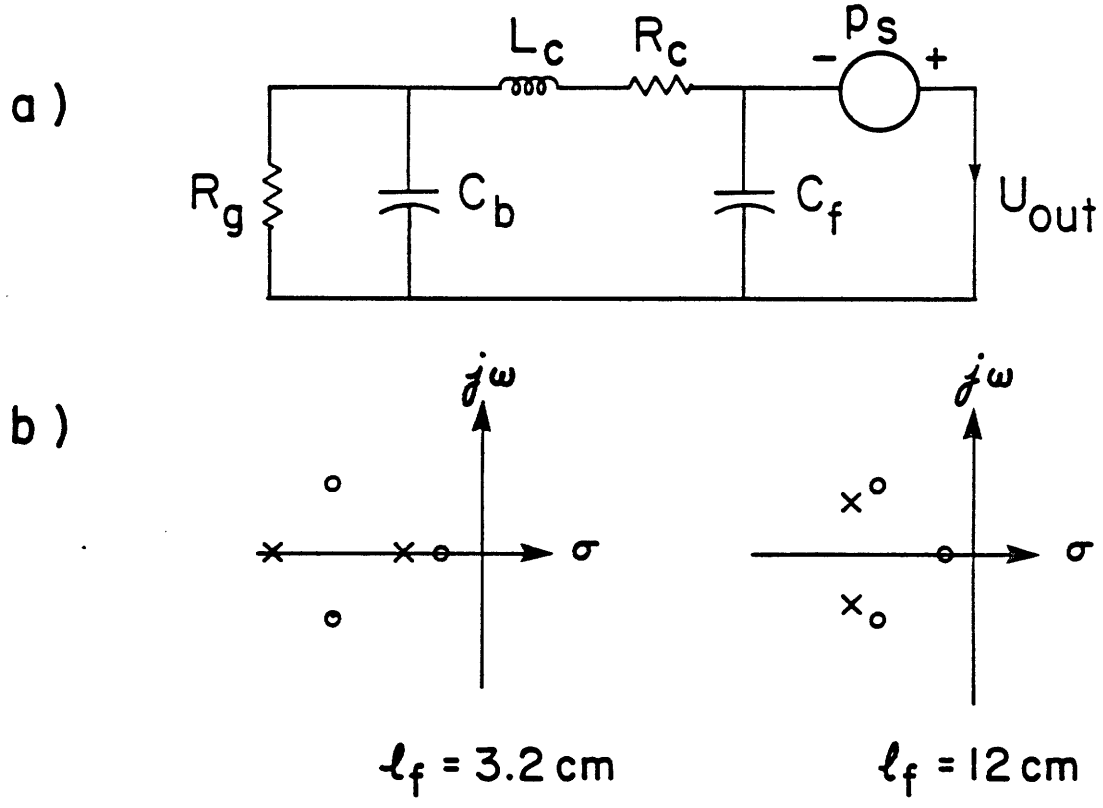


Figure 3.12: Low Frequency Model III.
a) Circuit
b) Pole-zero plot
c) Transfer function magnitude.

3.1.4 Higher Frequency Models

Model I.

We begin by considering the transmission-line models shown in Fig. 3.13, which are based on the same Model I approximation made in Section 3.1.3, that the constriction impedance is infinite. When we no longer make the low-frequency approximation that Z_R is short-circuiting the front cavity section downstream of the pressure source, we obtain two different models, one for each front cavity length. However, both transfer functions are similar in form. Since the distance from the back wall to the pressure source is the same in both cases, the numerators are identical. The denominators have the same form but depend on different values of l_f. The most efficient way to find the transfer functions is therefore to solve for the poles and zeros separately.

Let us first ignore the radiation impedance, that is, assume that $Z_R = 0$. We can solve for the poles by finding a transfer function of any two variables at opposite ends of the circuit, since all of these combinations will have the same poles. Following Flanagan's method (1972), we choose the ratio of volume velocities at the mouth (U_m) and glottis (U_g), which gives us

$$\begin{aligned} \frac{U_m}{U_g} &= \frac{z_b}{z_b + z_a} \\ &= \frac{1}{\cosh \gamma_f l_f} \end{aligned}$$

where

$$\begin{aligned} &l_f = \text{length of front cavity} \\ &\gamma_f = \alpha_f + j\beta \\ &\alpha_f \simeq \frac{R}{2}\sqrt{\frac{L}{C}} + \frac{G}{2}\sqrt{\frac{C}{L}} \\ &\beta \simeq \frac{\omega}{c} \end{aligned}$$

The poles satisfy:

$$\begin{aligned} \cosh \gamma_f l_f &= 0 \qquad (3.9) \\ \gamma_f l_f &= (\alpha_f + j\beta) l_f \simeq \pm j\frac{(2n+1)\pi}{2}, \quad n = 0, 1, 2, \ldots \\ j\beta &= -\alpha_f \pm j\frac{(2n+1)\pi}{2l_f} \end{aligned}$$

since $\beta \simeq \frac{\omega}{c}$

$$s_n \simeq -\alpha_f c \pm j\frac{(2n+1)\pi c}{2l_f}, \quad n = 0, 1, 2, \ldots$$

and the frequencies and bandwidths are:

$$\begin{aligned} f &= \frac{(2n+1)c}{4l_f}, \quad n = 0, 1, 2, \ldots \qquad (3.10) \\ \Delta f &= \frac{\alpha_f c}{\pi} \end{aligned}$$

These expressions incorporate viscosity and heat conduction, but as yet no loss due to radiation.

To compute the zeros, we need to find the poles of the driving-point impedance Z_{back} looking back from the pressure source, as described in Section 3.1.3. The driving-point impedance is:

$$\begin{aligned} Z_{back} &= z_{ao} + z_{bo} \\ &= Z_{0o} \frac{\cosh \gamma_o l_o}{\sinh \gamma_o l_o} \end{aligned}$$

where $Z_{0o} = \rho / A_o$, the characteristic impedance of the section of the tube between constriction and obstacle. $Z_{back} = \infty$ when

$$\frac{1}{Z_{0o}} \sinh \gamma_o l_o = 0$$

or

$$s_m = \sigma_m \pm j\omega_m = -\alpha_o c \pm j \frac{m\pi c}{l_o}, \qquad m = 0, 1, 2 \ldots$$

Thus the zero frequencies and bandwidths are

$$\begin{aligned} f &= \frac{\omega_m}{2\pi} = \frac{mc}{2l_o} \text{ Hz}, \qquad m = 0, 1, 2 \ldots \\ \Delta f &= \frac{\sigma_m}{\pi} = \frac{\alpha_o c}{\pi} \end{aligned} \tag{3.11}$$

The complete transfer function for this model, excluding radiation loss, is:

$$\frac{U_m}{p_s} = \frac{\sinh \gamma_o l_o / Z_{0o}}{\cosh \gamma_f l_f}$$

Let us now consider a non-zero radiation impedance. This affects the poles only, since the zeros are due to the resonances of the tube behind the pressure source only. Following Flanagan's method (1972), we arrive at a modified transfer function

$$\frac{U_m}{U_g} = \frac{1}{\cosh \gamma l + \frac{Z_R}{Z_0} \sinh \gamma l} = \frac{\cosh \gamma_r l}{\cosh (\gamma + \gamma_r) l}$$

where $\gamma_R l = \tanh^{-1}(Z_R / Z_0)$. As in Eqns. 3.9, $\cosh (\gamma + \gamma_R) l = 0$ when

$$(\gamma + \gamma_R) l = \pm j \frac{(2n+1)\pi}{2}$$

Taking the first term of the series

$$\tanh^{-1} x = x + \frac{x^3}{3} + \frac{x^5}{5} + \cdots,$$

a procedure that is valid as long as $Z_R / Z_0 \ll 1$, we have

$$\begin{aligned} \gamma_R &= \alpha_R + j\beta_R = \frac{1}{l} \frac{Z_R}{Z_0} \\ \alpha_R &= \frac{1}{l} \frac{A_f}{\rho c} \, Re(Z_R) \\ \beta_R &= \frac{1}{l} \frac{A_f}{\rho c} \, Im(Z_R) \end{aligned}$$

Substituting in, this gives us

$$
\begin{aligned}
j\beta_n\left(1+\frac{\beta_R}{\beta_n}\right) &\simeq -(\alpha+\alpha_R)\pm j\frac{(2n+1)\pi}{2l} \\
s_n &\simeq \frac{1}{1+\frac{\beta_R}{\beta_n}}\left[-(\alpha+\alpha_R)c\pm j\frac{(2n+1)\pi c}{2l}\right], \qquad (3.12)
\end{aligned}
$$

$$n = 0, 1, 2, \ldots$$

Flanagan used a relatively simple model for the radiation and thus was able to derive a concise analytical expression for s_n. Although we cannot do that here, the same observations he made still apply. For a nonzero radiation impedance, both α_R and β_R are greater than zero. Therefore, the effect of the radiation impedance will be to make the term outside the brackets in Eqn. 3.12 less than 1, decreasing the imaginary parts of the poles, and to make the real part of s_n more negative, thus increasing the damping.

When is the approximation to the value of $\gamma_R l$ valid? $Z_R \approx Z_0$ when $ka \approx 1.0$, where a is the radius of a piston mounted in an infinite baffle. $ka = 1.0$ corresponds to 4320 Hz for the $l_f = 12$ cm configuration, and 8640 Hz for $l_f = 3.2$ cm (the difference is due to the difference in mouth sizes). In the vicinity of these frequencies, using only the first term of the hyperbolic tangent series is not as good an approximation as elsewhere in the frequency range. If we include the first two terms of the series, the expressions for α_R and β_R become:

$$
\begin{aligned}
\alpha_R &= \frac{1}{l}\frac{A_f}{\rho c}\, Re(Z_R)\left(1+\frac{1}{3}\left(\frac{Re(Z_R)}{Z_0}\right)^2-\left(\frac{Im(Z_R)}{Z_0}\right)^2\right) \\
\beta_R &= \frac{1}{l}\frac{A_f}{\rho c}\, Im(Z_R)\left(1+\left(\frac{Re(Z_R)}{Z_0}\right)^2-\frac{1}{3}\left(\frac{Im(Z_R)}{Z_0}\right)^2\right)
\end{aligned}
$$

If we assume that both $Re(Z_R)$ and $Im(Z_R)$ equal 0.9, the value of α_R is half as big if the second term is included, and the value of β_R is half again as large. This means that leaving out the second term will result in both real and imaginary parts that are too large, in the vicinity of $ka = 1.0$. We will use this fact to assess the effect of this approximation, since in the transfer functions calculated here, only the first term of the series was used.

We can quantify our predictions about the effect of the radiation by tabulating the frequencies and bandwidths derived for this model. Equations 3.10, 3.12, and 3.11 were used to calculate the values for, respectively, the poles without radiation, poles with radiation, and the zeros. Eqn. 3.12 was used by solving for each pole frequency without radiation, computing γ_R at that frequency, and using that value to solve for s_n. Figure 3.13b shows a plot of the bandwidths as a function of frequency for this model. It is clear that the viscosity and heat conduction losses are insignificant compared to the radiation losses for the poles of the transfer function.

When the front cavity is 3.2 cm long, the obstacle occludes half of the mouth of the tube. The plane-wave model does not allow for the probable occurrence of evanescent modes in the vicinity of the obstacle, and thus the predicted frequencies of the poles are likely to be inaccurate. Since a theoretical consideration of such effects would mean abandoning the transmission line model which is of such great usefulness, we will proceed as planned theoretically but check and adjust results according to empirical data.

POLES $l_f = 3.2$ cm

Sources of Loss:	Viscosity Heat		Viscosity Heat Radiation	
	f	Δf	f	Δf
	2690	13	1850	220
	8080	23	6400	2540

POLES $l_f = 12$ cm

Sources of Loss:	Viscosity Heat		Viscosity Heat Radiation	
	f	Δf	f	Δf
	718	7	660	16
	2155	12	1990	100
	3590	15	3350	260
	5030	18	4760	470
	6465	20	6210	660
	7900	23	7680	790
	9340	25	9150	860

ZEROS

Sources of Loss:	Viscosity Heat	
	f	Δf
	0	0
	5750	19
	(11490)	27

Table 3.1: Poles and Zeros for Higher Frequency Model I. If the complex frequency of each singularity is defined as $s = \alpha \pm j\omega$, then the values listed here are $f = \omega/2\pi$, and $\Delta f = -\alpha/\pi$.

Model II.

We can now extend the model to include the constriction, terminated with a short circuit at the inlet as in Section 3.1.3. The constriction affects both frequencies and bandwidths of the resonances, since it has inductance as well as additional sources of loss, that is, viscosity and flow resistance. Since we already have the results of Model I, we will simply compute the approximate change in the resonance frequencies due to the constriction by neglecting all sources of loss, and then separately compute the change in the bandwidths by considering the energy distribution in the front cavity to be the same as for the no-loss situation.

We begin with the resonance frequency computation, using the circuits shown in Fig. 3.14a. Neglecting losses in the front cavity reduces the impedances from hyperbolic to circular functions, as shown, and sets the radiation impedance to zero. Neglecting losses in the constriction allows it to be modeled as an inductance. Resonances occur when the driving-point admittances looking to right and left at the junction sum to zero, or when

$$\begin{aligned} 0 &= Y_c + Y_f \\ &= -\frac{j}{\omega L_c} - j\frac{A_f}{\rho c}\cot kl_f, \qquad L_c = \frac{\rho l_c}{A_c} \end{aligned}$$

The graph in Fig. 3.14b shows that one of the effects of the constriction is to move resonances up in frequency. By iterating the above equation, the following frequency changes were found:

$$\begin{aligned} l_f = 3.2 : f \text{ increment} &= +54, +18, + \ldots \text{ Hz} \\ l_f = 12 : f \text{ increment} &= +51, +18, +11, +8, + \ldots \text{ Hz} \end{aligned}$$

These increments remained the same when an end effect for the front cavity, inducing the frequency-lowering effect of radiation, was included. In other words, the iterative method used to compute the amount of frequency shifting that results from the constriction is the same whether the initial frequencies include the effect of radiation or not.

For the zeros, the sum of the admittances is:

$$0 = \frac{-j}{\omega L_c} + j\frac{A_f}{\rho c}\tan kl_f$$

Iteration with this equation shows that the zero frequencies increase from 0, 5746, 11493... to 393, 5774, 11510... Hz. For the lowest zero it is actually more accurate to use the value of 250 Hz, as predicted for the lossy low-frequency (Model II) condition in Section 3.1.3. For the higher-frequency zeros, bandwidths will be computed based on the frequency increments given above.

With these resonant frequency changes in hand, we can now add loss to the circuit and compute bandwidths. Loss in the constriction is due to viscosity and flow resistance. The relative sizes of these two resistances can be compared by using Eqns. 3.7 and 3.5 for R_i, the incremental flow resistance, and R_{visc}, the frequency-dependent resistance due to viscosity, respectively. For the 0.318 cm diameter constriction, $R_{visc} = 0.132\sqrt{f} = 2.6$ g/s-cm^4 at 400 Hz, and 13.2 g/s-cm^4 at 10000 Hz. $R_i = 0.193 \times U + 8.9 = 39.8$ g/s-cm^4 when $U = 160$ cc/sec, and 96.4 g/s-cm^4 when $U = 500$ cc/sec. Thus R_i is always greater than R_{visc} across the entire range of frequency and flowrate used. Therefore, only R_i is used for the constriction.

We are now in a position to find the bandwidths of the higher-frequency poles and zeros. In order to do this, we compute

$$Q = 2\pi \frac{\text{energy stored in front cavity}}{\text{energy lost per cycle}} = \frac{f}{\Delta f} \tag{3.13}$$

for the front cavity resonances. For a given configuration the stored energy remains the same regardless of which losses are considered. Consequently, the bandwidths corresponding to different sources of loss can be computed separately and added together.

If we consider for the moment only losses due to the constriction, we can express both energy quantities in terms of P_0, the pressure at the constriction. Assuming that P_0 is a maximum at the constriction outlet (reasonable for $d_c \ll d_f$ and $l_c \ll \lambda$ such that the constriction impedance $Z_c \approx j\omega L_c + R_c$), and that we are considering a resonant frequency so that $l_f \simeq \lambda/4$, we have:

$$\begin{aligned} \text{peak energy stored, pole} &= \int_0^{l_f} \frac{1}{2} C |P|^2 \, dx \\ &= \frac{1}{2} C \int_0^{l_f} |P_0 \cos kx|^2 \, dx \\ &= \frac{1}{2} \frac{A_f}{\rho c^2} |P_0|^2 \frac{l_f}{2} \end{aligned}$$

where C is the per-unit-length compliance of the front cavity. For a zero, the integration must be performed from 0 to l_o instead, where l_o is the distance to the pressure source, at resonance frequencies for which $l_o = \lambda/2$. The change in the final expression turns out to be slight:

$$\text{peak energy stored, zero} = \frac{1}{2} \frac{A_f}{\rho c^2} |P_0|^2 \frac{l_o}{2}$$

If we now model the constriction as a resistance R_c in series with an inductance L_c, we have:

$$\begin{aligned} \text{energy lost/cycle} &= \frac{1}{2} \frac{|U|^2 R_c}{f} \\ &= \frac{1}{2} \left| \frac{P_0}{R_c + j\omega L_c} \right|^2 \frac{R_c}{f} \end{aligned}$$

where

$$\begin{aligned} R_c &= K_2 \frac{\rho U}{A_c^2} + K_3 + R_{visc} \simeq K_2 \frac{\rho U}{A_c^2} + K_3 \\ L_c &= \frac{\rho l_c}{A_c} \end{aligned}$$

We then have

$$\begin{aligned} Q &= 2\pi \left(\frac{1}{4} \frac{A_f}{\rho c^2} |P_0|^2 l_f \right) \left(\frac{1}{2} \left| \frac{P_0}{R_c + j\omega L_c} \right|^2 \frac{R_c}{f} \right)^{-1} \\ &= \pi \frac{A_f l}{\rho c^2} \frac{R_c^2 + (\omega L_c)^2}{R_c} f \\ \Delta f &= \frac{f}{Q} = \frac{\rho c^2 R_c}{\pi A_f \, l \, (R_c^2 + (\omega L_c)^2)} \end{aligned} \tag{3.14}$$

where $l = l_f$ in the final expression when the pole bandwidths are being computed, and $l = l_o$ for the zeros.

A graph of Δf vs. f according to Eq. 3.14 is shown in Fig. 3.14c for $R_c = 55.97$ g/sec-cm^4 ($U = 250$ cc/sec). A comparison of this figure with Fig. 3.13 reveals the effect of the constriction at all frequencies. The pole and zero frequencies and bandwidths are given in Table 3.2.

Model III

We now incorporate the back cavity and loss at the glottis in the model. Inclusion of the back cavity should affect the front cavity resonances, and introduce new back-cavity resonances. Since the constriction area is small relative to the front and back-cavity areas, we expect that there will be little coupling between front and back cavities, so that the effect on the front-cavity resonances will be slight and the back cavity resonance frequencies will be approximately equal to $(nc/2l_b)$, $n = 0, 1, 2 \ldots$. In the next section we will use the program TBFDA to calculate the effects of coupling between the cavities explicitly. In this section we will model the losses in the back cavity and compute their effect on the back cavity resonances.

Losses for the back cavity resonances arise from three sources: viscosity and heat conduction along the walls of the cavity, loss at the forward constriction, and loss at the glottis. From the front-cavity calculations we know that the viscosity and heat-conduction losses depend on frequency but not on cavity length, and therefore we have an idea already of the size of those losses. Therefore, let us first consider losses at the two constrictions.

The glottal pipe exits abruptly into the back cavity, but has a tapered inlet from the muffler. Since the muffler presumably absorbs all sound (it actually attenuates sounds about 20 dB in the 500 to 1500 Hz range, and about 30 dB in the 1.5 kHz to 6 kHz range), the glottal pipe acts as a reflectionless tube. The abrupt outlet and the change in area from 1 to 5 cm^2 indicate that the impedance looking upstream from the outlet is

$$Z_g = \frac{\rho c}{A_g}$$

where A_g = area of the glottal opening. The abruptness of the exit also means that the pressure is approximately a maximum there. We can thus compute Q as before, with $l_b = \lambda/2$:

$$\begin{aligned} \text{peak energy stored} &= \int_0^{l_b} \frac{1}{2} C |P|^2 \, dx \\ &= \frac{1}{2} C \int_0^{l_b} |P_0 \cos kx|^2 \, dx \\ &= \frac{1}{2} \frac{A_b}{\rho c^2} |P_0|^2 \frac{l_b}{2} \end{aligned}$$

where C is the per-unit-length compliance of the back cavity.

$$\begin{aligned} \text{energy lost at glottis/cycle} &= \frac{1}{2} |U|^2 \frac{R}{f} \\ &= \frac{1}{2} \left| \frac{P_0}{\rho c / A_g} \right|^2 \frac{\rho c}{A_g} \frac{1}{f} \\ &= \frac{1}{2} \frac{|P_0|^2 A_g}{\rho c \, f} \end{aligned}$$

$$
\begin{aligned}
Q_g &= 2\pi \left(\frac{1}{4} \frac{A_b}{\rho c^2} |P_0|^2 l_b \right) \left(\frac{1}{2} \frac{|P_0|^2 A_g}{\rho c f} \right)^{-1} \\
&= \pi \frac{A_b l_b f}{c A_g} \\
\Delta f_g &= \frac{c A_g}{\pi A_b l_b} \\
l_f = 3.2: \quad \Delta f &= 164 \text{ Hz} \\
l_f = 12: \quad \Delta f &= 525 \text{ Hz}
\end{aligned}
$$

Since Z_g is constant with respect to frequency, the contribution of glottal losses to Q is also a constant across the entire frequency range for each l_b.

We can now compute the losses at the constriction, using the same expression for stored energy as above.

$$
\begin{aligned}
\text{energy lost at constriction/cycle} &= \frac{1}{2} \left| \frac{P_0}{Z_c} \right|^2 \frac{R_c}{f} \\
&= \frac{1}{2} \frac{|P_0|^2}{f} \frac{R_c}{R_c^2 + (\omega L_c)^2}
\end{aligned}
$$

So

$$
\begin{aligned}
\Delta f_c &= \frac{\rho c^2}{\pi A_b l_b} \left(\frac{R_c}{R_c^2 + (\omega L_c)^2} \right) \\
\Delta f_{tot} &= \frac{c}{\pi A_b l_b} \left(A_g + \frac{\rho c \, R_c}{R_c^2 + (\omega L_c)^2} \right)
\end{aligned}
$$

Figure 3.15 shows bandwidth vs. frequency for the two back cavity lengths. The loss at the constriction causes the slight rise in bandwidth below 1 kHz. Losses due to viscosity and heat conduction are negligible compared to the losses at the entrance and exit of the back cavity.

The area function of the entire tract, including the back cavity, was input to TBFDA in order to generate the frequencies of the back cavity poles and zeros. Since the back cavity resonances show little effect of radiation impedance, they can be distinguished from the front cavity resonances by their low bandwidths. Their frequencies, generated by TBFDA, are shown in Table 3.3. Note that the pole and zero frequencies alternate. This can be proven to be necessary by the separation property (Potter and Fich, 1963), which states that the poles and zeros of a driving-point impedance or admittance function must alternate. Due to the position of the back cavity relative to the input and output of the transfer function, it influences the system via its driving-point impedance.

Since the back-cavity pole-zero pairs are all closer in frequency than the predicted back-cavity bandwidths, the small ripples they produce in the transfer function will be obscured. In the final transfer function the back cavity resonances are included with the bandwidths shown in Fig. 3.15, but their effect is much less pronounced than that of the front cavity resonances.

POLES $l_f = 3.2$ cm

Sources of Loss:	Viscosity Heat		Viscosity Heat Radiation		Viscosity Heat Radiation Constr.	
	f	Δf	f	Δf	f	Δf
					610	real
	2690	13	1850	220	1900	270
	8080	23	6400	2540	6420	2540

POLES $l_f = 12$ cm

Sources of Loss:	Viscosity Heat		Viscosity Heat Radiation		Viscosity Heat Radiation Constr.	
	f	Δf	f	Δf	f	Δf
					610	real
	718	7	660	16	710	75
	2155	12	1990	100	2000	110
	3590	15	3350	260	3360	270
	5030	18	4760	470	4760	470
	6465	20	6210	660	6210	660
	7900	23	7680	790	7680	790
	9340	25	9150	860	9150	860

ZEROS

Sources of Loss:	Viscosity Heat		Viscosity Heat Constr.	
	f	Δf	f	Δf
	0	0	250	610
	5750	19	5770	26
	(11490)	27	(11500)	29

Table 3.2: Poles and Zeros for Higher Frequency Model II

l_f = 3.2 cm		l_f = 12 cm	
Poles	Zeros	Poles	Zeros
1365	1377	4390	4395
2718	2706	8710	8750
4057	4046		
5395	5387		
6815	6735		
8112	8078		
9453	9422		

Table 3.3: Frequencies of the poles and zeros predicted when the back cavity is included in Higher Frequency Model III. The poles and zeros are arranged in pairs to demonstrate the small differences in frequencies relative to the bandwidths, derived in the text, of approximately 164 Hz for l_f = 3.2 cm, and 525 Hz, for l_f = 12 cm.

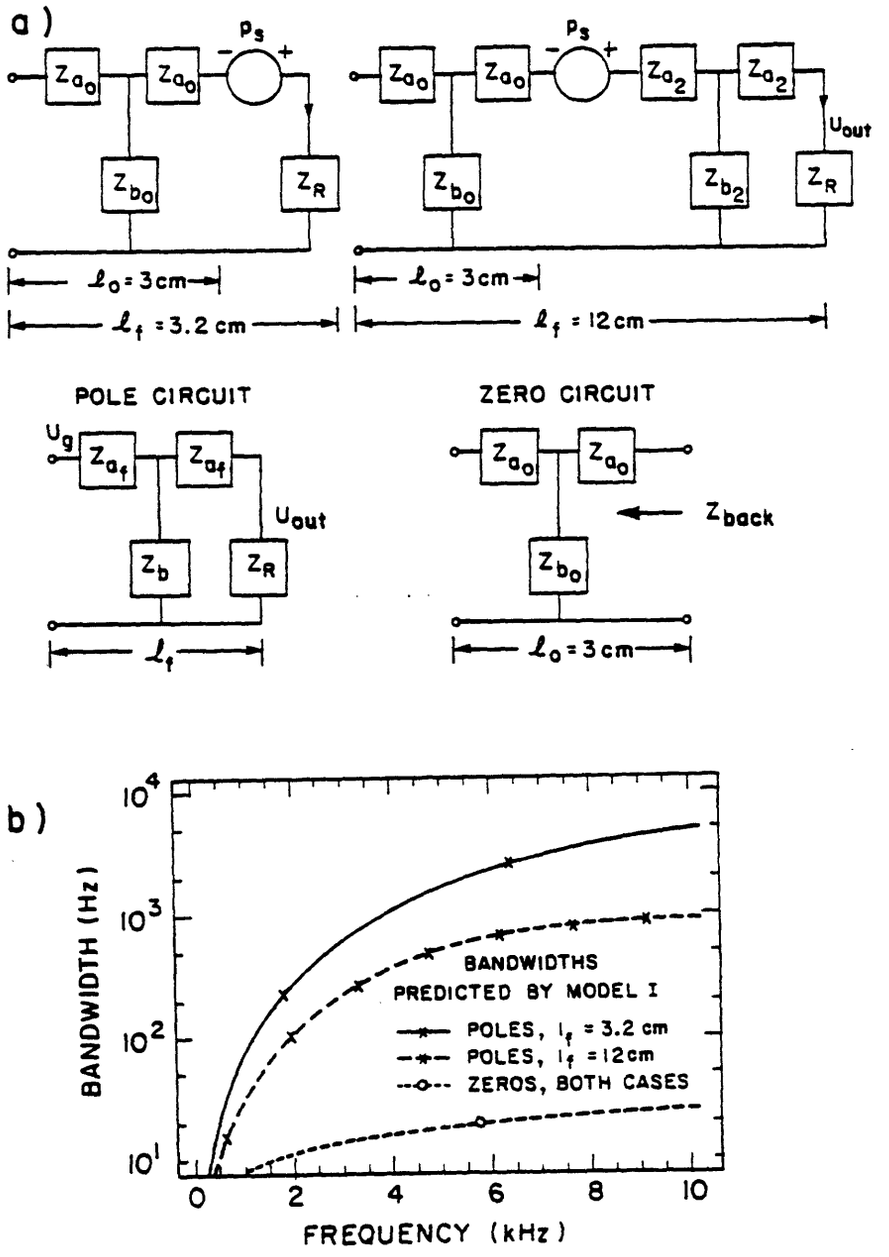


Figure 3.13: Higher Frequency Model I.
a) Networks of the front cavity, modeling constriction as a rigid wall.
b) Bandwidth vs. frequency for Higher Frequency Model I. Predicted poles are shown by crosses, predicted zeros by circles. For poles, bandwidths include losses for radiation, viscosity, and heat conduction; for zeros, bandwidths include losses due to viscosity and heat conduction only.

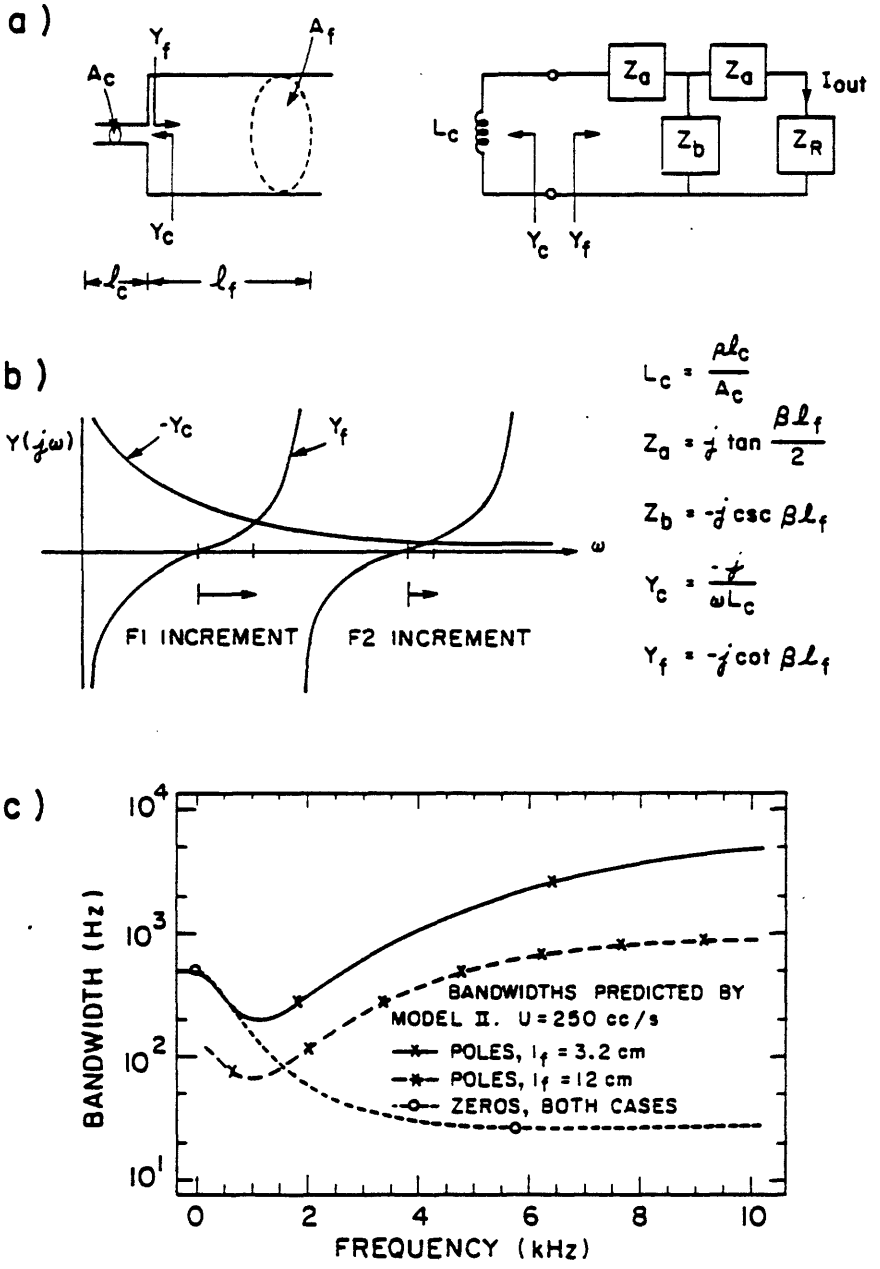


Figure 3.14: Higher Frequency Model II.
a) Diagram identifying admittances in relation to the tube (front cavity plus constriction), and lossless circuit model, including lumped element for constriction.
b) Admittances vs. frequency, showing the change in resonant frequencies due to constriction.
c) Bandwidth vs. frequency when loss is introduced in Model II. Predicted poles are shown by crosses, predicted zeros by circles. For poles, bandwidths include losses due to radiation, viscosity, heat conduction, and flow resistance; for zeros, bandwidths include losses due to viscosity, heat conduction and flow resistance.

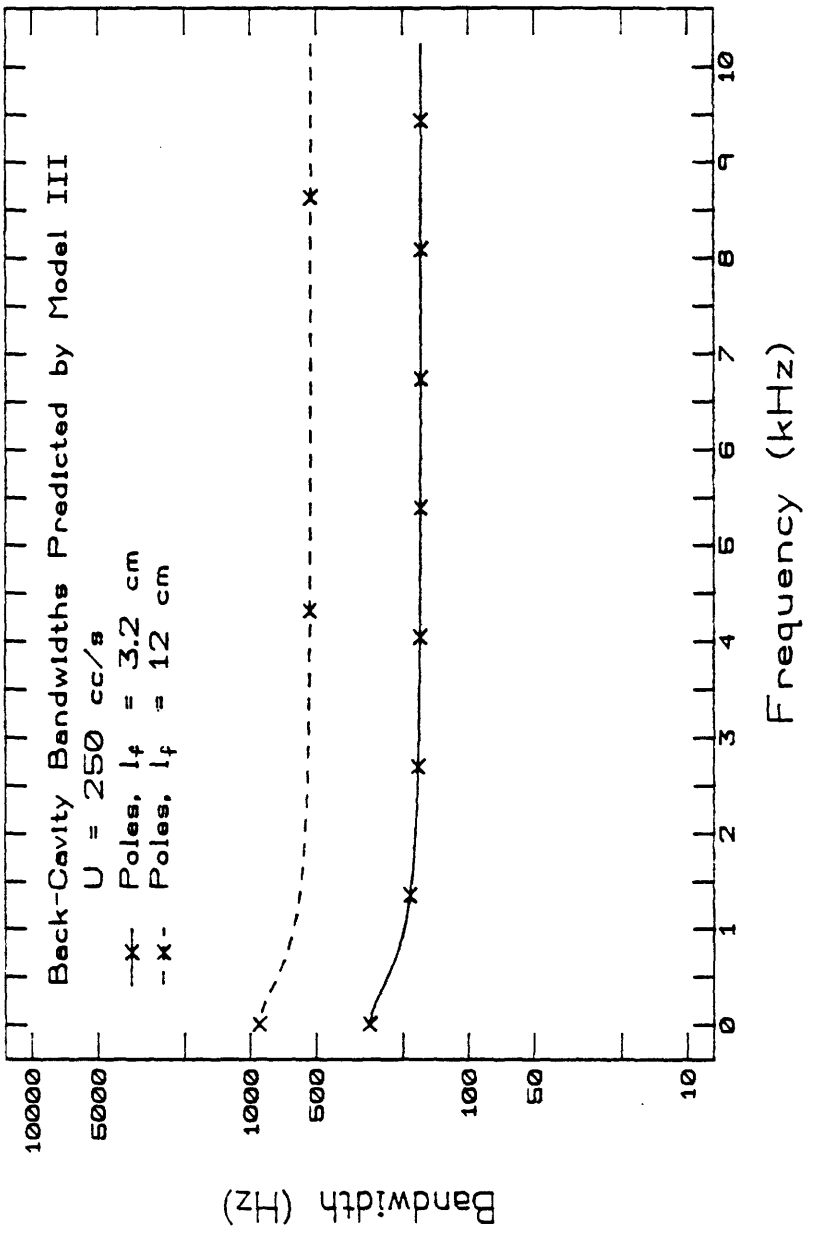


Figure 3.15: Bandwidth vs. frequency for Higher Frequency Model III for back cavity resonances. Bandwidths include losses due to glottis and constriction.

$l_f = 3.2$ cm					$l_f = 12$ cm				
Poles		Type	Zeros		Poles		Type	Zeros	
f	Δf		f	Δf	f	Δf		f	Δf
150	real	Bound Pair	50	real	280	820	Bound Pair,	100	real
530	real	Free Zero	320	625			Free Zero	450	720
1365	185	Bound Pair	1375	185	730	70	Free Pole		
1790	265	Free Pole			2010	115	Free Pole		
2720	170	Bound Pair	2705	170	3370	270	Free Pole		
4055	165	Bound Pair	4045	165	4390	535	Bound Pair	4395	535
5395	165	Bound Pair	5387	165	4870	490	Free Pole		
		Free Zero	5850	26			Free Zero	5850	26
6110	2340	Free Pole			6330	675	Free Pole		
6815	165	Bound Pair	6735	165	7680	790	Free Pole		
8112	165	Bound Pair	8080	165	8710	525	Bound Pair	8750	525
9453	165	Bound Pair	9420	165	9220	870	Free Pole		

Table 3.4: Complete set of poles and zeros for the final model, obstacle case, $U = 250$ cc/sec.

3.1.5 Final Models, Obstacle Case

The resonance frequencies for the various higher-frequency models have been found by deriving frequencies for the lossless front or back cavity, and then adjusting the values to account for radiation, the effect of the constriction, and so on. By proceeding in this manner it has been possible to see the detailed effects of changes in the configuration and of the different types of loss. However, the momentary change in area at the location of the obstacle has not been taken into account for the $l_f = 12$ cm case, and the influence of the back cavity on front cavity resonances, though likely to be small, has not been explicitly calculated.

Both of these problems can be handled by judicious use of TBFDA, where the specification of many short sections of different cross-sectional areas remains computationally feasible. Thus, TBFDA was used to generate the final transfer functions for the obstacle case, giving us slightly more accurate results than are possible with Model III. Since the way losses are modeled in TBFDA runs counter to our application, a no-loss tract – no radiation, no wall loss – was specified as input to TBFDA. The resonant frequencies it computed were used as input to a separate program (BAND) that calculated the final set of frequencies and bandwidths using the Model III losses (as specified by the equations of Section 3.1.4). The values generated by BAND for $U = 250$ cc/sec are given in Table 3.4. In the table, the back-cavity pole-zero pairs are designated as "Bound Pairs", and the poles and zeros that remain uncanceled are designated "Free Pole" or "Free Zero". Values for other flowrates are given in Appendix C.

The transfer functions for the two configurations have an infinite set of poles and zeros; the tables above specify only those falling within the 0 to 10 kHz frequency range. A transfer

function generated from only the listed singularities would have a pronounced upward (or downward) drift depending on whether there were more zeros than poles (or vice versa) specified. This drift can be counteracted by either specifying some singularities above 10 kHz, or applying a correction factor. Since a very large number of singularities must be specified in order to appreciably alter the drift, correction factors were computed for the zeros and for the two sets of poles in the following manner.

Since the back cavity poles and zeros very nearly cancel, the correction factors for each would cancel also. Likewise, all of the low-frequency poles and zeros cancel except for one zero. Therefore, we need only compensate for the higher frequency front-cavity poles, and the front cavity zeros including one zero near 0 Hz. Let us consider the poles first. A normalized allpole function consisting of an infinite number of poles evenly spaced in frequency with identical bandwidths will have the same amplitude at all the troughs, A_t, of 0 dB. The peaks will likewise all have the same value – call it A_p – which depends on the bandwidth. The desired correction function is then determined by the amounts by which the troughs of an allpole function, consisting of a *finite* number of poles, deviate from 0 dB, or the amount by which the peaks deviate from A_p.

Let us rephrase that more precisely. Call the sequence of peaks, from low to infinite frequency, p_i, for $i = 1, 2, \ldots, \infty$, and the sequence of troughs t_i, $i = 1, 2, \ldots, \infty$. Since both peaks and troughs should have constant (though different) amplitudes, we can use the more convenient amplitude as the basis of comparison between the correct function and the function to be corrected. A correction function can thus be computed by generating the transfer function $T_N(f)$ of N poles, where N is finite, and finding the trough amplitudes $B(f(t_i))$ at each trough frequency $f(t_i)$, $i = 1, \ldots, N$. These trough amplitudes are then subtracted from the correct amplitude $A_t = 0$ dB, generating amplitude differences $D_t(f(t_i)) = A_t - B(f(t_i))$. The correction function is generated by linearly interpolating between the amplitude differences $D_t(f(t_i))$, $i = 1, \ldots, N$, and is valid for frequencies less than $f(t_N)$.

Since the poles of the actual transfer function were not evenly spaced and had unequal bandwidths, the amplitude of the troughs would not necessarily be zero dB, or even necessarily equal, and the correction function is much more difficult to determine. An approximation to the correction function was therefore computed by generating a synthetic transfer function which did have evenly spaced poles of equal bandwidths and using it as described above. The synthetic transfer function used the actual bandwidth of the lowest frequency pole as the bandwidths for all poles, and the frequency of the first pole determined the frequency spacing of all higher poles by:

$$f_n = (2n - 1)f_1, \quad n = 2, 3, \ldots$$

Here f_i is the peak frequency of the ith pole. For example, for the $l_f = 3.2$ cm case, the pole frequencies used to compute the correction function were 1790, 5370, and 8950 Hz, and all had the bandwidth 275 Hz. If the poles are spaced far apart, the correction curve will have noticeable corners due to the linear interpolation. These can be "smoothed" by computing corrections at the frequencies of the peaks (which occur between each trough) also. The correct peak value A_p was found by specifying a large number of poles ($N = 21$ for $l_f = 3.2$ cm, $N = 38$ for $l_f = 12$ cm) at frequencies greater than 10 kHz, which was sufficient to correct the lowest-frequency pole. (The lowest pole was considered corrected when the trough just above it was within 1 dB of the correct value, 0 dB.) The amplitude of that first peak was then used as the correct peak value A_p, and the procedure continued as for the troughs. The largest error possible by this method was therefore 1 dB.

For the zeros the situation was complicated by the unnormalized real zero near 0 Hz. which makes A_f unknown. With a single real zero at 0 Hz and an infinite number of complex conjugate pairs occurring at half-wavelengths with the same bandwidths. the zeros will all have the amplitude A_z and the broad maxima between them will all have the amplitude A_m. Both amplitudes, however, depend on the bandwidth used. Therefore, a large number of zeros ($N = 28$) were specified in order to generate A_z and A_m, and then the differences between those amplitudes and the amplitudes of the actual transfer function were interpolated to form the correction function. Figure 3.16a shows the three correction functions used. The pole corrections are to be added, the zero correction subtracted from the uncorrected transfer function magnitude expressed in dB. Although the correction functions become very large at high frequencies, they are accurate for the specified functions to within 1 dB for frequencies less than the highest-frequency singularity specified. The non-constant bandwidths and non-harmonic resonance frequencies of the actual transfer functions on which the correction functions are used decreases the accuracy of the corrections by an amount that is difficult to determine but is likely to be quite small. Part b of the same figure shows the corrected transfer functions generated by the values tabulated here.

In Section 3.1.1 we discussed a way of deriving a source function $\hat{p}_s$ from the measured far-field sound pressure p_0. In Sections 3.1.2 through this section we have derived the transfer functions U_{out}/p_s which when multiplied by the source function will generate $\hat{U}_{out}$. We now need to convert $\hat{U}_{out}$ to the predicted far-field sound pressures $\hat{p}_3$ and $\hat{p}_{12}$ so that we can compare them to the measured pressures. We do this by regarding the end of the baffled tube as a simple source of strength $\hat{U}_{out}$, which is radiating all of its energy into a half-space. The magnitude of the far-field sound pressure is then

$$\hat{p}(r,\omega) = \frac{f\rho}{r}\,\hat{U}_{out}$$

So we have

$$\begin{aligned}
\hat{p}_3(r,\omega) &= \hat{p}_s(\omega)\,\frac{U_{out3}(\omega)}{p_s(\omega)}\,\frac{\widehat{p_3}(r,\omega)}{\hat{U}_{out3}(\omega)} \\
&= \hat{p}_s(\omega)\,\frac{U_{out3}(\omega)}{p_s(\omega)}\,\frac{f\rho}{r} \\
&= \frac{\omega\rho}{A}\,\frac{p_0(\omega)}{\hat{p}_0/S}\cdot T_3(\omega)\cdot\frac{f\rho}{r}
\end{aligned}$$

where $S = Ud$ of the dipole source, as defined in Section 3.1.1, and $T_3(\omega)$ is the transfer function when $l_f = 3.2$ cm. The expression for $\hat{p}_{12}$ is analogous to that for $\hat{p}_3$.

As frequency increases, the end of the tube radiates sound less uniformly with respect to angle. This is characterized by the directivity index for a baffled piston, which is a ratio of the actual sound intensity at a given angle to the intensity of a simple source. It would be possible to use the directivity index to correct the high frequency output; however, at 10kHz the difference on axis is only 3 dB, so this effect will be neglected.

3.1.6 The No-Obstacle Case

We now turn to the second idealized case, that of the constriction in a duct without a downstream obstacle. Morse and Ingard (1968) proposed that from far away, a jet may be

viewed as an acoustic line source of length L, with a source strength that varies with position, where L is the mean distance an eddy travels in its lifetime. As pictured in Fig. 3.17, the line source would be located on the axis of the jet, and would pulse with velocity $v_r(t)$. They solved for the radiated sound power by assuming that the jet was a frozen pattern of flow irregularity convected with the mean velocity V_0, that the velocity fluctuations were correlated along the entire length L of the jet, and that there was a simple harmonic velocity relation describing v_r along the length L. However, they indicated that a detailed analysis is impossible without the correlation function for the velocities along the length of the source. Such information is not currently available.

We could assume a correlation function and solve for the expected output when the assumed source is inside a tube. However, there are many effects occurring, such as refraction of the sound by the jet and the detailed distributions of eddy sizes and lifetimes, that make it very difficult analytically. We could replace the source by a very simple approximation, such as a distribution of plane sources in a one-dimensional model. But with the still relatively large number of parameters, such as the strength, spectral characteristics, position, and relative coherence of the source elements, that cannot be measured (given the present setup) or deduced, differences between the model and experimental data could have any of a number of causes. Furthermore, for the no-obstacle case, we do not have far-field measurements of the jet for $l_f = 0$ cm, i.e., for a constriction located at the mouth of the model. Deriving a source necessarily described by more parameters from a measurement already confounded by the effect of the tube seems so much more speculative than the obstacle case as to be meaningless.

Therefore, no attempt was made to model the source in this case. Nevertheless, we can investigate the extent of soruce-tract interaction in the following way. If there is no interaction, then the source depends only on the constriction shape and size, and the flowrate. Therefore, at a given flowrate we should have identical unknown sources for the two configurations, $l_f = 3.2$ and $l_f = 12$ cm. As in the obstacle case, these sources will excite the natural frequencies of the entire tract, which are different for the two configurations, and make up the denominator of the transfer function from source to output. Let us call this denominator $D(s)$. The roots of the numerator will depend on source location and type, but in any case will be related to the part of the tract behind each source. If there is no interaction, the source locations and types are identical for the two configurations, and therefore the numerators are substantially the same. Since the source is distributed, we really have many numerators, each corresponding to a single localized source element. Denoting the source elements by S_i and each numerator by N_i, the predicted output for the configuration with $l_f = l$ would have the form

$$P_l(s) = \frac{S_1(s)N_1(s) + S_2(s)N_2(s) + \cdots + S_n(s)N_n(s)}{D_l(s)} \tag{3.15}$$

Thus, if we divide the outputs from both configurations by the appropriate allpole functions, we should find

$$P_3(s)\,D_3(s) = P_{12}(s)\,D_{12}(s) = S_1(s)N_1(s) + S_2(s)N_2(s) + \cdots + S_n(s)N_n(s)$$

if the sources are identical in the two cases, that is, if there is no source-tract interaction.

We need merely to assemble the appropriate allpole transfer functions. Since this model is much cruder than in the obstacle case, we leave out the back-cavity poles that are very nearly canceled by zeros. We also leave out the poles derived from the low-frequency models, since the net number of low-frequency singularities depends on the unknown source type. The procedure

$l_f = 3.2$ cm		$l_f = 12$ cm	
f	Δf	f	Δf
2160	380	730	72
7410	2600	2030	117
		3410	280
		4840	490
		6310	670
		7800	800
		9290	870

Table 3.5: Poles for the no-obstacle case, $U = 275$ cc/sec.

for deriving the remaining front-cavity poles is essentially the same as that developed in Sections 3.1.3 through 3.1.5. TBFDA is used to obtain the resonance frequencies for the lossless case, but this time without the perturbation in the area function induced by the obstacle. The frequencies are then modified and their bandwidths computed by including radiation, viscosity, heat conduction and constriction resistance, as before. The resulting poles, computed for the lowest flowrate of 275 cc/sec, are given in Table 3.5. The corresponding allpole transfer functions are graphed in Fig. 3.18. Values for the other flowrates are given in Appendix D. In Section 3.2 transfer functions generated from these values will be divided into the spectra of the measured sound pressures as discussed above.

We now turn once more to measurements of the sound produced by mechanical models. Comparison of these predictions to the measured spectra will demonstrate how well the source models work, and whether there is evidence of interaction of the source with the tube.

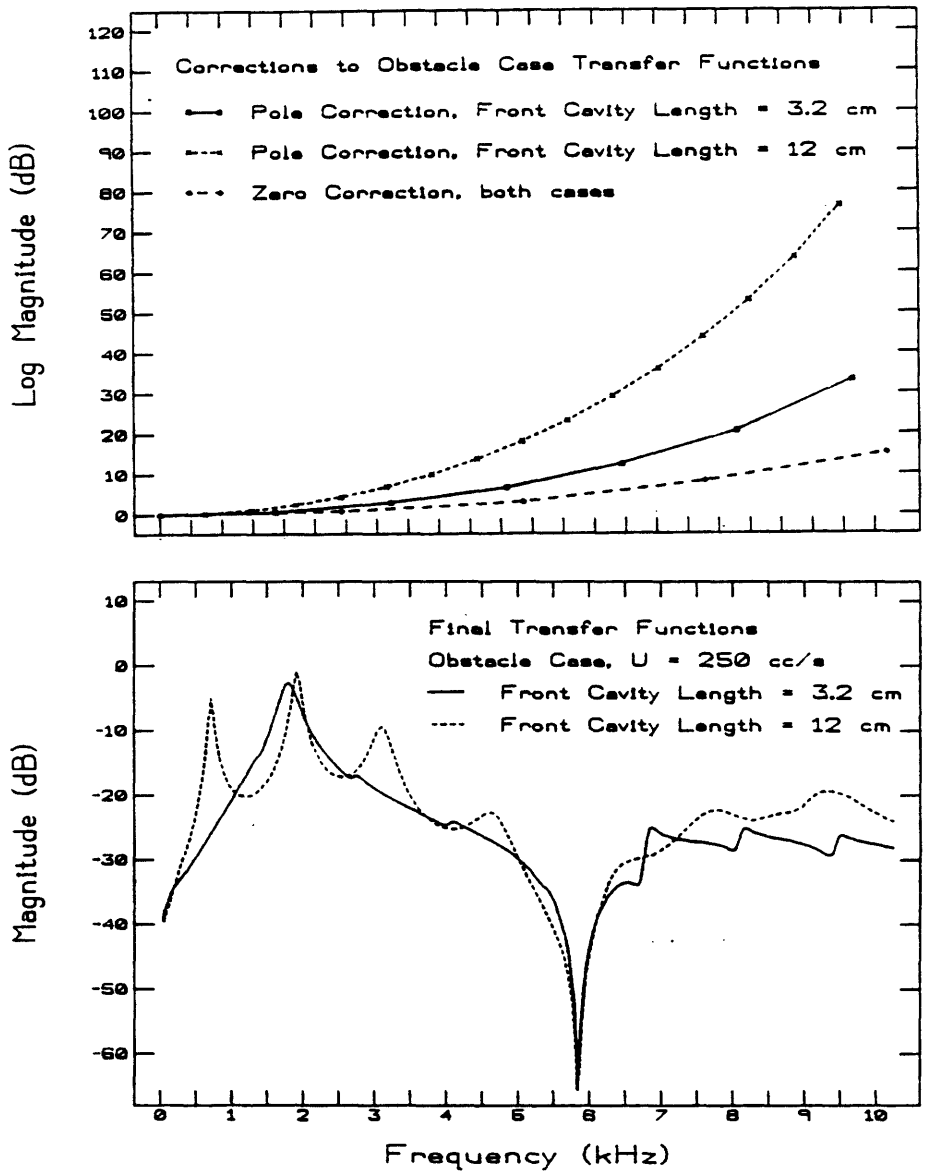
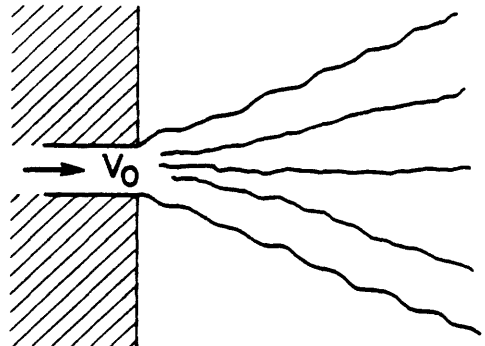
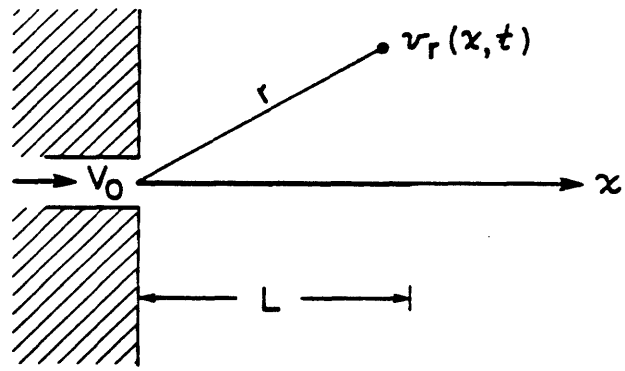


Figure 3.16: a) Higher-pole and -zero correction functions. The pole corrections are added, zero correction is subtracted from the transfer function.
b) Final transfer functions for the obstacle case, predicted U_{out}/p_s , at typical flowrate of 250 cc/sec.



JET IN FREE SPACE



$$v_r(x,t) = \begin{cases} 0 & x<0, x>L \\ v_0 \sin \frac{2\pi}{\Lambda}(x - V_0 t) & 0<x<L \end{cases}$$

Λ = "WAVELENGTH" OF FLOW IRREGULARITY

ACOUSTIC LINE SOURCE

Figure 3.17: Diagram of the acoustic line source proposed by Morse and Ingard (1968).

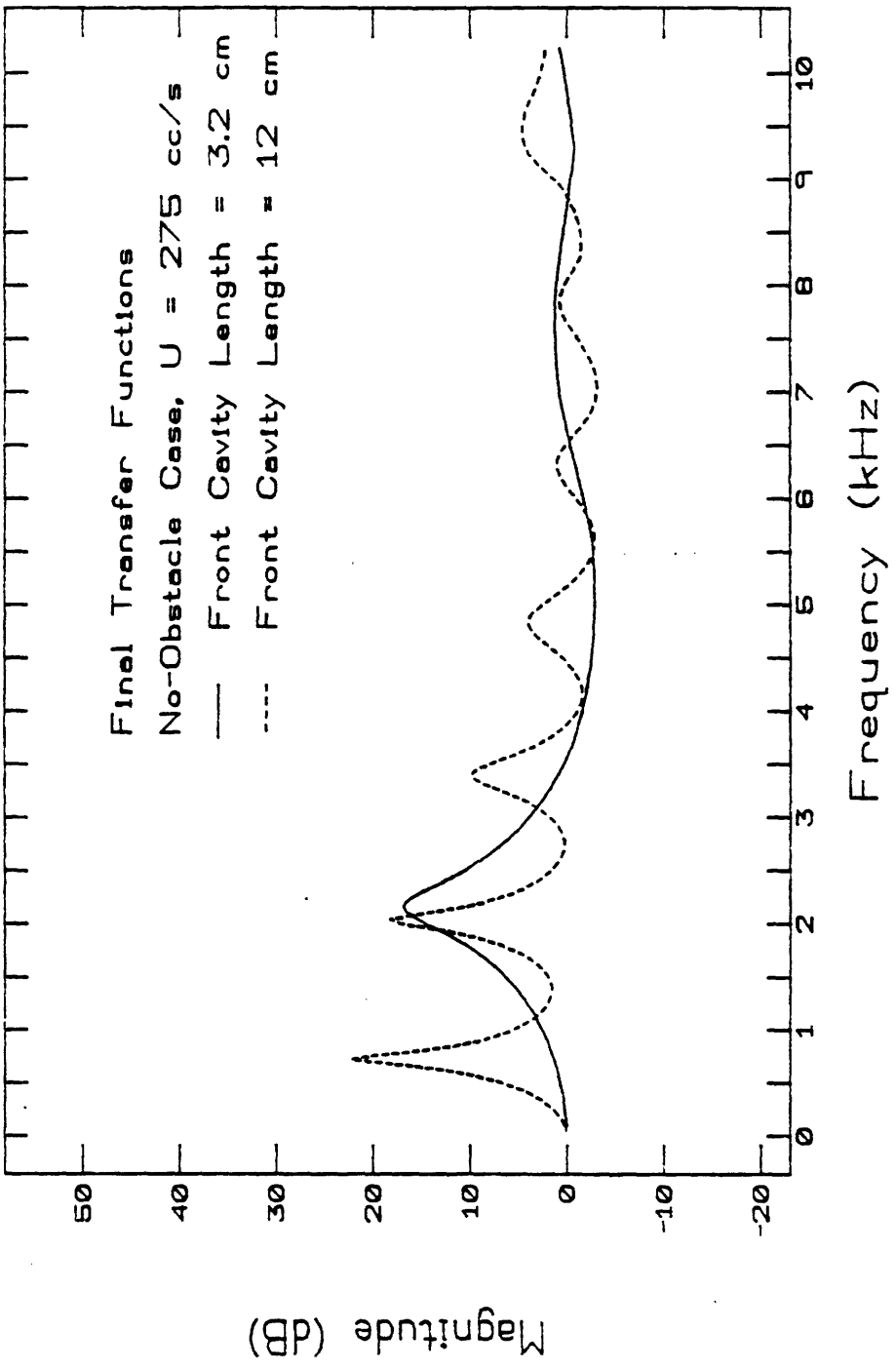


Figure 3.18: Final allpole transfer functions for the no-obstacle case at typical flowrate of 275 cc/sec.

3.2 Comparison of Experimental Data and Theoretical Predictions

3.2.1 Method

The experimental data presented in this section were generated and recorded using essentially the same method as discussed in Chapter 2. That is, air from a pressurized air tank passed through a flowmeter and muffler and into the tube with the configuration under study; the sound was picked up by a Bruel and Kjaer microphone, analyzed by an HP spectrum analyzer, and transfered to a VAX computer for further analysis. However, many more parameters of the setup were kept constant than in Chapter 2; they are summarized here. As mentioned in Section 3.1.1, the only constriction used was that with a tapered inlet, 1 cm long, and 0.32 cm in diameter. It was used in one of three positions inside a 17 cm long, 2.54 cm inner diameter plastic tube; either flush with the mouth of the tube, 3.2 cm back from the mouth, or 12 cm back from the mouth. The plexiglass baffle was always used, as was the fiberglass. When the obstacle was used, it was always positioned 3 cm downstream of the constriction (i.e. $l_o = 3$ cm) with the aid of the traverse mechanism. The flat edge of the obstacle was horizontal and aligned so that the center of that edge intersected the axis of the constriction.

The microphone was attached to a pivoting arm of length $r = 26$ cm with the pivot point located directly under either the mouth of the tube (for cases with a non-zero front-cavity length) or the obstacle (for the case where the constriction was flush with the mouth and the obstacle was 3 cm downstream of the constriction, hanging in free space). Directivity patterns were measured by recording the spectrum at fifteen different angles, ranging from 12° to 92°. Only data taken at the angle of 28° were used in the comparisons of filtered and raw data (in this case, filtered data refers to the data that was inverse filtered to remove the effect of the baffle and arrive at the estimated source function, and then filtered to predict the effect of the surrounding tube).

In the no-obstacle case, sound generated by the constriction when it was at the mouth of the tube was below the background noise until the microphone was brought to within 6 cm of the constriction. Since this is within the near field for frequencies under about 5000 Hz, these data were used only as a general reference and not for detailed analytical comparisons.

The ambient room noise was measured at the beginning of every experimental session. That noise spectrum was smoothed by performing a logarithmic curve-fitting procedure on it, in which a and b parameters were found giving the best least-mean-square fit to the following line:

$$\hat{p} = a \log_{10} f + b$$

The smoothed spectrum was then subtracted from all of the other raw autopower spectra before graphing them or using them to derive other functions. Figure 3.19 shows a typical room noise spectrum and its smoothed version. Although the curve-fitting procedure does not generate a perfect fit, it is excellent in the frequency ranges where the measured spectra tend to have their lowest values and thus require the largest room noise correction. Naturally, when the raw spectrum dips below the smoothed room noise (quite possible since the raw autopower spectrum is the average of a finite number of samples), it is pointless to subtract the room noise: in these cases the raw value is retained. This can result in an unrealistically jagged spectrum in the lower amplitude ranges.

3.2.2 Obstacle Case

Source Models

For the 0 cm front cavity plus obstacle configuration, the far-field sound pressure spectrum was measured at ten flowrates between 160 and 420 cc/sec. Figure 3.20 shows four of these spectra spanning the range of flowrates, together with a spectrum of the smoothed ambient noise. All four p_0 curves have the same general shape, with minima near 500 and 6000 Hz (the lower minimum is presumably at 0 Hz, but is obscured by room noise below 500 Hz). As with the data presented in Ch. 2 showing the effects of variations in l_o and θ, these spectra are consistent with the dipole model of the source.

For the baffle-obstacle distance of 3 cm, the minimum is predicted to be 6360 Hz, but in these data it actually falls at 6120 Hz. The difference in the frequency of the minima could be due to slight errors in measuring l_o or θ (1 mm or 4 degrees, respectively, would shift the frequency by 240 Hz), or to some other factor not incorporated in the source model (e.g. the principle sound generation does not occur at the upstream face of the obstacle, or the mean flow affects resonant frequencies slightly). We therefore note that a discrepancy exists, and proceed to use an inverse filter based on a source localized at $l_o = 3.1$ cm that matches the frequency of the minimum observed in the data.

After the inverse filtering procedure of Eqn. 3.3 and multiplication by $\omega\rho/A$, we obtain the predicted pressure source spectra, $\hat{p}_s$, shown in Fig. 3.21a. All spectra show a glitch around 6 kHz, which indicates that the actual minima are broader than that predicted. Possible explanations for this are that the noise from the jet itself fills in the zeros, since that (neglected) source of sound is not localized and is therefore not subject to cancelation at a specific frequency due to reflection in the baffle; that the discrepancy is due to the smearing effect of the window used by the spectrum analyzer; that sound generation at the obstacle is not as perfectly localized as was assumed; that some scattering occurs even with the baffle; or that the flow dipoles are not all longitudinal. The first explanation, that noise from the jet fills in the zeros, can be ruled out by reference to the no-obstacle experiments. The free-jet noise is simply too quiet to be observable when the microphone is 26 cm from the mouth of the tube. The second explanation is valid, since the Hanning window will have the effect of smearing the minimum, making it appear to be less sharp than it is. However, the predicted minimum has a bandwidth of approximately 300 Hz; since this is considerably larger than the window bandwidth of 60 Hz, the smearing effect is not large enough to explain the discrepancy. The experiments with non-zero front cavity lengths should help us decide among some of the other possibilities, since all of the explanations except that of longitudinal source distribution involve significant amounts of non-longitudinal propagation of sound, and evanescent modes die out quickly in a tube. Thus, if the trough due to reflections is also broader than predicted when l_f is *greater* than 0 cm, it becomes likely that the sound source is somewhat distributed.

Recognizing that the slight mismatch of data and hypothesized source model exists, it seems pointless to allow that mismatch to propagate through the remaining filtering operations. Therefore, a curve-fitting procedure that would estimate likely values in the region of the glitch is in order. But what type of curve is appropriate?

Data from numerous experiments have been published showing the overall sound power spectra for free jets, with a broad peak at $0.2V/d$ Hz (V = jet velocity at the nozzle, d = nozzle diameter) and falling off at an increasing rate of roughly -3, -4, and -5 dB/octave in successive octaves above the peak (Goldstein, 1976). Spectra for jets impinging on an obstacle,

scaled by size of jet and distance to obstacle, are not available in the literature, since the spectral characteristic of the sound produced depends heavily on the obstacle shape. However, for an edgetone configuration, $0.2V/x$ (x = distance between the jet nozzle and the edge the jet impinges on) is often given as the characteristic oscillation frequency. In other words, the flow instabilities will produce a very sharp fundamental at the characteristic frequency when there are no system resonances to couple into. It is not clear if the spectrum of such a configuration can be expected to exhibit a peak at the characteristic frequency when the flow is not in an unstable regime. Therefore, we will first try comparing our data to the free jet spectra. The predicted peak frequency, $0.2V/d$, ranges from 1280 to 3350 Hz, depending on the flowrate, for the constriction used here. We have built into $\hat{p}_s$ an ω^{-1} dependence relative to p_0, which would produce a slope of -6 dB/octave, simply from the way $\hat{p}_s$ is defined. Thus we should expect slopes of -9, -10, and -11 dB/oct in the successive octaves above the peak frequencies.

These predictions were checked (somewhat crudely) by plotting $\hat{p}_s$ on a log frequency scale, as shown in Fig. 3.21b, and fitting straight lines to the data within each relevant octave by eye. (The small number of points in each band and the variation intrinsic in the data made a more automatic procedure inadvisable.) The results for four flowrates are shown in Table 3.6. The minimum and maximum frequencies (f_1, f_2) and the change in $\hat{p}_s$ (Δp) are given for each octave.

U	$.2V/d$	1st octave			2nd octave			3rd octave		
		f_1	f_2	Δp	f_1	f_2	Δp	f_1	f_2	Δp
cc/s	Hz	kHz	kHz	dB	kHz	kHz	dB	kHz	kHz	dB
160	1275	1.3	2.6	-7	2.6	5.2	-13	5.0	10.0	-14
250	1990	2.0	4.0	-9	4.0	8.0	-13			
330	2630	2.5	5.0	-9	5.0	10.0	-14			
420	3350	3.0	6.0	-9						

Table 3.6: Source spectra slopes measured in successive octaves above predicted peak frequencies, from obstacle data.

The slopes definitely increase with increasing frequency, although somewhat faster than expected. The frequencies of the actual maxima, defined as the maximum amplitude (within the range of 500 to 10200 Hz) in each source spectrum derived from a measured sound pressure spectrum, are roughly half the predicted peak values. For the lowest flowrate the true peaks may well be obscured in room noise.

If we use the distance to the obstacle, 3 cm, as the value for l_o, the peak frequencies predicted by $0.2V/l_o$ are all reduced by a factor of 10 and move so low that they are obscured by the ambient noise. For $U = 420$ cc/s, the peak at 1440 Hz is then higher than predicted. It appears that the obstacle configuration falls inbetween the empirically derived $0.2V/d$ and $0.2V/l_o$ scaling rules. The relatively scarce data here are not sufficient to derive a more applicable rule. Nevertheless, the spectra do exhibit broad peaks which give way to increasingly steep slopes. Thus it is clear that fitting either a single straight line or a single logarithmic curve will

ignore significant features of the spectra.

Therefore, an exponential curve fit was used, which found a and b parameters giving the best least-mean-square fit for the following line:

$$\hat{p}_s = ae^{bf}$$

The regression was performed excluding values below the frequency of 500 Hz (since that contained the major part of the room noise) and between 5500 and 6700 Hz (since that contained the glitch). Values were then predicted for the entire frequency range, and a correlation factor of r^2 computed. The dashed lines in Fig. 3.21b and c show the derived curves. The regression parameter a varies from 80 at the lowest flowrate to 95 at the highest; b varies from -0.0007 to -0.0004; r^2 varies between 0.953 and 0.976. The correlation factor was lowest at the two lowest flowrates (160 and 190 cc/sec), due to the intrusion of room noise at low frequencies.

Source plus Tract Models

The exponentially-fit source functions were multiplied by the transfer functions obtained in Section 3.1 to arrive at predicted functions $\hat{p}_3$ and $\hat{p}_{12}$. Figure 3.22 shows the predicted and measured curves for a front cavity of 3.2 cm (i.e. p_3 and $\hat{p}_3$) at the same four flow rates used in the previous two figures. The free zero near 6 kHz is never as deep as predicted, although it deepens somewhat as the flowrate increases and raises the amplitude of p_3 relative to room noise. In this case, the predicted bandwidth of the zero is, at 24 Hz, significantly smaller than the analysis window bandwidth, which would tend to smear the minimum. Use of a restricted frequency range with an analysis bandwidth of 15 Hz (instead of 60 Hz) at the maximum flowrate of 420 cc/sec lowered the amplitude of the minimum by 9 dB and decreased the bandwidth accordingly, from approximately 350 Hz to 210 Hz. Thus, for the reduced frequency range, the observed minimum has an amplitude about 9 dB higher and a bandwidth 9 times larger than predicted. Smaller analysis bandwidths required such a restricted frequency range that the minimum was no longer observable. It appears, then, that much, but not all, of the discrepancy in the fit of the free zero is due to the analysis procedure itself.

The functions are not shown below 500 Hz since room noise predominates there. In the region between 500 and 1500 Hz the two functions generally differ by 3 or 4 dB. Since the source model and the expression for sound propagation from the mouth of the tube to the microphone depend on far-field approximations, the predicted function is less accurate in this range (progressively so as frequency decreases), where the microphone is in the near field (it is in the near field in order to have the signal not obscured by room noise; the near field portion is shown in spite of being the near field because it includes the first resonance of the $l_f = 12$ cm case). Also, the dimensions of the baffle are on the order of a wavelength in this frequency range, making its approximation in the model as an infinite baffle less accurate. The functions also do not match well above the free zero. It is expected that they will diverge around 8 kHz due to the cross mode, and that the neglected change in the directivity index will make the predicted curve slightly low in this region. In addition, the approximation made in deriving the effect of the radiation impedance can be expected to predict too much damping and too large frequency shifts in the vicinity of 8600 Hz. These factors taken together probably explain the discrepancy, which begins in the vicinity of the second front-cavity resonance around 6.5 kHz.

In spite of these mismatches, the prediction is actually quite accurate in terms of spectral shape and absolute level. At the two higher flowrates, the peak value of the first resonance is

underestimated by 4 to 6 dB; this could indicate a small interaction effect. We will consider this effect in more detail shortly.

We now turn to the 12 cm front cavity case. The predicted-measured comparison for this case is shown in Fig. 3.23 for a single flowrate of 250 cc/sec. In this case it is clear that the predicted resonant frequencies are generally too high. The causes for this discrepancy remain unclear. The approximation to the radiation impedance should be least accurate around 4300 Hz, but it is not clear that the use of additional terms in the series would tend to shift frequencies downward. If l_f were actually 12.5 cm, the resonant frequencies would be perfectly predicted, but such a large error in the length of the front cavity is highly unlikely. The abrupt change of area in the tube in the vicinity of the obstacle produces non-planar wave fronts which are not modeled by the one-dimensional transmission line. It may be that the frequency-lowering effect of the obstacle is actually larger than predicted. Quite possibly several small errors accumulated, producing a significant discrepancy that can not be attributed to any single cause.

In any event, in order to check the predicted spectra for absolute level, the four lowest peak frequencies observed in the measured data were substituted for the derived resonance frequencies, with the bandwidths left at their derived values. This resulted in the graphs shown in Fig. 3.24. Here, the same things can be said as of the 3.2 cm data: the zeros are not nearly as sharp as predicted, the high frequency behavior is not accurate, but the overall shape and level are quite accurate. Note that the predicted bandwidth of the resonance at 4500 Hz is too large: this is definitely the result of the approximation made in solving for the radiation impedance. Again, the highest peak's amplitude is not always matched by the prediction, but the discrepancy in this case – about 2 dB – is too small to consider significant when compared to the confidence limits of the autopower spectrum.

Comparison of the predicted and measured curves tests the source model, the tract model, and the assumption of no source-tract interaction simultaneously. We can investigate the question of source-tract interaction separately by examining, in the context of the three configurations, a parameter that is an attribute of the source only. In this way we can test whether the source remains the same regardless of the presence of a tube enclosing the jet and obstacle.

The source spectra increase in overall level and change shape somewhat as the flow velocity increases. Although the transfer functions are somewhat affected by flow velocity due to changes in the flow resistance of the constriction, these changes are minor compared to those of the source spectrum. Thus, the sound pressure-flow velocity relationship mainly reflects changes in the source. If the source is identical in the three configurations, the pressure-velocity relationship should be also.

As discussed earlier, one way to characterize the change in sound pressure due to flow velocity is by the power exponent n, as defined by $P \propto V^n$, where P = the total sound power generated by a flow of velocity V. The exponent n depends on the type of source: $n = 4$ for a flow monopole, 6 for a flow dipole, and 8 for a flow quadrupole. Recall from Ch. 2 that

$$\frac{n}{2} = \frac{20 \log_{10} p_2 - 20 \log_{10} p_1}{20 \log_{10}(V_2/V_1)}$$

where p_i is the sound pressure at flow velocity V_i. The quantity $n/2$ will be referred to as the pressure exponent henceforth.

The total sound power can be computed by measuring the sound pressure over a broad frequency range at several angles with respect to the source, and integrating the spectrum levels over the surface area of the sphere whose surface intersects the microphone positions.

The procedure used here was modified in two respects: measurements were made at only one angle, 28°, and only over the 0 to 10 kHz range.

For the l_f = 3.2 cm and 12 cm configurations, the directivity pattern was very nearly uniform, so that one angle was representative of the entire 180°. For the l_f = 0 cm case, the sound was quite directional, having a minimum for all frequencies around 90°, and the minimum whose frequency depended on θ. Clearly the total power computation is not accurate unless all minima are included. However, we want a characterization of the source, which need not require a total power computation. All that is necessary is that the variation of sound pressure with angle should be essentially the same for all flows.

For a free jet, the size of the power exponent varies with frequency, taking on its maximum value in the region above the broad spectral peak. Restricting the frequency range of the computation thus means that the higher-frequency contribution to the exponent, which is significant, will be ignored. Therefore the measured values of n will be less than the theoretical values by an unknown amount, and any conclusions about source type become much more tentative. However, as with the restriction on the angle of measurement, this restriction does not affect the comparison between configurations, or our ability to draw conclusions about source-tract interaction.

Spectral levels were calculated for seven different frequency bins as well as the entire spectrum in order to observe general trends with frequency and to be able to relate the exponent to specific features of the spectra. The frequency range of the bins were chosen so that the primary peaks and dips in the spectra would fall within a single bin. The first bin, 500 to 1500 Hz, includes the first resonance for the l_f = 12 cm configuration; the second, 1500 to 2500 Hz, includes the major resonances for both l_f = 3.2 and 12 cm; the fifth, 5000-6500 Hz, includes the zero for all three configurations. Finally, an overall SPL for the range 500 to 10200 Hz was computed.

The pressure exponents were found by doing a least-mean-squares fit of the form

$$20 \log_{10}|p(V)| = a \log_{10} V + b$$

to the data for each frequency bin-configuration combination. The parameter a then corresponds to $20 \cdot n/2$ for that particular bin-configuration combination. The pressure exponents for all combinations are plotted as a function of frequency in Fig. 3.25. Each point in the figure was computed from spectra at ten flowrates, ranging from 160 to 420 cc/sec (2000 to 5330 cm/sec at the constriction exit).

Note that all three curves increase with frequency in a pattern similar to that for free-jet spectra. The exponents for the three cases are quite similar, and fairly close to 3.0 as expected. The curves corresponding to l_f = 3.2 and 12 cm are almost identical except at low frequencies, where the values for l_f = 3.2 are higher. This fact reinforces the observation made of the p_3 versus $\dot{p}_3$ graphs, that the first resonance amplitude increased faster with flowrate than predicted. For l_f = 0 cm, both the shape of the curve and the magnitude are slightly different. At low frequencies, this is apparently due to the slightly lower amplitude of p_0 compared to p_3 or p_{12}, causing some of the ambient noise to be incorporated in the pressure exponent computation for p_0. This does not explain the slight difference in the pressure exponents in the 5 to 6.5 kHz range, where the spectra for all three configurations dip down to the ambient noise level. Nor does it explain the difference between the exponents for l_f = 3.2 and 12 cm at low frequencies.

The slight but inexplicable differences are evidence of source-tract interaction, as we have defined the source and tract. It is possible that a somewhat different description of the source,

for example, including quadrupole sources along with the dipole sources, would prove to be more independent of the tract configuration. Consideration of the no-obstacle case will provide additional data with which to consider this possibility. It *is* clear already that the effect is very small, creating at most a 6 dB error in the prediction of the sound pressure spectrum.

In summary, the plane-wave pressure source is a good model for the sound generated at the obstacle, the transmission line works well with fairly predictable problems, and there appears to be very little source-tract interaction. The source model predicts the correct zero frequency (for nonzero front cavity) which means source location and type are essentially correct. The zero bandwidth is too low in all cases, indicating that the source should be slightly distributed (on the order of 1 or 2 mm) longitudinally.

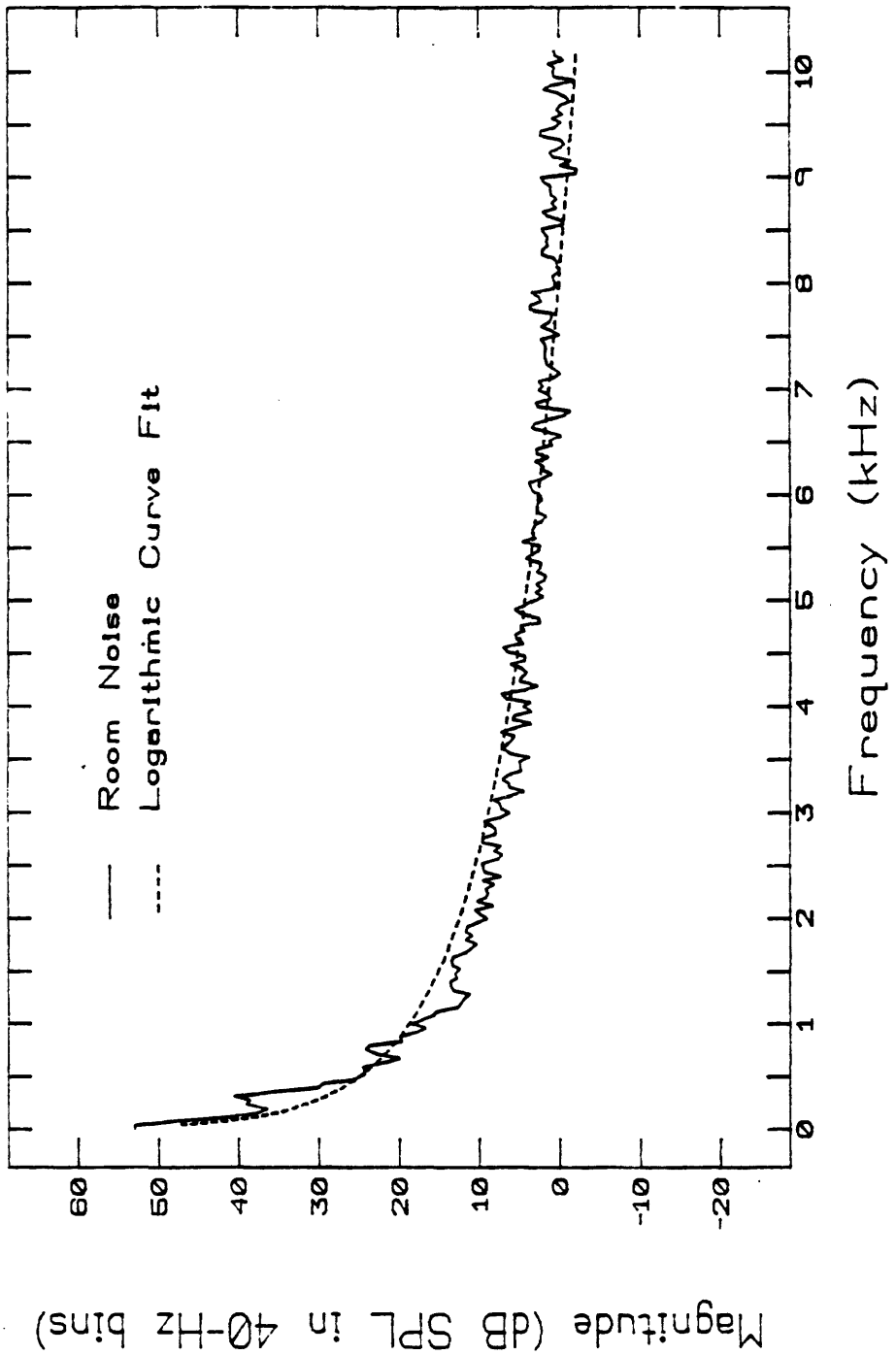


Figure 3.19: Power spectrum of the ambient noise, with the logarithmic curve fit used to remove room noise from all data. Sixteen 25 msec windows were averaged.

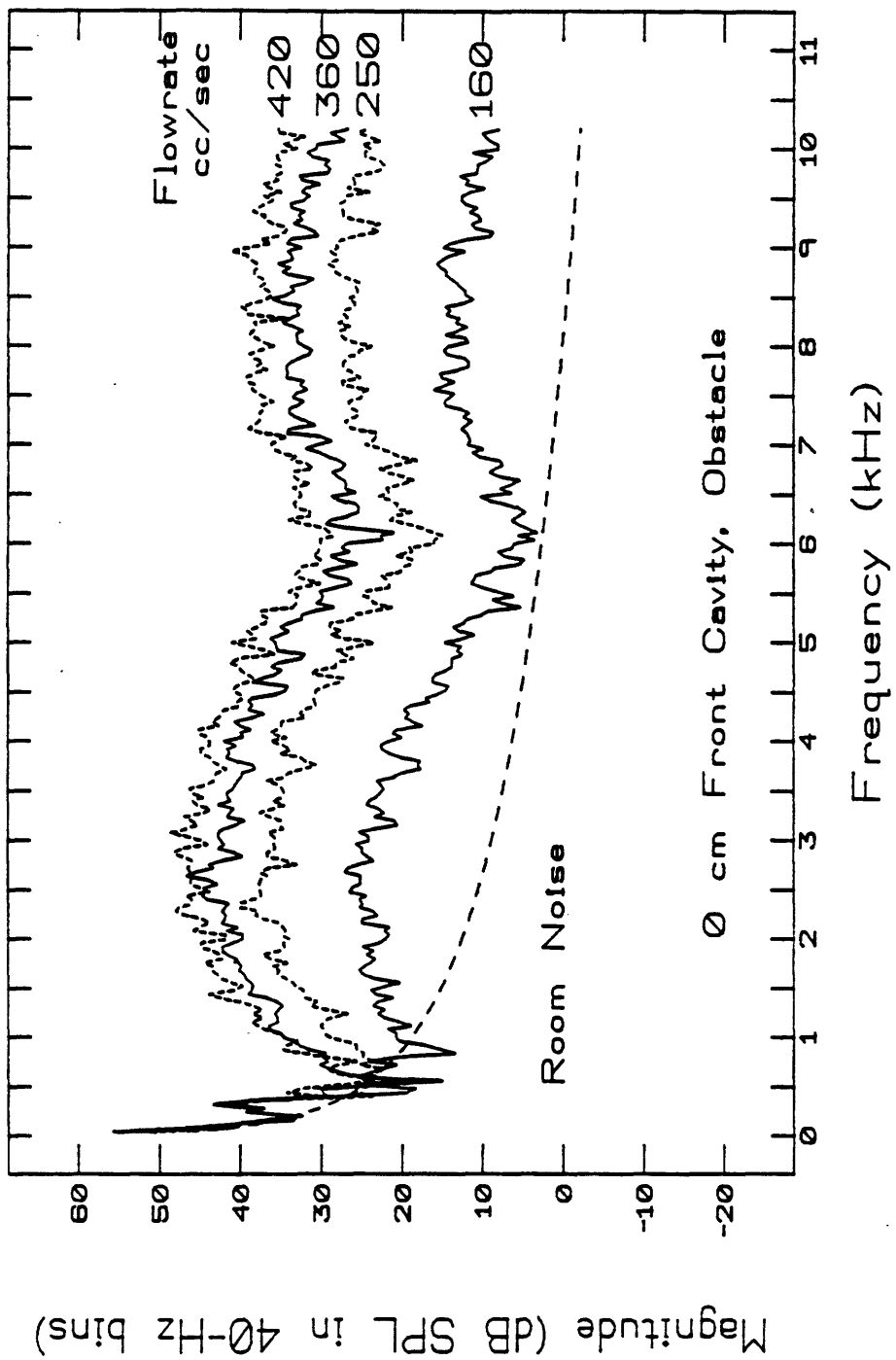


Figure 3.20: Far-field sound pressure p_0 at four flowrates, with regressed room noise. Front cavity length $l_f = 0$ cm; obstacle 3 cm downstream of constriction.

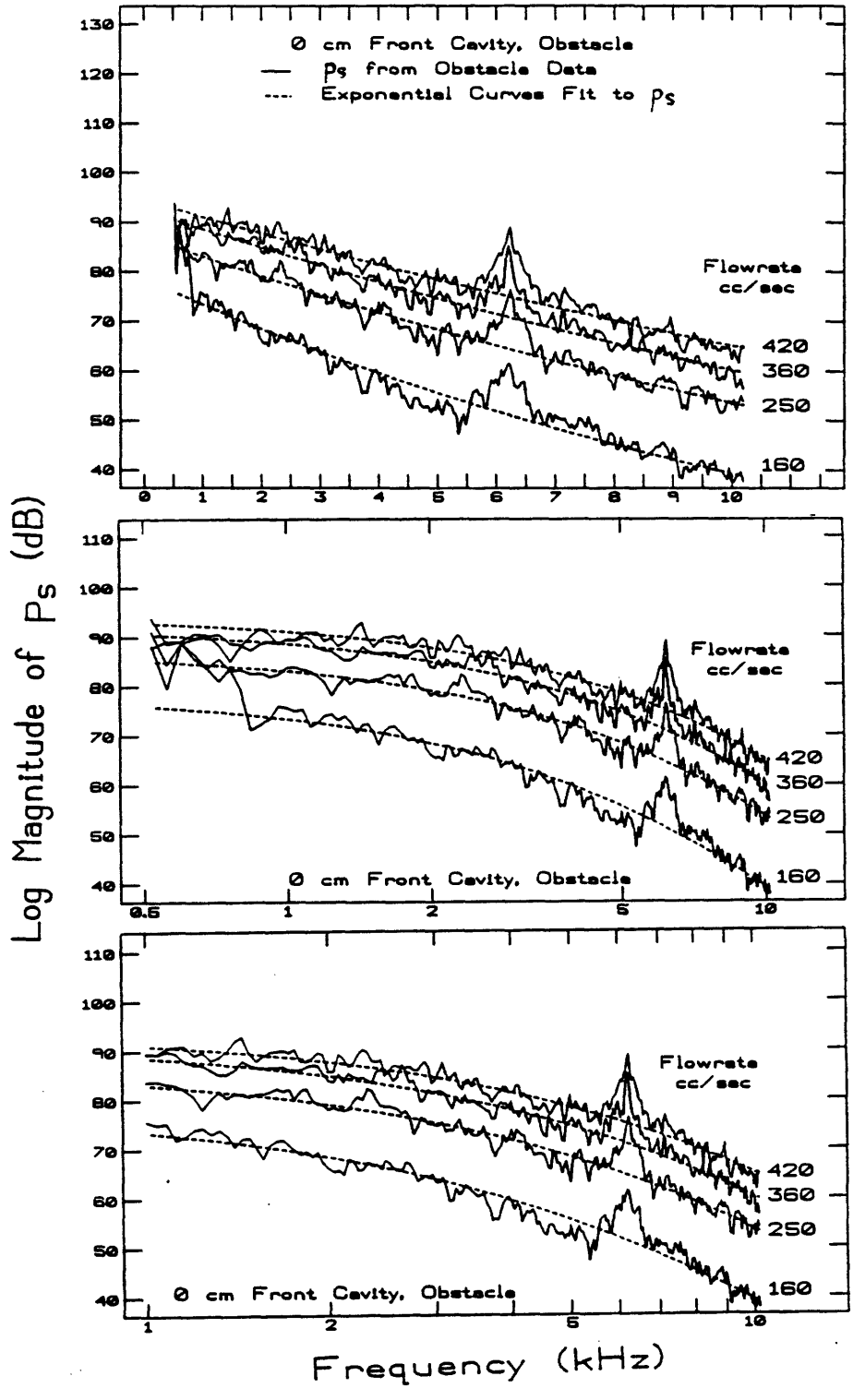


Figure 3.21: a) Pressure source $\hat{p}_s$ (derived from p_0) at four flowrates, on a linear frequency scale.
b),c) $\hat{p}_s$ at four flowrates, log frequency scale, with exponential curve fits shown by dashed lines.

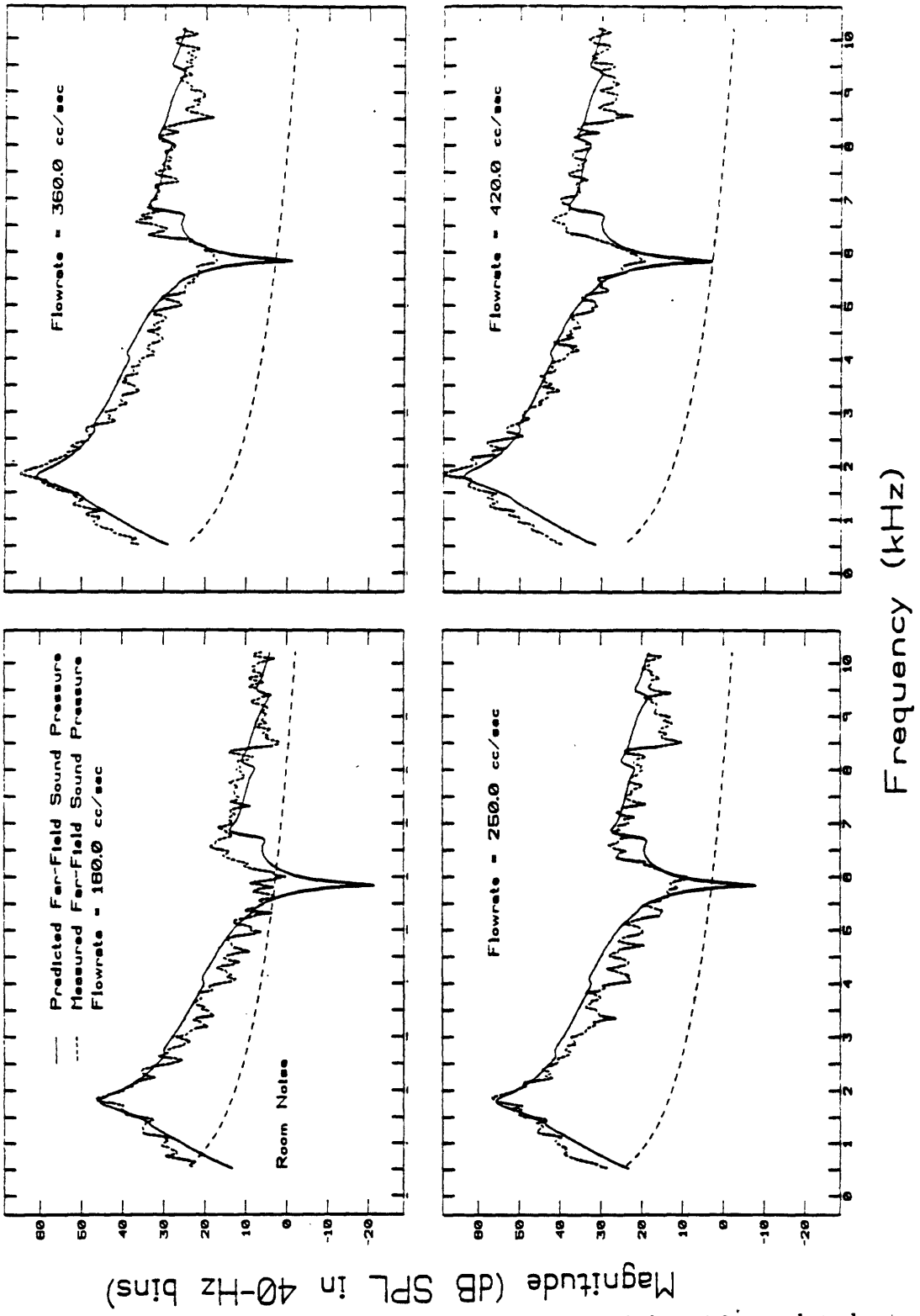


Figure 3.22: Far-field sound pressures for front cavity length $l_f = 3.2$ cm, obstacle at mouth: measured, p_3, versus predicted, $\hat{p}_3$, at four flowrates.

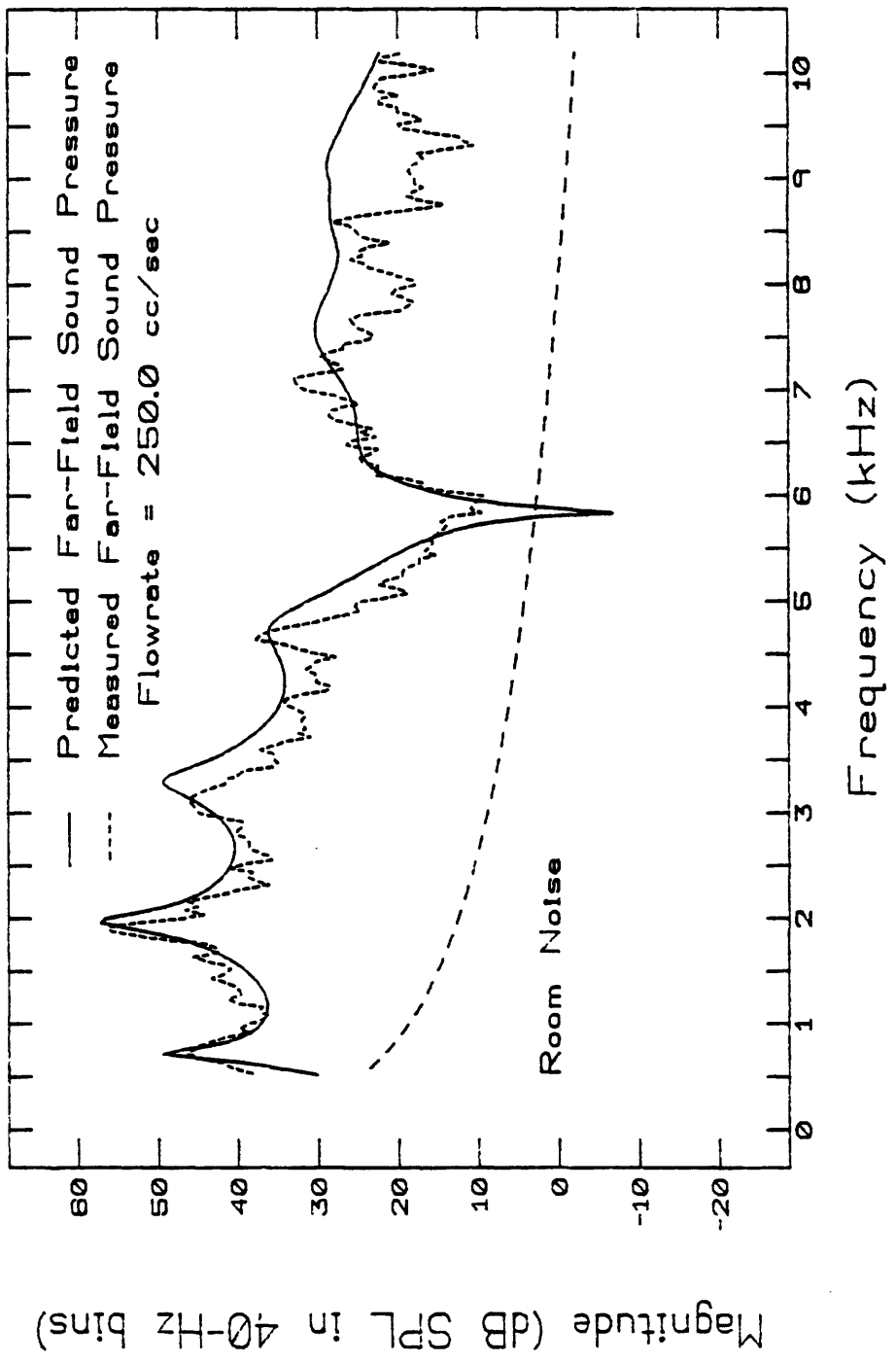


Figure 3.23: Far-field sound pressures for front cavity length $l_f = 12$ cm, obstacle 9 cm from mouth: measured, p_{12}, versus predicted, $\dot{p}_{12}$, at $U = 250$ cc/sec.

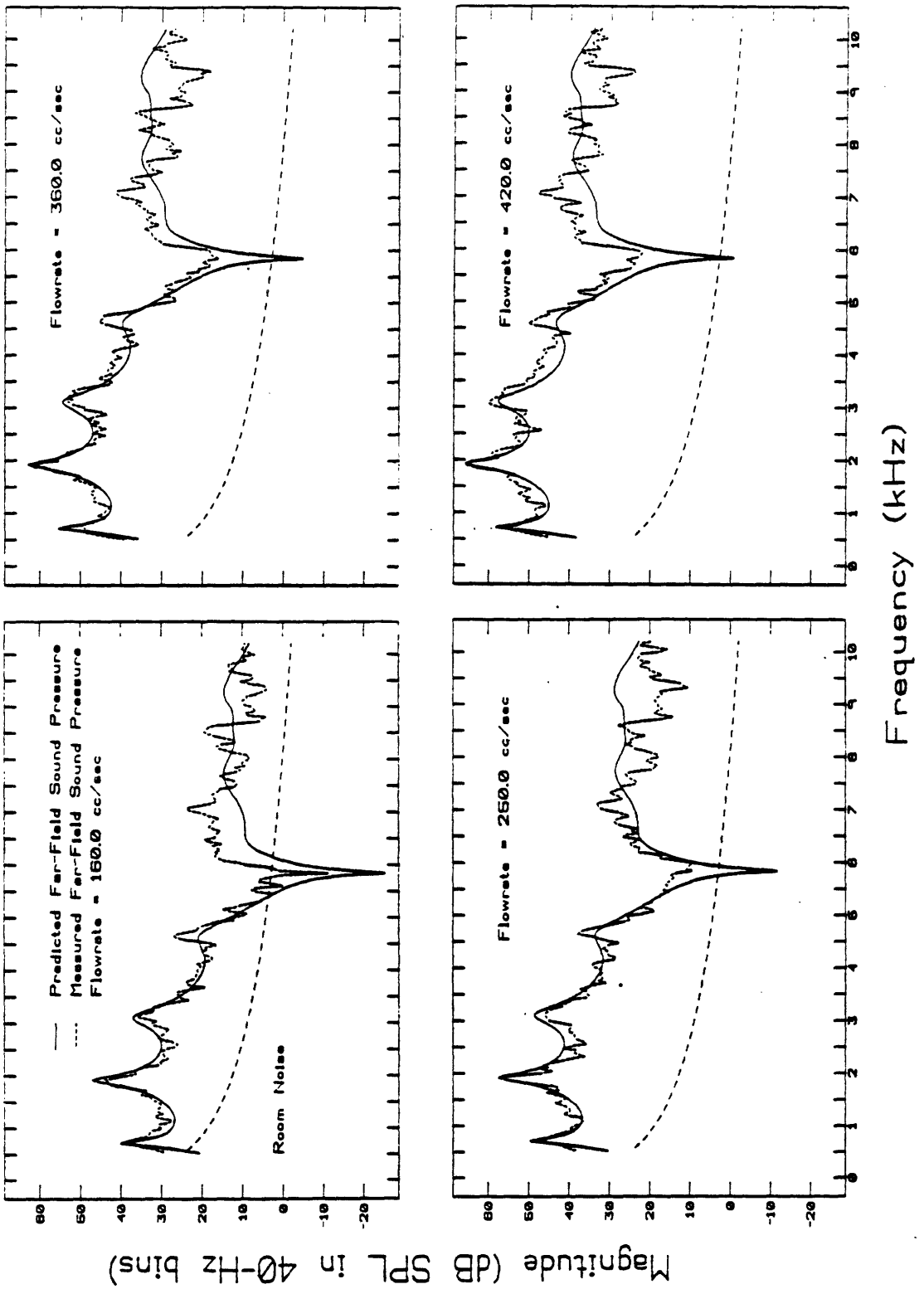


Figure 3.24: Far-field sound pressures for front cavity length $l_f = 12$ cm, obstacle 9 cm from mouth: measured, p_{12}, versus predicted, $\hat{p}_{12}$, with predicted resonance frequencies altered to match data, at four flowrates.

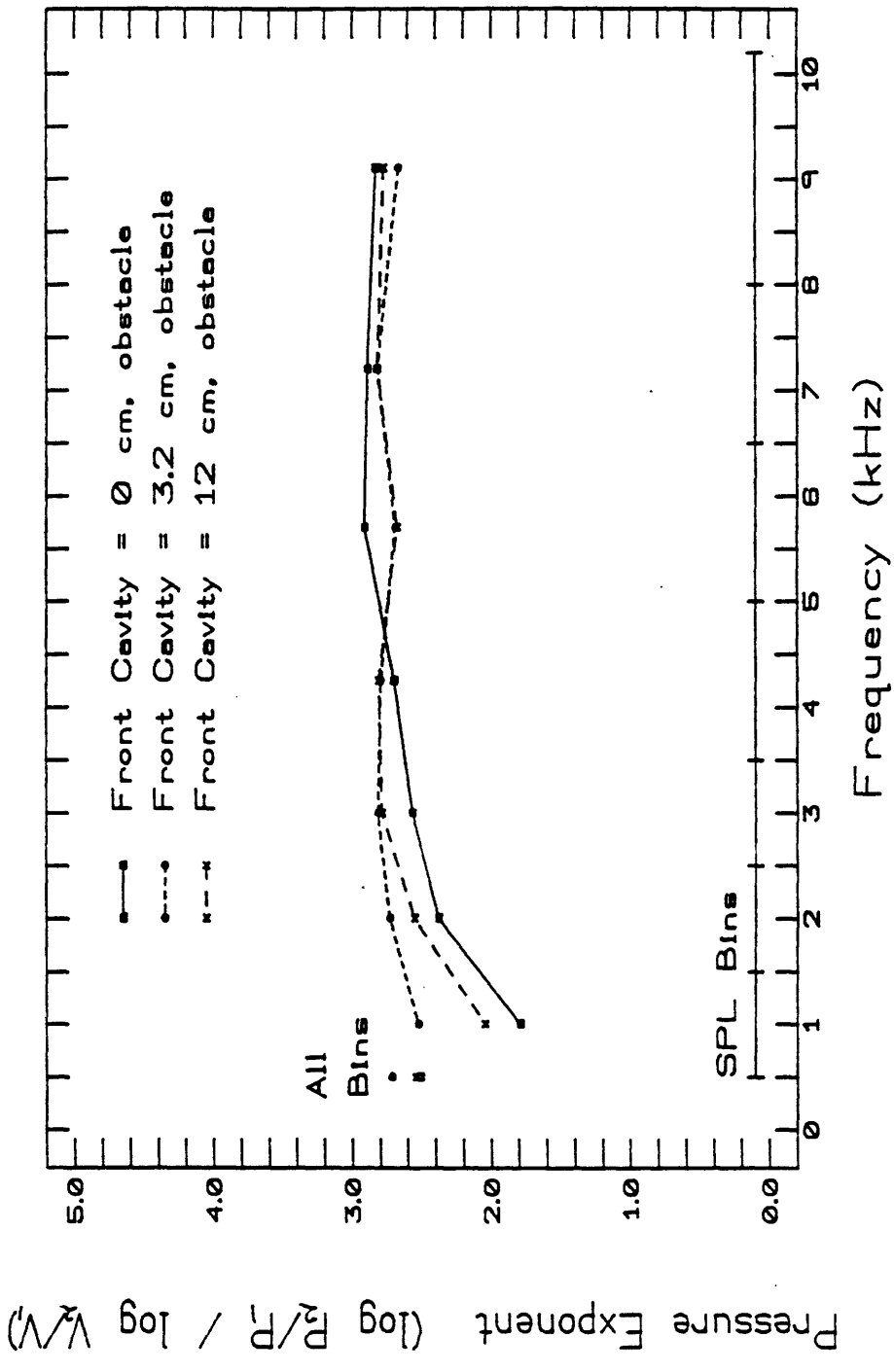


Figure 3.25: Pressure exponents for the three obstacle configurations. Each point represents the change in sound pressure with respect to flow velocity over the frequency range indicated by the markings on the line labeled "SPL Bins." The points labeled "All Bins" are computed over the entire range of 500 to 10200 Hz.

3.2.3 No-Obstacle Case

Figure 3.26 shows the spectra generated at three flowrates for the no-obstacle configurations with $l_f = 0$, 3.2 and 12 cm. Since the sound produced was at such a low level for the $l_f = 0$ cm case, the microphone was positioned 6 rather than 26 cm from the mouth of the tube. This meant that the microphone was in the near field for frequencies below about 5 kHz (as opposed to 1300 Hz for the 26 cm measurements).

Several things are immediately apparent from the graphs. First, for each front-cavity length, much less sound is produced than in the obstacle case. Second, there are no obvious zeros. This includes the region around 0 Hz; a low-frequency zero would cause not only an amplitude decrease there, but would boost all higher-frequency spectral levels. Third, spectra for the $l_f = 12$ cm case are consistently lower in amplitude than the corresponding 3.2 cm spectra.

The allpole filters computed in Section 3.1.6 were used to inverse filter the spectra, p_{3n} and p_{12n}. The results at a single flowrate are shown in Fig. 3.27. The predicted peak frequencies were slightly inaccurate, resulting in the jaggedness at low frequencies. As in the obstacle case for $l_f = 12$ cm, the resonance frequencies (but not bandwidths) were tweaked to match the data, resulting in the graphs shown in Fig. 3.28. If the inverse filters are correct and the sources are identical in the two configurations, we would expect the two curves at each flowrate to be the same. In the mid range of roughly 3500 to 7000 Hz, they do indeed line up. Above 7000 Hz, the discrepancies can be attributed to the inaccuracies of the transmission line model in that range. Below 3500 Hz, the $l_f = 3.2$ cm case consistently has higher amplitude. Once again, it appears that there is more low-frequency energy generated for this configuration.

Figure 3.29 shows the pressure exponents, computed as for the obstacle case, for each configuration. For $l_f = 0$ cm, each point is computed from spectra at eight different flowrates, ranging from 192 to 383 cc/sec (2430 to 4850 cm/sec). For the other two configurations, each point is computed from spectra at seven different flowrates, ranging from 275 to 467 cc/sec (3480 to 5910 cm/sec).

The exponents are not nearly as similar as in the obstacle case. They are also higher at low frequencies, as is expected, since the theoretical overall pressure exponent is higher for quadrupole than dipole sources. As in the obstacle case, the exponent for $l_f = 3.2$ cm is highest at low frequencies.

It seems at first glance that this is due to the troughs of the p_{12n} spectra sinking below the ambient noise level at the lower flowrates. This begs the question, though, since there is no obvious reason why the spectra should be of higher amplitude when the front cavity is shorter. We must conclude that more low-frequency sound is generated when the front cavity is 3.2 cm long than when it is 12 or 0 cm, and that this interaction is heightened when there is no obstacle.

Although we cannot derive source spectra from measurements as we did in the obstacle case, we can say a few things about the source in general. The extreme flatness of the spectra with $l_f = 0$ cm support the expectation of a distributed source. A localized source would show evidence of cancelation due to reflection in the baffle, as was the case when the obstacle was present. With a distributed source, cancelation is induced at a range of frequencies, and there is no obvious trough in the spectrum. Further, if plane-wave sources were to be considered, they would have to be volume-velocity, not pressure sources, due to the lack of a marked minimum near 0 Hz.

The source of sound is noticeably weaker when no obstacle is present, and has a bigger pressure exponent. Both characteristics are consistent with the quadrupole sources that we

would expect in a free jet.

3.2.4 Conclusion

In this chapter we have modeled the sound-production mechanisms for two basic configurations, a constriction in a duct with and without a downstream obstacle. The radiated spectrum produced when the obstacle is present is well-modeled by measurements of sound produced by the obstacle in free space used together with a linear plane-wave model of the tube system. This model could be improved somewhat by distributing the source slightly. A slight degree of source-tract interaction was observed: the presence of the tube around the obstacle increased the amplitude of the sound generated at low frequencies.

The no-obstacle case was not explicitly source-modeled due to the extremely low level of the sound generated. However, inverse filtering by the corresponding allpole transfer functions revealed that there was again source-tract interaction, more than for the obstacle case, and it verified that a source model should not include pressure sources and should be distributed.

The mechanical models used are simplified, idealized representations of configurations actually occurring in speech. In particular, real fricatives are likely to involve constrictions and obstacles which are spatially less distinct, that is, more distributed, than used in the models. We therefore turn now to an analysis of real speech and to results from more complex models designed to test our understanding of the controlling parameters in speech.

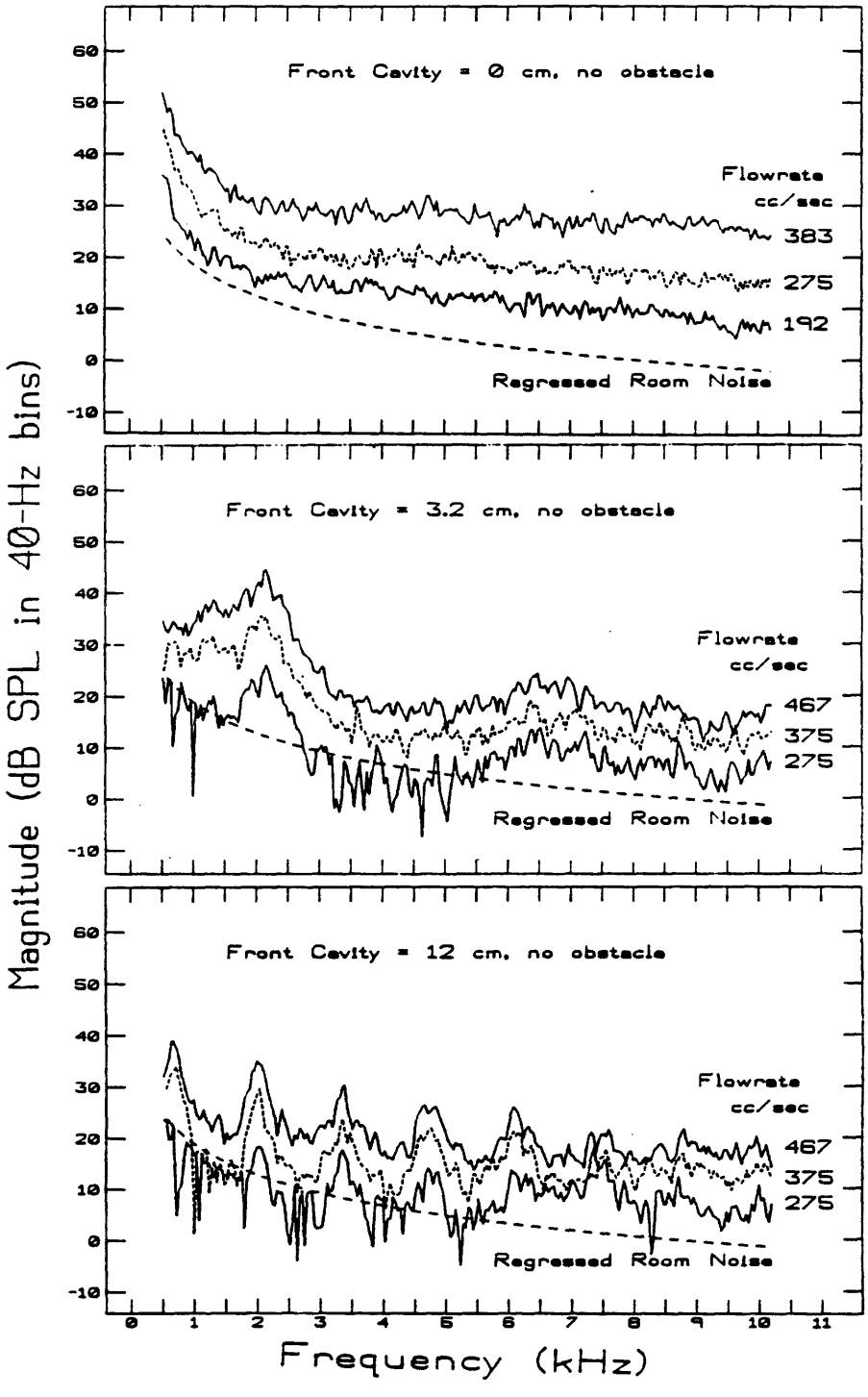


Figure 3.26: Sound pressure, at a distance r from tube mouth, for the no-obstacle cases. $r = 6$ cm for $l_f = 0$, 26 cm for $l_f = 3.2$ and 12 cm.

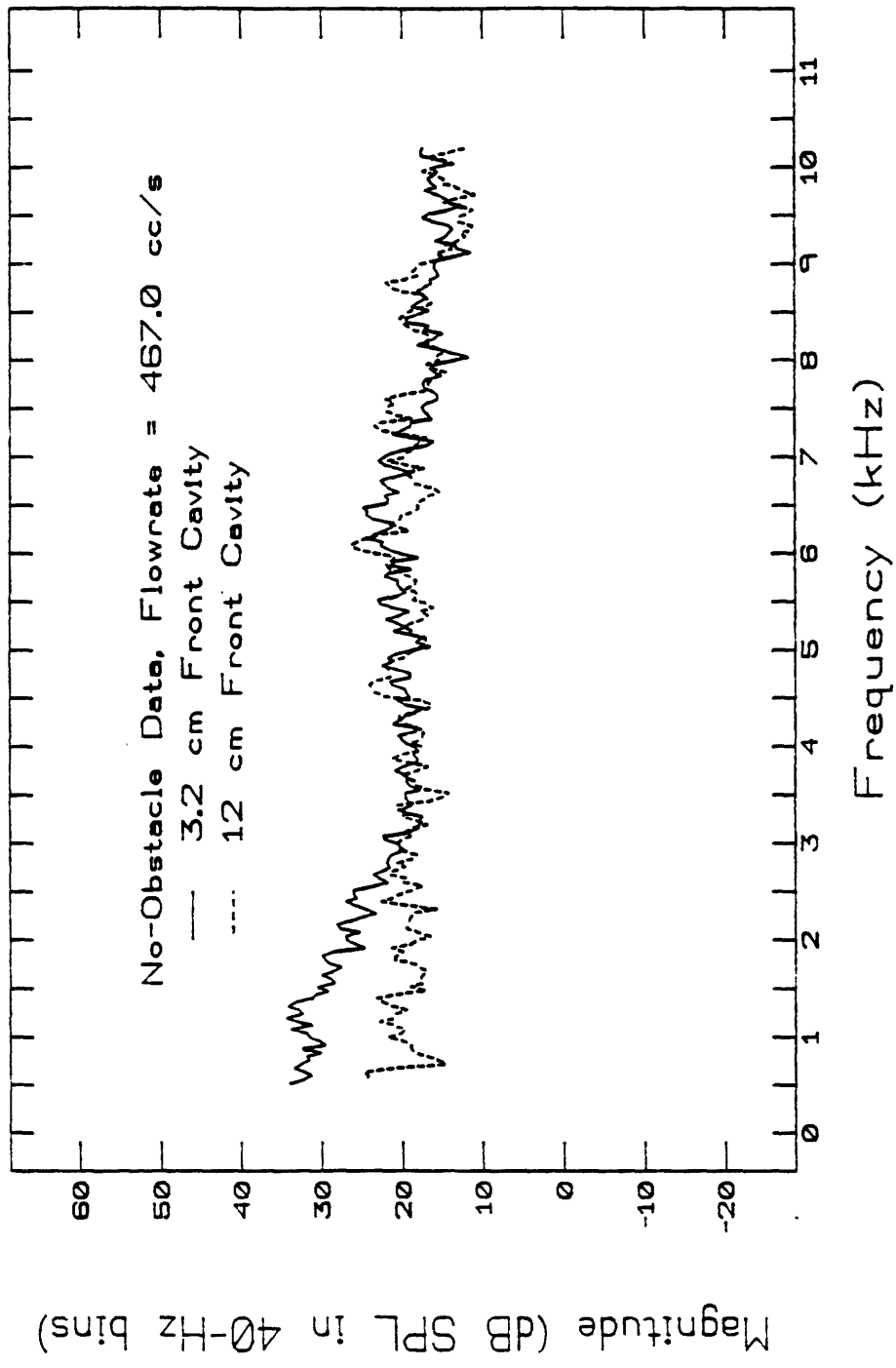


Figure 3.27: No-obstacle spectra for $l_f = 3.2$ and 12 cm, at $U = 467$ cc/sec, inverse filtered to remove poles.

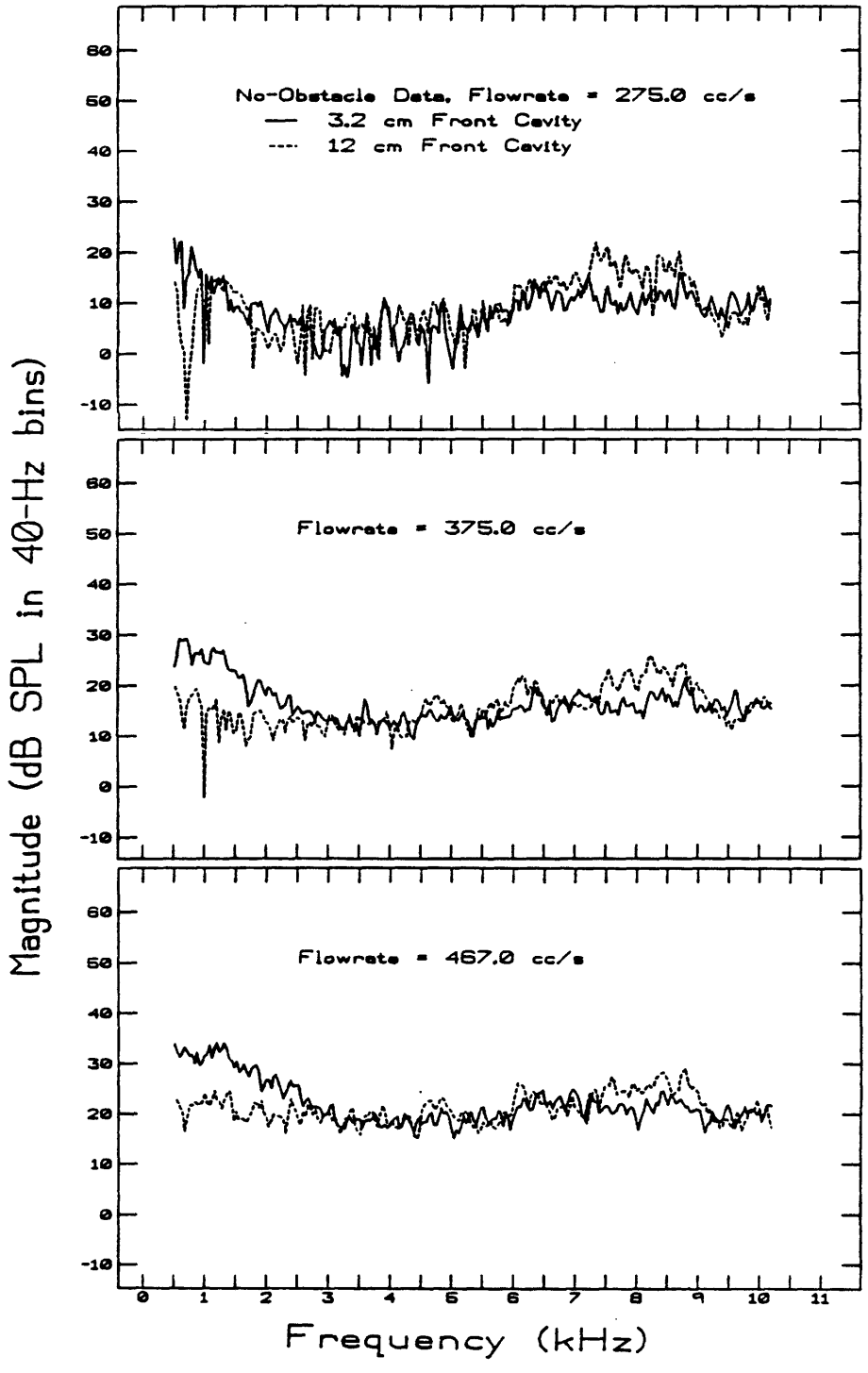


Figure 3.28: No-obstacle spectra for $l_f = 3.2$ and 12 cm, inverse filtered with resonance frequencies of filters adjusted to match data, at three flowrates.

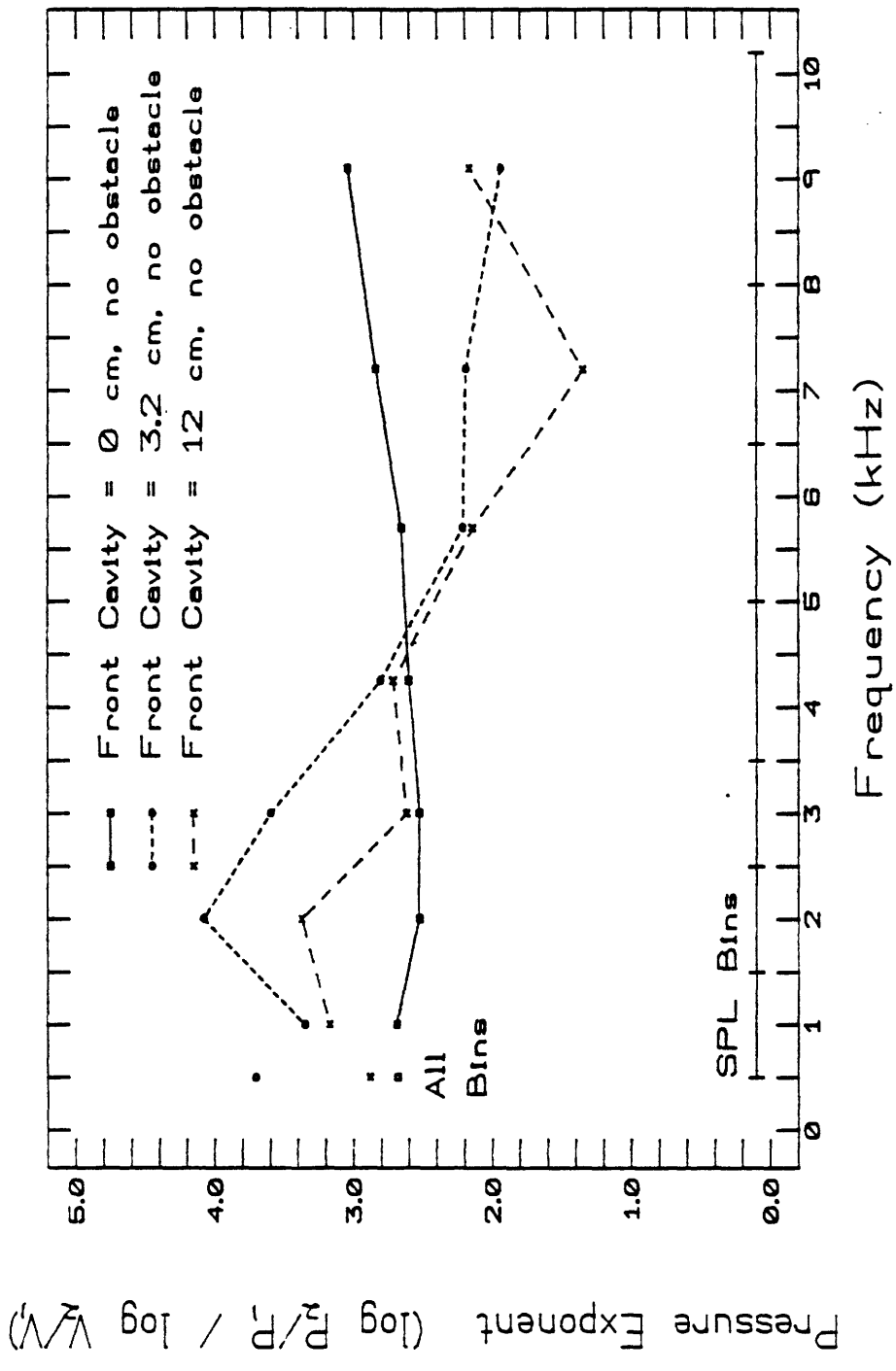


Figure 3.29: Pressure exponents for the no-obstacle configurations.

Chapter 4

Speech and Speech-Like Models

4.1 Speech Analysis

As stated in the beginning of this thesis, fricatives are thought to be produced by turbulence generated in the vicinity of a constriction in the vocal tract, and also, in some cases, at an obstacle downstream of the constriction, such as the teeth. In Chapter 2, experiments with various shape and dimension parameters in a physical model having the dimensions of the average vocal tract established that the presence or absence of an obstacle is the single most significant determiner of the sound produced. Accordingly, in Chapter 3 quantitative models were developed for two configurations differing only in presence or absence of an obstacle. These models therefore capture what we believe are the most significant features of fricative-producing configurations. It is time to check whether this generalization holds.

In this chapter we therefore analyze real speech, and then use those results together with the information gained in previous chapters to experiment with more complex mechanical models. These models, designed to explore the parameters that differentiate the various fricatives, test both the applicability of mechanical model results to speech and our understanding of the acoustic parameters that control fricative production.

4.1.1 Speech Recording Method

Several studies exist reporting extensive analyses of fricatives, as discussed in Chapter 1. Fricatives were analyzed in this study in order to be able to subject the real speech to the same types of analysis procedures as the sounds produced by the mechanical models, so that the two types of data would be directly comparable.

Five speakers were recorded uttering the six sustained fricatives /ϕ, θ, f, s, š, x/ at both normal and intense levels. This particular set of fricatives was chosen because they encompass a significant range of place of articulation, from the lips for /ϕ/ to the velum for /x/. [The phonetic symbols represent, respectively, the consonantal sound in italics in the following words: *wh*ew (approximately), *th*in, *f*in, *s*in, *sh*in, and the German word a*ch*.] Sustained productions were elicited in order to obtain data as similar as possible to the mechanical model data. Normal and intense levels were requested in order to obtain, as nearly as possible, two different flowrates for the same configuration. Undoubtedly the two productions encompass other changes as well.

The speakers (three female: PP, EM, and the author, referred to henceforth as CS; two male: KS and G) were of varying linguistic background. One was from India (G), one from Canada (KS), and the rest were from various parts of the United States. No one subject made all of the fricatives natively, but all were either trained phoneticians or speech researchers familiar with the phonemes. The fricatives recorded were thus not phonetically perfect renditions, but did reflect a systematic place variation for each subject.

The subjects were recorded on a Nakamichi LX-5 cassette tape recorder in a soundproof booth. An Altec microphone was suspended about 20 to 22 cm from the mouth (25 cm for G) and slightly to one side so it wouldn't pick up wind noise. Since the subjects' heads were not restrained, the differences in distance to microphone were not deemed significant. At the

start of the recording session for all subjects except G, two calibration signals were recorded. The signals, a 640 Hz sine wave and white noise, were amplified and played over a loudspeaker. While being recorded they were also measured with a General Radio 1551-C sound level meter which was held so that its microphone was next to the Altec. For subject G, a normal /s/ was used as calibration instead, in a similar manner. This procedure allowed the speakers' recordings to retain absolute intensity level information. Details of the procedure are discussed in Appendix A. The recordings were analyzed by playing them directly into the HP 3582A spectrum analyzer, which averaged eight 25-ms windows to generate a 0 to 10 kHz power spectrum of the input signal (see Ch. 2.1 for a complete discussion of the analysis). This number of averages requires an input signal at least five seconds long.

Phonation in each fricative position was desired in order to obtain a crude estimate of the allpole vocal tract transfer function for that fricative. Therefore, the subjects' instructions were to take a breath, hold the fricative as nearly the same as possible for five seconds or more, and then to freeze their position, draw a breath, and phonate in that position for five or more seconds. In some cases, the desired result was achieved: the flowrate necessary for voicing was low enough that there was no evidence of frication, and the spectrum was, in fact, a good estimate of the allpole transfer function from glottis to lips. In other cases (depending on both the fricative and the subject), even gentle voicing produced audible frication, so the resulting spectrum was that of a voiced fricative, which would presumably incorporate zeros.

The "sound-proof" booth had significant amounts of low-frequency noise due to the close proximity of the air-conditioning system for a computer room. Therefore, overall SPL values for the different tokens were computed using amplitudes between 500 and 10200 Hz, since this excluded the (uninteresting) room noise. The room noise varied over time as fans and compressors switched on and off, and as the recording level changed. Consequently it was analyzed separately for each subject. In order to facilitate comparisons, logarithmic curve-fitting procedures were used on the room noise above 500 Hz in order to smooth out some of the irregularities. The fitted curve for the room noise of the appropriate subject is shown in every graph of speech spectra.

4.1.2 Speech Analysis Method

We know that the fricatives are differentiated from each other articulatorily. We assume that these articulatory differences cause acoustic differences that are observable in the speech spectra. We will characterize the acoustic differences observed between spectra of different fricatives by a set of parameters that will be developed in this section. Since we hope that the acoustic effects of the articulatory differences will be related in some way to the parameters of the models we have considered to this point, we choose a set of parameters with which to describe the speech spectra that are relevant in terms of the models but are also related to parameters used by previous fricative analysis studies. The current data, supplemented by the previous studies, will allow us to establish a normal range of variation of each of the parameters, which will later provide a basis with which to judge the success of the mechanical models developed to imitate each of the fricatives.

The main studies to which these results will be compared are those of Jassem (1962), Strevens (1960), and Hughes and Halle (1956). Jassem recorded native speakers of American English, Swedish and Polish uttering the fricatives of their respective languages in consonant-vowel contexts. Both Swedish and Polish include the fricatives /ç, x/. The fricatives were analyzed by a spectrum section analyzer over the 0 to 9 kHz frequency range. Strevens examined

nine fricatives, including the six studied here, by making spectrograms of thirteen subjects uttering sustained unvoiced fricatives in isolation. He then used the spectrograms to identify regions of significant energy. Since the original spectrograms are not shown, his classifications provide less information than we would wish. However, his is the only study that includes the fricative /ϕ/. Hughes and Halle studied the four English fricatives /f, θ, s, š/ and their voiced counterparts in context: speakers read a list of words placing each fricative before and after the major classes of vowels. The recordings were analyzed by passing a 50-msec gated portion out of the center of each fricative through a wave analyzer and computing the energy density spectra. The spectra, composed of typically 25 samples spaced over the 0 to 10 kHz range, are shown separately for each speaker.

The parameters used by these studies have been found to be of use in distinguishing the fricatives both acoustically and perceptually, and, where the original spectra are not shown, provide our only basis of comparison with their data. We have an additional aim in this study of providing a basis for comparison between speech and mechanical model spectra, and for this purpose two additional parameters are here introduced.

Let us first consider the articulatory differences between the six fricatives. Figure 4.1 shows the midsagittal cross-section of the vocal tract during production of each of the six fricatives. The differences in the place of articulation will change the resonant frequencies. Varying the place of articulation will also change the resonance amplitudes, due both to shifts of the resonance frequencies relative to any anti-resonances and to the effect of the front-cavity length on the bandwidths of the front-cavity resonances. These differences can be captured by measuring the frequencies and amplitudes of spectral peaks, and the overall amplitude.

The source generation mechanism is likely to differ among fricatives, as well. Only /s, š/ have configurations clearly similar to the obstacle case of Ch. 3. /x, ç/ would seem at first glance to behave like the no-obstacle cases, but the path of the air flow along the roof of the mouth, diagrammed in the second part of Fig. 4.1, seems likely to generate some sound, of a nature not yet completely known. For /ϕ, f, θ/, the places of articulation are all in close proximity; the acoustic differences are therefore effected by changes in the shape of the constriction and position of surfaces (such as the lips) that the flow may come in contact with for a short distance upon exiting from the constriction.

The source differences encountered in Chs. 2 and 3 were controlled by the presence and location of an obstacle. Although the actual fricatives appear to involve even greater complexities, parameters that captured the acoustic differences due to an obstacle would clearly contribute to delineation of the fricatives. From Ch. 3 it is clear that an obstacle introduces free zeros into the spectrum, always including one near 0 Hz, and increases the overall amplitude significantly. The free zeros tend to increase the "dynamic range" of the spectrum, that is, the total amplitude excursion of an averaged power spectrum across all frequencies. Two parameters, chosen to measure this dynamic range, are illustrated in Fig. 4.2. They are

A_T = Total dynamic range
= dB between minimum and maximum amplitude of the entire spectrum

A_0 = Low-frequency dynamic range
= dB between low-frequency amplitude and maximum amplitude of the entire spectrum

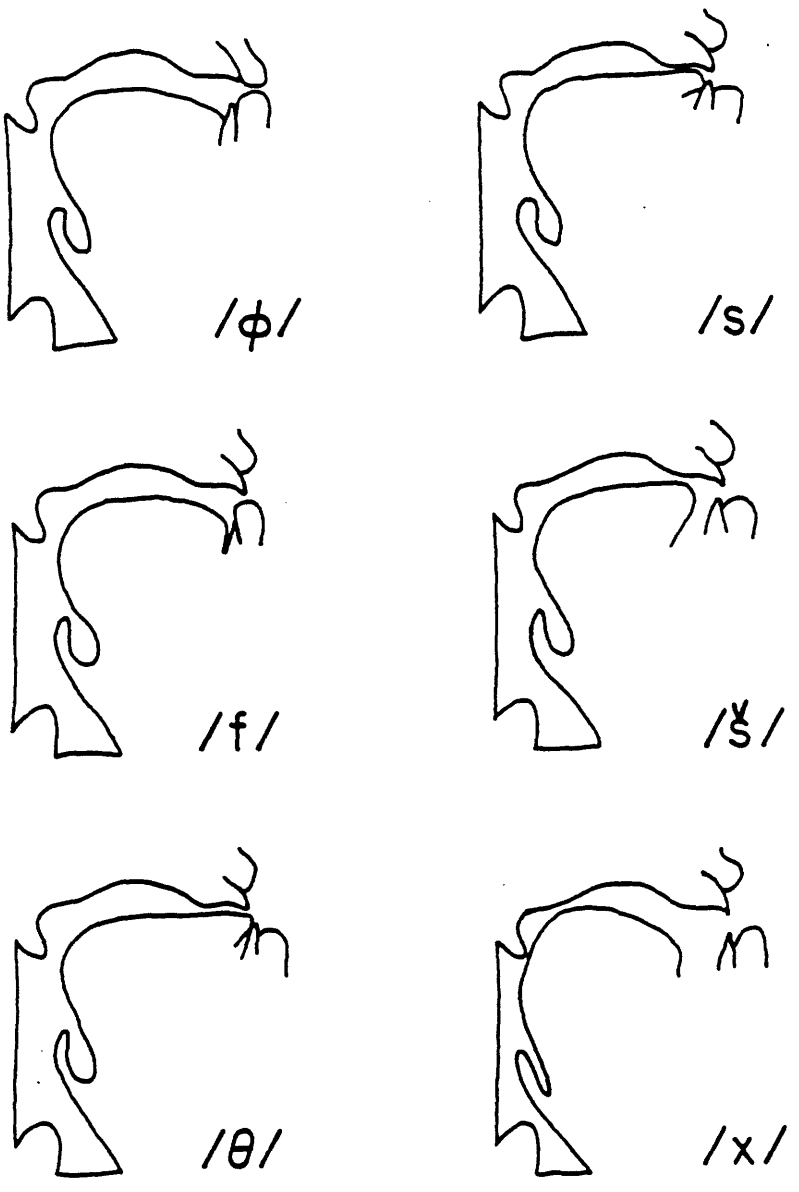


Figure 4.1: Midsagittal view of the vocal tract for the recorded fricatives. After Flanagan (1972), Fant (1960).

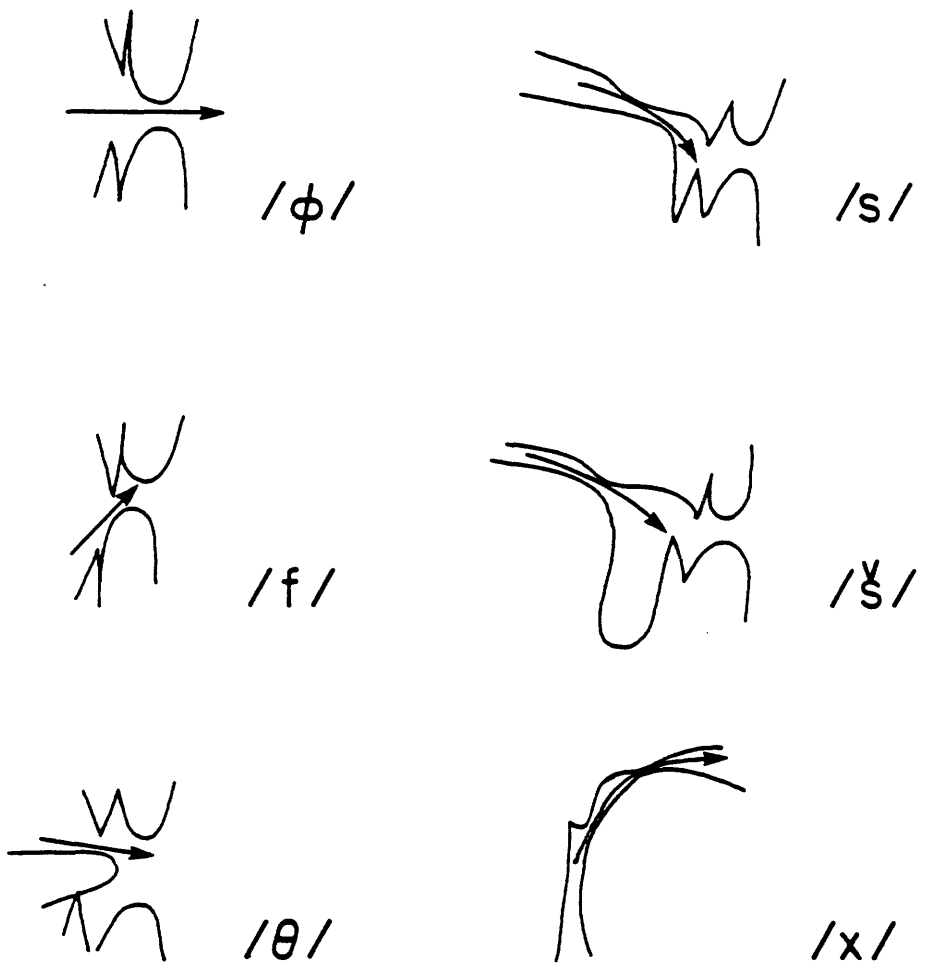
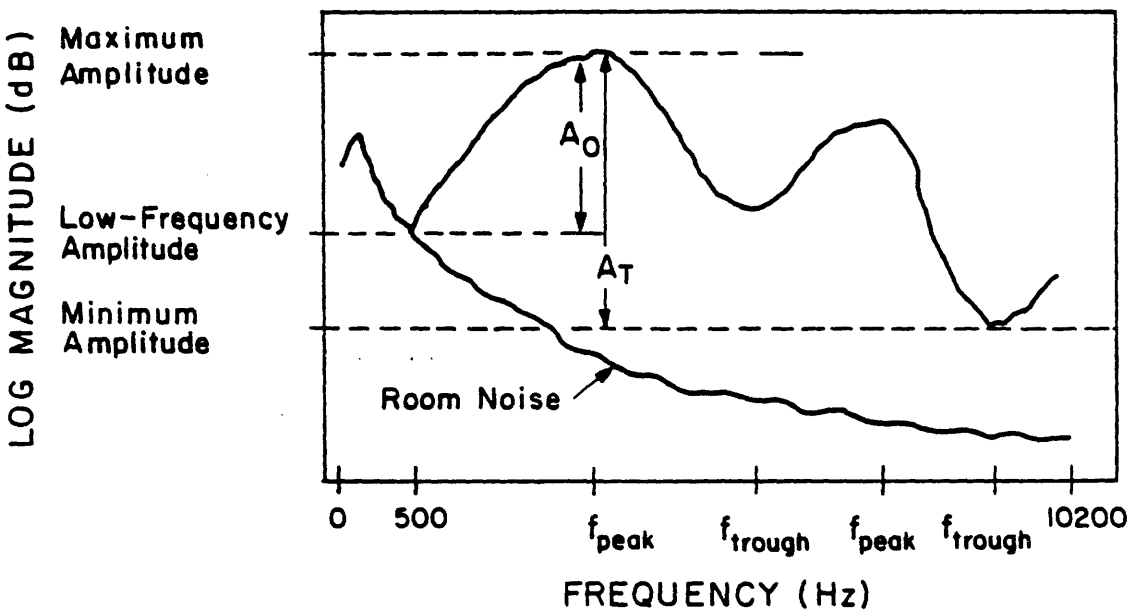


Figure 4.1: (continued) Midsagittal view of the region of the vocal tract in the vicinity of the constriction for the recorded fricatives.



SPECTRAL PARAMETERS

Overall Amplitude, 500-10200 Hz dB SPL

f_{peak} , f_{trough}

A_T

A_0

Figure 4.2: Schematic spectrum, showing dynamic range parameters A_T and A_0.

For both parameters, the relevant amplitudes were measured after the spectrum was smoothed by hand to eliminate local variation. For A_0, we would ideally measure the low frequency amplitude very near 0 Hz; since the room noise obscures the spectrum between 0 and 500 Hz, the amplitude at 500 Hz was used. In cases where the amplitude at even this frequency was identical to the room noise, a higher frequency (but in any case less than 1000 Hz), where the signal clearly rose above room noise, was used.

As shown in Ch. 3, the model for the obstacle case, a pressure source located downstream of a constriction, will always result in a transfer function with a net zero near 0 Hz, and free zeros at harmonically spaced higher frequencies. By a similar argument, a volume-velocity source at the same location will result in a net zero at infinity, and free zeros at different, but still harmonically-spaced, frequencies. For either source type, slight changes in location cause large changes in the frequencies of the harmonically related zeros. Thus, distribution of sources of either type, likely models for the no-obstacle case, can be expected to result in broader, shallower zeros. In terms of the dynamic range parameters, we therefore expect both A_T and A_0 to be larger when there are isolated free zeros within the 0 to 10 kHz range than when there are not, implying a localized source 2 cm or more (for a pressure source; 1 cm for a volume-velocity source) from the constriction. Likewise, we expect A_0 to be larger when there is a zero near 0 Hz, implying a pressure source anywhere downstream of the constriction. We do not expect either parameter to be affected by overall amplitude of the spectrum, except for cases in which the overall amplitude influences the type of source. (For instance, a whistle would produce extremely large values of both A_T and A_0, and is more likely to occur at lower flowrates.) We can first check these expectations and, thus, the validity of the parameters, by obtaining A_T and A_0 values for the spectra of Ch. 3, that is, Figs. 3.22, 3.24 and 3.26. (Values of A_0 have been omitted for the lowest flowrates of the no-obstacle cases, in which the low-frequency portions of the signal are obscured in room noise.) The results are given in Table 4.1; to enable comparisons later on, the overall amplitudes, A_S, are included as well.

The means averaged across all flowrates are indeed higher for the obstacle cases for both A_T and A_0, as expected. The differences between the means were tested for their significance by performing one-tail t-tests, both within and across the two front-cavity lengths l_f. The t-tests showed that the hypotheses

$$
\begin{aligned}
H_1: &\quad A_T(\text{obstacle}, l_f = 3 \text{ cm}) > A_T(\text{no obstacle}, l_f = 3 \text{ cm}) \\
H_2: &\quad A_T(\text{obstacle}, l_f = 12 \text{ cm}) > A_T(\text{no obstacle}, l_f = 12 \text{ cm}) \\
H_3: &\quad A_T(\text{obstacle}) > A_T(\text{no obstacle}) \\
H_4: &\quad A_0(\text{obstacle}, l_f = 3 \text{ cm}) > A_0(\text{no obstacle}, l_f = 3 \text{ cm}) \\
H_5: &\quad A_0(\text{obstacle}, l_f = 12 \text{ cm}) > A_0(\text{no obstacle}, l_f = 12 \text{ cm}) \\
H_6: &\quad A_0(\text{obstacle}) > A_0(\text{no obstacle})
\end{aligned}
$$

are all true with $p < .01$. This means that both A_T and A_0 are significantly higher for the obstacle than the no-obstacle case, for either value of l_f alone or for both considered together. In addition, A_T is consistently larger than A_0 (based on two-tail t-tests, $p < .01$): this may be due to our inability to measure below 500 Hz, rather than to an intrinsic difference between the two parameters. A_T did not vary significantly with l_f. For the obstacle case, A_0 decreases as l_f increases, with a modest significance ($.05 < p < .10$); this is expected since a longer l_f moves the front-cavity resonances down in frequency towards the zero near 0 Hz, decreasing the amplitude differential between the zero and the lowest pole.

	U (cc/sec) =	160	250	360	420	μ	s
$l_f = 3$ cm	A_T (dB)	45	47	47	48	46.8	1.3
Obstacle	A_0 (dB)	25	27	30	28	27.5	2.1
	A_S (dB SPL)	54.9	65.5	73.5	77.2		
$l_f = 12$ cm	A_T (dB)	43	46	43	43	43.8	1.5
Obstacle	A_0 (dB)	13	18	21	20	18.0	3.6
	A_S (dB SPL)	53.6	64.1	70.4	75.4		
	U (cc/sec) =		275	375	470		
$l_f = 3$ cm	A_T (dB)		23	25	30	26.0	3.6
No Obstacle	A_0 (dB)			10	12	11.0	1.4
	A_S (dB SPL)		38.2	48.2	55.6		
$l_f = 12$ cm	A_T (dB)		15	25	25	21.7	5.8
No Obstacle	A_0 (dB)			3	6	4.5	2.1
	A_S (dB SPL)		35.8	43.9	49.2		

Table 4.1: Measures of spectral amplitude, A_S, A_T and A_0, applied to Ch. 3 data. A_S is the overall amplitude, found by summing the squares of the sound pressure spectrum over the range 500 to 10200 Hz. A_T and A_0 are the dynamic range measures described in text. μ = mean (dB), averaged over all flowrates; s = estimated standard deviation in dB.

Presumably the dynamic range would decrease if the constriction-obstacle distance l_o decreased to the point where the first free zero occurred above 10 kHz. That effect is not observable in the data of Ch. 3, but as we shall see, it is important for fricatives with a more forward place of articulation, such as /s/.

In summary, for the model data analyzed above, the parameters A_T and A_0 do detect the differences in source type encountered, and are only slightly affected by (a) the changes in resonance frequency due to a varying l_f, and (b) overall intensity due to variation in flow rate. We will now therefore use these parameters to compare the speech spectra to the obstacle and no-obstacle cases.

4.1.3 Speech Results

Sets of fricative spectra for each of the five subjects are shown in Figs. 4.3 through 4.7. The amplitude parameters A_S, A_T, and A_0, measured for each of these spectra, are given in Table 4.2. Averages of the amplitude parameters across all subjects, for both normal and intense productions, are given separately in Table 4.3. As we discuss each fricative in turn, we will be referring to these figures and tables repeatedly.

/ϕ/

As shown in Fig. 4.1, /ϕ/ is produced by forming a constriction, in the shape of a long narrow slit, at the lips. The spectra all possess a broad peak at low frequencies, centered at about 2 kHz, and decrease sharply at frequencies below the peak. For some subjects (CS, PP, KS, EM's intense version) there is a second lower-amplitude peak at high frequencies, roughly 7 to 9 kHz. The overall level, averaged across all subjects, is 44/52 dB SPL for normal/intense productions. Compared to the values for the other fricatives, as listed in Table 4.2, /ϕ/ is both the weakest fricative and the one with the greatest variation in sound level. The variation may be because /ϕ/ is not in the native language of any of the speakers, or because it is produced with the least and most gradual constriction of the tract. The total amplitude excursion, A_T, is about 24 dB, averaged across all subjects. This is in the range measured for the no-obstacle case, which matches the way in which ϕ is produced (e.g., see model values for no obstacle in Table 4.1). For /ϕ/, A_0 is around 12 dB, or again as we would expect for the no-obstacle case.

For some subjects (CS, KS) there are other significant peaks spaced about 1.5 kHz apart. Comparison with the spectra for the phonated versions (not plotted) shows that these correspond to the first three formants. The pattern remains essentially the same in the intense versions as well, indicating that neither the vocal tract configuration nor the excitation of these formants by the turbulence noise have altered significantly.

For three subjects (KS, CS, PP) the difference between the intense and normal versions is greatest at high frequencies. Since the lower formants do not shift in frequency, it seems likely that the tract shape has not changed substantially, and therefore the energy distribution of the sound produced at the bilabial constriction shifts upwards in frequency as flowrate increases.

Strevens found /ϕ/ to be the weakest of the nine fricatives he tested. He reported that energy was visible on the spectrograms between 1.6 and 6.5 kHz, with three peaks visible in that region. These results are in agreement with the tokens shown here.

/f/

The fricative /f/ is produced by forcing air between upper teeth and lower lip so that it strikes the upper lip before exiting the mouth (see Fig. 4.1). The spectral shapes evident here are similar to those for /ϕ/: two broad peaks, with a trough between them and at low frequencies. Although the overall amplitude of /f/ is low relative to all of the fricatives, it is of higher amplitude than /ϕ/ for every speaker except CS, as seen in Table 4.2. The mean A_T is 21.5 dB, which is comparable to that of /ϕ/ (23.6 dB). The mean A_0 is 12.0 dB, which is nearly identical to that of /ϕ/ (11.9 dB).

Formant-like peaks are again visible in the spectra, but this time they are most noticeable for subject PP and less so for subjects KS and CS. The spectral shape varies a good deal between speakers, but not within a speaker across their own normal and intense versions.

These results are in agreement with those of Strevens, who found energy in the frequency ranges of 1500 to 1700 and 7000 to 7500 Hz. Jassem showed spectra similar to those presented here, with broad peaks and a central trough centered between 4 and 6 kHz. The spectra given by Hughes and Halle exhibit a greater between-speaker variation. Three of the four speakers show two broad peaks of approximately equal amplitude. The fourth speaker has a pattern more like the subjects in this study.

		A_S		A_T		A_0	
		Norm	Int	Norm	Int	Norm	Int
/ϕ/	CS	55.4	66.2	28	24	12	18
	PP	54.2	55.5	26	20	20	13
	EM	27.5	34.0	20	18	0	0
	KS	34.2	52.7	29	38	14	16
	G	49.1	53.1	17	16	15	11
	μ	44.08	52.30	24.0	23.2	12.2	11.6
	s	12.53	11.61	5.24	8.79	7.43	7.02
/f/	CS	47.8	59.6	21	19	16	20
	PP	59.0	58.3	32	19	19	14
	EM	46.5	57.6	21	20	6	4
	KS	33.8	56.7	22	26	10	10
	G	53.9	59.1	19	16	12	9
	μ	48.19	58.26	23.0	20.0	12.6	11.4
	s	9.48	1.16	5.15	3.67	5.08	5.98
/θ/	CS	38.3	52.6	22	24	6	0
	PP	50.5	61.5	20	30	15	18
	EM	47.8	53.1	25	21	9	0
	KS	35.3	51.9	16	14	3	0
	G	51.8	55.9	26	26	16	4
	μ	44.74	58.26	21.8	23.0	9.8	4.4
	s	7.47	3.94	4.02	6.0	5.63	7.80

Table 4.2: Amplitude Measures A_S, A_T and A_0 of Spoken Fricatives /ϕ, f, θ/. A_S is the overall Sound Pressure Level, given in dB SPL computed over the range 500 - 10200 Hz in 40 Hz bins, measured at a distance of approximately 20 cm from speaker's mouth and approximately 30° off axis. A_T is the total dynamic range, and A_0 is the low-frequency dynamic range of the averaged sound pressure spectrum, both measured over 500-10200 Hz, and given in dB. "Norm" and "Int" refer to the normal and intense productions of each fricative. CS, PP, EM are female subjects, KS and G are male subjects. μ = mean in dB or dB SPL, averaged over all subjects; s = estimated standard deviation in dB.

/θ/

The fricative */θ/* is produced by forcing air between the upper teeth and tongue tip (see Fig. 4.1). The spectra tend to be very low amplitude, similar to /f/ and */φ/* in having two broad peaks with significant dips. All speakers show the trough between peaks, though at different frequencies: around 4 kHz for KS, CS, and G, and 6 kHz for EM and PP. Some subjects (CS, KS, EM) do not show a trough at low frequencies. All speakers show additional formant-like peaks, which are most jagged for CS, G. The high-frequency part of the spectrum has significant energy for EM, G, KS.

The mean overall levels place */θ/* as second lowest in amplitude of the six fricatives, and only slightly higher than */φ/*. The mean A_T is 22.4 dB, comparable to that for */φ/* and /f/. The mean A_0 is 7.1 dB, the lowest value of all the fricatives and the one with the greatest variation.

Strevens listed a few more peaks than were visible here, and defined regions of energy that are in agreement with the results shown here. He also noted that */θ/* has the second lowest energy level. Jassem presented only one token for */θ/*, which is nearly flat (A_T = 10 dB) and at a low level relative to the other fricatives.

/s/

/s/ is produced by holding the blade or tip of the tongue near the alveolar ridge, and aiming the air jet thus produced towards the lower teeth (see Fig. 4.1). The spectra analyzed here have in common a general increase in amplitude from low to high frequencies. This is reflected in the mean A_0, which, at 18.4 dB, is well above the value for the previous three fricatives. For subject G, the result is a smoothly rising spectrum with a broad peak at 7 kHz. For the other subjects, formants intrude at different frequencies. All subjects exhibit a relatively high amplitude for /s/. The mean A_T is 20.0 dB, similar to that of the previous three fricatives. The intense /s/ is simply a higher-amplitude version of the normal /s/ except for subject PP, for whom the spectrum of the intense version is much smoother than the normal version and is nearly flat. This atypical token may be a consequence of the difficulty that PP, alone of the subjects, had with sustaining a fricative for a full five seconds.

Strevens gives 3.5 kHz as the lowest frequency at which energy is visible on a spectrogram. PP has a significant peak at 2 kHz, but otherwise the results here are in agreement with Strevens' description. Hughes and Halle's subjects show the same kind of variation evident here: some subjects have a smoothly rising spectrum, while others have peaks scattered throughout. The minimum amplitude does always occur at low frequencies, though, as observed with the present data. Jassem's subjects all exhibit a relatively sharp peak at 4 kHz, but otherwise a continuous rise in amplitude from low to high frequencies.

/š/

/š/ is produced by placing the blade of the tongue against the posterior part of the alveolar ridge (see Fig. 4.1). All subjects show a trough at a low frequency, roughly 1 kHz, a peak or two peaks close together around 2 to 5 kHz, and a broad lower amplitude peak around 8 kHz. Phonation shows that the second, third and fourth formants cluster together to form the major peak. The overall amplitude is the highest of all the fricatives, as is A_0: the mean A_T, at 31.7 dB, is second only to that of /x/.

		A_S		A_T		A_0	
		Norm	Int	Norm	Int	Norm	Int
/s/	CS	58.5	66.2	22	17	22	16
	PP	61.5	66.1	17	15	15	16
	EM	58.6	61.1	18	16	16	12
	KS	50.9	64.1	18	22	16	20
	G	60.2	71.6	28	27	25	26
	μ	57.94	65.82	20.6	19.4	18.8	18.0
	s	4.13	3.84	4.56	5.03	4.44	5.29
/š/	CS	60.5	70.3	36	30	25	22
	PP	56.1	73.2	29	25	23	20
	EM	61.8	67.5	34	24	30	22
	KS	68.3	72.1	37	36	32	36
	G	70.6	79.0	31	35	31	35
	μ	63.46	72.42	33.4	30.0	28.2	27.0
	s	5.92	4.26	3.36	5.52	3.96	7.81
/x/	CS	58.9	64.4	33	29	17	14
	PP	63.8	63.6	50	48	15	15
	EM	59.7	64.2	40	37	20	16
	KS	53.9	64.2	27	30	16	26
	G	60.5	68.5	34	41	22	26
	μ	59.36	64.98	36.8	37.0	18.0	19.4
	s	3.58	1.99	8.7	7.91	2.92	6.07
/ç/	EM	54.7	65.2	36	30	24	27

Table 4.2: (continued). Amplitude Measures A_S, A_T and A_0 of Spoken Fricatives /s, š, x/. See previous page for explanation of parameters. μ = mean value in dB or dB SPL, averaged over all subjects; s = estimated standard deviation (dB).

	A_S		A_T		A_0	
	μ_T	s_T	μ_T	s_T	μ_T	s_T
/ɸ/	48.19	12.18	23.6	6.83	11.9	6.82
/f/	53.23	8.29	21.5	4.50	12.0	5.27
/θ/	49.87	7.80	22.4	4.86	7.1	7.02
/s/	61.88	5.60	20.0	4.57	18.4	4.62
/š/	67.94	6.78	31.7	4.67	27.6	5.87
/x/	62.17	4.03	36.9	7.84	18.7	4.55

Table 4.3: Amplitude Measures A_S, A_T and A_0 of Spoken Fricatives, given in dB or dB SPL computed over the range 500 - 10200 Hz, averaged across normal and intense productions of all subjects. CS, PP, EM are female subjects, KS and G are male subjects. μ_T = mean (dB or dB SPL), averaged across all productions of that fricative; s_T = estimated standard deviation (dB).

The onset and offset frequencies given by Strevens are in rough agreement with the speech analyzed here. Hughes and Halle's spectra are more enlightening: one subject has a single sharp peak at 2 kHz, but the others have a cluster of lower-frequency peaks, and some have broad, lower-amplitude high-frequency peaks. Jassem's subjects have similar characteristics and range of variation. Hence, variation of the same sort as exists here is evident in Hughes and Halle's and Jassem's data.

/x/

/x/ is a dorso-velar fricative, which means that it is produced by bringing the posterior part of the tongue near the velum (see Fig. 4.1). Subject EM produced tokens for /ç/ as well, a dorso-palatal fricative, which is produced with the same part of the tongue held farther forward in the mouth, near the palate. Subject KS produced an /x/ token very similar to EM's /ç/ and dissimilar to the other subjects' /x/ tokens, raising the question of whether there was some confusion as to which fricative was being requested at the time KS was recorded. For KS's /x/ and EM's /ç/, there is one broad peak centered at 3 kHz, a pronounced dip at 6 kHz, and a broad upper peak around 8.5 kHz.

The remaining four subjects, CS, PP, EM and G, exhibit similar /x/ tokens. These consist of three peaks which decrease in amplitude with frequency, described by Strevens as a "formant-like" structure. They occur at different frequencies for the four subjects: lowest of all for G (the only male in this grouping), and next lowest for PP. Dips between the peaks are pronounced at low frequencies and around 6 kHz, with the exception of G, whose spectra do not show the 6 kHz dip.

Overall sound level is the second highest of the six fricatives studied here. The mean A_T, at 36.9 dB, is the highest of the six. The mean A_0, at 18.7 dB, is comparable to that of /s/.

Strevens found /ç/ to be the most intense of nine fricatives, and /x/ to be the fourth most

intense. He found considerable variation in both the sound quality and upper frequency limit of energy distribution of /x/ tokens. Those that sounded like they had a more back place of articulation tended to have a more formant-like structure. That observation is borne out by a comparison of EM's /ç/ and /x/ spectra. Jassem's data likewise show peaks steadily decreasing in amplitude with frequency for /x/, and two peaks of nearly the same amplitude centered at 4 kHz for /ç/.

Summing Up

In spite of the somewhat unnatural task of sustaining an isolated fricative for several seconds, the spectra were similar to those in the literature in terms of their relative sound level, the general spectral shape characteristic for each fricative, and the tremendous variation in spectral shape to be found between speakers. The speech spectra may thus safely be labeled "typical", and may be compared to the mechanical model spectra, both in terms of general spectral shape and level as well as the more detailed parameters A_T and A_0.

The purpose of requesting both normal and intense productions was to obtain a variation in airflow with hopefully all else remaining the same. Subject PP found it difficult to sustain intense fricatives for five seconds; she tended to use lots of air very quickly and then as little as possible for the remainder of the five seconds. Not surprisingly, her fricatives showed very little difference between normal and intense productions. Although the other subjects had no such difficulty, it is quite likely that their pairs do not represent purely a change of air flow rate; in other words, they may well be making other articulatory accommodations, such as changing the constriction area or tensing the articulators, in order to withstand the increased intraoral pressures necessary to perform this task. In view of the similarities, for most speakers and fricatives, between the normal and intense productions, such articulatory changes were not significant acoustically.

The intense version of each fricative was always accompanied by a higher amplitude than the normal version, though that difference was not always evenly distributed across all frequencies. For some fricatives, intense production simply increased the intensity of the entire spectrum, while for others the spectral shape was modified, e.g. high frequencies increased more than low frequencies.

In terms of overall sound pressure level, the fricatives clustered into two groups. The low-amplitude group, consisting of /ϕ, f, θ/, had mean levels that were 10 to 15 dB below those of the high-amplitude group, consisting of /s, š, x/. The absolute levels of these two groups fall within the ranges of the no-obstacle and obstacle values of A_S, respectively. Generally, the farther back the point of articulation, the more the formant structure was revealed, and the higher the amplitude was.

In terms of A_T, the fricatives clustered into two groups. /ϕ, f, θ, s/ had low values of A_T, similar to those of the no-obstacle cases of Ch. 3. /š, x, ç/ had higher values, similar to the obstacle cases. It is somewhat surprising that /s/ should be with the first group, since we expect the teeth to behave as an obstacle. Evidently this occurs because the front cavity is so short as to cause the first free zero to occur above 10 kHz, outside the range included in the calculation of A_T. Certainly for A_0 we see a grouping of /ϕ, f, θ/ and /s, š, x, ç/, with the latter group having the higher values, closer to those measured for the obstacle cases.

By applying the things we learned from the idealized mechanical models in Ch. 3 to the articulatory configurations of the fricatives analyzed here, we have been able to predict many of the observed acoustic differences. However, the simple models do not explain fully why A_T is

lower than we expect for /s/, or why /x/ and /ç/, without an obvious obstacle, have parameter values similar to the obstacle cases. We therefore turn to more complex mechanical models in an effort to address these seeming inconsistencies.

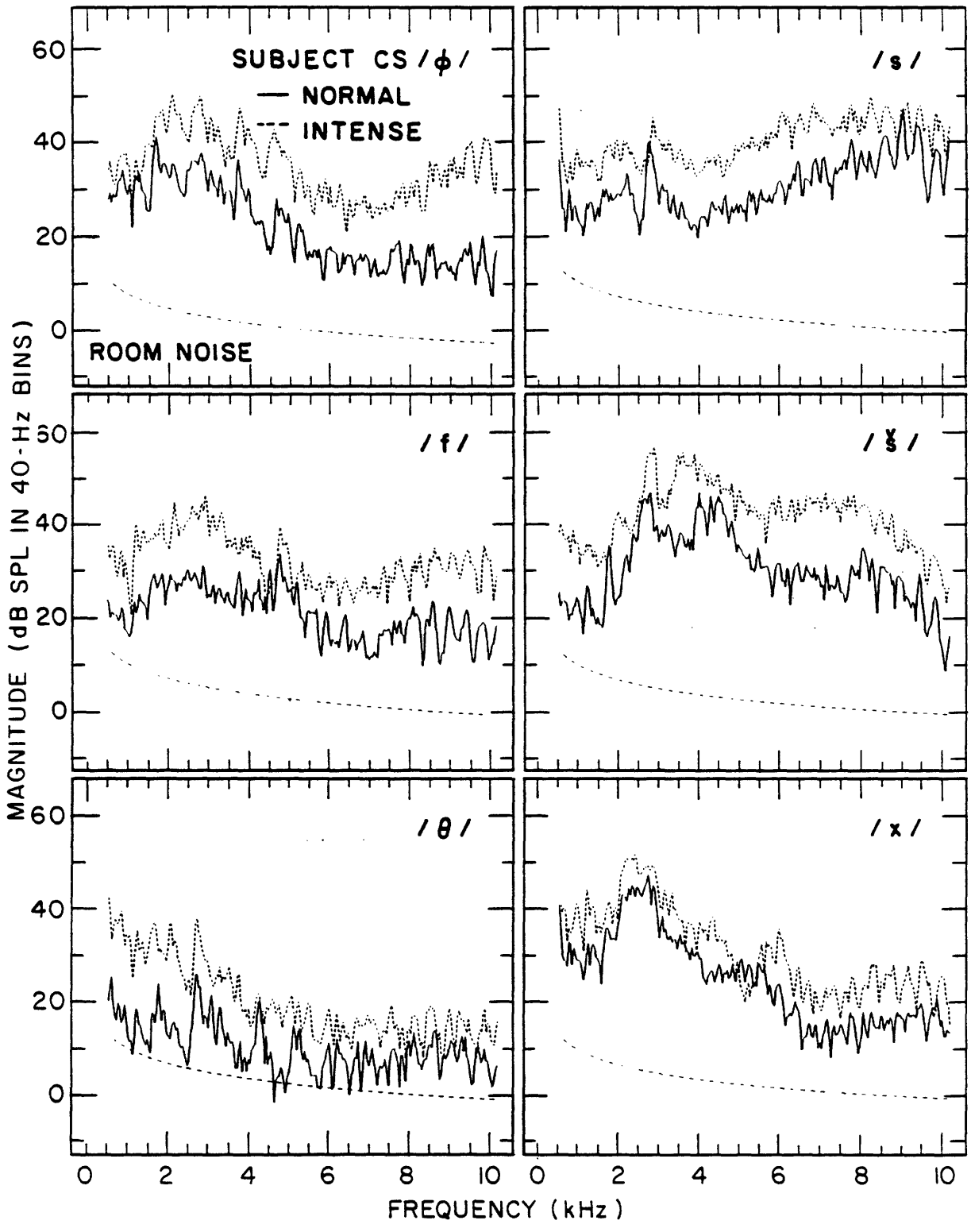


Figure 4.3: Fricative spectra for subject CS (female). Each graph contains the normal (solid line) and intense (dotted line) productions of the fricative shown in that graph, and regressed room noise (dashed line).

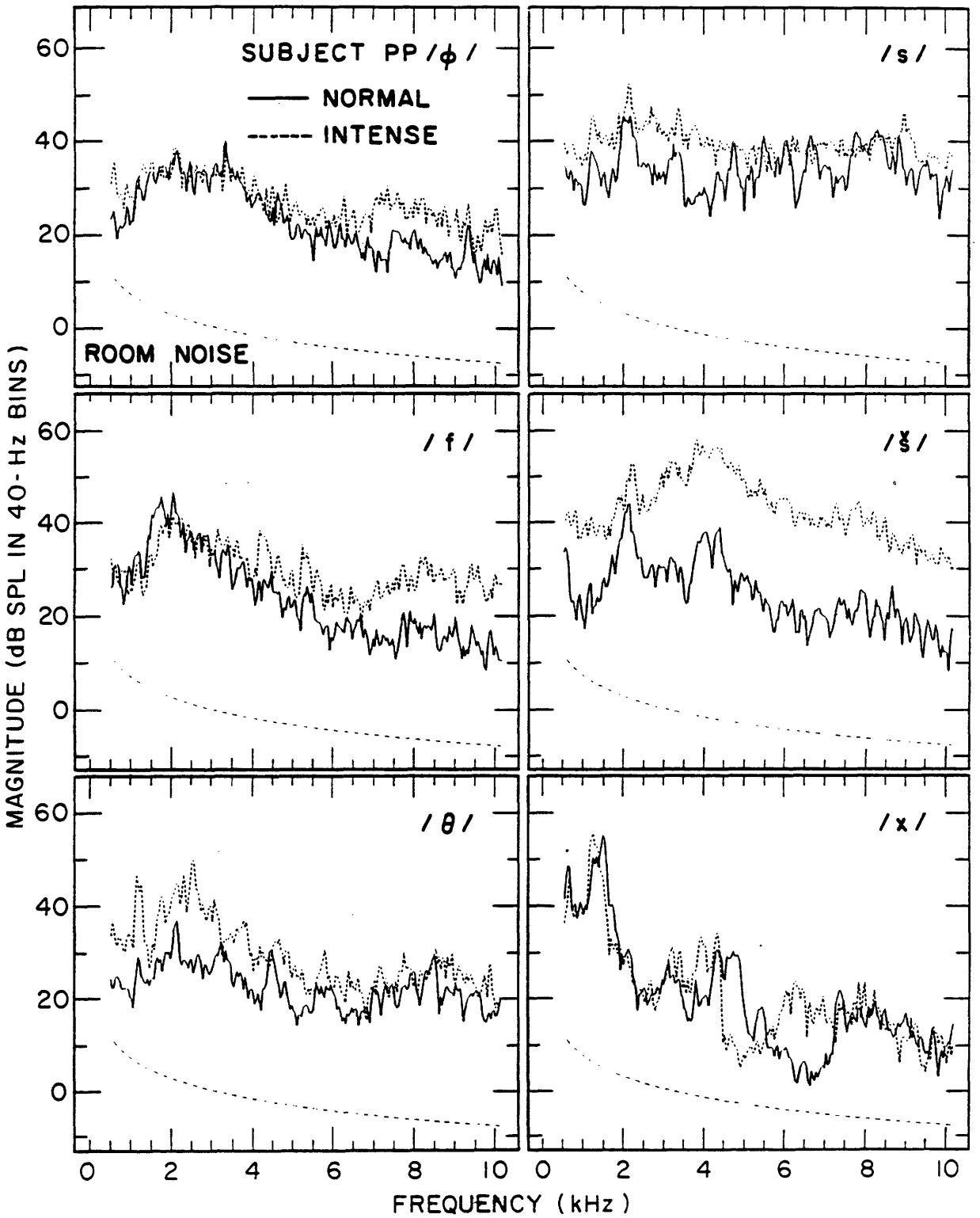


Figure 4.4: Fricative spectra for subject PP (female).

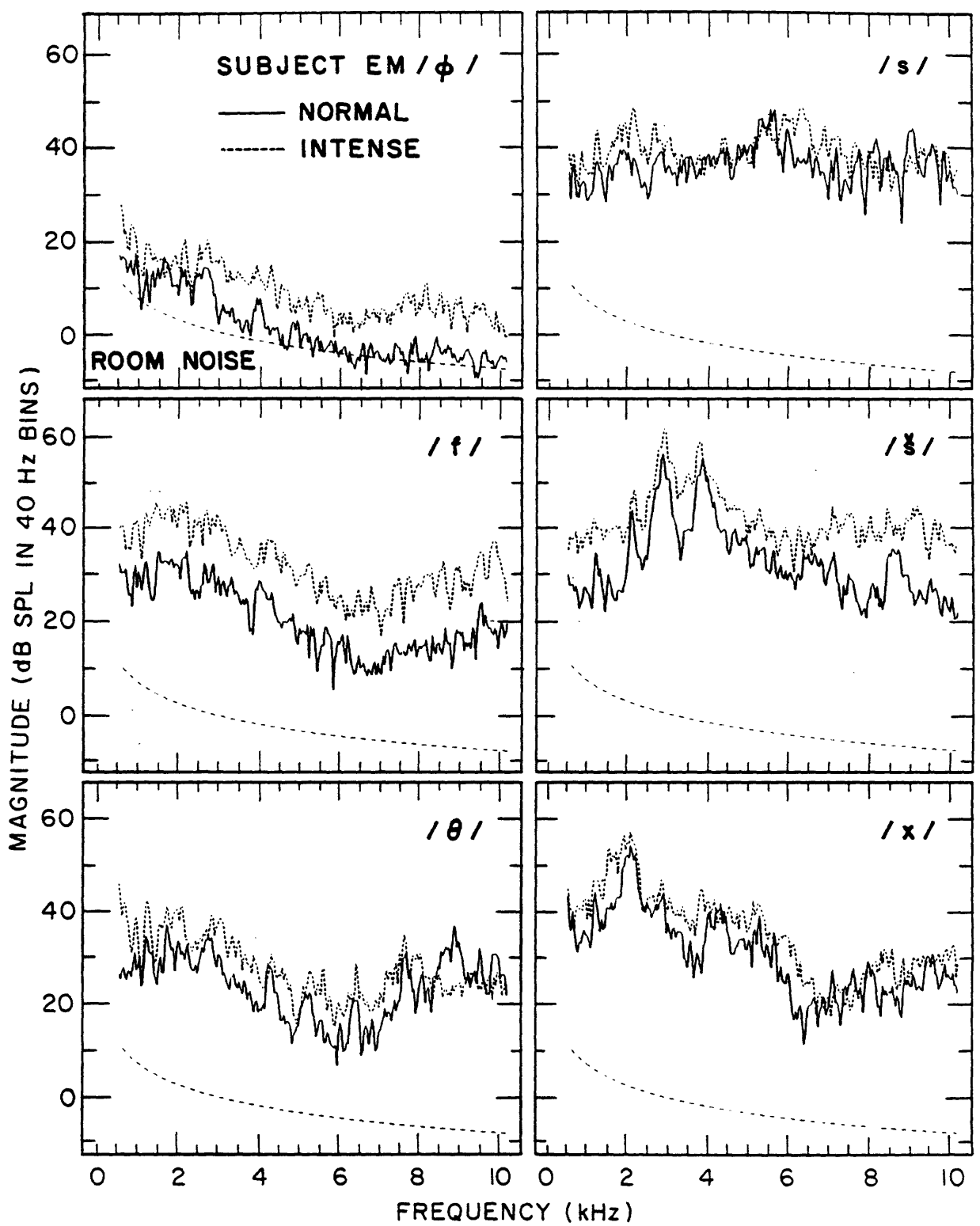


Figure 4.5: Fricative spectra for subject EM (female).

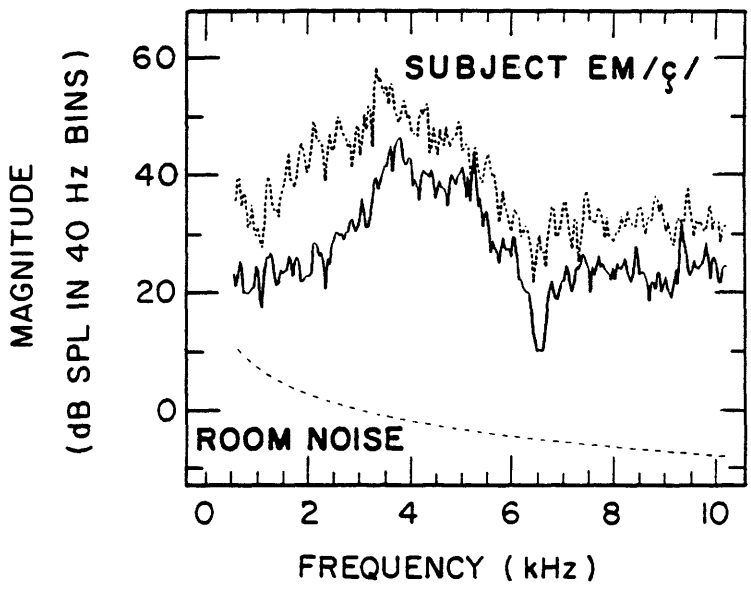


Figure 4.5: (continued) Fricative spectra for subject EM (female).

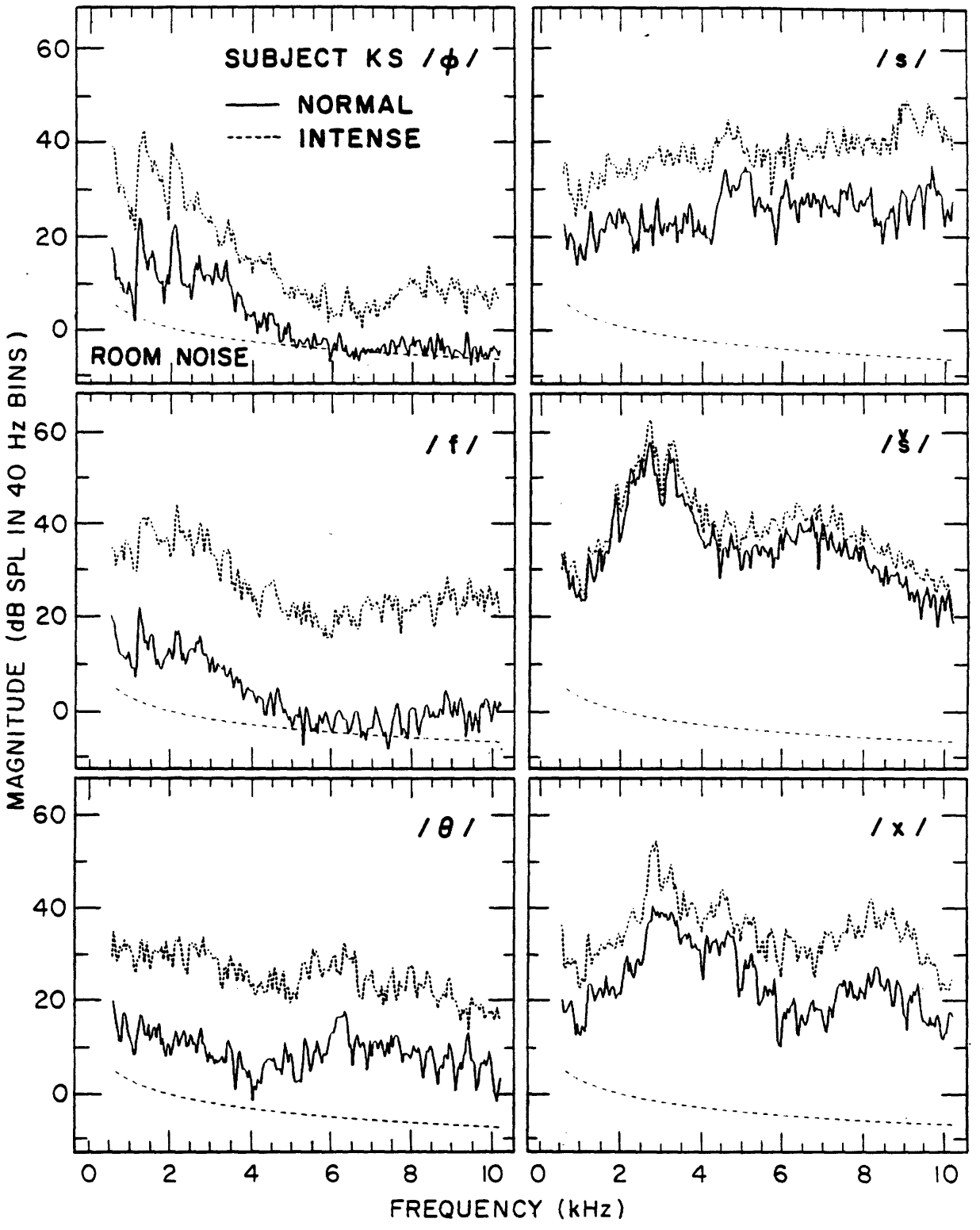


Figure 4.6: Fricative spectra for subject KS (male).

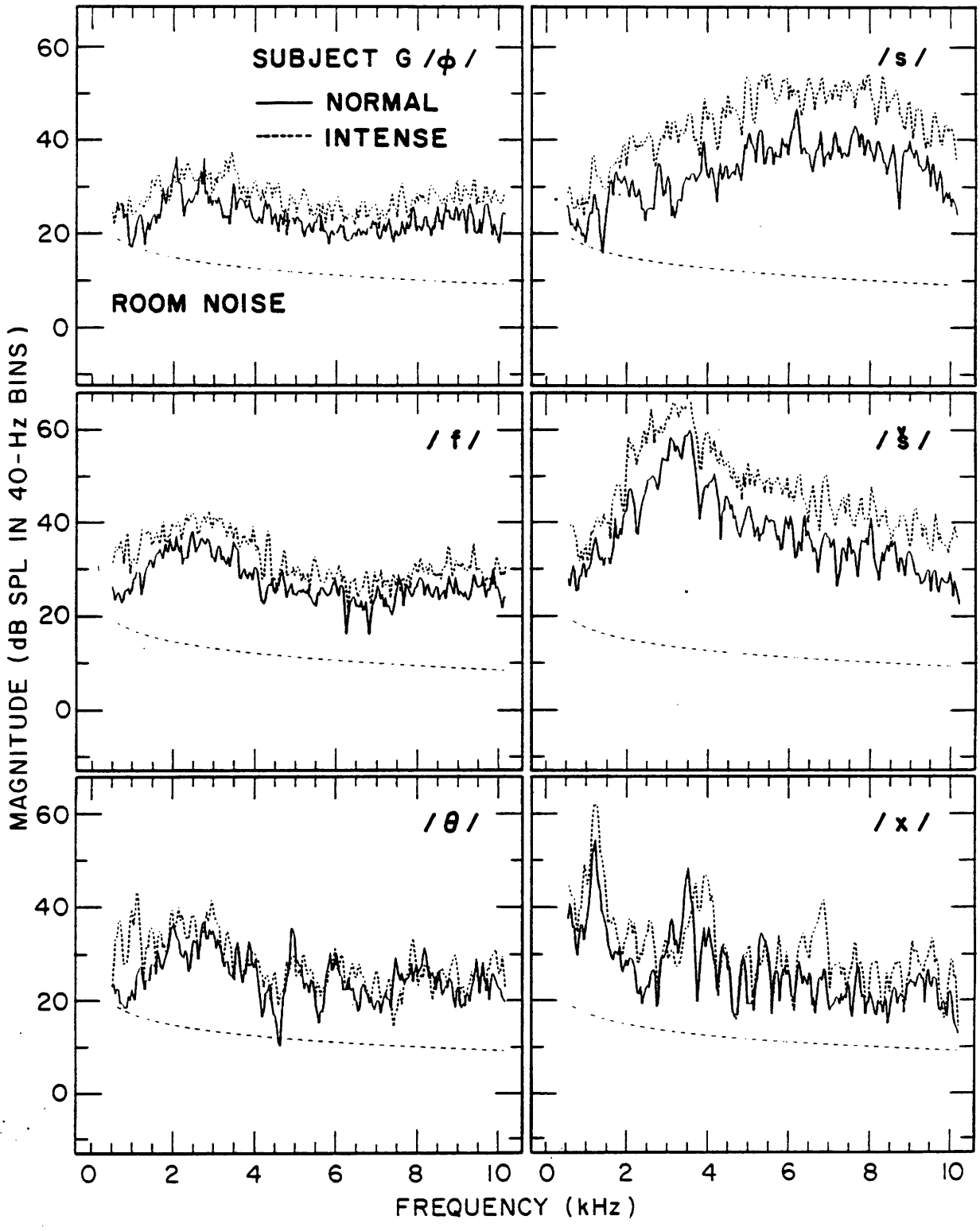


Figure 4.7: Fricative spectra for subject G (male).

4.2 Speech-like Models

Mechanical models were assembled that incorporated the articulatory features believed to differentiate the various fricatives. Due to the extremely idealized geometries and the non-yielding tube walls, it was not expected that these models would produce spectra identical to real speech. The object was rather to imitate (if somewhat crudely) the spectral shape of real fricatives by slight alterations in the models' configurations. A certain level of understanding of the acoustic mechanisms at work could then be achieved, and the results could demonstrate that study of even such simple models is not that far removed from the complex configurations of the vocal tract.

As will be seen in this section, we can refine the obstacle/no-obstacle dichotomy by considering sound generation at surfaces nearly parallel to the jet. Inclusion of this additional mode of sound generation allows us to group models of the fricatives (and, by implication, the fricatives themselves) into three classes: /s, š/, where the predominant sources are generated by an obstacle at right angles to the jet; /ϕ, f, θ/, which have sound generation at a surface nearly parallel to the jet, and a very short front cavity; and /ç, x/, which have sound generation along the wall of the tube, and a long front cavity. Whistles produced by fricative configurations involve an entirely different mechanism, and are therefore considered separately. However, the fricative classes are useful in distinguishing among the whistle mechanisms as well.

The mechanical models discussed here were assembled and measured according to the methods presented in Chapters 2 and 3. For all spectra shown here, the baffle was always mounted flush with the mouth of the tube, and the microphone location relative to the mouth of the tube was always given by $r = 26$ cm, $\theta = 28^{\circ}$ (see Fig. 2.3 for the definitions of these dimensions).

4.2.1 Source due to Obstacle: /s/ and /š/

The X-ray tracings given by Fant (1960) for his subject indicate that the major articulatory difference between /s/ and /š/ lies in the amount of space between the tongue and teeth. For /s/, the tongue tip is just behind the teeth, sometimes resting against the lower teeth, and the tongue blade is held against the alveolar ridge, forming a narrow channel. Only a small cavity remains between the tongue and teeth. For /š/, the tongue tip is raised, forming a slightly wider channel against the posterior part of the alveolar ridge. A significantly larger cavity exists between the tongue and teeth. From the X-ray tracings, it seems possible that the tongue directs the air towards the lower teeth. Catford (1977) showed spectra of the fricatives /θ, s, š/ produced by subjects with and without their false teeth (both upper and lower sets). The spectra, reproduced in Fig. 4.8, show clearly that /s/ and /š/ change drastically, losing much high-frequency energy when the teeth are removed. In contrast, /θ/ changes only slightly. Catford's data show that the teeth contribute substantially to the sound for /s/ and /š/, and the speech spectra measurements of the previous section showed that, in terms of the parameters A_S and A_0, /s/ and /š/ behave most like the idealized obstacle case of Ch. 3 of all the fricatives. The values of A_T grouped /š/, but not /s/, with the obstacle case as well. With some reservations concerning /s/, these two sets of data therefore constitute greater evidence for the assumption made earlier, that the teeth are behaving as an obstacle.

This idea was tested further by constructing models, pictured in Fig. 4.9, which included an obstacle 1.0 cm from the mouth of the tube, representing the teeth. The dimensions of the models represent rough approximations to the X-ray data presented by Fant (1960). For /s/, a circular constriction 1.0 cm in length was located 0.5 cm upstream of the obstacle, with a

plug, semicircular in cross-section, filling in some of the space between the constriction and obstacle. For /š/, the same constriction was located 1.5 cm upstream of the obstacle, and the space between constriction and obstacle was left empty. The effect of the plug and of the constriction location were assessed by examining twelve different configurations, consisting of the four cases of obstacle only, plug only, both, and neither, each at three values of l_o, the constriction-obstacle distance.

Flanagan (1972) showed that if /s/ is modeled by a two-tube model, that is, a back tube of large cross-sectional area, a shorter front tube of length 2.5 cm having the cross-sectional area of the constriction, and a pressure source at the mouth, the predicted transfer function includes a free zero at roughly 3500 Hz, a free pole at 6500 Hz, and bound pole-zero pairs. The spectrum therefore rises steadily from 3500 to 6500 Hz. The /s/ configuration proposed here is a four-tube model, with a pressure source located within a short front cavity. Inclusion of the front cavity introduces front-cavity resonances; for $l_f = 1.5$ cm, these quarter-wavelength resonances will fall at approximately the frequencies $f = nc/4l_e = 3.3, 10.17, \ldots$ kHz, where $l_e = l_f + 8r/3\pi$ is the effective length including the end effect (Morse, 1976). This is more in keeping with the measured speech spectra, which showed a rather prominent formant around 2 to 4 kHz.

Spectra of the sound generated when air flows through the models are shown in Fig. 4.10 for $l_o = 1.5$, 1.0, and 0.5 cm. First, as in Ch. 3, inserting the obstacle always increases the amplitude, whether or not the plug is present. It also induces a minimum at low frequencies, so that even with the difference in amplitude, both obstacle and no-obstacle cases appear to coincide near 0 Hz. A trough at higher frequencies, such as was ascribed to the presence of a free zero in Ch. 3, is only visible for the longer front cavities; when $l_o = 0.5$ cm, the predicted free zero is considerably above 10 kHz.

As the front cavity is lengthened, the lowest resonance of the front cavity decreases in frequency. Since the lengthening has the effect of increasing the cavity volume while keeping the area of the mouth the same, the formant's bandwidth decreases, making the formant more prominent (alternatively, the damping of a resonance due to radiation decreases as the resonance frequency decreases). Over the span of $l_f = 1.5$ to 2.5 cm, these changes are sufficient to change the spectral shape from that of smoothly rising, with one small formant, to one possessing a centralized region of high energy.

For the same reasons, adding the plug increases the bandwidth of the front cavity resonances, but the main effect of the plug is to generate additional sound. For the 0.5 cm plug the effect is small: when the plug is used without the obstacle, the amplitude above 6 kHz increases slightly. When the obstacle is present, the presence or absence of the plug makes virtually no difference in the spectrum. In speech, likewise, the tongue (which the plug represents) may touch the lower teeth or not during production of an /s/. By contrast, the tongue tip does not touch the lower teeth during an /š/; and, indeed, the spectra show that the 1.0 and 1.5 cm plugs used alone generate significantly more sound than the equivalent no-obstacle, no-plug configurations, and affect the sound significantly in the obstacle-plus-plug cases. A reasonable assumption is that the jet has not widened out enough in the first 0.5 cm to generate much sound by impinging on the plug. Although the high-energy regions of the model spectra never become as narrow as that of the speech spectra, it is clear that the model spectra are more similar to /š/ when the plug is absent.

The peak obvious in Fig. 4.10a, the obstacle-only configuration, is the lowest front-cavity resonance; the /s/ tokens of CS, EM and PP show a peak of similar bandwidth and prominence,

l_f cm	Configuration		A_S dB SPL	A_T dB	A_0 dB
1.5	Obstacle		62.8	20	20
	Obstacle + Plug	/s/-like	63.2	22	22
	Plug		47.4	14	10
	Nothing		45.7	18	10
2.0	Obstacle		66.9	27	27
	Obstacle + Plug		66.6	26	26
	Plug		61.3	21	21
	Nothing		49.8	23	13
2.5	Obstacle	/š/-like	72.2	30	30
	Obstacle + Plug		70.2	31	31
	Plug		60.8	32	21
	Nothing		51.4	26	12
	Speech	/s/	61.9	20.0	18.4
		/š/	67.9	31.7	27.6

Table 4.4: Overall Amplitude (A_S), total dynamic range (A_T), and low-frequency dynamic range (A_0) of the spectra shown in Fig. 4.10, at $U = 420$ cc/sec. The codes in the "Configuration" column indicate whether there is an obstacle or a plug in the front cavity. When the obstacle is present, it is 1.0 cm from the mouth of the tube. When the semicircular plug is present, it fits against the constriction, and is of length $(l_f - 1.0)$ cm. The values given for actual speech are repeated from Table 4.3 for ease of comparison.

about 1 kHz lower, and KS shows one about .5 kHz higher. These variations are well within the range expected for different mouth shapes and sizes, and different production strategies.

Values of A_S, A_T, and A_0 are listed in Table 4.4 for each of the spectra shown in Fig. 4.10.

Comparison with Tables 4.2 and 4.3 show that the two models expected to be most /s/- or /š/-like do, in fact, produce spectra whose parameters fall within the first standard deviations of the parameter values measured for speech. In terms of these parameters, then, the models provide very good matches to the articulatory configurations. The flowrate of 420 cc/sec, which affects the overall amplitude, falls within the 300 to 500 cc/sec range typical for fricatives. The absolute level for the /š/-like model is higher than that of the /s/-like model for the same flowrate, which agrees with the relation of /s/ and /š/ in speech spectra also. Since the /s/-like model exhibits a relatively low value of A_T, which agrees well with that observed in speech, we must conclude that our initial reservations about the validity of the obstacle-case model for /s/ were groundless. Instead it appears that A_T is influenced by front cavity length in a way that could not be established with the models of Ch. 3.

4.2.2 Short Front Cavity, Source due to Surface: /ϕ/, /f/, and /θ/

The articulatory differences between /ϕ/, /f/, and /θ/ are more subtle. For /ϕ/ the constriction is in the shape of a wide slit between the lips. For /f/, the upper teeth and lower lip direct the air towards the upper lip. For /θ/, the tongue is held just under or just behind the upper teeth, leaving a wide slit for air to flow through. The slight shift in the place of articulation can be modeled by positioning a constriction at the mouth of a 17-cm tube for /ϕ/ and 1.0 cm from the mouth (according to Fant's X-ray data) for /f/ and /θ/. We know from Ch. 3 that a circular constriction at the mouth will generate a nearly flat spectrum at too low a level to be observable above the background noise at 26 cm. Presumably even a 1.0 cm front cavity will not increase the level substantially. However, these three fricatives, though of low amplitude, are above the background noise. Therefore, we need to mimic the articulation a little more closely.

Figure 4.11 shows the different constriction shapes and combinations used. The centered rectangular slot differs from the centered circular hole of the same area only in the shape of the constriction. The role of the lips and teeth in providing surfaces at which additional sound generation may take place is investigated with two different constrictions, the flat-topped plug where air flows along the wall of the tube, and a two-slot configuration, where a larger rectangular slot is placed downstream of the smaller slot. These are both contrasted with the semicircular obstacle, located 0.5 cm from the centered circular constriction.

Figure 4.12 contrasts the rectangular and circular constrictions positioned 1.0 cm from the mouth of the tube, at flowrates 330, 420 and 520 cc/sec. The two cases are substantially the same: both have a broad peak around 5 kHz, corresponding to the quarter-wavelength resonance of the front cavity (including the end effect), have very low amplitudes overall, and a very small dynamic range ($A_T \simeq 10$ dB). The rectangular slot spectra increase less evenly with flowrate than do the circular spectra. The spectral shape is slightly different, with more low frequency energy generated by the circular constriction.

The spectra change significantly when the jet impinges on a surface in some way. Figure 4.13 contrasts the two-slot, the flat-topped-plug, and the obstacle configurations at the flowrates 190, 330, and 520 cc/sec. All three configurations produce higher amplitude sound than the rectangular or circular constriction alone. The flat-topped plug behaves similarly to the obstacle, increasing A_0 and A_T. The two-slot configuration is of comparable amplitude, but has a greater spectral amplitude at low frequencies. Apparently this configuration does not generate sound in a way that resembles the action of a pressure source, at least at low frequencies.

We could continue with other shapes, more closely approximating the shape of lips, tongue and teeth, but it is already clear that the type of articulatory variation that distinguishes these fricatives produces substantial variation in the spectrum. Further, we have an explanation for why there is so much variation between subjects on these weak fricatives: clearly when there is no strong effect, such as that of an obstacle, very small shifts in constriction shape and location of surfaces near the flow can make a substantial difference in spectral shape.

Considered on an articulatory basis, the flat-topped plug should be closest to /f/, the single or double rectangular slot(s) should model /θ/, and /ϕ/ should be either a slot at the mouth or a thinner two-slot configuration. Comparison with the speech data shows no such clear pattern. Table 4.5 lists the values of the three spectral parameters for the spectra presented here, at $U = 330$ cc/sec. No single configuration has a set of parameter values that fall within the first standard deviation of the mean values measured for the fricatives. Due to the large between-speaker variability for these fricatives, it is possible to find model configurations that match

l_f cm	Configuration		A_S dB SPL	A_T dB	A_0 dB
1.0	Circle + Obst.		58.3	25	25
	FTP		60.4	24	24
	2 Rectangles	/θ/-like	49.4	14	0
	Circle		37.6	5	0
	Rectangle		39.5	11	10
	Speech	/ϕ/	48.2	23.6	11.9
		/f/	53.2	21.5	12.0
		/θ/	49.9	22.4	7.1

Table 4.5: Overall Amplitude (A_S), total dynamic range (A_T), and low-frequency dynamic range (A_0) of the spectra shown in Fig. 4.13, at $U = 330$ cc/sec. The codes in the "Configuration" column indicate whether an obstacle is present in the front cavity of the configuration, and the cross-sectional shape of the constriction. The obstacle, when present, is located 0.5 cm from the mouth of the tube. FTP refers to the "flat-topped plug" constriction. "2 Rectangles" refers to the configuration using two rectangular constrictions at the mouth of the tube. The values given for actual speech are repeated from Table 4.3 for ease of comparison.

individual tokens somewhat better in terms of the parameters (see Table 4.2). For instance, the /θ/ tokens of CS, EM and KS have parameter values that are much closer to those of the "/θ/-like" model than are the values averaged over all subjects' /θ/ tokens.

Catford's data, showing little difference in spectra of /θ/ produced with and without teeth, indicates that the difference between shape and rigidity of teeth and gum is not highly significant. The data here show that the presence of surfaces near the jet is more important than the exact shape of the constriction.

4.2.3 Long Front Cavity, Source due to Surface: /ç/ and /x/

Fant's X-ray tracings show that /x/ is produced by pulling the tongue back against the velum and soft palate, leaving a long (about 8 cm for his subject) and relatively large-volumed front cavity. For /ç/, the tongue is more forward, reducing both length and volume of the front cavity. In both fricatives, the curve of the hard palate towards the alveolar ridge occurs anterior to the constriction. The air is therefore directed along the roof of the mouth, providing a large surface for additional sound generation.

Accordingly, the mechanical models used to mimic /x/ and /ç/, shown in Fig. 4.14, explore the differences between two constriction shapes (a centered circular constriction and a flat-topped plug) and two front cavity lengths (4 and 6 cm). Spectra of the sound generated when air passes through these models are shown in Fig. 4.15; for purposes of comparison, spectra generated when $l_f = 1$ cm are also included. The constriction shape affects the spectrum more when it is closer to the mouth. For $l_f = 6$ cm, the flat-topped plug spectrum is 15 dB higher in amplitude than the centered circular spectrum in the vicinity of the first resonance, but tapers down until it is equal in amplitude above 5500 Hz. The frequencies and bandwidths of the resonances are approximately the same. For $l_f = 4$ cm, the amplitude difference has grown to 20 dB at the first resonance, and extends up to 9000 Hz. The frequencies for the resonances are still similar, but the bandwidths are smaller for the flat-topped plug. For $l_f = 1$ cm, the circular constriction produces a nearly flat spectrum, while the flat-topped plug produces sound that is 25 dB higher in amplitude over most of the frequency range. The spectral shapes are obviously quite different.

The spectra produced by the centered circular constriction decrease in amplitude as the shortening front cavity results in larger losses due to radiation. For all three front cavity lengths, the flat-topped plug spectra have similar peak amplitudes of about 45 to 50 dB SPL. The position of the constriction, which presumably generates sources along the wall, will set up asymmetric wavefronts radiating outward from the wall. When the front cavity is long, these wavefronts flatten out into plane waves. For the 1 cm front cavity, the recorded sound is mainly that of the source, whose wavefronts have not become planar. This is another case where the area function alone is not sufficient to predict the spectrum of the sound that will be produced.

If we now compare the model spectra to the analyzed speech, either constriction located at 6 cm generates resonances with frequencies fairly close to those of the /x/ tokens of G and PP. EM and CS have resonances at somewhat higher frequencies, making the 4 cm location a better match. Table 4.6 lists the three amplitude parameters for these models. All three parameters match much more closely when the flat-topped plug is used.

Finally, KS's /x/ and EM's /ç/, which have a single high-amplitude, high-bandwidth peak at 3-4 kHz rather than several narrow peaks as do the other /x/ tokens, seem to combine qualities of the flat-topped plug spectra for $l_f = 4$ and 1 cm. Given the variation between subjects, it seems pointless to draw firm conclusions on the basis of one or possibly two tokens

l_f cm	Configuration		A_S dB SPL	A_T dB	A_0 dB
6.0	Circle		48.3	29	10
	FTP	/x/-like	60.6	41	20
4.0	Circle		44.4	21	5
	FTP	/ç/-like	62.2	36	23
1.0	Circle		41.5	8	0
	FTP		65.7	28	28
	Speech	/ç/	60.0	34.5	25.5
		/x/	62.2	36.9	18.7

Table 4.6: Overall Amplitude (A_S), total dynamic range (A_T), and low-frequency dynamic range (A_0) of the spectra shown in Fig. 4.15, at $U = 370$ cc/sec. The codes in the "Configuration" column refer to the cross-sectional shape of the constriction, whether circular or that of the flat-topped plug. The values given for actual speech are repeated from Table 4.3 for ease of comparison.

of /ç/. We observe, therefore, that the sorts of changes induced in the spectrum by having the flat-topped plug located so that l_f is somewhat less than 4 cm are consistent with the /ç/ tokens analyzed.

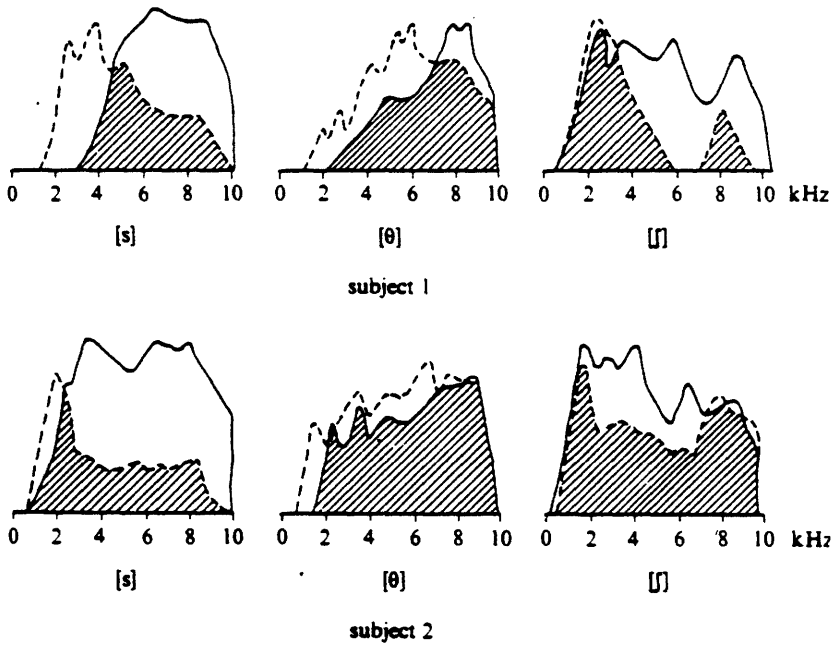


Figure 4.8: Spectra of the fricatives /s, θ, š/, recorded by two subjects with and without their false teeth. Solid line = with all teeth, dotted = with no teeth. Shaded area = area of overlap of normal and toothless spectrum. (From Catford, 1977)
Note that the missing y-axes (*sic*) are most likely the log magnitude of the sound pressure, measured in dB, using a relative scale.

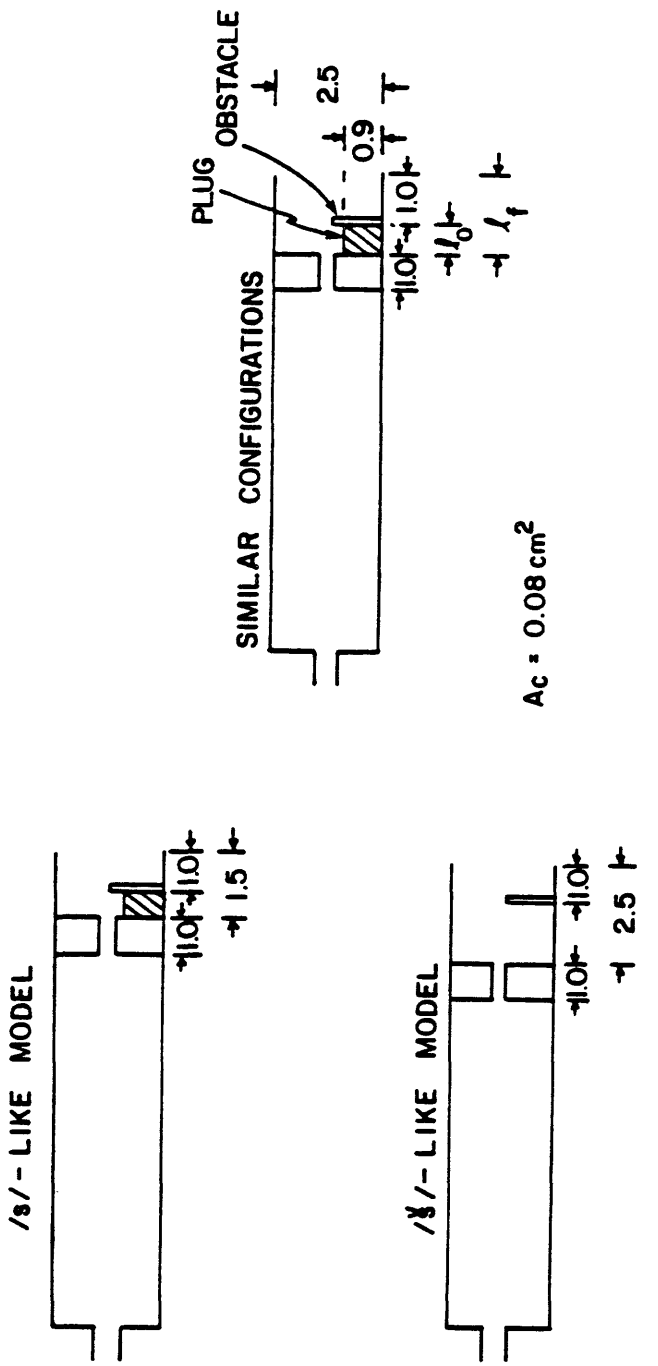


Figure 4.9: Mechanical models used to mimic /s/ and /š/. All dimensions are in cm.

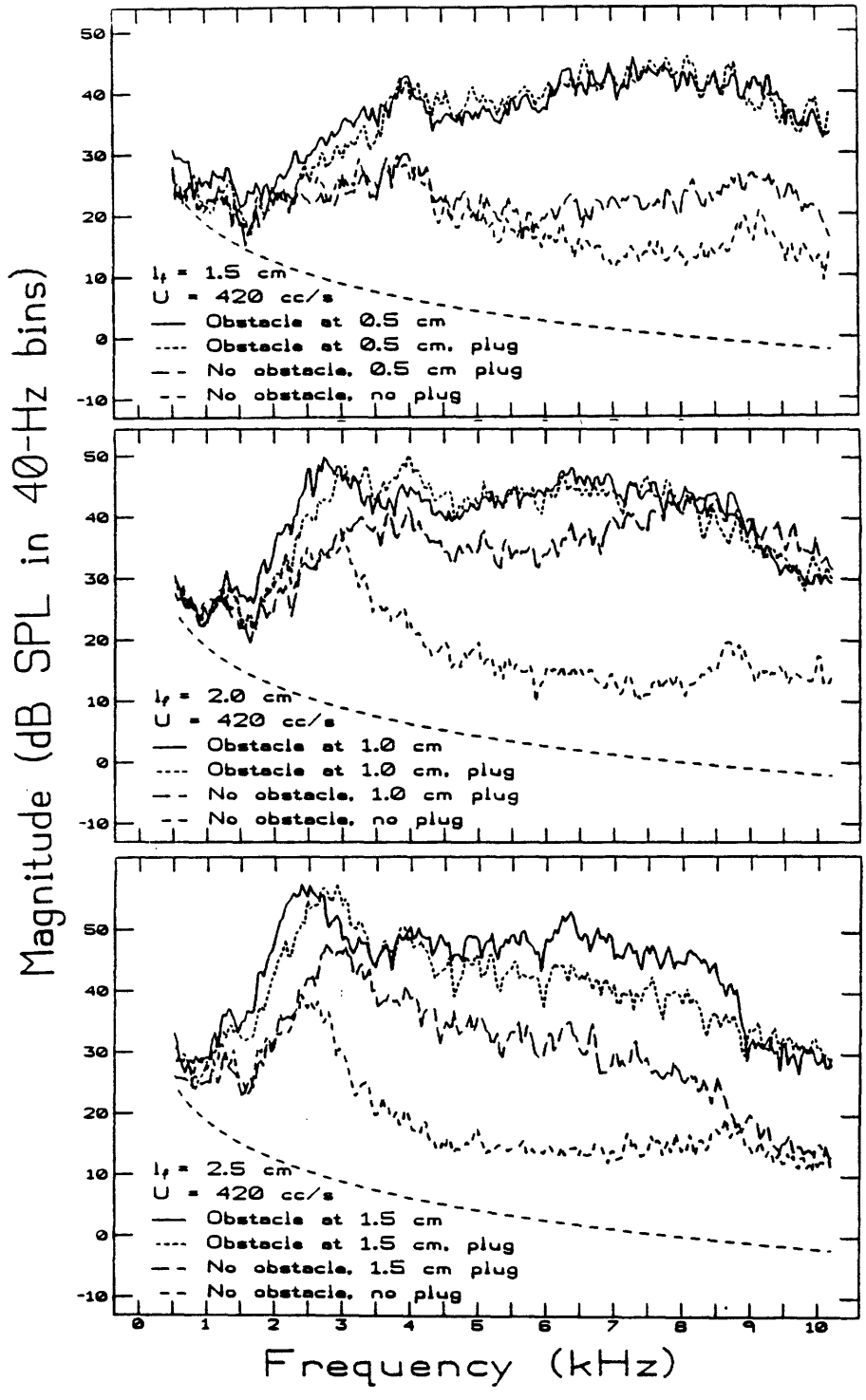


Figure 4.10: Spectra of the sound generated by air flowing through the models of Fig. 4.9, at flowrate 420 cc/sec.
a). Front-cavity length $l_f = 1.5$ cm, distance from constriction to obstacle $l_o = 0.5$ cm.
b). $l_f = 2.0$ cm, $l_o = 1.0$ cm.
c). $l_f = 2.5$ cm, $l_o = 1.5$ cm.

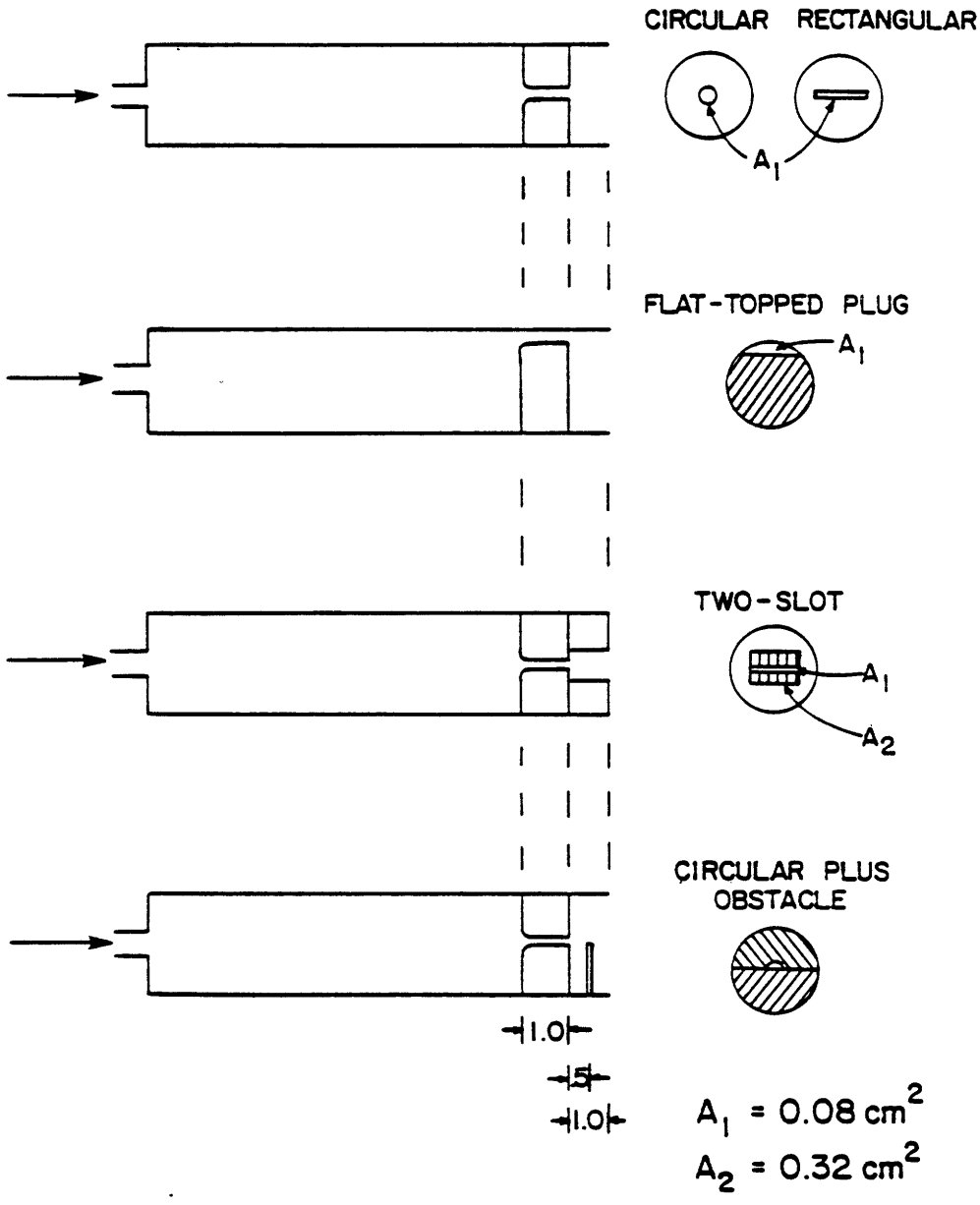


Figure 4.11: Mechanical models used to mimic /ϕ, f, θ/.

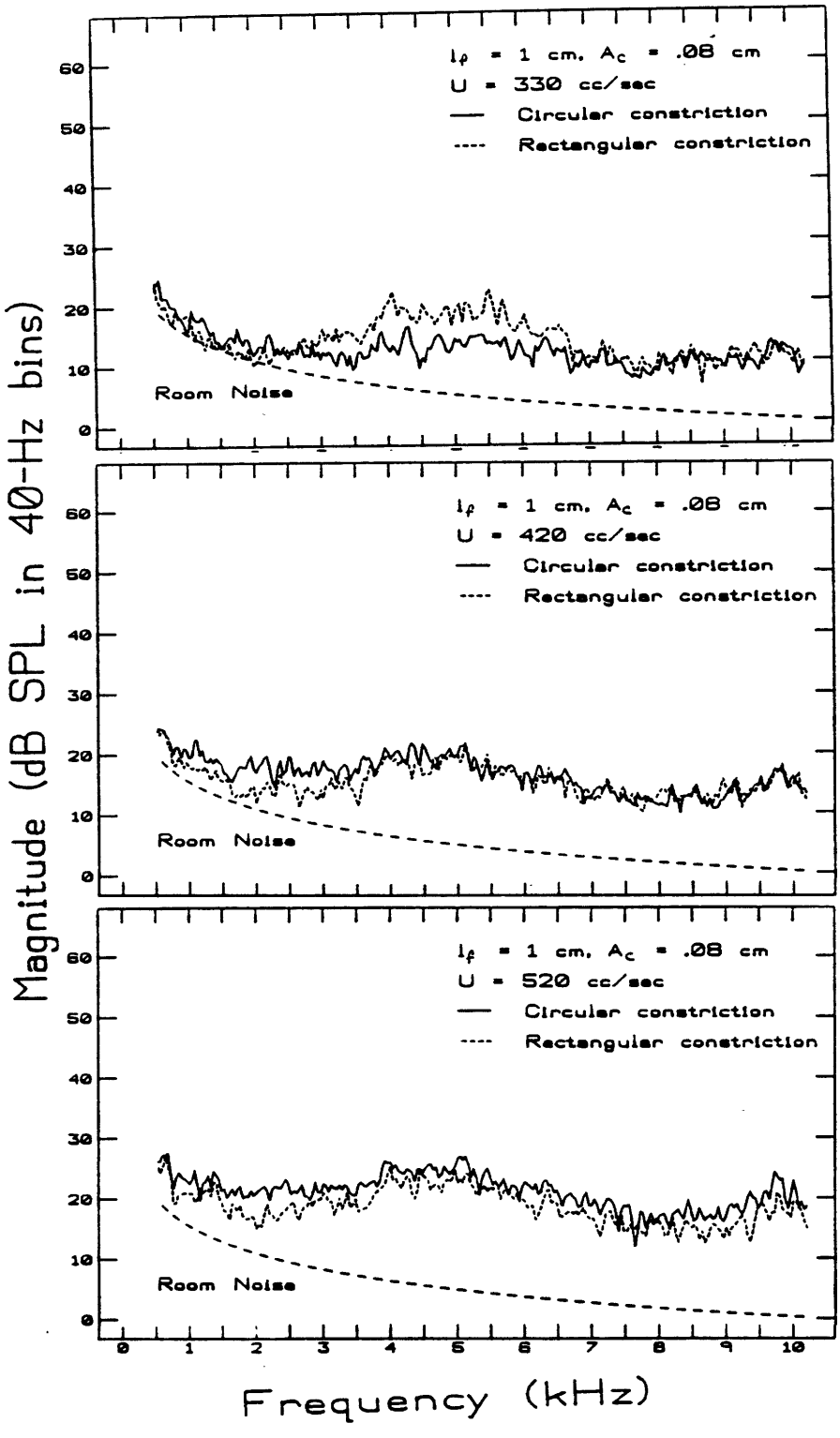


Figure 4.12: Spectra produced by rectangular and circular constrictions 1.0 cm from mouth, at the flowrates 330, 420, and 520 cc/sec.

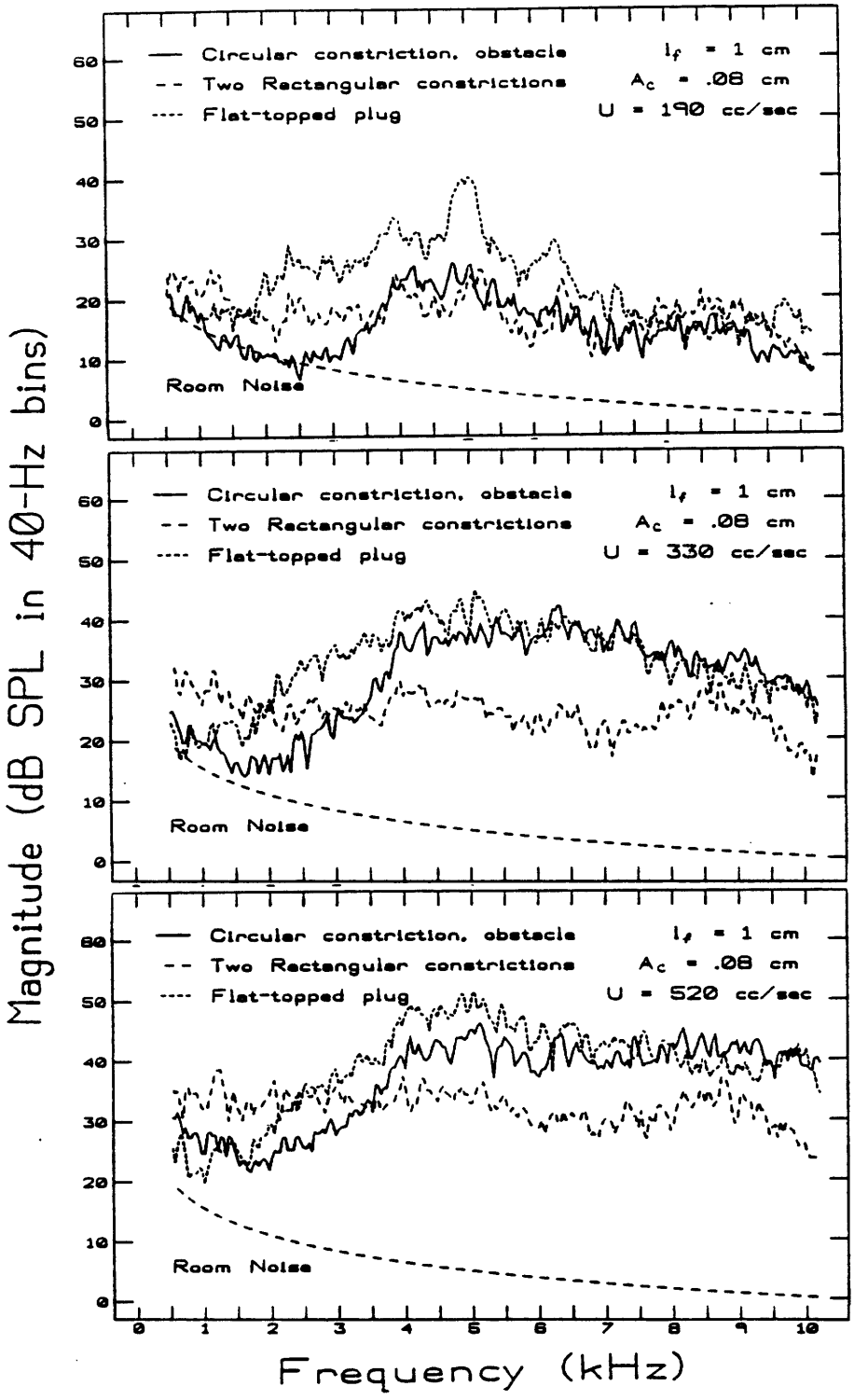
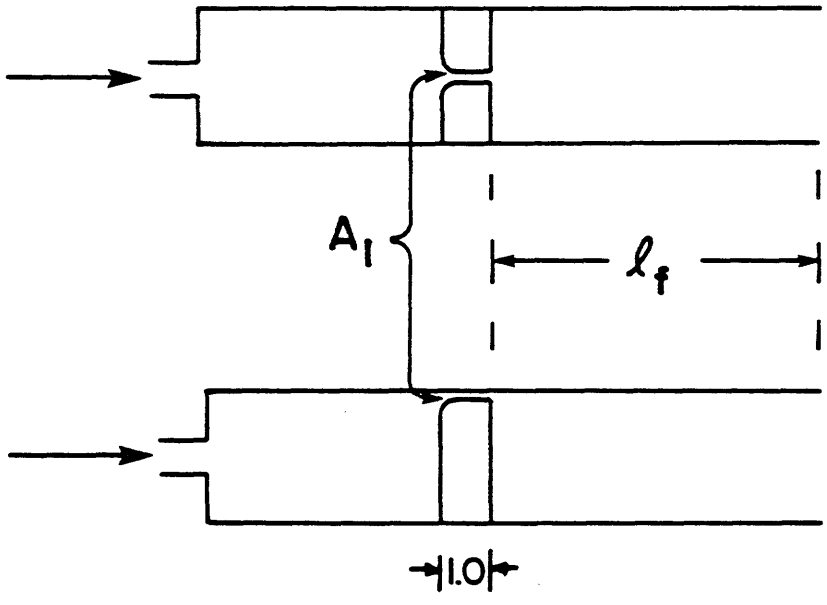


Figure 4.13: Spectra produced by the two-slot, the flat-topped plug, and the circular constriction plus obstacle, each positioned 1.0 cm from mouth, at flowrates 190, 330, and 520 cc/sec.



$$A_1 = 0.08\ cm^2$$

$$l_f = \begin{cases} 4\ cm \\ 6 \end{cases}$$

Figure 4.14: Mechanical models used to mimic /x, ç/.

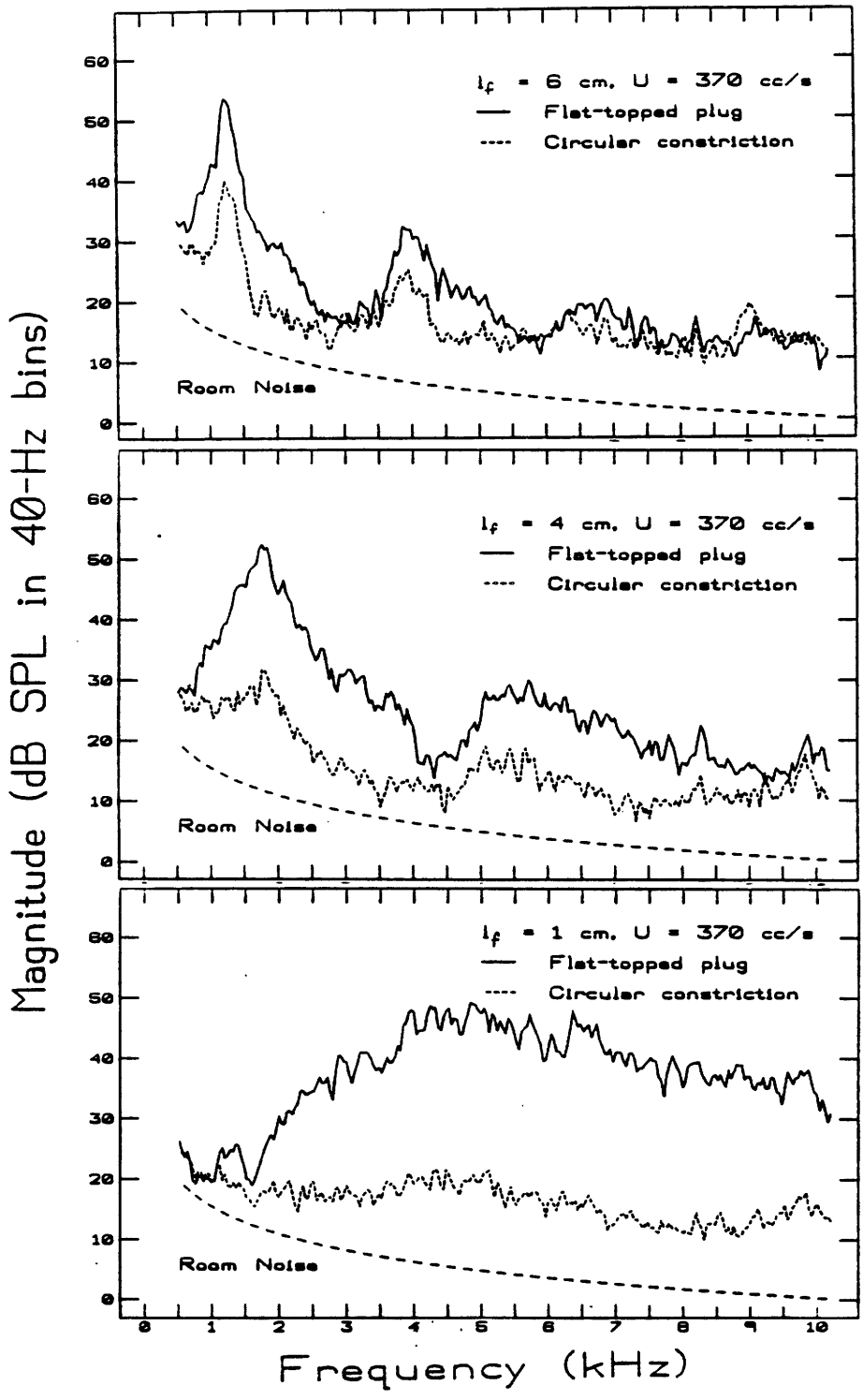


Figure 4.15: Spectra produced by the circular and flat-topped constrictions, located 6, 4, and 1 cm from mouth, at flowrate 370 cc/sec.

4.2.4 Whistles

The transition from a fricative to a whistle produced with nearly the same articulatory configuration provides additional information about the acoustic mechanisms in operation. Only /ϕ, s, š/ can merge into whistles with minor adjustments of airflow and articulator positions. For /ϕ/ the adjustment is the most profound: for CS (the only subject to record fricative-whistle transitions) it consisted of pursing the lips to form a more circular opening, and tightening and pulling them in, while raising the tongue slightly in the mouth. Figure 4.16 shows spectra of /ϕ/ and a bilabial whistle produced during the same breath by making these articulatory adjustments; the phonation recorded following the sustained /ϕ/ is also shown. The whistle couples into the second formant, which is defined mainly by the tongue position in the anterior-posterior direction.

In order to test our understanding of the acoustic mechanism that produces the whistle and to create an experimental model, a mechanical model was designed to mimic the bilabial whistle. The articulatory changes needed to produce a whistle from a /ϕ/ position can be translated into mechanical model terms by making the constriction at the mouth of the configuration more circularly shaped and decreasing the tapere of its inlet, and introducing a second constriction of similar cross-sectional area behind the first to mimic the raised tongue. Three such models, which were assembled from constrictions of three different areas, are shown in Fig. 4.17. In the following discussion, the constriction areas are referred to as A_t, for the constriction due to the tongue, and A_m, for the constriction at the mouth. Typical values for A_m during actual whistling range from 0.20 to 0.39 cm^2 (Shadle, 1983). Since according to the same study the transfer functions of the whistle configurations resemble those of a highly rounded /i/ or /u/, area functions of those vowels can be used to give a typical value for A_t of 0.3 cm^2. If a smaller tongue constriction is used, a whistle can still be produced, but it sounds more hissy. We would thus expect the middle-sized constriction (with an area of 0.32 cm^2) used in the mechanical models, to mimic human whistling more accurately than the 0.08 or 0.71 cm^2 constrictions.

A mechanical model with A_m = 0.08 cm^2 will not whistle at any flowrate in the range of 0 to 520 cc/sec. Any other combination of the three constriction areas will whistle for at least some flowrates in that range. Figure 4.18 shows a typical whistle produced when A_t = 0.71 cm^2 and A_m = 0.32 cm^2. The sharp peak at 1120 Hz is evidence of a quiet but aurally unmistakable whistle. The distance between constrictions affects the whistle frequency somewhat: for front cavity lengths of 1.0, 2.0 and 3.0 cm, the whistle occurred at 1160, 1120, and 1040 Hz, respectively. When the overall tract length was reduced to 16 cm, the 1.0 cm front cavity length whistled at 1240 instead of 1160 Hz. It seems clear then that the whistles occur at half-wavelength resonances of the entire tract, which are slightly affected by the location of the upstream constriction.

For this configuration, the whistle will not occur unless the inlet to the downstream constriction is abrupt. If A_t is smaller than 0.71 cm^2, whistles will occur even when the downstream constriction inlet is tapered. The whistle is somewhat clearer when the downstream inlet is abrupt.

Figure 4.19 shows the effect of changing the distance between constrictions, l_f, when A_t = 0.08 and A_m = 0.32 cm^2. The whistle frequency decreases as l_f increases, though not in a simple geometric proportionality. The resonance that the whistle has coupled into is essentially

the Helmholtz resonance of the front cavity and the mouth constriction:

$$f = \frac{c}{2\pi}\sqrt{\frac{A_m}{l_m A_f l_f}}$$

where A_f is the cross-sectional area of the tube between the constrictions. The amplitude of the whistle and the degree to which it is distinct from the rest of the spectrum depend on both flowrate and l_f. Changes in flowrate for a given configuration cause the whistle frequency to fluctuate around the resonance, increasing and decreasing the amplitude of the whistle as a consequence. The differences between the spectra for $l_f = 1.5$ cm are evidence of such changes; clearly the same two flowrates do not produce the same effect when $l_f = 3.0$ cm.

When $A_t = 0.32$ and $A_m = 0.71$ cm^2, two resonances occur only 300 Hz apart. As shown in Fig. 4.20, the result is a double-peaked (and double-pitched) whistle, over a range of flowrates. The figure shows a changeover when an increase in flowrate causes the higher-frequency peak to assume the higher amplitude. Such frequency jumps have been documented for all types of mechanical whistles (Chanaud and Powell, 1965), and for humans whistling as well (Shadle, 1983).

The ability of these models to whistle at a wide range of flowrates and constriction sizes, the relative unimportance of tapering on the downstream constriction, the effect of front cavity length on the whistle frequency, and the presence of frequency jumps, are all examples of behavior that are also exhibited by humans whistling. Humans, however, can whistle up to much higher frequencies; the models top out at 1500 Hz, but humans can whistle as high as 4000 Hz. This is most likely accomplished by decreasing the cross-sectional area of the front cavity as the tongue is brought forward, thus raising the resonance frequency.

Since the existence and behavior of the whistles are controlled by the same parameters in the models and in humans, the acoustic mechanism producing whistles in both cases is very likely the same. As described by Chanaud and Powell (1965), this type of configuration, with two colinear orifices, gives rise to the hole tone. The jet produced by the upstream constriction is unstable; the instability is reinforced by reflections from the downstream constriction; a train of ring vortices is set up, which generates sound that excites and couples into the cavity resonances.

For /s/ the adjustment necessary to produce a whistle was more subtle than for /ϕ/. For CS it consisted of reducing the airflow rate and moving the tip of the tongue very slightly until a critical whistle-producing location was found. Although such whistles could be repeated and manipulated in frequency somewhat, they were always quite hissy. Once a good location for the tongue tip was found, the lips could be moved independently to alter the frequency. Predictably, a bigger mouth area resulted in a higher whistle frequency. Spectra of a whistled and regular /s/, and a phonated /s/, are shown in Fig. 4.21. In this case the whistle coupled into the third formant, at about 2500 Hz.

The /s/-like models developed earlier were tested for their ability to whistle, either with or without the plug. At low flowrates, when the obstacle was present, somewhat hissy whistles were generated at the frequency of the first visible resonance, as shown in Fig. 4.22. It is contrasted with a non-whistling higher flowrate through the same configuration. The whistle frequency and its hissy quality correspond quite closely to the /s/-whistle produced by subject CS.

/š/ needed the least adjustment of all: a very slight tongue-tip adjustment and airflow reduction were sufficient to produce a fairly clear whistle. Figure 4.23 gives an example in

which the fourth formant was coupled into by the whistle. This was a front-cavity resonance, as became clear when the tongue was held still but the position of the lips was altered. The formant structure changed somewhat from that of a normal /š/ due to the adjustments necessary to produce a whistle.

Since the /š/-like mechanical model whistled at low flowrates, no attempt was made to duplicate the slight change in position of the tongue tip. Figure 4.24 shows spectra for the same model at two flowrates, the lower of which produced a whistle. Although the /š/ and /š/-like whistle spectra are dissimilar, both whistles occur at a front-cavity resonance, and therefore the whistle frequencies are controlled mainly by the tongue position. Possibly the additional peaks in the speech spectrum are due to a larger constriction area in that case, which would allow back-cavity resonances to become more visible. The whistle is more distinct for both /š/ and the /š/ model than for /s/ or the /s/ model. This is most likely due to the greater damping of the whistle formant in the /s/ case.

The whistle mechanism for the /s/ and /š/ models, and thus apparently for the mouth, is an edge tone. In this case, feedback is set up between the origin of the jet at the tongue and the edge created by the obstacle/lower teeth. The whistle couples into the nearest cavity resonance.

Transition from the bilabial fricative to the bilabial whistle involves significant articulatory change, and results in an acoustically dramatic difference. For the /s/ and /š/ whistles, neither the articulatory nor acoustic changes are as profound. A lower flowrate through the same configuration can produce a fricative that is slightly whistly. It is therefore more likely that /s/- and /š/-whistles will occur during normal speech, as, for instance, during the transition into or out of a fricative.

4.3 Discussion

In this chapter, sustained fricatives as produced by five speakers were analyzed and compared to previous studies, to the mechanical model data of Ch. 3, and to more complex mechanical models designed to imitate each fricative. Similarities between the tokens analyzed in this and in previous studies in terms of overall amplitude, peak and trough frequencies, and the amount of between-speaker variation of these parameters, demonstrated that neither the unnaturalness of sustained production nor the analysis method impaired the essential characteristics of the fricatives.

An additional set of parameters that measured the dynamic range of the spectra were developed, based on the mechanical model work in Chapters 2 and 3, as a way of characterizing the acoustic differences caused by source generation at an obstacle. Comparison of the speech and the mechanical model data of Ch. 3 revealed that the fricatives /ϕ, f, θ/ were most similar to the no-obstacle case, judging by the overall amplitude and the dynamic range parameters; /š, x/ were most similar to the obstacle case; and /s/, somewhat unexpectedly in view of its obstacle-like articulatory configuration, had elements of both.

This initial set of comparisons, together with a knowledge of the articulatory configurations of each of the fricatives, was used to design more complex mechanical models intended to mimic the fricatives. These models fell into three classes, according to the dominant source mechanism. For /s, š/, the jet impinges on an obstacle, producing a high-amplitude spectrum with a few obvious poles and zeros. For /ϕ, f, θ/, surfaces nearly parallel to the jet generate additional sound. The result is a relatively low-amplitude spectrum whose shape depends heavily on the

exact angle and shape of the constriction and surfaces. For /x, ç/, the front cavity is relatively long, and the sound is generated by the action of the jet on the wall of the tube, mimicking the role of the palate and alveolar ridge. The spectra have a relatively high amplitude with well-defined resonances. The two latter classes clearly illustrate that knowledge of the area function alone is not sufficient to predict the resulting sound spectrum.

Comparison of the spectra produced by these more complex models to the fricative spectra in terms of the various spectral parameters showed that the models for /s, š/ are very good, as are those for /ç, x/. The anomaly noted when /s/ was compared to the models of Ch. 3 disappeared when it was compared to the more realistic model. For /ϕ, f, θ/ no models were found for which all parameter values fell within the first standard deviation of the values found for speech. However, the partial fits provided by the models that were tried made it clear that a simple no-obstacle model would not suffice.

From these results we may infer that theoretical models for /s, š/ should consist of a series pressure source located at the obstacle/teeth. Theoretical models for /ç, x/ should include a distributed pressure source, representing the sound generated along the roof of the mouth. Theoretical models for /ϕ, f, θ/ are likely to include a distributed source as well, of uncertain type, and may need to include evanescent as well as longitudinal modes due to the shortness of the front cavity.

Consideration of whistles produced in the /ϕ, s, š/ positions resulted in mechanical models that whistled at similar frequencies and flowrates, and with the same frequency-controlling parameters. It is thus possible to infer the whistle mechanism from the mechanical models: /s, š/ whistles are produced by edgetones, and /ϕ/ whistles are due to holetones.

The principle conclusion is that even highly simplified mechanical models can mimic speech spectra with a fair degree of success, and thus provide a useful domain in which to explore the parameters controlling sound production due to turbulence in the vocal tract.

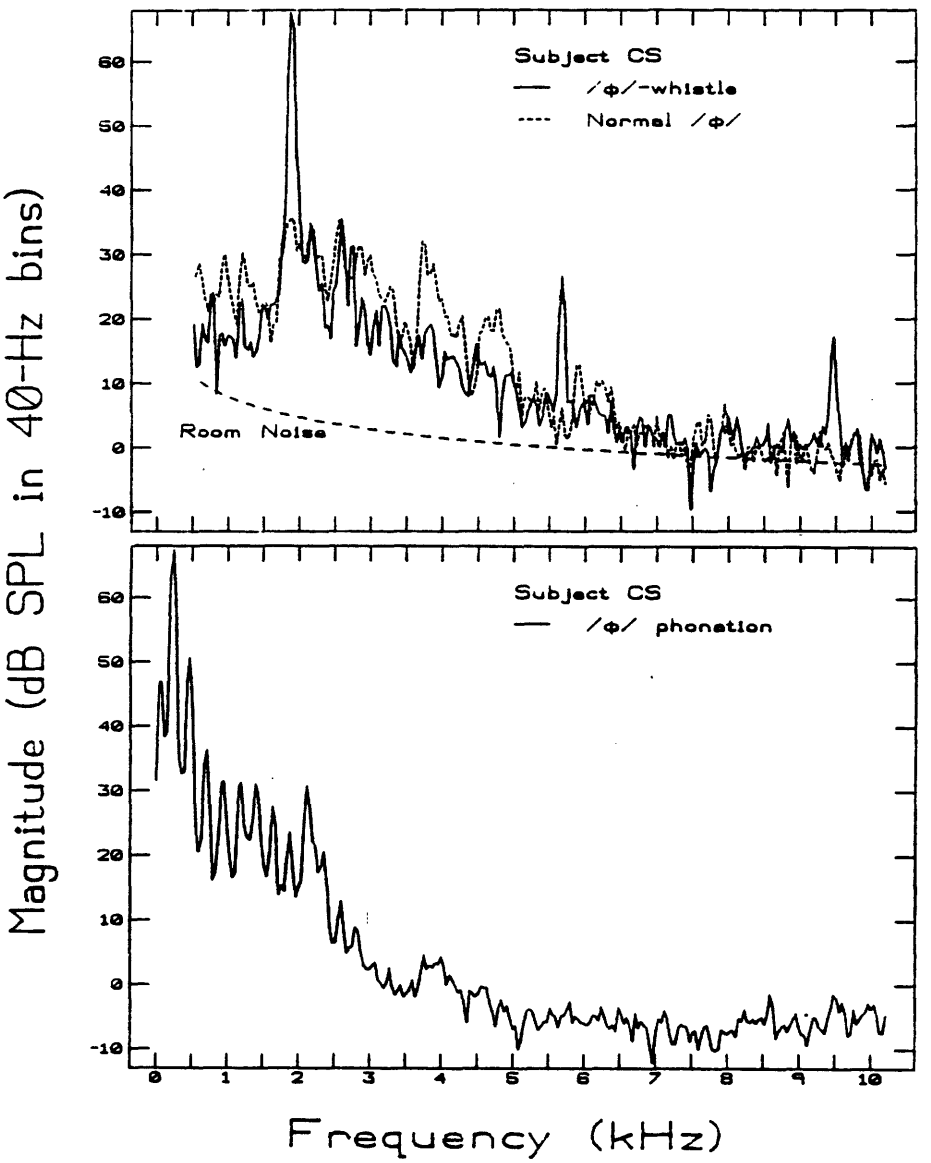
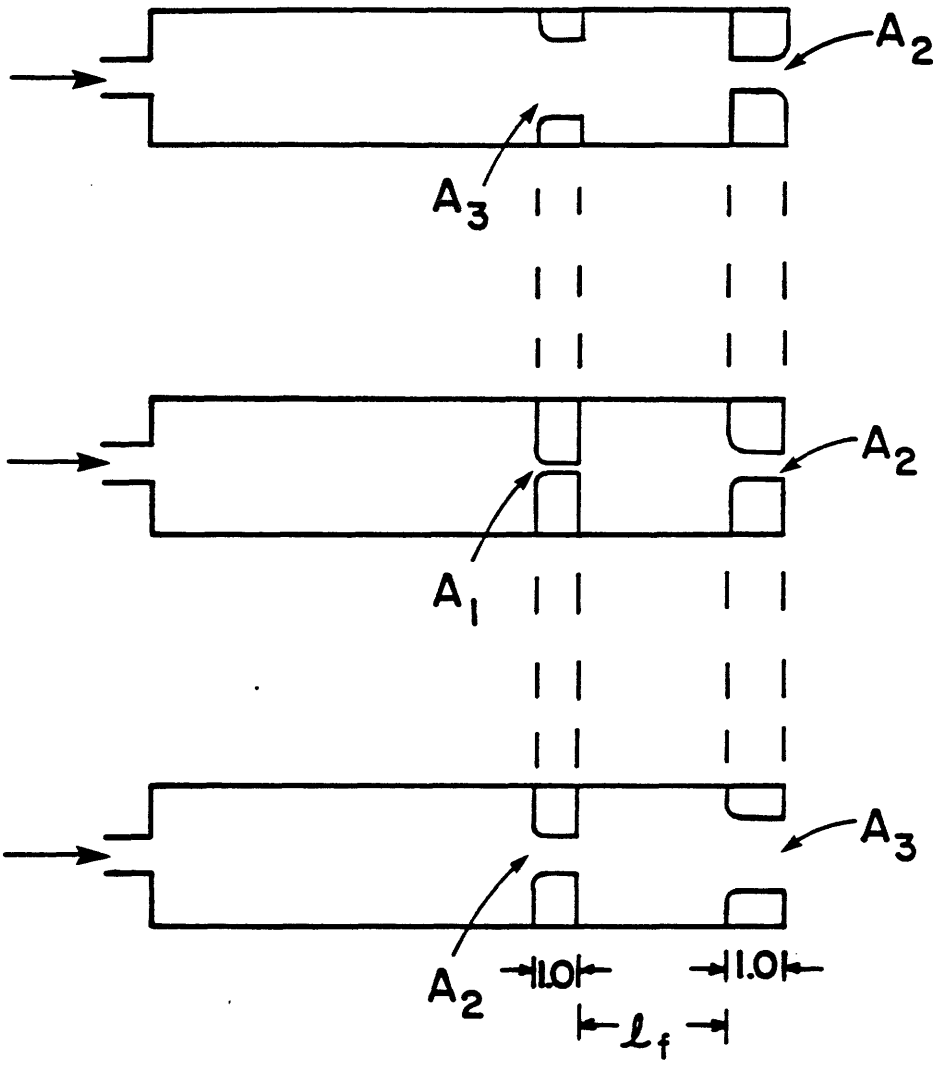


Figure 4.16: Spectra of bilabial fricative /ϕ/ and whistle, and accompanying phonation, produced by subject CS.



$A_1 = 0.08\ cm^2$

$A_2 = 0.32\ cm^2$

$A_3 = 0.71\ cm^2$

Figure 4.17: Mechanical models used to mimic the bilabial whistle.

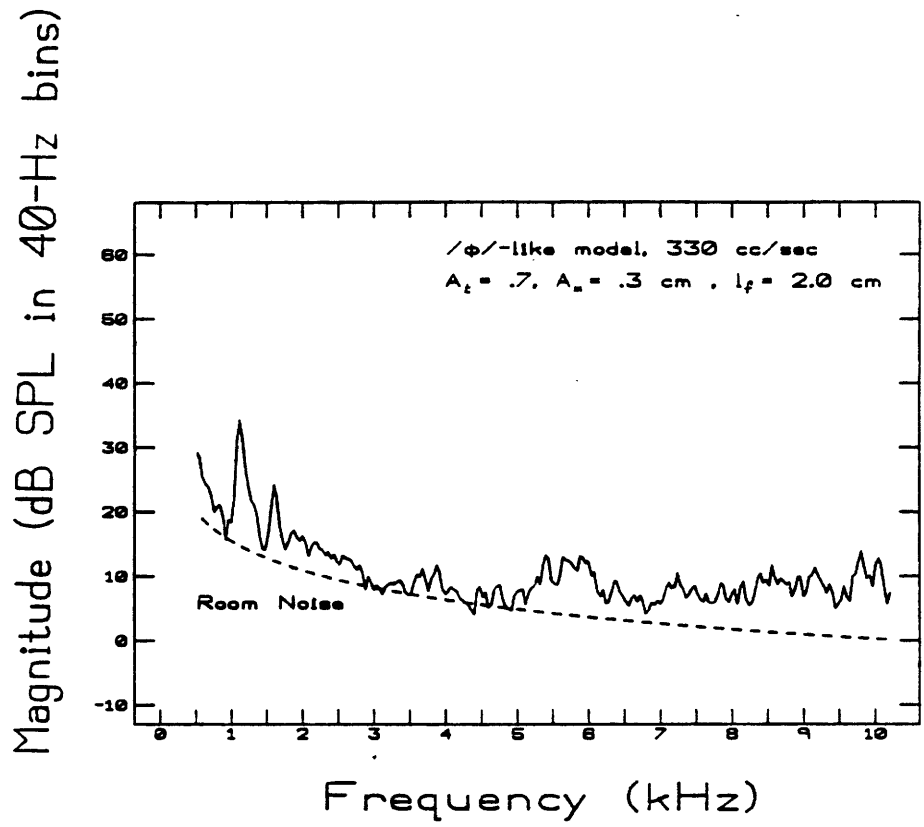


Figure 4.18: Spectra produced by mechanical model of bilabial whistle pictured in Fig. 4.17, with $A_t = 0.71$, $A_m = 0.32$ cm^2, $l_f = 2$ cm, at flowrate 330 cc/sec.

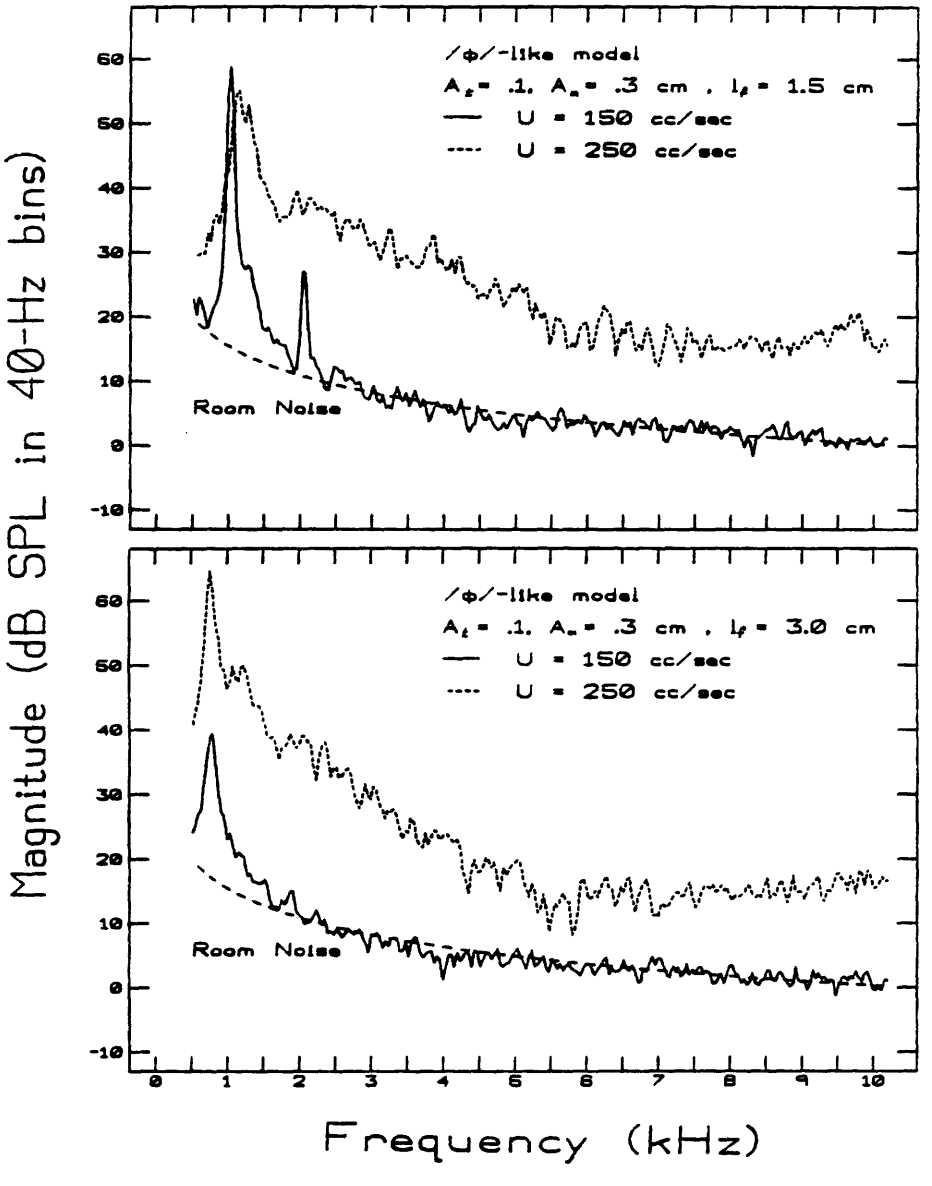


Figure 4.19: Spectra produced by mechanical model of bilabial whistle pictured in Fig. 4.17, with $A_t = 0.08$, $A_m = 0.32$ cm^2, at flowrates 150 and 250 cc/sec.
a) $l_f = 1.5$ cm.
b) $l_f = 3.0$ cm.

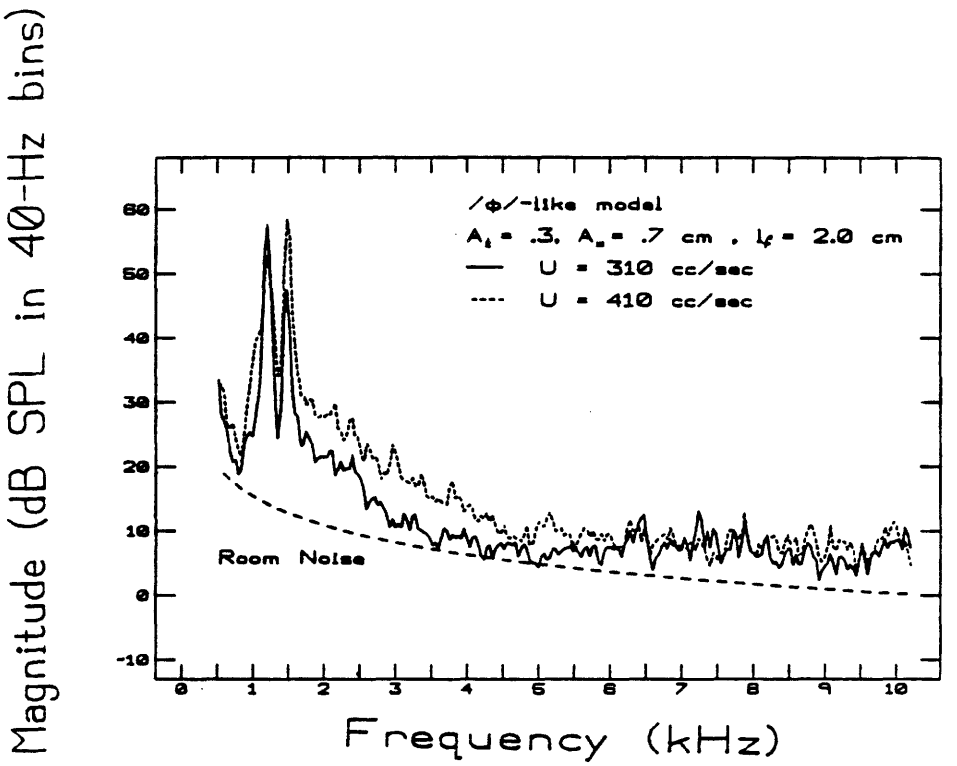


Figure 4.20: Spectra produced by mechanical model of bilabial whistle pictured in Fig. 4.17, with $A_t = 0.32$, $A_m = 0.71$ cm^2, $l_f = 2$ cm, at flowrates 310 and 410 cc/sec.

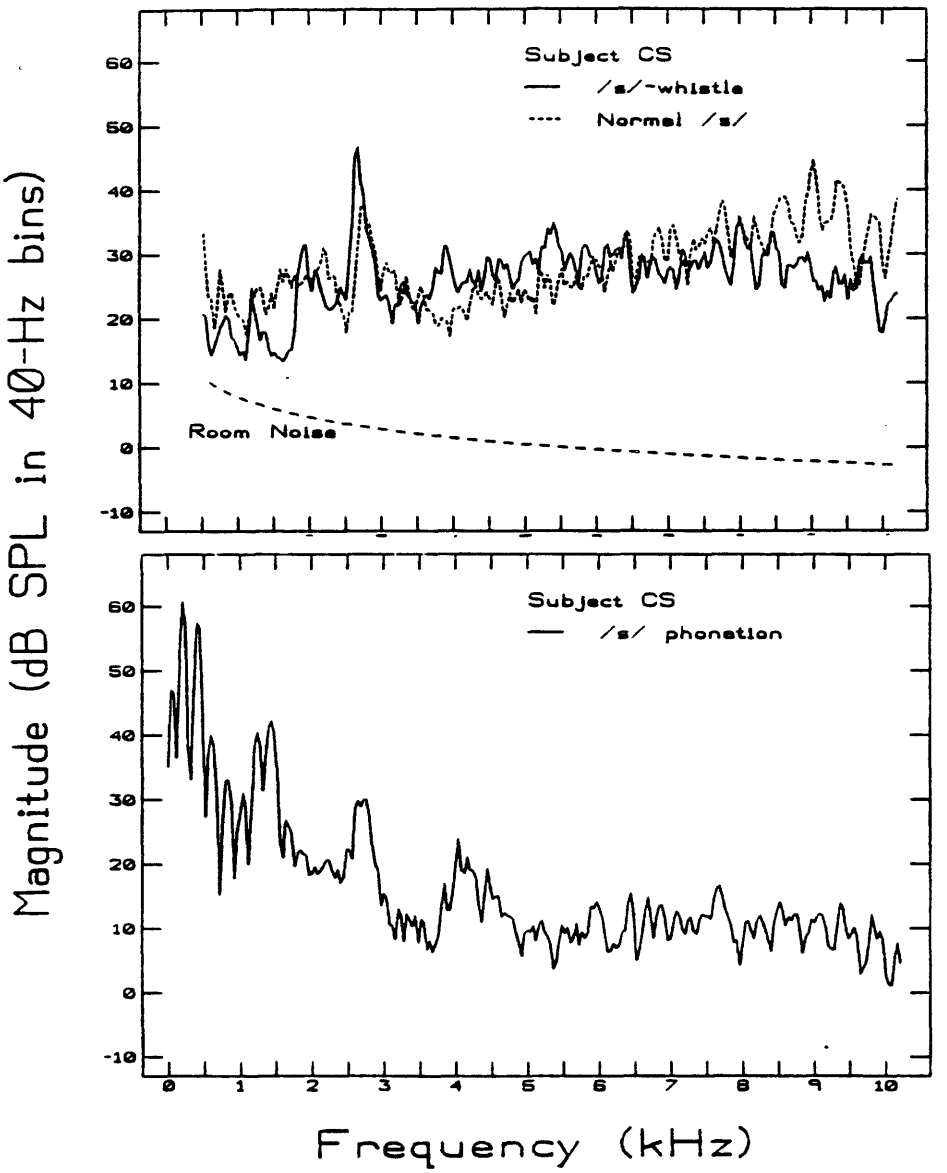


Figure 4.21: Spectra of /s/ and /s/-whistle, and accompanying phonation, produced by subject CS.

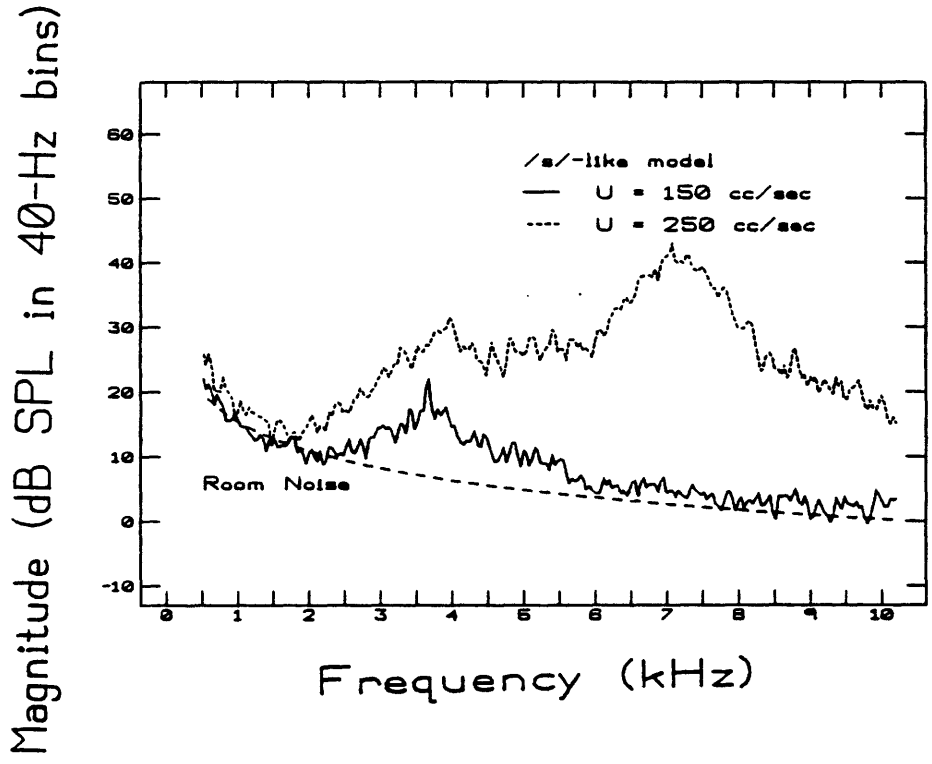


Figure 4.22: Spectra of /s/-like mechanical model without plug at flowrates 150 and 250 cc/sec; at 150 cc/sec, the model whistles.

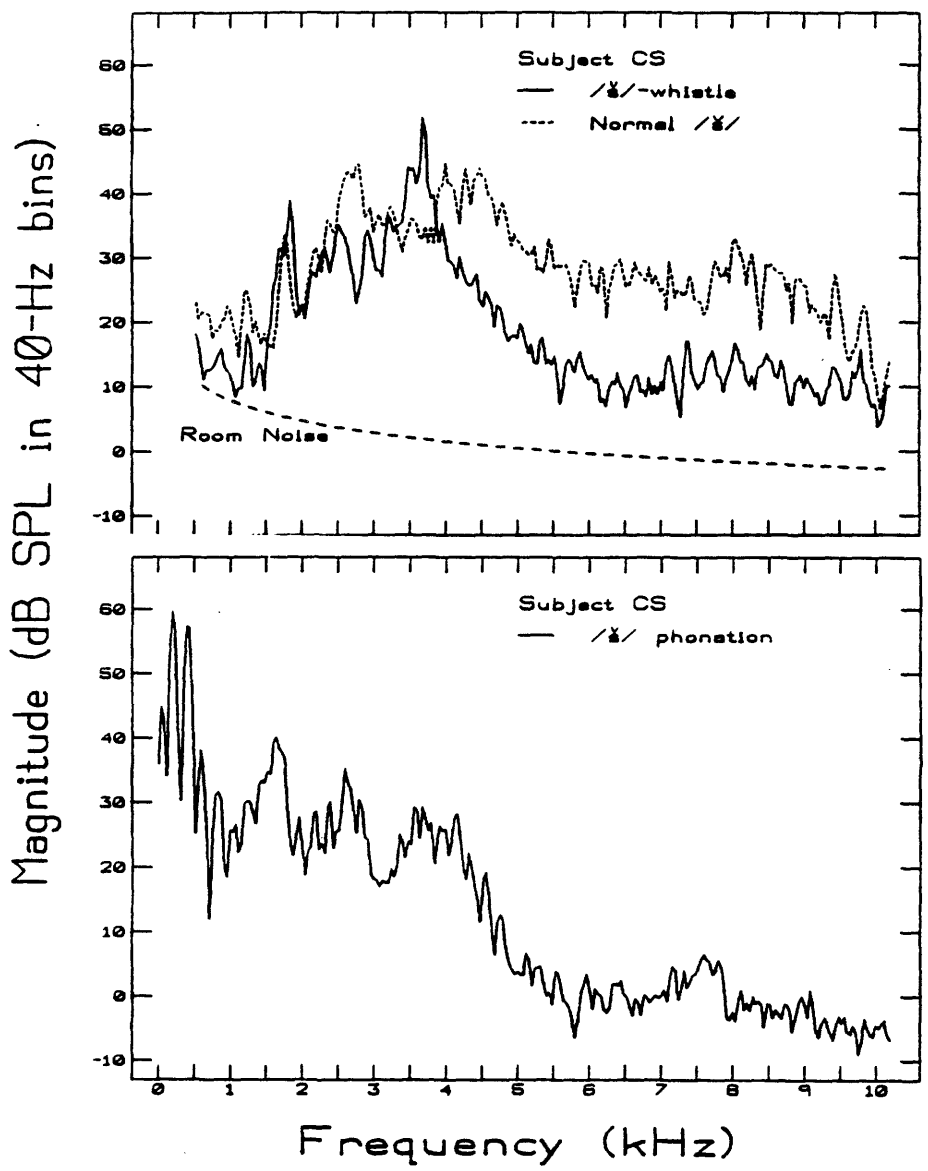


Figure 4.23: Spectra of /š/ and /š/-whistle, and accompanying phonation, produced by subject CS.

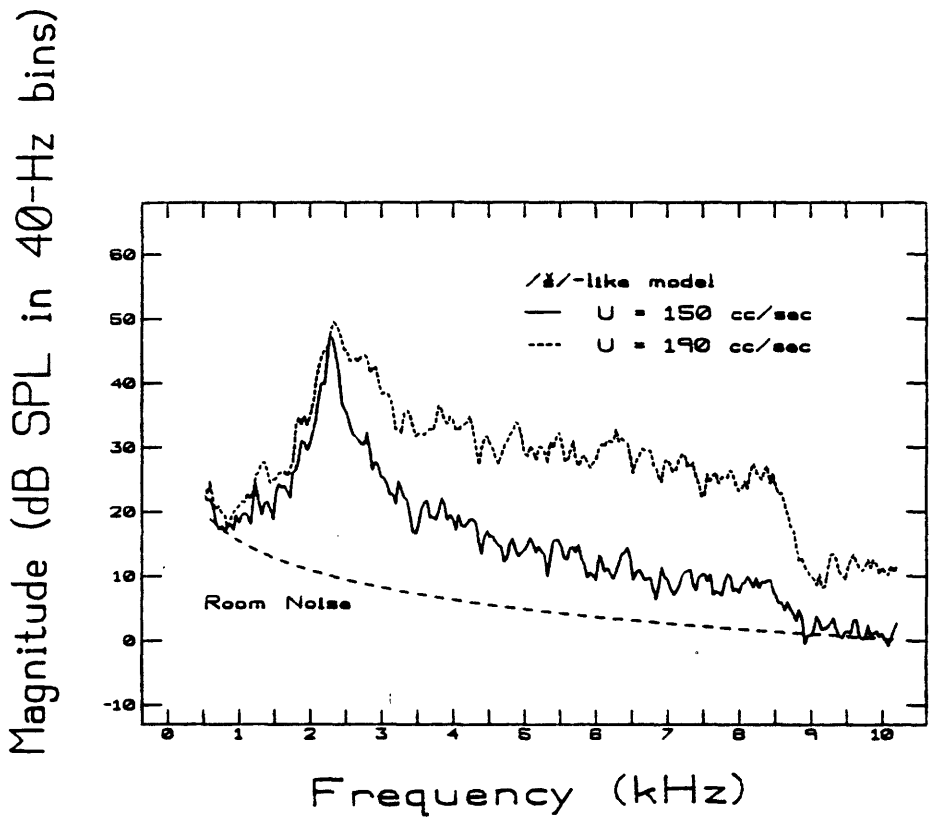


Figure 4.24: Spectra of /š/-like mechanical model at flowrates 150 and 190 cc/sec; at 150 cc/sec, the model whistles.

Chapter 5

Conclusion

5.1 Summary and Discussion of Results

The central goal of this thesis has been to understand the acoustic mechanisms for the production of fricative consonants. The approach of working in three domains, namely speech, mechanical models, and theoretical models, has allowed us to choose an appropriate context for each stage of the work, as well as providing independent checks of many of the basic assumptions. In particular, the applicability of mechanical model results to speech has been established, and the assumption of no source-filter interaction has been explicitly tested.

All fricative configurations have in common a small turbulence-producing constriction within the vocal tract. This basic configuration of a constriction in a tube was investigated by use of a mechanical model, in which parameters such as constriction area, length, location, and degree of inlet tapering, and presence of an obstacle, were varied. It was found that acoustically the most significant parameters are the presence of an obstacle, the length of the front cavity, and the flowrate. Therefore, configurations in which only these parameters were varied were examined more thoroughly and modeled theoretically.

A source function for the obstacle case was derived from the far-field sound pressure measured when the obstacle was located in space, downstream of a constriction which was mounted flush with a baffle. The directivity pattern produced by the obstacle in this position was similar to that of a dipole, as expected. A dipole source located inside a duct is equivalent to a pressure source in a transmission-line model, when only the longitudinal modes of a duct are considered. The filter function, corresponding to the effect of the duct on such a pressure source, was derived for the transmission-line representations of two configurations in which the obstacle was located inside a front cavity of nonzero length. The transmission-line model included losses due to viscosity, heat conduction, radiation and flow resistance. The flow resistance was computed from measurements made for the actual constriction used; the other sources of loss were predicted theoretically. The spectra predicted by this source-filter model, when compared to the far-field pressure measurements of the equivalent mechanical models, provided a very close match in both absolute sound pressure level and spectral shape. Thus a major finding of this thesis is that the simple linear source-filter model works well in the obstacle case.

For the no-obstacle case, it was not possible to derive a source from a free-field measurement, so absolute level of the prediction could not be checked. Investigation of the pressure versus flow velocity power laws showed evidence of source-tract interaction for the no-obstacle case, but none for the obstacle case. This is in agreement with Heinz' results (1958): he demonstrated that the power exponent for the no-obstacle case decreased when the constriction was moved back from the mouth of the tube.

Spectral measures were developed that captured the acoustic differences between the obstacle and no-obstacle cases. Analysis of real speech in terms of these spectral measures revealed that fricatives are more similar to the obstacle than to the no-obstacle case, again a major finding. More complex speech-like mechanical models were developed and compared to the speech, again in terms of the spectral measures. Very good models of /s/ and /š/ were obtained by

using an obstacle at right angles to the flow and varying the constriction location. For the fricatives /ϕ, f, θ/, in which the constriction is located forward in the tract, the shape of the constriction was crucial. Constrictions that allowed the jet to come in contact with a surface produced sounds that most closely resembled the analyzed examples of these fricatives. For /x/ and /ç/, a surface also caused the mechanical model spectra to become the most like the speech spectra, but in general the effect of the constriction shape lessened as the length of the front cavity increased.

Because of the theoretical work done on the simple models, some conclusions can be drawn regarding the best source model for each class of fricatives. For /s, š/ a series pressure source at the location of the teeth is needed, with a source function similar to those derived for the obstacle case. For the fricatives other than /s, š/, a "distributed obstacle", modeled as a distributed pressure source, may be the dominant acoustic mechanism. Presumably such a source would be weaker than a localized pressure source at a given flowrate. Note that although a simple rule can be specified predicting the source distribution from the angle that the obstacle makes with the jet, the information necessary to use such a rule is not contained in the conventional cross-sectional area description of the tract.

We have thus established that the series pressure source used by Fant (1960) is correct for the obstacle case, and presumably also therefore for /s, š/, when placed at the location of the obstacle. What about the source characteristics? Fant found that the following source descriptions provided the best match to speech spectra:

/f/ -6 dB/oct between 800 and 10000 Hz, or
-3 dB/oct between 800 and 10000 Hz (better fit)
/s/ 0 dB/oct, 800 to 4000 Hz; -6 dB/oct, 4000 to 10000 Hz
/š/ apical source of 0 dB/oct from 300 to 6000 Hz, or
dental source of 0 dB/oct from 300 to 2000 Hz,
-12 dB/oct from 2000 to 10000 Hz
/x/ -6 dB/oct between 300 and 4000 Hz

If we compare these to the source functions derived in Ch. 3, that is, to Table 3.6, some striking similarities appear. First, we now know that /š/ should have a dental, not apical source. The slopes specified by Fant are quite similar to those listed in Table 3.6. For /s/, the constriction is closer to the teeth than in /š/, which should cause the peak frequency to be higher. Indeed, the breakpoint between flat and decreasing slopes, as specified by Fant, is at a higher frequency for /s/. We cannot at this point comment on Fant's choices for /f/ and /x/.

The whistles generated by several configurations were also investigated. Orifice tones occurred for untapered constrictions, at a greater range of constriction sizes than predicted by Succi (1977). The frequencies of the tones corresponded to the half-wavelength resonance of the constriction and its harmonics. Edgetones occurred for configurations including an obstacle, at frequencies related to the flowrate and distance to the obstacle. They occurred more readily for a rectangular than circular constriction cross-section. For either shape, placing the obstacle inside a tube altered the conditions (specified by flowrate and constriction-obstacle separation) under which whistles occurred, and the frequency and amplitude of the whistles.

Edgetones were generated by the /s/- and /š/-like mechanical models that were similar acoustically to the whistles produced by a subject with the /s/ and /š/ configurations. An even more striking parallel between the whistles produced by speech-like models and humans was found for the typical bilabial whistles produced by /ϕ/-like configurations. Two constrictions,

reproducing the role of tongue and lips, were necessary to model these whistles. The agreement in flow range, frequency, and control parameters of the holetones thus produced and the whistles generated by humans was good enough to make it likely that the same acoustic mechanism occurred in both the models and the human vocal tract.

5.2 Future Work

We are now in a position to reconsider the issue of source-tract interaction. Recall that in the obstacle case, the source was determined by the location of the obstacle relative to the constriction, and the flowrate, which is presumably a function of the constriction diameter and linear flow velocity. That source showed no evidence of interaction with the tract. The no-obstacle source, which was presumed to be controlled by only the constriction diameter and flowrate, exhibited some interaction according to the methods used in Ch. 3. The work with more complex models in Ch. 4, and the experiments using constrictions of different areas in Ch. 2, indicated that sound generation occurs at the walls and sometimes at the mouth of the tube for the no-obstacle case. For the flat-topped plug, used in Ch. 4, such wall sources are clearly very significant. Some (though not necessarily all) of the interaction is due to such additional sound generation.

If we had a way of including in the source specificiation the jet parameters (U, d, and shape of constriction) and the location of all surfaces contacted by the jet, it appears that that source would be independent of the tract for all non-whistling cases. The geometry of the tract would then be included in descriptions of both source and filter, but the acoustic behavior would be non-overlapping in the two specifications. In this way we can consider separately the tendency of a wall to react on turbulent fluid leading to the generation of flow dipoles, and its tendency to act as a resonator for longitudinal waves. Source specification of this sort would eliminate source-tract interaction of a type that might be termed rigid-body interaction, to indicate the cause of the additional source generation, or model-dependent interaction, to indicate that the presence of interaction depends on the completeness of the source model.

For whistles, on the other hand, the sound is generated by an instability combined with acoustic or aerodynamic feedback. When the whistle couples into a resonance of the cavity, the features of the geometry giving rise to that resonance are clearly an integral part of both "source" and "filter". It is thus probably not useful conceptually to use independent source-filter models for whistles; an oscillator with a dependent source is likely to match the physical reality better. Developing such an oscillator-based model for the whistle alone, and then extending the typical obstacle-case source-filter model to include the whistle model, would be a significant step forwards. A single theoretical model would then be able to predict whistly /s/'s, or demonstrate the transition from /ϕ/ to a bilabial whistle, just as human vocal tracts do. This type of interaction should clearly be termed acoustic interaction.

These distinctions should provide a useful framework within which to pursue an obvious next step, that of developing theoretical models for the more complex speech-like mechanical models. As in Ch. 3, they should be created in such a way that they can be tested directly against experimental data for both spectral shape and absolute level. A key element of such a development will be the source model for the flat-topped plug. It is likely to be expressible as a distributed pressure source, which means that the location, strength and spectral characteristic must be determined for each source element, as well as the coherence between elements. For the /s, š/ models, the single pressure source at the location of the obstacle can still be used. Since

the constriction-obstacle distance l_o is shorter for /s, š/ than in Ch. 3, the source characteristic should be rederived. Aside from providing necessary information for the /s, š/ model, such data will allow the establishment of more general relationships between the source function, l_o, and the flowrate.

For /ϕ, f, θ/ the mechanical models should be improved. Due to the variability of the subjects and the need for more detailed information, more thorough measurements of speech should be made first. Correspondences between the speech and mechanical models would be easier to establish if the sound was recorded simultaneously with measurements of the vocal tract areas (at least in the region of the lips and teeth) and flowrate; if more tokens were measured; and if the recording took place in an anechoic chamber. Altering the method of signal analysis so that long sustained fricatives were not necessary might reduce the between-subject variability as well.

On whistles, much remains to be done. As with the fricatives, more precise measurements on humans giving, simultaneously, articulator positions, flowrate, and the sound pressure, would allow us to predict the frequency, amplitude, and occurrence of whistles produced by more complex configurations more precisely. Further work with models to predict the region of occurrence of the orifice tone, and scale up the edgetone so it could be studied more carefully, particularly with constriction shapes and obstacle orientations found in the vocal tract, are indicated. Flow visualization or detailed velocity distribution data would be useful in developing a theoretical model.

Another direction in which to extend this work is to consider all instances of sound production due to turbulence in the vocal tract, instead of restricting ourselves to the static case of sustained fricatives. A reasonable way in which to make such an extension is to study increasingly complex dynamic modes of turbulent sound production. First, we have assumed that a sustained fricative is acoustically nearly identical throughout. It may be that this is not so, due to changes in the flowrate as the end of the breath nears, to small adjustments of the configuration, or to, for example, a mechanical vibration of an articulator such as the uvula induced by the particular combination of flowrate and articulatory configuration. For this as for all dynamics work, better control of the analysis procedure is essential, so that the researcher may have complete control of the samples chosen for analysis.

Slightly greater complexities are involved in a consideration of the transition into and out of fricatives. A detailed look at the sequence of articulatory events would indicate whether the vocal tract passes through configurations that are whistle-prone, and the degree to which the articulatory events (e.g., decreasing the area of the constriction and increasing the flowrate) are consistently coordinated. An obvious candidate for a theoretical model would be a sequence of models of static configurations. In developing such a model, one must determine the combination of parameters controlling the amount of sound that is generated, the level above which this sound is deemed significant, and, in particular, whether the speed of the transition affects this level.

Stops, affricates, and voiced fricatives involve still more complex dynamic situations. In the case of stops, turbulence is generated briefly due to a sudden rush of air out of the abruptly opened mouth. For affricates, the constricted area is not released completely, so that the initial stop becomes a steady-state fricative. For voiced fricatives, the constriction remains constant in area, but the amount of air flowing through it varies periodically. As with transitions into unvoiced fricatives, it becomes important to identify the combination of parameters determining the amount of sound generation due to turbulence. It may well prove that for stops or voiced

fricatives, in which the flowrate is not constant. more sound is generated for a given configuration than would be for a static fricative. For voiced fricatives, in which sound is generated at two distinct locations in the vocal tract, the back-cavity resonances now become important.

Perceptual questions have been avoided in this thesis. They are unnecessary for understanding the acoustic phenomena, but are naturally relevant when considering speech synthesis. For instance, if the zero due to an obstacle is not audible, it is not particularly important that the synthesizer reproduce it exactly. A study of confusions between fricatives may well reveal that the very fricatives whose spectra seemed so similar, and were consequently difficult to model, are often confused by listeners as well. Likewise, fricatives produced by different speakers that are acoustically dissimilar may well be perceived as different versions of the same fricative. The development of a perceptual distance metric for fricative sounds and further refinement of the parameters used in this thesis for acoustic comparisons would allow us to quantify such distinctions. Finally, the dynamic characteristics may well aid the perception of each fricative. Thus, perception studies would tell us what to pay attention to in modeling.

In summary, this thesis has increased our understanding of the acoustic mechanisms of fricative consonants by the use of mechanical models, theoretical models, and analysis of speech and human whistling. A search for evidence of source-tract interaction led to a recognition of the differences between model-dependent and acoustic source-tract interaction, which should provide a useful conceptual framework with which to evaluate and develop theoretical models. In the future, this work should be extended in three directions. First, development of theoretical models, specifically for sound generation along a wall and of whistles, should be continued. Second, the subject of study, sustained fricatives, can be broadened to include all instances of turbulence in the vocal tract by consideration of increasingly complex dynamic models. Mechanical models constitute a useful technique with which to investigate these first two areas. Third, consideration of perceptual questions would increase the relevance for applications such as speech synthesis and recognition, and allow us to consider not only the articulatory-to-acoustic, but also the acoustic-to-perceptual transformation for sounds produced by turbulence in the vocal tract.

Appendix A

Establishment of an Absolute Reference Level for the Microphones

The Bruel and Kjaer microphone was calibrated so that its output could be converted to absolute sound pressure levels by using a General Radio Sound Level meter as a reference. The sound level meter itself was calibrated by use of a General Radio 1562-A Sound Level Calibrator. The calibration signals, all of the same amplitude, produced meter readings within .4 dB of each other over the 125 to 1000 Hz frequency range.

An oscillator and a white noise generator were used in turn to drive an AR-2A loudspeaker. The Bruel and Kjaer microphone and a General Radio Sound Level meter were positioned 10 cm from the AR-2A, taped together so the centers of the microphones were 2.5 cm apart. The outputs of the sound level meter and of the Bruel and Kjaer preamplifier fed into the two inputs of the spectrum analyzer. Then various signals were played and analyzed, and the averaged power spectra were transfered to the VAX. These signals were: white noise at two amplitudes 13 dB apart, and sine waves (one at a time) at frequencies between 160 and 1200 Hz, at two amplitudes for each frequency. The amplitudes were chosen to be above the level of the room noise. Meter readings of the Sound Level meter were taken at the time of the recording to provide a rough check of the output of the meter.

The two inputs were run through the whole system and a raw overall sound pressure level, from 0 to 10240 Hz, was found after compensating for the preamplifier and attenuator. Then the differences between the two inputs for each signal were averaged. Two correction factors were added to the average difference to arrive at the total correction factor: that resulting from the Sound Level meter calibration, and a correction for the overlapping bandwidths of the Hanning window used in the analysis. The latter correction was computed in order to enable comparisons among spectra generated with different windows or frequency ranges (Beranek, 1954). Since the ratio of the bandwidths of the Hanning (overlapping) to uniform (non-overlapping) window remains the same for any frequency range, this correction factor is a constant -1.76 dB, and the resulting sound pressure level will be based on the width of the bins for the frequency range used.

The final correction factor had a standard deviation of $s = 1.1$ dB. The correction factor is a constant across all frequencies, based upon a calibration performed only at low frequencies. This is forced upon us by the frequency range of the Sound Level meter calibrator, but is justified since the Bruel and Kjaer characteristic is extremely flat over the entire frequency range of interest.

For the speech recordings (see Ch. 4) a similar procedure was used to calibrate the output of the Altec and tape recorder. In this case only the meter reading of the Sound Level meter rather than the electrical output was used, and a single calibration signal of 600 Hz sufficed.

Appendix B

Relationship of Transfer Function Form to Network Topology

Consider a transmission line represented by a ladder network in which the series elements have finite impedance and the shunt elements have finite admittance. We wish to show that, with no side branches, the transfer function between input and output of like quantities (current to current or voltage to voltage) is allpole, that is, it is of the form

$$\frac{I_{out}}{I_{in}} = \frac{K}{s^m + b_{m-1}s^{m-1} + \cdots + b_0}$$

A schematic of the transmission line defining variables is shown in Fig. B.1a. It consists of a terminating impedance Z_{out} and n topologically identical sections, where the ith section contains a series impedance Z_i and a shunt admittance Y_i. We are considering the eigenstates of the circuit only. An eigenstate is a state such that, for some single value of s, every current in the circuit can be expressed as $I_i(s)\,e^{st}$ and every voltage can be expressed as $V_i(s)\,e^{st}$, with only the complex amplitudes $I_i(s)$, $V_i(s)$ varying with s. The e^{st} term can then be divided out of every current and voltage. In the following discussion, only the complex amplitudes are used, and their "(s)" designation is dropped for simplicity.

An allpole transfer function has no zeros, which means that an input that is non-zero at a given complex frequency must produce an output that is non-zero at that same frequency. One way to prove that a given transfer function is allpole is to prove the equivalent statement: assume that the output I_{out} is zero, and prove that that assumption implies that the input I_{in} must then also be zero. In mathematical notation, when we can prove that (if $I_{out} = 0$ then $I_{in} = 0$) is true for a given circuit at all finite frequencies, that is equivalent to proving that (if $I_{in} \neq 0$ then $I_{out} \neq 0$) is true for the same circuit, which is the same thing as saying that the transfer function I_{out}/I_{in} is allpole.

The assumption stated at the outset, that $Z_i \neq \infty$, $Y_i \neq \infty$ at any finite complex frequency, is certainly true if all of the series elements are inductors and shunt elements are capacitors, which forms an equivalent circuit for a lossless transmission line. It is also true for a lossy transmission line, which we will discuss later. Note that the output element is considered to be a series element if the output quantity is a current, or a shunt element if the output quantity is a voltage. Thus, for Fig. B.1a, Z_{out} could be represented by an inductor, which obeys the constraint of $Z_{out} \neq \infty$.

Working backwards through the network shown in Fig. B.1a, we can make the following statements. If $I_{out} = 0$ then $V_n = 0$, since $V_n = Z_{out}\,I_{out}$ and $Z_{out} \neq \infty$. $I_{Y_n} = Y_n\,V_n$; since $Y_n \neq \infty$, if $V_n = 0$, then $I_{Y_n} = 0$. Since $I_{Y_n} = I_{out} = 0$, $I_n = 0$. The voltage drop across Z_n is equal to $V_{n-1} - V_n = V_{n-1} = Z_n\,I_n$. Since $Z_n \neq \infty$, if $I_n = 0$, then $V_{n-1} = 0$.

We have shown that if $I_{out} = 0$, then $I_n = 0$ and $V_{n-1} = 0$. Since all of the sections are of the same form as the nth section, all of the I_i and V_i equal zero if $I_{out} = 0$. In other words, the only way to have a zero output for this network is for the entire network to be at rest. Therefore it is not possible for $I_{in} \neq 0$ when $I_{out} = 0$, and therefore I_{out}/I_{in} cannot have zeros.

For a lossy transmission line both series and shunt impedances acquire a real component. An equivalent circuit could be modeled as a resistance and inductance in series for the series elements, and a conductance and capacitance in parallel for the shunt elements. The assumption of $Z_i \neq \infty$. $Y_i \neq \infty$ is still met, and thus the proofs above still hold.

When can zeros arise? Suppose that we place an impedance Z_{back} to the left of I_{in}, as in Fig. B.1b. Note that Z_{back} is not a single series or shunt element: it is the driving-point impedance of the entire network in back of the source. A network whose elements obey the constraint of $Z_i \neq \infty$ and $Y_i \neq \infty$ may still have an overall driving-point impedance of

$$Z_{drive} = Z_1 + \cfrac{1}{Y_1 + \cfrac{1}{Z_2 + \cfrac{1}{Y_2 + \cdots}}} = 0$$

at some frequency. Then $I_{out} = 0$ does imply that $I_1 = 0$, but $I_1 \neq I_{in}$; now, $I_{in} = I_1 + I_{back}$. $V_{in} = I_1 Z_1 = 0$ since $I_1 = 0$, and $V_{in} = I_{back} Z_{back} = I_{in} Z_{back} = 0$. Therefore we can satisfy all conditions and have $I_{in} \neq 0$ whenever $Z_{back} = 0$. In other words, we have a zero in the transfer function I_{out}/I_{in} at frequencies where Z_{back} is a short circuit.

If we consider a cross-variable transfer function, such as I_{out}/V_{in}, we have the situation shown in Fig. B.1c. As before, the assumptions that $I_{out} = 0$ and none of the Z_i, Z_{out}, Y_i have poles lead to the conclusion that $I_1 = 0$. The voltage across Z_1 must then equal zero also since $Z_1 \neq \infty$, and the voltage across Z_1 is the source voltage V_{in}. Therefore $V_{in} = 0$.

If we add a back network in parallel with the source as we did for the current source, this result is unchanged: $V_{in} = 0$. However, if we add a back network in series with the voltage source, we have the situation shown in Fig. B.1d. Then the voltage across Z_1, which must still equal zero, is equal to $V_{in} + V_{back}$. Thus $V_{in} = -V_{back} = I_{back} Z_{back}$. Since $I_{back} = I_1 = 0$, the voltage across the back network and therefore the source voltage is nonzero only when $Z_{back} = \infty$. A current-voltage transfer ratio therefore has the opposite result of the current-current transfer ratio: a voltage source generates zeros in the transfer function at the frequencies of the poles of the back network, whereas a current source generates zeros at the frequencies of the zeros of the back network.

The same reasoning applies to a side branch at any point in the circuit. At the point of juncture, there are two paths for the current to take. In the case of the current-current transfer ratio, at any frequency at which $Z_{side} = 0$, $I_{out} = 0$ no matter what value I_{in} has. The transfer function then has zeros as well as poles.

If we wish to consider an output voltage rather than current, the voltage must be defined across a shunt rather than series element. The necessary assumption for the proof above becomes $Y_{out} \neq \infty$, and the proof proceeds as before. In general, a change in the output variable changes the assumption about the output impedance; a change in the input variable changes the frequencies at which zeros occur in the cases where the system is no longer allpole.

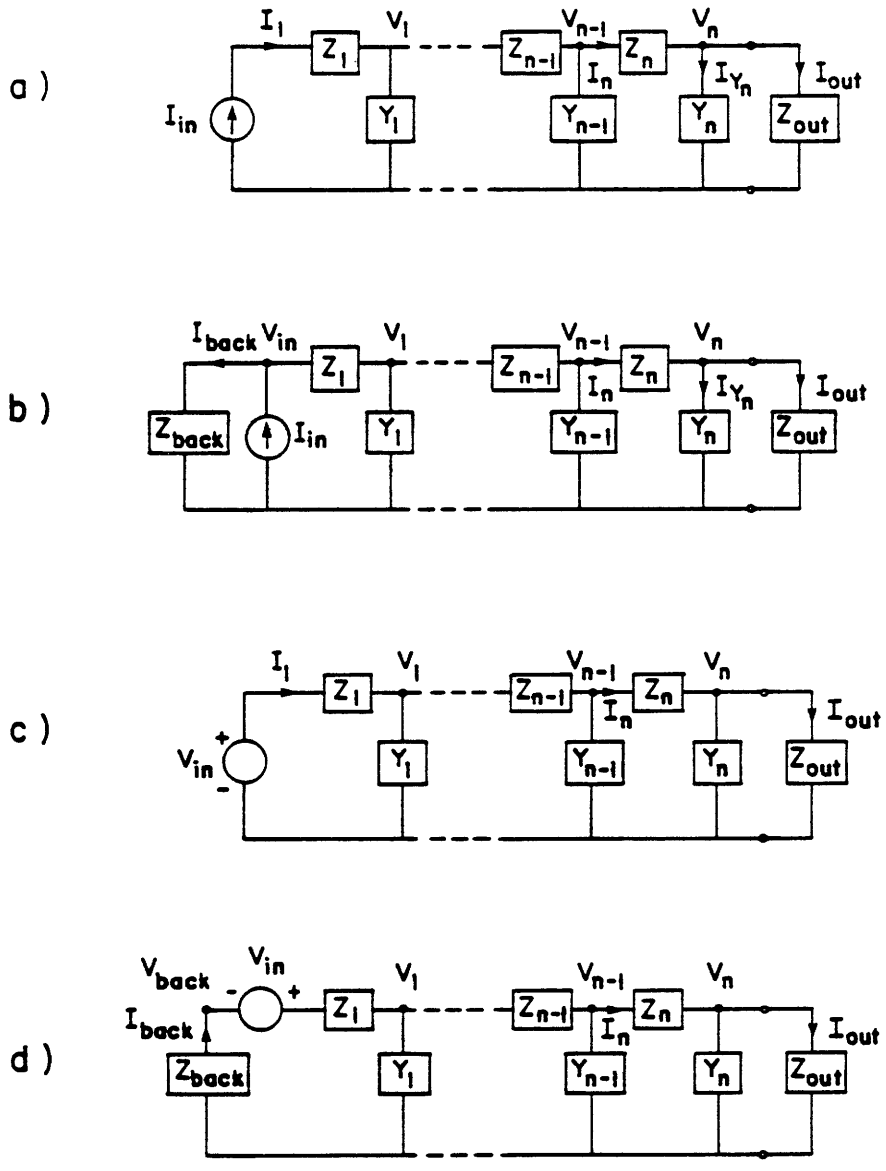


Figure B.1: Four network topologies, leading to different transfer function forms. Transfer functions between the indicated *in* and *out* quantities for networks a) and c) are allpole, for b) and d) have zeros as well as poles.

Appendix C

Complete Set of Transfer Functions for the Obstacle Case

The method of deriving the predicted transfer functions for the two tube configurations including an obstacle was explained in Chapter 3. Included in Tables C.1. C.1, and C.2 , are the final values specifying the transfer functions at the three flowrates of 160, 360, and 420 cc/sec. Specifications for 250 cc/sec were given throughout Section 3.1, with the final values tabulated in 3.1.5. The four flowrates span the range of flowrates used in the experiments, and are used consistently for comparisons and figures in Section 3.2.

As in Section 3.1, the values here represent the complex coordinates of the poles and zeros. For a complex singularity s_1 located at $\sigma_1 + j\omega_1$, the listed value of $f = \omega_1/2\pi$, and $\Delta f = \sigma_1/\pi$. For a real singularity, $f = \sigma_1/2\pi$.

As described in Section 3.2. some of the resonance frequencies for the $l_f = 12$ cm configuration were altered to match the peak values in the experimental data. The values changed were the first four front-cavity poles, which are distinguishable in the above tables by the lack of corresponding zeros. At all flowrates, these values were changed to 720, 1920, 3120 and 4680 Hz in order to compute the transfer functions.

l_f = 3.2 cm				l_f = 12 cm			
Poles		Zeros		Poles		Zeros	
f	Δf	f	Δf	f	Δf	f	Δf
75	490	53	real	330	640	110	real
		430	390			525	490
1365	180	1375	180	730	65		
1790	250			2010	110		
2720	170	2705	170	3370	270		
4055	165	4045	165	4390	530	4395	530
5395	165	5387	165	4870	490		
		5850	24			5850	24
6110	2340			6330	670		
6815	165	6735	165	7680	790		
8110	165	8080	165	8710	525	8750	525
9450	165	9420	165	9220	870		

Table C.1: Poles and Zeros for the Obstacle Case, U = 160 cc/sec.

l_f = 3.2 cm				l_f = 12 cm			
Poles		Zeros		Poles		Zeros	
f	Δf	f	Δf	f	Δf	f	Δf
120	real	50	real	150	1050	90	real
780	real	160	850			370	950
1365	189	1375	189	730	70		
1790	275			2010	115		
2720	170	2705	170	3370	270		
4055	170	4045	170	4390	535	4395	535
5395	165	5390	165	4870	490		
		5850	28			5850	28
6110	2340			6330	675		
6815	165	6735	165	7680	790		
8112	165	8080	165	8710	530	8750	530
9450	165	9420	165	9220	870		

Table C.2: Poles and Zeros for the Obstacle Case, U = 360 cc/sec.

$l_f = 3.2$ cm				$l_f = 12$ cm			
Poles		Zeros		Poles		Zeros	
f	Δf	f	Δf	f	Δf	f	Δf
110	real	50	real	440	real	90	real
920	real	660	real	730	real	300	1090
		315	real				
1365	190	1375	190	730	70		
1790	280			2010	120		
2720	175	2705	175	3370	270		
4055	170	4045	170	4390	540	4395	540
5395	165	5387	165	4870	490		
		5850	29			5850	29
6110	2340			6330	675		
6815	165	6735	165	7680	790		
8112	165	8080	165	8710	530	8750	530
9453	165	9420	165	9220	870		

Table C.3: Poles and Zeros for the Obstacle Case, $U = 420$ cc/sec.

Appendix D

Complete Set of Transfer Functions for the No-Obstacle Case

The method of deriving the predicted transfer functions for the two tube configurations without an obstacle was explained in Chapter 3. Included in Tables D.1 and D.2 are the final values specifying the allpole transfer functions at the two flowrates of 375 and 470 cc/sec. Specifications for 275 cc/sec were given in Section 3.1.6. These three flowrates span the range of flowrates used in the experiments, and are used consistently for comparisons and figures in Section 3.2.

As in Section 3.1, the values here represent the complex coordinates of the poles of the transfer functions. For a complex poles s_1 located at $\sigma_1 + j\omega_1$, the listed value of $f = \omega_1/2\pi$, and $\Delta f = \sigma_1/\pi$. For a real pole, $f = \sigma_1/2\pi$.

As described in Section 3.2, some of the resonance frequencies for the configurations were altered to match the peak values in the experimental data. For the $l_f = 3.2$ cm configuration, the frequency of the second pole was altered at all flowrates from 7410 to 6400 Hz. For the $l_f = 12$ cm configuration, the frequencies of the first six poles were altered at all flowrates from the values listed above to 680, 2040, 3360,4720, 6040, and 7360 Hz.

$l_f = 3.2$ cm		$l_f = 12$ cm	
f	Δf	f	Δf
2160	387	730	70
7410	2600	2030	119
		3410	280
		4840	487
		6310	670
		7800	800
		9290	870

Table D.1: Poles for the No-Obstacle Case, $U = 375$ cc/sec.

$l_f = 3.2$ cm		$l_f = 12$ cm	
f	Δf	f	Δf
2160	393	730	67
7400	2605	2030	121
		3410	279
		4840	490
		6310	670
		7800	800
		9290	870

Table D.2: Poles for the No-Obstacle Case, $U = 470$ cc/sec.

References

Anderson, A.B.C. (1954) "A Jet Tone Orifice Study for Orifices of Small Thickness-Diameter Ratio," *J. Acoust. Soc. Am.* 26:1, 21-25.

Bendat, J.S. and A.G. Piersol (1971) *Random Data: Analysis and Measurement Procedures.* New York, John Wiley.

Beranek, Leo L. (1954) *Acoustics.* New York: McGraw-Hill Book Company.

Beranek, Leo L. (1949) *Acoustical Measurements.* New York: John Wiley and Sons.

Binder, Raymond C. (1962) *Fluid Mechanics.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

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

Curle, N. (1955) "The influence of solid boundaries on aerodynamical sound," *Proc. Roy. Soc.*, (London), Ser. A, Vol. 231, No. 1187, p.505-514.

Chanaud, R.C. and A. Powell (1965). "Some experiments concerning the hole and ring tone", *J. Acoust. Soc. Am.* 37 no. 5, 902-911.

Elder, S.A., Farabee, T.M., and Demetz, F.C. (1982). "Mechanisms of flow-excited cavity tones at low Mach number", *J. Acoust. Soc. Am.* 72, 532-549.

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

Fant, C.G.M. (1983) "Preliminaries to Analysis of the Human Voice Source," *Speech Communications Group Working Papers* Vol. 3, 17-40.

Flanagan, J. (1972) *Speech Analysis Synthesis and Perception.* New York: Springer-Verlag.

Flanagan, J.L. and Ishizaka, K. (1976). "Automatic generation of voiceless excitation in a vocal cord-vocal tract speech synthesizer", *IEEE-ASSP* 24 no. 2, 163-170.

Flanagan, J.L., Ishizaka, K., and Shipley, K.L. (1980). "Signal models for low bit-rate encoding of speech", *J. Acoust. Soc. Am.* 68, 780-791.

Flanagan, J.L., Ishizaka, K., and Shipley, K.L. (1975). "Synthesis of speech from a dynamic model of the vocal cords and vocal tract", *Bell System Technical Journal* 54 no.3, 485-506.

Fletcher, N.H. and Thwaites, S. (1983). "The Physics of Organ Pipes", *Scientific American*, 248(1), 94-103.

Goldstein, M.E. (1976) *Aeroacoustics.* McGraw-Hill, New York.

Heinz, J.M. (1956) "Fricative Consonants," M.I.T. Research Laboratory of Electronics *Quarterly Report*, October-December, 5-7.

Heinz, J.M. (1958) "Sound generation by turbulent flow in an acoustic resonator", unpublished S.M. thesis, Department of Electrical Engineering and Computer Science, M.I.T.

Heinz, J.M., and Stevens, K.N. (1961). "On the properties of voiceless fricative consonants", *J. Acoust. Soc. Am.* 33 no. 5, 589-596.

Henke, W. (1966) "Dynamic Articulatory Model of Speech Production Using Computer Simulation", Appendix B, Unpublished Ph.D. thesis, Department of Electrical Engineering and Computer Science, M.I.T.

Hewlett-Packard Co. (1978) "Signal Averaging with the HP 3582A Spectrum Analyzer", Application Note 245-1.

Holger, D.K., T.A. Wilson and G.S. Beavers (1980). "The amplitude of edgetone sound", *J. Acoust. Soc. Am.* 67 no. 5, 1507-1511.

Holger, D.K., T.A. Wilson and G.S. Beavers (1977). "Fluid mechanics of the edgetone," *J. Acoust. Soc. Am.* 62 no. 2, 1116-1128.

Hosein, P. (1983) "Computer Simulation of a Model for the Vocal Tract," Unpublished B.S. thesis, Department of Electrical Engineering and Computer Science, M.I.T.

Hughes, G. and M. Halle (1956) "Spectral Properties of Fricative Consonants," *J. Acoust. Soc. Am.* 28:2, 303-310.

Jassem, W. (1962) "The Formant Patterns of Fricative Consonants," *STL-QPSR* 3, p. 6-15.

Jakobson, R., C.G.M. Fant and M. Halle (1963) *Preliminaries to Speech Analysis.* Massachusetts Institute of Technology Press, Cambridge, Mass.

Lighthill, M.J. (1952) "On Sound Generated Aerodynamically: I. General Theory," *Proc. Roy. Soc.* A211, 564.

Lighthill, M.J. (1954) "On Sound Generated Aerodynamically: II. Turbulence as a Source of Sound," *Proc. Roy. Soc.* A222, 1.

Lindblad, Per (1980) "Some Swedish Sibilants," Travaux de l'Institut de Linguistique de Lund v. 16.

Meyer-Eppler, W. (1953). "Zum Erzeugungsmechanismus der Gerauschlaute", *Z. Phonetik* 7, 196-212.

Morse, P. (1976) *Vibration and Sound.* American Institute of Physics, New York.

Morse, P.M. and K.U. Ingard (1968) *Theoretical Acoustics.* McGraw-Hill, New York.

Pendergrass, N.A. and J.S. Farnbach (1978) "A High-Resolution, Low-Frequency Spectrum Analyzer," *Hewlett-Packard Journal* September 1978, 2-14.

Pollack, M.L. (1980). "Flow-induced tones in side-branch pipe resonators", *J. Acoust. Soc. Am.* 67 no. 4, 1153-1156.

Potter, J.L. and S.J. Fich (1963). *Theory of Networks and Lines.* Prentice-Hall, Inc., New Jersey.

Powell, A. (1961) "On the edgetone", *J. Acoust. Soc. Am.* 33 no. 4, 395-409.

Powell, A. (1962). "Vortex action in edgetones", *J. Acoust. Soc. Am.* 34 no. 2, 163-166.

Rothenberg, M. (1981) "Acoustic Interactions between the Glottal Source and the Vocal Tract," in K.N. Stevens and M. Hirano (eds.) *Vocal Fold Physiology*, University of Tokyo Press, 305-328.

Schlichting, H. (1979). *Boundary Layer Theory.* 7th ed., McGraw-Hill, New York.

Shadle, C.H. (1983). "Experiments on the Acoustics of Whistling", *The Physics Teacher* 21 no. 3, March 1983, 148-154.

Stevens, K.N. (1971) "Airflow and Turbulence Noise for Fricative and Stop Consonants: Static Considerations," *J. Acoust. Soc. Am.* 50:4:2, 1180-1192.

Strevens, P. (1960) "Spectra of Fricative Noise in Human Speech," *Language and Speech* 3, 32-49.

Succi, G.P. (1977) "The interaction of sound with turbulent flow", M.I.T. Ph.D. thesis, published as Gas Turbine and Plasma Dynamics Lab Report no. 140.

Thomas, N. (1955) "On the production of sound by jets", *J. Acoust. Soc. Am.* 27 no. 3, 446-448.

Thwaites, S. and N.H. Fletcher (1982). "Wave Propagation on Turbulent Jets: II. Growth," *Acustica* 51, 44-49.