In The Handbook of Phonetic Sciences 2nd ed.
eds. Hardcastle, Laver & Gibbon Wiley-Blackwell
2010

2 The Aerodynamics of Speech

1624

CHRISTINE H. SHADLE

1 Introduction

Aerodynamics is the study of the motion of air. It is a subset of fluid mechanics, since air is only one possible fluid; it is a subset in another sense because mechanics includes statics as well as dynamics, but one must understand something about fluid statics in order to consider dynamics. Acoustics is the study of sound, and sound involves a particular type of wave traveling through a medium. Acoustics in air is therefore a part of aerodynamics.

These distinctions may seem pedantic, but they are blurred often in speech research and result in some confusion. Aerodynamics in speech tends to be thought of as "everything the air is doing that isn't sound." In speech we ultimately care only about the sound that is radiated to the far field, well outside the vocal tract. Here, near the microphone or someone's ear, the air is essentially at rest except for the sound wave, and describing that wave is an acoustics-only problem. However, inside the vocal tract, the air is not at rest; we speak, for the most part, while exhaling, and the sound waves travel through that moving airstream. Further, most speech sounds are generated by that airstream: it sets the vocal folds vibrating which in turn chop up the steady airstream, and it can become turbulent and generate noise.

The chief difficulty in considering acoustics and aerodynamics of the vocal tract together is that they operate at different time and spatial scales. Convection velocities – how fast air moves from glottis to lips – are very slow compared to the velocity of sound. Conversely, the spatial resolution needed to model turbulence and its sound-generating mechanisms is much greater than that needed to model sound propagation. The usual approach is thus to consider the larger picture -- i.e., the nonacoustic aerodynamics – in order to define the acoustic sources that are operating, include these sources in a model that considers only acoustic waves, and thereafter ignore the moving stream of air. However, our understanding of the various types of sources is not very far advanced in some cases, and the limitations of these definitions need to be understood. Further, some sources

continue to interact with the moving air, and thus are less suited to such a separation of "acoustic" and "aerodynamic" function.

We also need to consider the wider aspects of aerodynamics when we measure speech or any aspect of speech production. There can be obvious effects, like the need to avoid breath noise on a microphone, or more subtle effects, like the limitations of inverse filtering. One can devise certain methods of recording various parameters in speech that avoid pitfalls, but new measurement techniques are developed all the time. It is important to be aware of the issues involved.

Aerodynamics texts are rarely written with speech applications in mind, and tend also to be highly mathematical. In spite of the high level of mathematics required, there are topics that currently resist any analytical solution, and must be dealt with empirically. In this chapter mathematics is not avoided altogether, but the chief aim is to convey an appreciation for the physical mechanisms involved, provide a pointer to more detailed treatments of each subject, and describe some of the limitations in our current understanding of the aerodynamics of speech.

In section 2 we describe some basic aerodynamic concepts and define the variables and nondimensional parameters needed. In section 3 we use these basic concepts to consider mechanisms of speech production, grouped in terms of the aerodynamic behavior(s) present. In section 4 we consider measurement methods and their limitations, including methods in general use and those adapted for speech research. Finally, in section 5 we discuss some models of speech production that incorporate aerodynamics.

2 Basic Considerations

We will first consider fluid statics, that is, the behavior of a fluid at rest, and the properties of a sound wave moving through it. Then we consider fluid dynamics. The motion of a fluid can alter sound passing through it; it can also generate sound, with the properties of that sound depending on the fluid motion and its interaction with its boundaries.

Air has a mass and a springiness, or compressibility. It takes energy to move air or to compress it, and the air imparts energy to an object that stops it from moving or confines it to its container when it expands. In a static situation – a set number of air molecules sealed in a container – the behavior of the air is described by its pressure, volume, and temperature, by the relation

$$PV = nRT \tag{1}$$

where P = pressure, V = volume, T = temperature, R = the universal gas constant, and n = mass of gas in moles (Halliday & Resnick, 1966). So, for this sealed-up gas where n cannot vary, if the temperature increases, the pressure or the volume or both must also increase; if the temperature stays the same, any increase in pressure must be offset by a corresponding decrease in volume, and vice versa.

The temperature, T, affects the density, ρ, viscosity, ν, and speed of sound, c, in a gas. Equation (1) can be used to derive the equations relating T to ρ, ν, and c. Values of these parameters for humid air at body temperature have been computed and are listed in the Appendix.

We are treating the enclosed mass of gas as though the pressure everywhere within it were constant. This is not strictly true: the gas at the bottom of the container has the weight of the gas above pressing on it, so its pressure is slightly greater. Because the density of air is low, it takes a very tall container for this effect to be noticeable: an increase in altitude of 1 km decreases atmospheric pressure by only 10 percent, for instance (Halliday & Resnick, 1966). But in a liquid, which is more dense, the effect is more noticeable, and this is exploited in the operation of the manometer, a basic instrument for measuring static pressure. In the manometer, a U-tube of constant inner diameter contains a liquid of known density ρ'. One end of the tube is attached to the gas with the pressure P to be measured; the other end is attached to a gas at a reference pressure P_0 (if that end is left open, the reference pressure is atmospheric pressure). The difference in the height of the liquid in each arm of the tube, h, is proportional to the difference in pressure:

$$P - P_0 = \rho' g h \tag{2}$$

where g is the gravitational acceleration. A denser liquid, with higher ρ', will show a smaller difference in height for the same pressure difference. Thus atmospheric pressure at sea level is 76.0 cm of mercury and 1,033 cm of water. The subglottal pressure during speech can range from 3 to 30 cm H_2O above atmospheric pressure; for such a relatively small value, the pressure can be measured more accurately by using water.

A sound wave traveling through a fluid that is otherwise at rest consists of a longitudinal pressure-rarefaction wave. This means that particles of the fluid are alternately pressed together more tightly than normal and pulled apart further than normal. As the wave travels through the fluid, individual particles oscillate about their original positions, but do not have a net movement. The molecules in the compressed regions tend to move towards the rarefied regions, so that particles in the rarefied regions have higher velocity. This tendency towards re-establishing equilibrium moves the high- and low-pressure regions along at a speed regulated by the properties of the fluid: the speed of sound.

The ideal gas law given in equation (1) can be simplified when we are talking about the pressure and volume changes induced by a sound wave traveling through air. In this case the gas is undergoing an adiabatic process, which means that no heat flows into or out of the system. Then

$$PV^{\gamma} = \text{constant} \tag{3}$$

where V = volume and $\gamma = 1.4$ for air. Note that this is not the same as saying that the temperature remains constant; instead, it says that if the temperature changes, it must change back again quickly before any heat exchange can take

place. When a sound wave travels through air, the pressure at a given location increases and then decreases. The temperature locally rises and falls, but the sound wave passes through so quickly that it behaves adiabatically.

Pressure and particle or volume velocity of the fluid as a function of time and location in space are the basic quantities used to describe a sound wave. They can also be used to describe a fluid in motion without a sound wave traveling through it. As the name indicates, particle velocity, v, is the velocity at a specific point in a fluid, and is expressed in units of distance per unit time; a particle at that location will have that velocity. The volume velocity, U, instead describes the rate of volume flow per unit time past a particular cross-sectional area. Any differences in particle velocity across that area will be averaged out by the description in terms of volume velocity.

There are many different types of fluid flow; recognizing which type occurs in a certain situation allows one to simplify the equations describing the fluid motion accordingly. One of the simplest types of flow to describe is steady, incompressible flow. Steady flow means that the flow does not change in time: if we measure pressure and particle velocity at a particular point, the values will remain the same even as the flow continues past our measurement point. This means that the flow cannot be turbulent, since turbulence implies that pressure and velocity will vary randomly in space and time. But nonrandom changes over time are excluded as well: if the overall flowrate is very slowly increased and then decreased without producing turbulence, it is still not a steady flow.

Liquids are very nearly incompressible; gases, with their lower density, are compressible. Sound waves cannot exist unless a fluid is compressible. However, describing a fluid flow as incompressible does not mean we are restricting ourselves to liquids: it means that we are ignoring the compressible effects in our model. So, assuming steady, incompressible flow in a duct allows one to derive a form of Bernoulli's Equation relating the pressure and velocity at two places along the flow, assuming no work, heat transfer, or change of elevation occurs between those two places:

$$-gH_L = \frac{p_2 - p_1}{\rho} + \frac{v_2^2 - v_1^2}{2} \tag{4}$$

where g = the gravitational acceleration, H_L = head loss (or energy per unit weight lost to friction) from point 1 to point 2, p_1, p_2 = static pressure at points 1 and 2, v_1, v_2 = particle velocity at points 1 and 2, and ρ = density. We can use the relation of volume velocity to particle velocity $U = vA$ and the fact that the volume velocity will be the same at any point along the duct to rearrange the equation. The head loss is related to the internal energy of the fluid; because the fluid has friction, some energy is converted to heat. If we assume that the flow is frictionless, $H_L = 0$, and do some rearranging, we get:

$$U = \frac{A_2}{\sqrt{1-(A_2/A_1)^2}} \sqrt{\frac{2(p_1 - p_2)}{\rho}} \tag{5}$$

where U = volume flowrate (m^3/s), A_1, A_2 = cross-sectional flow areas at points 1 and 2 (m^2), p_1, p_2 = static pressure at points 1 and 2 (Pa), and ρ = density of the fluid (kg/m^3). Although this equation strictly applies only to frictionless, incompressible, steady flow, it is used in practice where these restrictions are violated to measure volume flowrates. The calibration procedures and empirical coefficients that can render such practice more accurate are discussed briefly in section 4, and more extensively in Doebelin (1983).

All fluids are viscous; as a result, the head loss can become significant for flow along a length of pipe. It is proportional to the length of pipe and to the flow velocity squared, but the constant of proportionality is an empirically determined friction factor that depends on the nondimensional parameters of wall roughness and Reynolds number. The Reynolds number is defined as the ratio of inertial to viscous forces, and can be determined by:

$$Re = VD/\nu \qquad (6)$$

where V = a characteristic velocity, D = a characteristic dimension, and ν = the kinematic viscosity. For pipe flow, the V normally used is the average particle velocity in the center of the pipe (and, because of the averaging, is therefore typically capitalized in the literature, confusing it with volume) and D is the pipe diameter (Massey, 1984). Although we are not often called upon to compute the head loss in the vocal tract, the Reynolds number is used in models of speech production, and it is therefore important to understand what it means.

All fluid motion can be broadly classified into three regimes: laminar, unstable, and turbulent flow. For a particular geometry – take, for example, a constricted region in a duct – the flow progresses from one regime to the next as the Reynolds number is increased. For a particular size of that geometry, this could be observed simply by increasing the flow velocity. In laminar flow, at the lowest velocity range, individual particles follow paths that do not cross paths of other fluid particles. The particles nearest the walls of the duct will move the slowest, constrained by friction to stick to the non-moving walls. In the center of the duct the particles will move the fastest. In going through a constriction the flow will hug the walls of the duct, and the velocity gradient and therefore the velocity in the center of the duct will increase as the area decreases. Laminar flow is dominated by friction forces, and the empirical friction factor is highest for lowest Reynolds numbers.

As the flow velocity increases, inertial forces begin to dominate over friction forces. As the fluid enters the constriction, it overshoots a bit, and the moving flow separates from the walls. The vena contracta, thus formed, effectively reduces the area of the constriction. The region of transition from the fast-moving flow to the still flow near the walls is known as a boundary layer, and it can itself become unstable. In an unstable regime, any perturbations will tend to increase in amplitude.

If the Reynolds number is increased still further, the flow may pass through a sequence of unstable states, but eventually it becomes fully turbulent. Here

inertial effects dominate. Paths of fluid particles cross each other unpredictably, so the flow as a whole has a random fluctuating component superimposed on the mean flow. This is very effective at mixing the flow.

For a particular geometry the characteristic velocity and dimension can be defined, and then a critical Reynolds number Re_{crit} can be found that marks the change from laminar to unstable flow regimes. This means that the flow regime can be predicted for any velocity in any size of that geometry. The value of Re_{crit} may differ though for a square instead of circular pipe, for instance, or a rectangular instead of circular constriction. The behavior above Re_{crit} may also depend on geometry: for fully turbulent flow in smooth pipes the friction factor decreases with increasing Re, but for rough pipes it remains relatively constant (Massey, 1984).

Sound waves traveling through a fluid can be affected by the flow regime. First, turbulence can diffract and absorb sound waves, though it is questionable whether this is a significant effect for speech (see discussion in Davies et al., 1993). Second, the sound wave traveling through a moving medium will travel faster downstream than upstream relative to an observer at rest. We can gauge the strength of this effect by computing the average Mach number $M = V/c$, where V = the average particle velocity of the fluid. In a vowel, where average volume velocity $U = 200$ cm^3/s and the most constricted region has an area of approximately $A_c = 1$ cm^2, the Mach number in the constricted region will be $M = U/(A_c c) = 200/(1 \cdot 35{,}000) = 0.0057$. Since $M \ll 1$, this effect is not significant. However, for fricatives, a typical $U \approx 600$ cm^3/s and $A_c \approx 0.1$ cm^2, so $M = U/(A_c c) = 600/(0.1 \cdot 35{,}000) = 0.17$. Here the value of M relative to 1 indicates that the convection velocity is significant with respect to the speed of sound, and may have to be taken into account.

In addition to these effects that flow can have on sound traveling through it, flow can also generate sound, with different characteristics according to the type of flow that produced it.

An unstable flow regime can lead to a self-sustaining aerodynamic oscillation. One or more positive feedback paths must exist. The sound that can result is characteristically high-amplitude, narrow-bandwidth: a whistle. Its frequency and the parameters that control it are related to the underlying instability.

We spoke earlier of the boundary layer that can detach from the walls of an orifice. In fact, a boundary layer exists between any two regions with significantly different flow parameters: they may have different velocities (as with the fluid moving in the center of a duct and the still fluid clinging to the walls of the duct), different densities (as with the Gulf Stream, which is warmer and saltier than the surrounding water), or actually be two different as yet unmixed fluids (cream just poured into coffee). The boundary itself is unstable for certain ranges of the difference of the two parameter values. In this unstable range, any small perturbation of the boundary will tend to grow. At first this will appear as ripples on the boundary; the ripples grow larger and curl up into vortices, which continue to rotate while being convected downstream.

The length of time required to traverse the feedback path tends to determine the spacing between vortices, because the initial perturbations are reinforced at

that interval. In general an integral number of vortices will be found between abrupt discontinuities such as the two ends of a sharp-edged orifice, or the distance between an orifice exit and an edge. These patterns, and the sound generated, will couple into the resonances of a surrounding cavity. Increasing the flow velocity will tend to increase the frequency of the sound produced, but not uniformly; it will remain steadily coupled into one resonance, then jump abruptly to the next higher one, with the jumps exhibiting hysteresis (Chanaud & Powell, 1965; Holger et al., 1977).

Turbulence also generates sound, but since the motion is more random than an unstable state reinforced by feedback, the sound that results is noise with a relatively flat spectrum. Such noise cannot be predicted precisely from moment to moment, but can only be characterized statistically, and modeled by a collection of idealized flow-generated noise sources: the flow monopole, dipole, or quadrupole. These are analogous to idealized acoustic sources. The *acoustic* monopole can be thought of as produced by a pulsing sphere, which generates spherical sound waves. The acoustic dipole consists of two adjacent out-of-phase monopoles, which generate sound waves that interfere with each other; the result is a characteristic figure-eight directivity pattern. Solid objects such as a piston or a loudspeaker cone that act on the air can be modeled using these acoustic sources: in the far field, the directivity pattern observed is the same as that which would be produced by the idealized sources used to model it. In a *flow* source, the flow of air itself acts upon the surrounding air so that the far-field sound exhibits monopole, dipole, or quadrupole properties. Theoretically, the noise generated by turbulence away from any solid boundaries appears in the far field as if it were produced by flow quadrupoles; the noise generated by turbulence that results in a fluctuating force being applied to a solid object, by flow dipoles. As with acoustic sources, these can be thought of as collections of four and two flow monopoles, respectively, pulsing out of phase. In each case, the source strength depends upon the flow velocity. The total sound power of a flow quadrupole is proportional to V^8; that of a flow dipole, to V^6, and a flow monopole, to V^4. However, the flow quadrupole is much less efficient than a flow dipole, which in turn is less efficient than a flow monopole. It can be shown that the ratio of the total sound powers of the flow quadrupole to the flow dipole, or of the flow dipole to the flow monopole, is proportional to the Mach number squared (Goldstein, 1976). Thus, for $M < 1$, if a flow generates both dipole and quadrupole sources, the dipole sources will have higher sound power even though the sound power of the quadrupole sources increases faster with an increased flow velocity.

If the far-field sound pressure of a jet is recorded for a variety of jet sizes and mean velocities, the results can best be compared by plotting a normalized spectrum. A spectrum typically shows a measure of amplitude, such as sound pressure level, versus frequency. Every variation of V and D would result in a different curve. If we compared two circular jets with the same velocity V but different diameters, the larger jet will produce higher-amplitude noise with the peak at a lower frequency than that produced by the smaller jet. However, we can normalize the sound pressure level by dividing it by V^8D^2, which reflects the theoretically

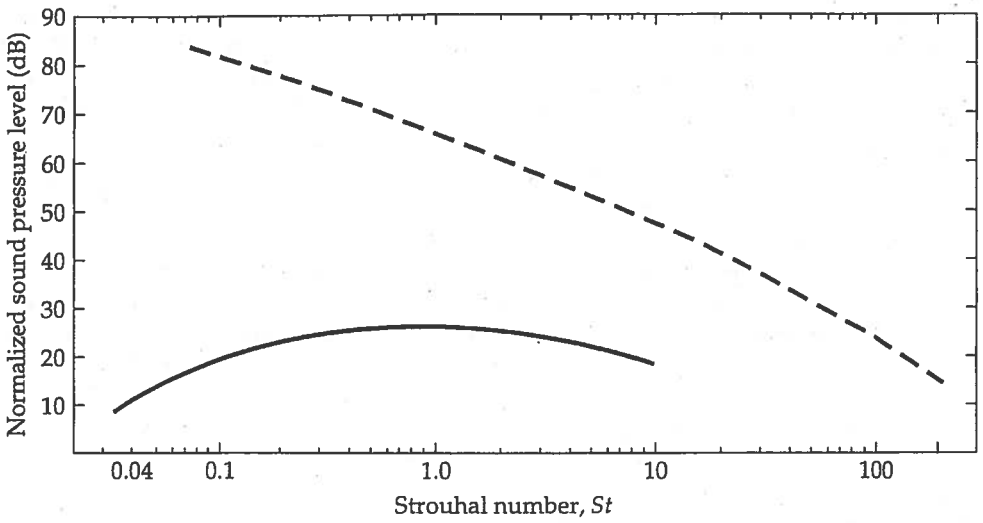

Figure 2.1 Normalized spectra of the noise generated by (a) free, subsonic jet noise (solid line) and (b) flow past a spoiler in a duct (dashed line), for various sizes of jets and spoilers. One-third octave sound pressure level is normalized by V^8D^2 for (a), by V^6 or V^4 for (b). Levels of the two curves relative to each other are arbitrary. (After Goldstein, 1976, and Nelson & Morfey, 1981)

predicted variation with V and D. We can also normalize the frequency axis by plotting instead the Strouhal number, St, where

$$St = \frac{fD}{V} \tag{7}$$

This will cause the peak frequencies to be aligned. As a result of these normalizations, all jet spectra for any size and velocity (as long as $V < c$) fit the same curve, as shown by the solid curve in Figure 2.1.

A similar collapse of data can be done for the noise produced by flow past a spoiler in a duct. The presence of the duct changes the dependence of source strength on V below the first cut-on frequency[1] (Nelson & Morfey, 1981). However, the principle of collapsing the data by using nondimensional parameters is the same. Here normalization by V^4 below cut-on and V^6 above cut-on frequency for the duct is used, and the resulting curve has a different shape from that of the free jet, as shown by the dashed curve in Figure 2.1.

The Strouhal number can be thought of in many ways. Equation (7) is derived by finding the ratio of the acceleration due to the unsteadiness of the flow to the convective acceleration due to the nonuniformity of flow. So, for small St, the unsteady component is relatively small. If $St < 10^{-2}$, the flow is quasisteady to a first approximation (Pelorson et al., 1994). The Strouhal number can also be used to characterize the shedding frequency of vortices from a jet. The frequency will

depend on the jet velocity and diameter, but similar jets (same shape and thus behavior, even though of different D and V) will have the same St corresponding to the shedding frequency (Sinder, 1999).

3 Aerodynamically Distinct Tract Behaviors

In this section we consider the different mechanisms of speech production, grouping them from an aerodynamics point of view and proceeding from the simplest to the most complex. In each section we describe the physical events, and give parameter values typical for speech.

3.1 Breathing

Respiration is the simplest tract behavior aerodynamically because sound generation is not essential to the process and the time scales are relatively long.

The trachea extends about 11 cm below the larynx and then branches into the bronchial tubes. The bronchi continue to branch until the small, elastic-walled alveolar sacs are reached. The entire spongy mass is encased within the pleural sacs, which are suspended in the rib cage and surrounded on all sides by muscles. There are two sets of muscles that decrease lung volume when tensed: the internal intercostals, attached to the ribs, and the abdominal muscles. There are two sets that increase lung volume when tensed: the external intercostals, and the diaphragm, suspended across the bottom of the rib cage. By tensing and relaxing these sets of muscles in turn we can actively breathe in and out (as described by Hixon et al., 1973, and Hixon et al., 1976). But we can also use the elastic recoil force of the lung tissue itself as a passive mechanism for exhalation: if we cease to actively hold the rib cage expanded, we will passively exhale until the lung volume is small enough that the elastic recoil force no longer operates (see Figure 2.2). The lungs will not be empty at this point; the volume of air still in them is termed the functional residual capacity (FRC). To empty our lungs further we must actively tense muscles, and even doing so, we cannot empty them below a residual volume (RV).

The total lung capacity (TLC) in an adult male is approximately 7 liters of air. The RV is approximately 2 liters. The FRC varies with posture, but is typically 4 liters. The vital capacity is the maximum amount of air that can be exchanged in one breath, and is the difference between total lung capacity and residual volume, or about 5 liters (Ohala, 1990).

Typical respiration involves actively expanding lung volume (and therefore inhaling) to about 0.5 liters above FRC, and passively letting elastic recoil deflate the lungs (and therefore exhaling) back to the FRC. A typical respiration rate is 15 to 20 breaths per minute (Thomas, 1973). We hold the vocal folds as far apart as possible during inspiration (maximum area is 52 percent of tracheal area according to Negus, 1949; tracheal area ranges from 3.0–4.9 cm^2 according to Catford, 1977) and keep the tongue relaxed and velum down to provide a relatively

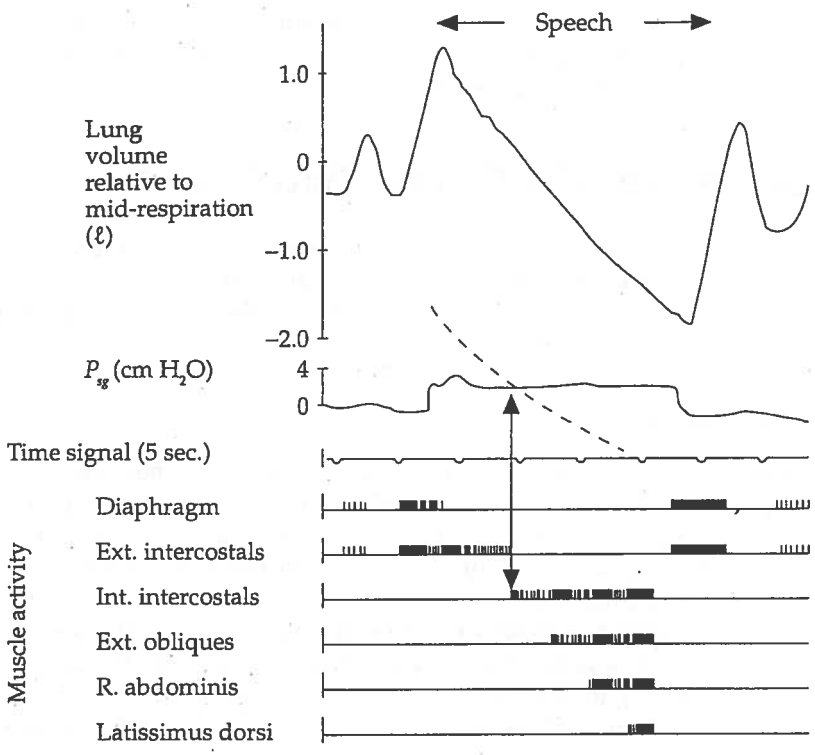

Figure 2.2 Lung volume versus time during speech and respiration, showing measured lung volume and subglottal pressure, and diagrammatic representation of the muscle activity. The dashed line indicates the relaxation pressure. (From Draper et al. (1959). Reprinted with permission from "Respiratory muscles in speech" by M. H. Draper, P. Ladefoged, and D. Whiteridge, *Journal of Speech and Hearing Research*, 7, 20. Copyright 1959 by American Speech-Language-Hearing Association. All rights reserved.)

unimpeded path for the air. During expiration the glottal area is smaller, but still of the order of 1 cm^2 (Sawashima, 1977); this is wide open compared to phonation, with an average glottal area of 0.05–0.1 cm^2.

For short utterances of speech at normal level, normal expiration is sufficient. For louder and/or longer speech, we need to use muscles actively to inhale more deeply, to offset the greater relaxation pressure, and to expel air below the FRC. During speech our goal appears to be to hold the subglottal pressure P_{sg} approximately constant, at a level corresponding to the loudness level of speech. It ranges from 3–30 cm H_2O (with normal speech typically 5–10 cm H_2O), as deduced by measuring esophageal pressure (Draper et al., 1959; Slifka, 2003) or by using tracheal puncture to measure the pressure directly (Isshiki, 1964). The lung volume then decrements fairly steadily; during stops the rate of decrement decreases momentarily, and during fricatives it increases (see Figure 2.3).

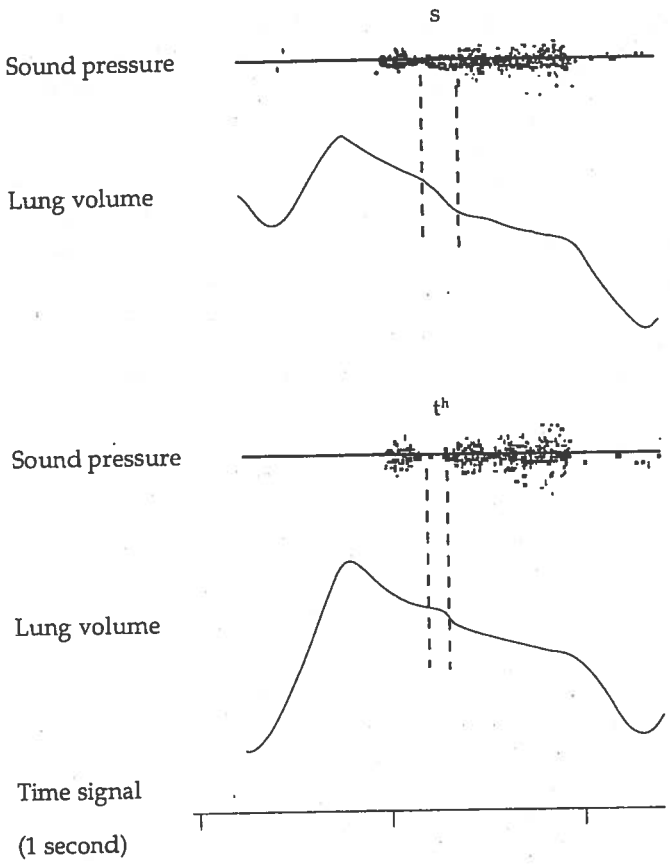

Figure 2.3 Lung volume versus time during the phrase "Deem–oon real," where the blank was filled in by [s] (top) and [tʰ] (bottom). (From Ohala (1990). Reprinted from J. Ohala, "Respiratory Activity in Speech," in *Speech Production and Speech Modelling*, eds. W. J. Hardcastle and A. Marchal, p. 36, copyright 1990, Kluwer Academic Publishers, with kind permission of Springer Science and Business Media.)

The respiratory system can be modeled uncontroversially as a simple mechanical system, as described by Draper et al. (1959): a set of bellows, with one active force (external intercostals and diaphragm) pulling outwards on the handles, one active (internal intercostals and abdominal muscles) and one passive force (elastic recoil) pulling inwards, and a variable-resistance opening in the bellows. What remains controversial, however, is the control mechanism for such a model. Ohala (1990) asserts that we either aim for a constant pressure to be applied to the lungs, or a long-term constant lung-volume decrement, and provides evidence to support the former. In particular, he argues that observed variations in subglottal pressure and in lung-volume decrement are due to variations in the downstream

flow resistance and to the inertia of the system, i.e., the time it takes to re-establish equilibrium. It is also true, however, that stressed syllables during an utterance are correlated with bursts of activity in the internal intercostal muscles (Draper et al., 1959). Both passive and active factors, then, may account for variations in the rate of lung-volume decrement.

The issue of passive versus active control mechanisms is also important in explanations of f_0 declination across the duration of an utterance. First, it is not clear whether declination is intended or a byproduct. Second, both respiratory and laryngeal muscles can affect f_0; it is not clear which produces declination. Variations in P_{sg} correlated with variations in f_0 are sometimes taken as evidence that respiratory activity is controlling f_0, but this is not necessarily the case since the tract impedance, including laryngeal posture, can affect P_{sg}. P_{sg} is a measurable quantity and is constant enough to seem to be a controlling parameter, but it is a result, not a pure source parameter. A fuller discussion of declination, which concludes that its cause remains unresolved, can be found in Ohala (1990).

3.2 *Frication*

Fricatives are produced by making a tight constriction, with area of the order of 0.1 cm^2, somewhere in the vocal tract. The air emerging from the constriction forms a turbulent jet, and this jet produces noise. For unvoiced fricatives the vocal folds are held apart, giving a typical glottal area of 1 cm^2. This means that most of the subglottal pressure is dropped across the supraglottal constriction, rather than across the glottis (the obvious exception to this is [h], where the glottal constriction can be the only constriction; as a result, the vowel context can make [h] into an approximant, as in /ihi/).

Although the area of the constriction is much smaller than any tract area used during a vowel, it is larger than the average glottal area during voicing and thus the volume flowrate is higher during unvoiced fricatives, ranging typically from 200 to 400 cm^3/s or more (for [h] it may be 1,000–1,200 cm^3/s). In a vowel–fricative transition usually the glottis opens before the supraglottal constriction is formed, resulting in a momentary maximum volume velocity (see Figure 2.4). Then, as $\hat{A}_c$ decreases, U decreases also and the pressure drop across the supraglottal constriction increases. At some point frication begins; it would be useful to be able to predict precisely when. For voiced fricatives, with a lower mean U, the situation is even more complicated: turbulence noise is usually generated more weakly than in the unvoiced equivalent, but it is also effectively modulated by the voicing. This was first described by Fant (1960); the changes to the noise source spectrum as a result of the modulation have been described more recently (Jackson & Shadle, 2000) and will be discussed further in section 3.4. Flanagan's model of fricatives (1972, pp. 248–59) incorporates modulation based on the Reynolds number, and is discussed further in section 5. For both voiced and unvoiced fricatives, the first question must be: for what dimensions and flowrate does a turbulent jet form? This can be rephrased as, what is the critical Reynolds number for vocal tract geometries?

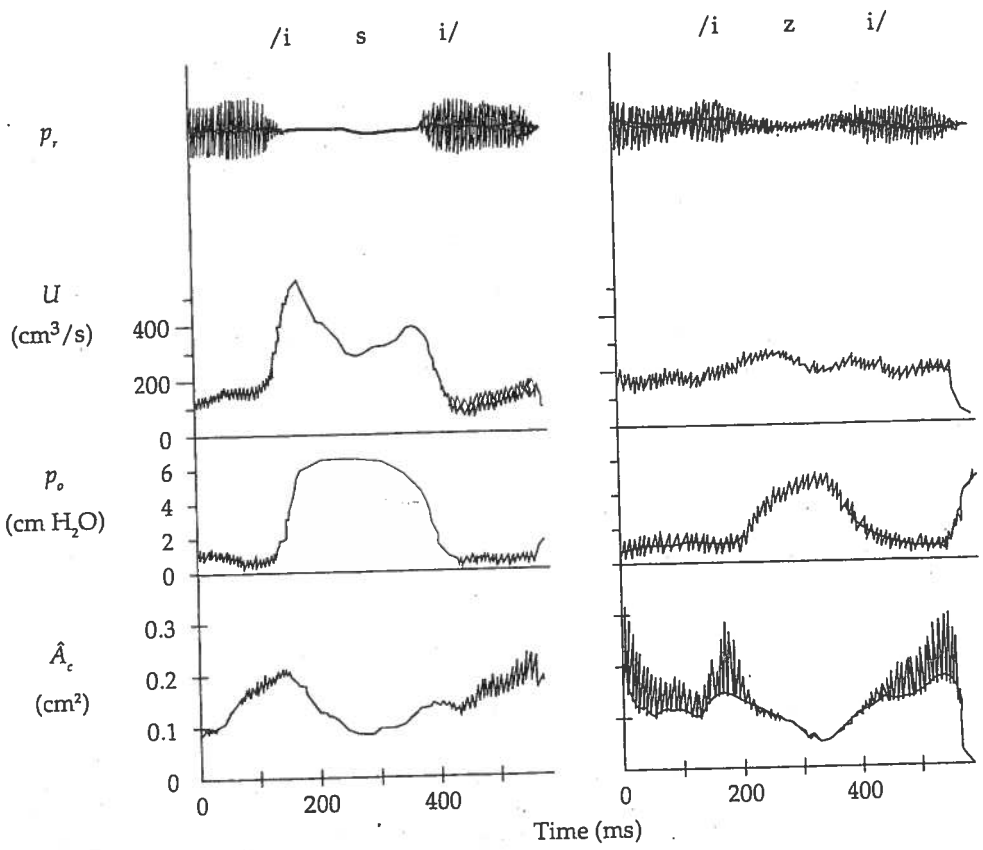

Figure 2.4 Time traces of measured radiated sound pressure p_r, volume velocity at the lips U, intraoral pressure p_o, estimated constriction area $\hat{A}_c$ for unvoiced and voiced fricatives. An adult male subject produced [pisi] (left) and [pizi] (right).

Meyer-Eppler (1953) conducted experiments to determine Re_{crit} for the fricatives [f, s, ʃ]. He measured radiated sound pressure (p_r) and oral pressure (p_o) for a speaker uttering the three fricatives and for air flowing through plastic tubes with three different elliptical constrictions. As shown in Figure 2.5, in each case a different minimum p_o was required to produce a measurable p_r; above this minimum, the rate of change of p_r with respect to p_o also varied. For the elliptical constrictions, he was able to arrive at a single line for p_r as a function of Re by using two different definitions of the effective width of the constriction for the three cases. For this line, he defined Re_{crit} to be the intercept where $p_r = 0$, and found Re_{crit} = 1,800. He then generalized this to speech, on the assumption that the same value of Re_{crit} would work for all fricatives provided the effective width was properly defined in each case.

This idea has gained wide acceptance. Studies using various ducts and orifices have led to a range of Re_{crit} values, from 1,700 to 2,300 (Ishizaka & Flanagan, 1972;

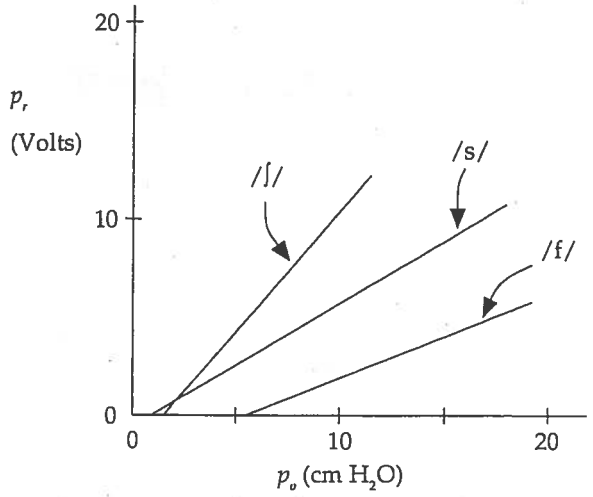

Figure 2.5 Radiated sound pressure p_r vs. intraoral pressure p_o for [f, s, ʃ]. (After Meyer-Eppler, 1953)

Catford, 1977). There are two problems, however. Since it is so difficult to measure the cross-sectional shape of the constriction, there is no independent check of *Re*. We do not know what the effective width should be for a particular constriction shape. Second, using the Reynolds number to collapse data carries with it the assumption that the geometries and therefore the source mechanism are the same, and thus allows comparison for different sizes and flowrates. But constriction shape is definitely not the same for different fricatives. Are we then losing or gaining by collapsing them together?

There is evidence that there are different source types operating to produce different fricatives. The noise produced by the jet alone, generated by relatively inefficient flow quadrupoles, is quite weak for the jet sizes encountered in the vocal tract. Anything solid in the path of the jet, however, produces a much more efficient noise-generation mechanism. Stevens (1971) recognized this difference, and adapted the work of Heller and Widnall (1970) on flow spoilers to frication. By treating the tongue-constriction as a spoiler, he found an equation giving source strength in terms of the pressure drop across the constriction. Although he acknowledged that the location of the constriction in the tract could affect the power-law relationship of the radiated sound power to the pressure drop, this was seen to be due to changes in the proximity of tract resonances to the source spectrum peak rather than an effect on the source mechanism.

Based on more recent analysis of speech and work with mechanical models, it appears that a flow dipole mechanism is operating, but not necessarily at the tongue constriction (Shadle, 1990, 1991). There are at least two distinctly different fricative geometries that result in different sources. The obstacle case has an obstacle such as the teeth at approximately right angles to the jet axis. The

source is localized at the upstream face of the obstacle. [s, ʃ] fall into this category. The wall case has an "obstacle" such as the hard palate at a more oblique angle. The jet generates noise all along the wall, resulting in a much more distributed source. The fricatives [ç, x] and presumably all pharyngeal fricatives fall into this category. The weak front fricatives [f, θ] should also possibly be grouped in this category, since noise is clearly generated along the lips (Shadle, 1990). The "wall" does not continue on very far, however, and so it may be that these sounds should be considered as a third category.

The geometry affects not only where noise is generated, but how much, that is, the spectral characteristics of the noise and the way they change with flow velocity and area of the constriction. Rather than absorb these differences by means of effective width formulae, it would seem useful to express the acoustic properties of the noise in terms of the aerodynamic and articulatory parameters for each category. Some work has been done on this, e.g., source curves as a function of volume velocity have been measured for models of [ʃ, ç, x], and power laws have been determined for human speakers (Badin, 1989). Much remains to be done. For instance, it seems clear that ΔP across a constriction depends principally on the volume velocity through it and the constriction's shape and area. The amount of noise generated by it can be related to ΔP, but the particulars of the relationship will depend very much on what is present downstream of the constriction exit.

Sinder, Krane, and Flanagan took a more theoretical approach, developing a jet model based on Howe's work showing that sound generation occurs when jet vorticity crosses streamlines, as can occur with a change in duct area. Their jet model depends on the location of flow separation, and the geometry and flow speed at that location (Krane et al., 1998; Sinder, 1999). Some comparisons to experimental measurements of mechanical models were made (Sinder, 1999). Krane (2005) further elaborated the model, showing that the jet could be modeled for aeroacoustic sound generation purposes as either a train of vortex rings or a train of inclined vortex pairs. The source spectrum can be considered to be the convolution of a harmonic and a broadband function; the arrival time of the vortices determines which function dominates.

Howe and McGowan (2005) and McGowan and Howe (2007) also took a theoretical approach. For [s], they noted that the upper and lower teeth overlap, creating a channel in which turbulence could diffract sound and thus increase it, and obtained predictions of the radiated sound that matched experimental data. Their use of the compact Green's function explains the interaction of the source and the sound field and thus can accurately predict the level and shape of the source spectrum.

All of these fricative models (Shadle's, Krane & Sinder's, and Howe & McGowan's) have used drastic simplifications of the vocal tract shape during fricative production. Although the resulting source models differ somewhat, all agree that the geometry, not only the area function, of the constriction and downstream of the constriction have an important effect on sound generation by turbulence. It may also be that for some fricatives the articulatory charateristics

most important for sound production have not all been identified. The ways in which the parameters of each model affect the sound generation need to be investigated further.

3.3 *Transient excitation: Stops*

Stops are intrinsically transient. Complete closure is effected somewhere in the vocal tract, from glottis to the lips. As shown in Figure 2.6, for a supraglottal unvoiced stop the pressure upstream of the closure typically builds up rapidly for a short time, and then continues to increase more slowly, possibly reaching a plateau. The neck and cheeks expand slightly in response to this pressure, and

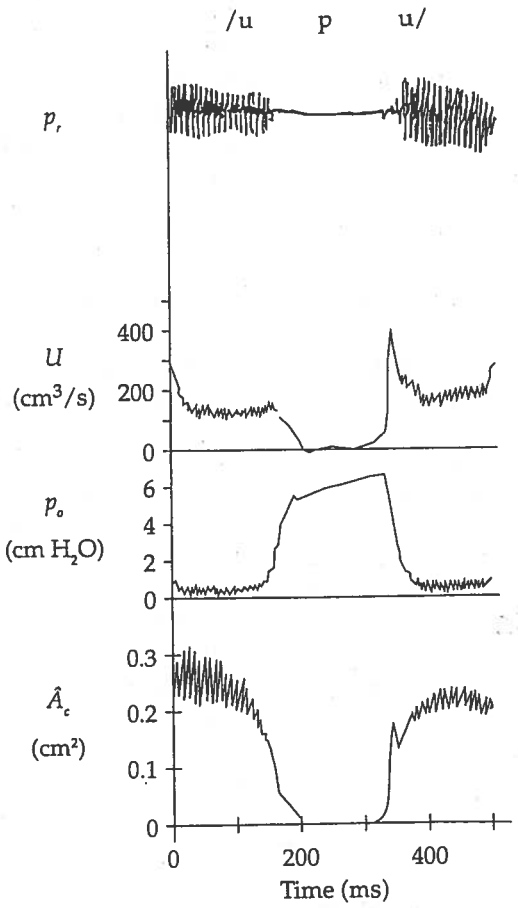

Figure 2.6 Time traces of measured radiated sound pressure p_r, volume velocity at the lips U, intraoral pressure p_o, and estimated constriction area $\hat{A}_c$ for the stop [p] in the context [upu]. An adult male was the subject.

the rate of decrease in lung volume eases slightly (Ohala, 1990). When the stop is released, either at the place of closure or at the velum, the oral pressure drops suddenly, lung volume suddenly begins to decrease more rapidly, and air is pushed out of the vocal tract explosively. The expelled air may become turbulent, and the patch of turbulence travels downstream, gradually dissipating. Depending on the position of the vocal folds, this brief period of high airflow may result in aspiration noise being generated.

The closure must be held for a perceptible amount of time, from a minimum of 20–30 ms to 100 ms or more. The release burst and ensuing frication last for a short time, of the order of 5 ms, and the aspiration, if it occurs, may last 50 ms or more before voicing begins. Indeed, the voice onset time will be longest if aspiration is present, and this is not a coincidence. Glottal area during stops is largest for unaspirated voiceless stops; for other cases, the glottal area depends somewhat on position of the stop within the word (Sawashima, 1977). Differences in voice onset times thus appear to be largely related to the time it takes to adduct the vocal folds (Catford, 1977). The wide-open glottis allows a high glottal volume velocity once the stop has been released, thus producing audible turbulence noise.

For a voiced stop, a pressure drop of at least 200 Pa must be maintained across the glottis for voicing to occur (Westbury, 1983). As a result the oral pressure does not increase as much as during an unvoiced stop. Fundamental frequency decreases as the pressure increases, and if closure is held long enough, vocal fold vibration may cease altogether. However, it appears that voicing during stops is extended by a combination of passive and active vocal tract expansion. The passive expansion occurs when cheek and neck tissues yield, puffing out slightly. We can control the degree of expansion somewhat by tensing or relaxing our cheek muscles; relaxed tissue yields more. The active expansion occurs by moving articulators: the larynx tends to move down, the soft palate up, and the tongue dorsum and blade down more during voiced than unvoiced stop closure. Both kinds of expansion serve to lower the pressure in the vocal tract, and therefore increase the transglottal pressure difference. Without such means, voicing can theoretically continue for approximately 60 ms after closure. With such means, voiced closure can extend theoretically to 200 ms or more – and in practice, voiced intervals of 100 ms or more are not uncommon (Westbury, 1983).

Sound production during and after the release has been modeled by Maeda (1987), and electrical analogs incorporating this developed by Stevens (1993, 1998). Maeda proposed a simple dynamic model that generates two different kinds of sources: an initial brief coherent source, followed by a longer frication source. The coherent source is caused by the assumption that when the closure is first opened, there is actually reverse flow into the sudden expansion, which causes a negative impulse of pressure. Although this flow monopole is predicted to last no more than 0.1 ms, it should be a very efficient sound source. The subsequent frication source is predicted to last from 1 to 5 or more milliseconds depending on the model parameter settings. Its strength would, of course, depend on the place of the constriction and the changing parameters.

Maeda demonstrated the coherent source with data from the utterance [mi]. The end of the [m] is released without a pressure buildup and therefore without the frication or aspiration sources, and shows a release bar on the spectrogram and extra negative-going radiated pressure. Flow visualization was done by seeding the flow with smoke particles. For a human subject, this is relatively easily accomplished by asking the subject to inhale cigarette smoke and using a high-speed camera to photograph the smoky jet leaving the lips as the subject says [mi]. Such a film showed a noticeable delay between opening the lips and the emergence of smoke, supporting Maeda's model (X. Pelorson, personal communication, 1994). More recently, Pelorson et al. (1997) studied bilabial plosives using flow visualization on human and mechanical models, numerical simulation, and theory. As Maeda found with [m], it takes approximately 20 ms for a jet to emerge once the lips are opened, and another 10–20 ms for vortex formation along the jet. The jet is essentially symmetric, in spite of the asymmetric lip horn, which might lead one to expect the jet to separate at different points from upper and lower lip. The frication noise that has been described to occur after the initial release (Stevens, 1993, 1998) appears likely to be produced by small-scale turbulence. Pressure–flow relationships are insufficient to describe this progression; the constriction shape as a function of time is needed. In the first few milliseconds after opening, viscous and boundary layer effects are important. After that, a low Reynolds number prevails, and boundary layer effects are not important.

3.4 *Mechanical oscillation: Trills and voicing*

Since the walls of the tract and the articulators are for the most part not rigid, it is possible for the airstream to set up a mechanical oscillation. This has been thought to be due to the Bernoulli force operating in the narrowed region such as the true or false vocal folds, the uvula, tongue tip or lips: here the air flows with a higher particle velocity and therefore the pressure drops. An inwards force is applied to the surrounding structure, and if that structure is flexible enough and the force strong enough, it may be pulled closed (see, for example, Catford, 1977). The closure of the "valve" formed by the vocal folds, tongue tip, etc. interrupts the airflow and allows pressure to be built up behind the closure, so it blows open and the process can repeat. The frequency of repetition is determined by both aerodynamic variables around the "valve," such as the original upstream pressure, the velocity through and area of the opening, and mechanical variables: the mass, compliance, and damping factors of the tissues making up the valve.

This simple model is no longer considered adequate. The quasistationary assumption on which the Bernoulli effect is based is good through much of the glottal cycle but does not hold when the glottis is very small, i.e., around closure. And the Bernoulli equation does not predict pressure–velocity relations well where the flow has separated (Titze, 2006).

The situation is somewhat similar to that of reed instruments such as the clarinet, in which the reed vibrates enough to close off the flow of air periodically, and those vibrations couple into and excite the resonances of the clarinet tube.

However, in the clarinet the natural frequency of the reed is well above the resonances of the tube, and so the pitch of the resulting sound is that of the lowest resonance (Benade, 1976). In the vocal tract, the natural frequency of the vocal folds is usually below that of the lowest formant, and so the pitch that results is that of the vocal fold vibration, ranging from 40 Hz (for creaky voice) to 1,000 Hz or more (for sopranos and children). For uvular and tongue-tip trills the mechanical oscillation is slower, in the range 20–35 Hz (Recasens, 1991; McGowan, 1992).

Vocal fold vibration has been extensively studied in both humans (see Hirose, this volume) and using excised canine larynges. There are many sets of muscles both in and around the vocal folds that can be adjusted to provide a very fine degree of control. The initial separation of the vocal folds, their length, and the tension of the three layers of the folds can all be separately controlled, in some cases by more than one mechanism. By these means the mode of vibration of the folds can be selected, and the frequency of vibration controlled within each mode.

The different modes of phonation are distinguished both by the pattern of movement of the vocal folds and by the resulting sound quality. The modes range from falsetto, in which the bulk of the folds are still and the margins vibrate, resulting in a relatively high-frequency sound with weak harmonics and a nearly sinusoidal glottal area function $A_g(t)$, to chest voice, in which a wave travels through the mucosa (the vocal fold cover) in the direction of the vocal tract's longitudinal axis, thus adding an extra component to the simple lateral motion of the folds (for more information see Gobl & Ní Chasaide, this volume). The closed phase for chest voice is a significant proportion of the total cycle, and upper harmonics of the fundamental carry a significant proportion of the total energy (Gauffin & Sundberg, 1989).

Within a mode, frequency of oscillation is primarily controlled by the length and tension of the folds and the subglottal pressure. The subglottal pressure is not an independent parameter in the way that the mechanical settings of the folds are: for instance, the minimum pressure required to achieve phonation appears to increase with f_0, and that relationship differs for singers and nonsingers (Titze, 1992, 2000).

There are numerous models of the vocal folds. Of the self-oscillating models, the best known are the one-mass and two-mass models (Flanagan & Landgraf, 1968; Ishizaka & Flanagan, 1972). Variations on the mechanical structure of the folds have included increasing the number of masses (Titze, 1973, 1974), using a distributed rather than lumped model (Titze & Talkin, 1979), a collapsible tube model (Conrad, 1985), and a translating and rotating one-mass model (Liljencrants, 1991a). In all of these, sufficient degrees of freedom are included to allow different modes of vibration. The different parts of each fold are coupled, either directly (e.g., via a spring) or indirectly (e.g., controlled by the same aerodynamic parameter). The effect on the flow of the current shape of the folds is handled generally by computing the point of flow separation within the glottis, and allowing pressure, velocity, and effective glottal area to vary accordingly. Pelorson

et al. (1994) improved the two-mass model's performance by systematically testing different ways of computing the separation point, and incorporating the best model.

More recently, finite-element models (FEM) have been used, which are more computationally expensive but have the power to represent the internal mechanical properties of the vocal folds and, potentially, pathological structures as well. Gunter (2003) highlights the differences among some of them: Alipour et al.'s model (2000) is self-oscillating and is not restricted to unrealistic geometries like earlier continuum mechanics models, but lacks the fine spatial resolution and mechanical stress distribution calculations needed to investigate vocal fold pathologies. Jiang et al.'s model (1998) has a finer spatial resolution but does not represent collision forces, which are important for some vocal fold pathologies and studies of voice quality. Gunter's model (2003), intended for use in studying vocal fold pathologies, includes fine temporal and spatial resolution and represents vocal fold collisions, but is not self-oscillating.

Two very recent papers indicate still more progress. Tao et al. (2006) discuss a self-oscillating finite-element model with which they studied vocal fold impact pressure, relating that pressure to lung pressure and glottal width. Unlike Gunter's model, they modeled the air as well as the vocal fold tissue in order to have not only the interaction of the folds with each other, but the folds with the fluid, represented. Aerodynamic properties but not acoustic wave propagation were included. This model was then set up with a stiffness asymmetry, which they showed resulted in biphonation (Tao & Jiang, 2006). This particular structural asymmetry may not be the, or the only, cause of such biphonation, as they point out, but it does demonstrate the potential uses of the model and thus justifies the model's complexity.

It is difficult to test such models since it is impossible to compare "output" for a human phonating with the same parameter "settings." It is accepted, however, that source–tract interactions occur in humans (Rothenberg, 1981; Guérin, 1983; Titze, 2000), and evidence of such interactions is sought for each model. For instance, one of the advantages of the two-mass over the one-mass model is that the two-mass model shows more realistic behavior when F_0 approaches and exceeds the frequency of the first formant. More recent FEM models have been tested by, for instance, comparing pressures and predicted impact forces of one fold on the other (Story & Titze, 1995) with such pressures and forces measured in a canine larynx (Jiang & Titze, 1994). Even though the model is not tailored precisely to the particular subject, when the collision forces match but predicted peak pressure is five times smaller than measured, it is clear that the model is not yet realistic in that regard. The more complex a model is, and the more input variables it has, the more difficult it is to validate it, as described clearly by Gunter (2003).

There have been numerous studies of the detailed aerodynamics of the glottis using mechanical models. First, static models were used to measure the pressure–flow relationships (Scherer, Titze, & Curtis, 1983; Scherer & Titze, 1983; Scherer & Guo, 1991). Three glottal profiles have been used – convergent, uniform, and divergent – to capture various stages in a single glottal cycle; the findings were

then assembled under a quasistationary assumption (Pelorson et al., 1994; Shinwari et al., 2003). Flow visualization of such models revealed the Coanda effect, in which a jet forms at the glottal exit and veers off to one side or the other, and remains attached to that side. However, an obstruction downstream, similar to the shape and position of the false vocal folds, could straighten the flow and prevent the Coanda effect (Shadle et al., 1991).

The static constraint on models has been lifted in a few different ways. One way is to use static models of the vocal folds, but start the flow impulsively (Hirschberg et al., 1996). This helped to establish where the separation point was at the glottal exit, how that varied according to the glottal profile, and the amount of time it took for a jet to appear and roll up into vortices. Hofmans et al. (2003) measured the time needed to establish the Coanda effect, which was much longer than a typical glottal cycle, and predicted therefore that it would not be able to be established in the more realistic situation when the vocal folds are moving. However, Erath and Plezniak (2006) did observe a Coanda effect for their static divergent glottal model with pulsatile flow.

The other class of mechanical model experiments involves a steady mean flow, but moving folds. These are usually driven, not self-oscillating, and have a fixed shape, but that shape can be varied between experiments in some setups. These have been used for various purposes, such as to visualize the flow downstream of the vibrating folds, with the results that the Coanda effect has been observed for a dynamic driven model with uniform glottis (Shadle et al., 1991). The quasi-stationary assumption was shown to hold apart from the early stages of the glottal cycle, when it departs significantly (Mongeau et al., 1997; Z. Zhang et al., 2002). Particle velocities and pressures have been measured in the tract in order to model the sound generation process (Barney et al., 1999).

The combination of all of these models with different constraints relaxed has reshaped our thinking about phonation. The separation point tends to be fixed when the glottal outlet is abrupt, as with a convergent or uniform glottis. With a divergent glottis, the separation point occurs before the glottal exit, at a point depending on the dimensions and angle of the glottis (if static). If the vocal folds are moving through all of these profiles, the separation point moves. A Coanda effect can then be observed within the glottis, with the jet attaching to one of the folds. The transition to turbulence is then asymmetric within the glottis, which changes the pressure–flow relationship significantly.

It has long been assumed that sound generation occurred at or near the glottal exit; this has been classically modeled as a monopole source, capturing the periodic appearance of the glottal jet. McGowan (1988) predicted theoretically that a downstream dipole source due to the vorticity–velocity interaction force, as well as a monopole source at the glottal exit, was necessary to characterize the phonation source. Since then, two experimental studies using driven folds have demonstrated other sources: Barney et al. (1999) showed that the glottal jet in their model developed a vortex street, and the vortices generated sound when they exited the tract. Z. Zhang et al. (2002) showed that the type of flow source varies during the glottal cycle; however, dipole sources dominate the tonal sound

below 2 kHz. This result was supported by numerical simulations of Suh and Frankel (2007).

Although the quadrupole source generated by the glottal jet has been shown to contribute insignificantly to the radiated sound for driven mechanical models (e.g., Z. Zhang et al., 2002), turbulence noise generation at the glottis can become significant in breathy and hoarse phonation. In both cases, the vocal folds oscillate but do not completely close. In breathy voicing a chink is left open near the arytenoid cartilages (Fritzell et al., 1986; Södersten & Lindestad, 1990; Södersten et al., 1991). The dc offset measured in an inverse filtered glottal waveform is used in many studies as evidence of such a chink, and is observable in both men and women subjects "almost universally," though women's voice qualities, on average, are breathier than men's (Holmberg et al., 1988). Karlsson (1986), however, did not always observe a dc offset, even in her women subjects; when it occurred, it was mainly at weak effort levels. She noted large subject variation, and also cited several methodological aspects that could explain the differences among studies (Karlsson, 1992). The inverse filtering method itself does not take account of the difference in the velocities, and therefore travel time from glottis to lips, of convection and sound, which may further confound such studies. (See the further discussion in section 5.)

Hoarseness is more variable; it may be caused by swollen folds resulting in slow oscillation, a node on one fold preventing a clean closure, or a paralyzed fold allowing a more significant gap (Hammarberg et al., 1984). In all of these cases there is a relatively inefficient conversion of the energy from the steady airstream into sound. Some work has been done to model hoarse phonation by, for instance, modifying the two-mass vocal fold model to generate a pathological model (Koizumi & Taniguchi, 1990). The finite-element models discussed earlier can do this in more detail but so far do not predict acoustic output.

In breathy or hoarse phonation the turbulence noise fluctuates with the glottal cycle. This occurs in voiced fricatives as well, though the turbulence in that case is generated well downstream of the glottis. It has long been recognized that the frication noise is modulated by the voicing source, but the mechanism was not clear: does the sound generated at the glottis affect the turbulent jet downstream, or does the unsteady flow field generated by the oscillating vocal folds convect downstream and result in a pulsing jet at the constriction? It was observed that the harmonic and inharmonic components of the radiated sound were out of phase with each other during voiced fricatives, but not during vowels. The mechanism appears to be that sound generated at the glottis travels to the constriction, and there influences jet formation; the phase difference is related to the travel time from glottis to constriction at the speed of sound, and through the constriction and front cavity to the main noise source location at the slower convection velocity (Jackson & Shadle, 2000, 2001).

For tongue-tip trills, the vibrating structure is not so finely controlled as the vocal folds, and partly as a consequence has a smaller range of frequency of vibration. Both unvoiced and voiced trills can be produced. In either case, the tongue blade and dorsum are held steadily in position and the tongue tip vibrates against

the hard palate at a rate of between 20 and 35 Hz. Closure is seldom complete, judging from electropalatography data of Catalan speakers and Rothenberg mask data (showing a nonzero minimum flow) of English speakers (Recasens, 1991; McGowan, 1992).

McGowan simulated the tongue-tip trill by modeling the tongue tip as a hinged trap door in the spirit of the one-mass vocal fold model. Wall compliance was included for the tract upstream of the tongue tip, and proved to be an essential part of the model. The oscillation of the tongue tip is only self-sustaining if net energy is transfered from the airflow to the motion of the tip during each cycle; this is accomplished if the pressure is greater during the opening phase than during the closing phase. This asymmetry occurs in the model because of the compliance of the walls. When the tongue-constriction is closed and the oral pressure rises, the walls expand. When the tip is released, they deflate, but they do so relatively slowly, thus maintaining a higher pressure for a time as the constriction opens. The wall effect is apparently more important for a smaller glottal area, since that limits the extent of variation in glottal volume velocity.

McGowan did not attempt to model the details of the flow near the tongue tip, and suggested that this might be important for two reasons. First, the simulated traces were much smoother than the measured ones. Second, flow separation in the constriction could also result in energy exchange tending to sustain the oscillation. Finally, although he included an adducted-glottis condition to approximate the average glottal opening during voicing, he did not actually allow the glottal area to vary, whether under direct control or via a self-oscillating vocal fold model.

3.5 *Aerodynamic oscillation: Whistling*

Whistling in speech occurs primarily in whistle languages (Busnel & Classe, 1976; Meyer & Gautheron, 2006), but may also occur in whistly fricatives, both deliberately as in Shona (Ladefoged & Maddieson, 1996, p. 171) and accidentally in languages that do not use a whistle for linguistic purposes (Shadle & Scully, 1995). Whistle languages can be used over distances of up to a few kilometres, and consist basically of a loud whistle that follows the *F*2 pattern of the whistler's ordinary language, or duplicates f_0 patterns of lexical tone. Whistly fricatives have whistles and frication noise occurring together; the whistle peak occurs generally in the high-amplitude region of the frequency spectrum, in the fricatives [s, ʃ, z]. Both kinds of whistling are best understood by considering "recreational" human whistling. Here there tends to be very little frication noise. The whistle may occur at *F*2 or *F*3, giving a frequency range of from 500 to 4,000 Hz (Shadle, 1983).

As described earlier, in order to produce a whistle sound there must be an unstable boundary layer and feedback that reinforces the instability. We would like to know when a whistle will occur and at what frequency, and therefore we need to know where the boundary layer forms and under what conditions it becomes unstable.

Because whistles are so geometry-dependent, the controlling parameters of a few classic geometries have been thoroughly investigated. Those that seem

most applicable to the vocal tract are the orifice tone, the edge tone, and the hole tone. The orifice tone, however, depends on sharp edges at the inlet causing the boundary layer to separate from the walls of the orifice. This is inconsistent with the shape of the lips, and the controlling parameter – length of the orifice – predicts too high a whistle frequency (Shadle, 1985).

The hole tone can be produced without sharp-edged inlets. It results from two orifices in a row. The first produces an unstable jet, which curls up into vortices in the region between the orifices. In the absence of surrounding walls, the distance between the orifices determines the feedback path length; with surrounding walls, the whistle couples into one of the resonances of that cavity (Chanaud & Powell, 1965). If the constriction formed by the tongue is the first orifice, and the rounded lips form the second orifice, the resonances of the cavity in between should control the whistle frequency; the lowest of these is in fact *F*2 (Shadle, 1985), consistent with whistle languages.

For whistly fricatives, the edge tone appears to be a more appropriate model. In this geometry, the unstable jet formed by an orifice strikes a solid object: a sharp edge of varying angle, or a cylinder. With laminar flow, the jet will divide smoothly around the object. When the jet becomes unstable, it tends to go to one side or the other of the object, alternating periodically and shedding vortices alternately. Here the orifice diameter and the distance to the edge are critical parameters (Powell, 1961, 1962; Holger et al., 1977). Elder et al. (1982) describe how a combination of tones and broadband noise can be produced simultaneously. Using a mechanical model consisting of a long pipe with a side cavity, and edges protruding over the cavity opening, they measured sound produced as flow velocity was gradually increased, and identified the parameters controlling the various tones and turbulence produced. In addition to the well-known whistle phenomena of high-amplitude narrow-bandwidth peaks coupling into resonances for which the phase relationships reinforce the instability, they demonstrated how whistles combine with turbulence excitation of the pipe resonances. It appears that this mechanism could be at work with the whistly fricatives, with the tongue again forming the jet-producing constriction and the teeth serving as the edge. This is consistent with the role of the teeth in noise production, and simply indicates that some structure can exist in a turbulent flow (Shadle & Scully, 1995).

Because whistles are so sensitive to small changes in the geometry or flowrate, it is difficult to model them for the vocal tract where dimensions are difficult to determine and easily varied. They are also difficult to model for another reason: the whistle mechanism exhibits a complete interaction of "source" and "filter."

4 Measurement Methods

4.1 *Basic methods*

A steady-state or slowly varying pressure can be measured by use of the manometer, which was described earlier. The tap can be placed in a sealed tank of gas,

or at a particular place of interest along a duct. In the latter situation, where there is a relatively steady flow along the duct, the tap must be designed so as to measure the desired static pressure without altering the flow by its presence. In general, having the tap flush with the wall, of a diameter much smaller than the duct diameter, and the edge of the tap abrupt rather than beveled, is sufficient. One must also pay attention to local variations in the pressure. For instance, there is a net loss in pressure across an orifice, and it is often of interest to measure this difference. However, in and near the orifice the pressure may show the opposite tendency, rising just upstream of the constriction, dropping significantly just downstream, then gradually recovering somewhat. To measure the pressure drop reliably, then, one must space the taps away from the orifice by an amount that depends on the orifice shape; for instance, for a thin orifice plate, the taps should be located at 2½ diameters upstream and 8 diameters downstream of the orifice (Doebelin, 1983).

Pressure drop across a known orifice is often used to deduce flowrate. In some cases an existing orifice is measured and calibrated; in others, an orifice of known shape and area is inserted into a duct. In either case, the flowrate U is derived from the pressure drop measured at two taps for an incompressible fluid by:

$$U = \frac{C_d A_2}{\sqrt{1-(A_2/A_1)^2}} \sqrt{\frac{2(p_1 - p_2)}{\rho}} \tag{8}$$

where U is flowrate in m^3/s, A_1 = pipe cross-section area (m^2), A_2 = orifice cross-section area (m^2), p_1, p_2 = the pressure measured at the two taps (Pa), ρ = the density of the fluid (kg/m^3), and C_d is a dimensionless discharge coefficient that depends on Reynolds number and the ratio of orifice to pipe diameter, as shown in Figure 2.7. Including C_d, an empirically-determined coefficient that varies with orifice shape and the locations of the pressure taps, allows actual areas to be used rather than flow areas as in equation (5), and includes frictional losses. Calibration to determine C_d for every new setup can be avoided by using standard dimensions for the orifice meter and relying on the extensive experimental data available (Doebelin, 1983).

Equation (8) can be modified for compressible fluids to give the weight flowrate rather than the volume flowrate. For a small pressure drop ($p_2/p_1 > 0.99$), this is sufficient. For isentropic (i.e., frictionless and adiabatic) flows with larger pressure drop, an equation for weight flowrate can be derived whose only empirical coefficient is C_d. For a sharp-edged orifice plate, however, enough turbulence is generated that the isentropic assumption is not a good one. In this case, an experimental compressibility factor Y must be incorporated in the equation; Y depends on the pressure drop and orifice diameter in a different way for different placement of the pressure taps. For a known and stable configuration the final equation, though complicated, can be quite accurate (Doebelin, 1983). If the configuration is not known or is known to change, however, it may be more practical to use equation (8) for incompressible flow and determine or estimate an empirical

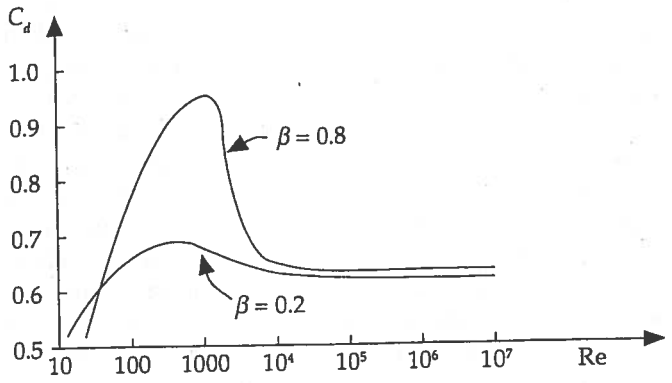

Figure 2.7 The dependence of discharge coefficient C_d on Reynolds number, Re, and on β, the ratio of orifice to pipe diameter. (After Doebelin, 1983, p. 531. Reproduced with permission of the McGraw-Hill Companies from E. Doebelin, *Measurement Systems*, 3rd ed., copyright 1983, McGrawHill)

coefficient for every change in geometry or significant change in flow (Massey, 1984).

Volume velocity can also be measured by a rotameter, which consists of a float in a vertical tube of varying cross-sectional area. The flow enters at the bottom of the tube and blows the float up to the point where the vertical forces of differential pressure, gravity, viscosity, and buoyancy are balanced. The same equations for flowrate as a function of area apply, but since the float position rather than pressure drop is the output measured, and flowrate is linearly related to float position (for the typical tube tapering) but related to the square root of pressure drop, the rotameter has a greater accurate range than orifice flowmeters (approximately 10:1 rather than 3:1 maximum:minimum flowrate, respectively) (Doebelin, 1983).

Particle velocity can be measured by a number of methods. The pitot tube is a probe that is placed directly into the flow, pointing upstream. It measures two pressures: the stagnation pressure, by a tap at its upstream end, and the static pressure, via taps along its sides. The difference between these two pressures can be used to derive the particle velocity at the location of the upstream tap.

Although the pitot tube is quite accurate, it cannot measure very low flow velocities, nor will it register quickly fluctuating velocities, as in turbulence. Higher frequency variations in particle velocity can be measured by using a hot-wire anemometer, which consists of a very fine wire with current passing through it. When held in a moving fluid, the flow cools it and changes the resistance slightly. In the constant-temperature form of the instrument, the current is adjusted to keep the wire temperature constant, as measured by its resistance. The square of the current is then related to the flow velocity. Because the wire is so fine, it responds quickly, and fluctuating flow velocities (up to as much as 100 kHz,

depending on the compensating circuit) can be measured. Also, the wire and its support can be made small enough to provide minimal disturbance to the flow. The difficulties with the technique are that the wires are very fragile; they will not register flow direction, but only its magnitude; and each hot-wire must be calibrated with known velocities in the fluid in which it is to be used (Doebelin, 1983).

High-frequency pressure fluctuations can be measured with a microphone, but in some situations this includes more than the sound wave when only the sound wave is wanted. The most familiar example is breath noise; the microphone can be moved further away or out of the breath stream, or a foam windscreen can be used that absorbs the mean flow before it deflects the microphone's diaphragm. Inside a duct with both sound waves and a nonzero mean flow, similar problems occur. Pressure transducers can be flush-mounted on the walls, or effectively extended into the flow by use of probes. A probe in a moving fluid can, however, in itself become a location for sound generation. A different way to measure fluctuating pressures is to use two hot-wires a known distance apart. Their two velocities can be used to compute a velocity gradient proportional to pressure. The cross-correlation of the two signals can be used to compute the time delay between the two sensors, and therefore the speed of propagation of a particular signal. By this means hydrodynamic and acoustic pressure disturbances can be separated out: the former travel at approximately the mean flow velocity, the latter at the speed of sound.

Fluctuating pressures can also be measured in terms of the force they exert on an object. Heller and Widnall (1970) used force transducers to deduce the source strength from the force applied by the flow to spoilers in a duct. Accelerometers can also be quite useful for measuring the effect of flow on solid bodies, provided they have a mass much less than that of the object they are attached to.

Flow visualization can be accomplished by many techniques. The flow can be seeded with visible particles such as smoke, and pictures taken of the patterns thus revealed (e.g., Shadle et al., 1991, or Pelorson et al., 1997). Alternatively, the difference in refractive index caused by differences in density can be made visible by three different optical techniques. The shadowgraph technique is the simplest, but registers only large density gradients such as in shock waves. The Schlieren technique is more sensitive, but cannot reliably be used for absolute measurements of density (Pelorson et al., 1994; Pelorson et al., 1995). Interferometry can be used for quantitative density measurements, but is quite complex to set up. All three methods depend on passing light through the flow (Massey, 1984).

If the time between successive photographs is known, the time of travel of vortices and rate of their growth can be computed. In general, flow visualization works best with flows that are essentially two-dimensional, for example, with rectangular rather than circular jets. Obviously, internal flows (i.e., flow in ducts) cannot be visualized unless at least one wall of the duct is clear and the flow is "lit" by a means appropriate to the visualization method.

4.2 *Speech-adapted methods*

Ideally, in speech as in any other system, aerodynamic parameters should be measured without disturbing the flow producing them. Likewise, parameters not directly measured should be derived with due regard for the type of flow. However, the difficulties of accessing the vocal tract mean that many parameters cannot be measured directly, and a certain degree of pragmatism is therefore essential.

The aerodynamic parameters needed to model respiration tend to be more slowly varying than those for the larynx and supraglottal system. The lung volume cannot be measured directly; it is inferred by measuring changes in body volume. Total body volume can be measured with a plethysmograph, in which the body is sealed in an airtight container. Changes in volume are deduced either by measuring changes in pressure within the container, or by measuring the flowrate through a single port into the container. Alternatively, the motion of the thorax and abdomen can be monitored by use of multiple position sensors, and the lung volume then deduced (Draper et al., 1959; Hixon et al., 1973, 1976; Ohala, 1990; Slifka, 2003).

Subglottal pressure can be measured directly by tracheal puncture (Isshiki, 1964) or by pressure transducers lowered through the glottis (Cranen & Boves, 1985, 1988). It can be inferred from esophageal pressure (Draper et al., 1959; Slifka, 2003). All of these methods are invasive medical procedures requiring the presence of a physician, and thus cannot be done routinely. They can be invaluable to validate and evaluate other less invasive procedures, however. For instance, Cranen and Boves placed two pressure transducers above and two below the glottis. This allowed not only measure of the subglottal pressure, but use of the pressure gradient to deduce flow through the glottis, which could be compared to the glottal flow derived from simultaneous laryngograph, photoglottograph, and inverse filtering.

Supraglottal pressure can be measured directly much more easily than subglottal pressure by introducing a thin plastic tube at the side of the mouth and bending it behind the rear molars so that its open end is midsagittal and perpendicular to the longitudinal axis of the vocal tract. The pressure measured, P_o, should thus be the static pressure upstream of all labial, dental, and alveolar constrictions. The tube is typically attached to a pressure transducer sensitive to 1–2 kHz, and referenced to atmospheric pressure (Scully, 1986). It can then be used as an estimate of the pressure drop across the constriction, ΔP_c, with the proviso that $P_o \geq \Delta P_c$. During the stop [p], P_o increases quickly as pressure in the tract behind the constriction equalizes with the lung pressure. The maximum value of P_o measured during [p] can thus be used to estimate subglottal pressure, and the estimate can be extrapolated to the surrounding speech sounds. The exact value used depends on the respiratory model accepted, however: an assumption of constant pressure applied to lungs, or of constant lung-volume decrement, gives slightly different results (C. Scully, personal communication, 1994).

Volume velocity at the lips can be measured by a variety of masks containing flow or pressure transducers, some with nasal and oral airflow separately measured.

The Rothenberg mask provides the least acoustic distortion: it measures the pressure drop across screens of known flow resistance (Rothenberg, 1973). Its frequency range is limited to 0–1.8 kHz partly by that of the transducers used, but also by acoustic resonances of the mask itself (Hertegård & Gauffin, 1992). Although it is relatively nondistorting acoustically within this range, the screening very likely massively disrupts any vortex pattern emerging from the mouth. Whether or not this is significant for the far-field sound is unknown at present.

The volume flow from the mouth measured by the Rothenberg mask is also commonly used to estimate the volume flow from the glottis, U_g, by inverse filtering; U_g is then related to activity of the vocal folds and the voicing source. Because it is not invasive and provides an essential source function, it has been used extensively. Possible limitations of the method are related to the source–filter model of speech production on which it is based, and are therefore considered in the next section.

Particle velocity has been measured within the vocal tract by using shrouded hot-wire anemometers during production of open vowels. Open vowels were necessary so that the hot-wire holder could be inserted and traversed across the tract (Teager, 1980; Teager & Teager, 1983). The shrouding was used to enable detection of flow reversal; whether it does that without undue distortion of the flow is a matter of some debate. The hot-wires can be expected to have a short life in such an environment, but the difficulties of calibration for low flow velocities and the inherent inability of a single hot-wire to detect flow reversal are more significant problems (see, for example, the extensive discussion printed as part of Teager & Teager 1983, pp. 394–401). It can be quite useful, however, to use hot-wires in human subjects for validating more extensive hot-wire measurements done on mechanical models (see, e.g., Shadle et al., 1999).

The technique used by Heller and Widnall (1970) of mounting the flow spoilers on force transducers in order to measure the force generated by the flow directly is clearly not possible with an articulator like the tongue. However, accelerometers and other motion-sensing devices have been used in the vocal tract to measure motion of the velum, jaw, vocal folds, and tongue. It is beyond the scope of this chapter to review such methods. However, when aerodynamic parameters must be inaccurately measured, or deduced from indirect measurement, or outright estimated, the presence of independently obtained articulatory data can help put such estimates on a firmer footing. We describe such a process below.

In fricative consonants, the area of the constriction is a key parameter that is clearly related to the properties of the noise generated, although perhaps not so simply as has been proposed by Stevens (1971). It is difficult to derive this area from vocal tract imaging methods because it is so small. However, we can use an oral pressure tube and a Rothenberg mask simultaneously, and measure P_o and U_m as a function of time during, say, a vowel-fricative-vowel transition. We can then estimate the area, A_c, by rearranging equation (8):

$$\hat{A}_c = KU_m\sqrt{\frac{\rho}{2P_o}} \qquad (9)$$

where K is an empirical constant, nominally a shape factor, taken equal to 1 by, for example, Scully (1986), and equal to $1/(0.65)$ by others. Since the constriction area has been assumed to be much less than the tract area ($A_c = A_2 \ll A_1$), K should correspond approximately to $1/C_d$.

An obvious limitation of this estimate is that P_o does not measure the pressure drop across the constriction only: lip rounding will increase it while not affecting constriction area or, presumably, frication noise. A less obvious problem is that this form of the equation is based on steady, incompressible, frictionless flow, which we clearly do not have. Although equation (9) is used for flow measurement in cases that also violate these assumptions, that is done for particular geometries for which extensive empirical data exist. Not only are such data nonexistent for the vocal tract, but the geometry is continually changing. The little we do know indicates that if we use Reynolds numbers and area ratios appropriate for the transition to and from a fricative, Figure 2.7 predicts that the discharge coefficient C_d in equation (8) will traverse a range of values, from approximately 0.9 to 0.6, yet K is typically held constant.

Pelorson (2001) tested this approximation and three variations on it by using mechanical models with three different constriction shapes, all possible shapes for speech. He showed that the best estimate of the area is found when the flow separation point can be estimated from a knowledge of the constriction shape. Since this is not always possible in speech, equation (9) using $K = 1$ is within 20 percent of the real area. The equation was also tested on unsteady flow, and is a reasonable approximation except near closure. It should be used with care for fricatives when turbulence is likely to occur within the constriction, as the losses will then be higher.

A related problem occurs in estimating the flow resistance of constrictions, which is relevant for the glottis as well as for fricatives. A typical procedure is to use the average volume velocity through and pressure drop across a constriction to define an operating point on an essentially parabolic function. The incremental resistance is then defined as the slope of the tangent to the curve at the operating point (Heinz, 1956). Pressure fluctuations due to sound waves are assumed to be small excursions about that point which can be modeled linearly; for small sound pressure amplitudes, this assumption is borne out by the measurements of Ingard and Ising (1967). However, the flow resistance in practice is often deduced from constriction area and volume velocity alone (Badin & Fant, 1984), whereas constriction shape can influence the pressure drop and, therefore, the operating point (Shadle, 1985).

5 Models Incorporating Aerodynamics

The classical acoustic theory of speech production models the acoustic properties of the vocal tract as an analogous electrical network. In so doing, several assumptions are made: sources and filter are independent, the filter is composed of passive elements and constitutes a linear system, sound propagation is one-dimensional, and in the most restrictive models, there is no mean flow. In this type of model,

all aerodynamic effects are essentially confined to the source functions. Because source and filter are independent, whistles or whistly fricatives cannot be generated, but this lack would not of itself be of undue significance for speech models for most languages. Unfortunately, problems of greater consequence do arise; it is instructive to consider the ways in which some existing models have approached greater physical realism by relaxing some of the assumptions.

All models of phonation must include mean flow as an input, and, classically, fluctuating volume velocity is generated as an output. However, tract models do not always include mean flow. How can fricatives then be generated? Scully (1990) includes mean flow in her synthesizer by having separate acoustics and aerodynamics blocks. The aerodynamics block computes static pressure and mean flow throughout the tract, including the lungs, and generates frication sources with strength related to the pressure drop across the constriction. These sources are then fed forward to the final source–filter model. The sources and filter cannot interact extensively, but some influence is possible via numerous interconnecting paths.

A somewhat different approach is taken by Flanagan and Ishizaka (1976), who derive the fluctuating glottal flow from the two-mass model and a mean flow from a dc atmospheric-pressure source. This arrangement allows respiration, as well as frication. Frication is then modeled by providing each transmission-line section with a noise pressure source parameterized by Reynolds number. A particular source would generate noise only if the area and volume velocity in that section resulted in $Re > Re_{crit}$. The amplitude of the noise source is modulated by a function proportional to Re^2, making the modulation observed in voiced fricatives possible (as demonstrated in earlier work based on a similar model, Flanagan, 1972). The noise source spectrum is flat, a reasonable simplification given the frequency range of the simulation (0–4 kHz).

A later synthesizer modified this scheme by using only one Re^2-dependent noise source per constriction (Sondhi & Schroeter, 1987). Location of the source was problematic, however: the internal impedance of the source was high enough that it restricted the volume flow unnaturally when placed at the constriction exit. Locating it downstream got around that problem, but each consonant required a different source location. This indicated greater physical realism, consistent with mechanical model studies (Shadle, 1985), but was "very inconvenient" in the context of an automatic text-to-speech synthesizer. The solution adopted was to place the source one section downstream of the narrowest part of the constriction, and represent it in parallel form, as a volume velocity source.

Narayanan and Alwan (2000) adopted a similar framework but worked to specify more physically realistic noise sources for a parametric synthesizer. They combined three-dimensional data derived from magnetic resonance imaging and source characteristics derived from mechanical model studies with an analysis-by-synthesis approach. All fricatives had a dipole source, corresponding to an obstacle downstream of the constriction, of either the teeth or the lips. All fricatives also had a monopole source, based on Pastel's findings (1987) from her experimental work modeling noise sources near the glottis. In addition, the palatoalveolar fricatives had another dipole source modeling wall noise. The monopole source

was found to be unimportant, and the dipole source locations could be chosen uniformly within each fricative class, thus agreeing with mechanical model results. By adjusting the source strengths and spectral characteristics of the different sources, best fits were found for each fricative; in general, the stridents were matched more successfully than nonstridents.

There are no experimentally-derived source models for interdentals, and the position of the noise sources so near the lip opening is likely both to change source characteristics and make radiation and other loss models more critically important. For sibilants, the known changes in source characteristics above and below the cut-on frequency of the duct have not been modeled by Narayanan and Alwan, perhaps because this would be inconsistent with their assumption of a plane-wave model. These simplifications, as well as allowing the source strengths to be adjusted relative to each other, decrease the physical realism of the parametric source models. However, this work represents the best effort to date at modeling fricatives within the classical framework.

The inverse filtering procedures using the Rothenberg mask are based on a similar model of an independent source and a filter that is linear, time-invariant, and composed of passive elements only (Rothenberg, 1973). The model allows for the glottal volume velocity, including a mean flow component, to be estimated from the pressure drop measured across the mask; if intraoral pressure is simultaneously measured, the source strength of a stop or fricative can also be estimated.

To estimate the voicing source, the vocal tract transfer function must be calculated; the difficulties of doing so are discussed elsewhere in this volume (see Gobl & Ní Chasaide, this volume). The glottal volume velocity so derived is hard to reconcile with what we now understand of the physics at the glottis. A dipole as well as a monopole source are needed; which is dominant varies during the glottal cycle, as discussed in section 3. Travel time from glottis to lips is much slower at convection velocity than at the speed of sound (e.g., 170 ms and 0.5 ms, respectively), so that flow passing through a glottal chink arrives at the mask as a dc component much later than does the sound that was generated at the glottis at the same time. This disparity will vary with the tract area function and subglottal pressure, and so cannot be easily estimated and compensated for. While inverse filtering is undoubtedly useful, it appears that the waveform it generates has a more complex relationship to actual velocities existing near the glottis than was originally appreciated.

A more recent model of sound propagation in the vocal tract (Davies et al., 1993) retains a separation of source and filter while relaxing many of the traditional assumptions. Sound propagation is not always one-dimensional, and it need not be isentropic near junctions; mean flow is allowed, and the speed of sound is adjusted accordingly, but flow sources are not generated by the model. Because the tract is not modeled as an electrical analog, but instead is divided up into different duct elements that affect sound propagation differently, more physical realism is possible while still remaining powerful conceptually.

Teager sought to relax the assumption of independent source and filter. His hot-wire data showed evidence of nonuniform velocity across the vocal tract during vowel production (Teager, 1980). He suggested that source–filter interaction was therefore essential to a speech production model, and described a jet-cavity

interaction paradigm (Teager & Teager, 1983). However, he did not propose a quantitative model, as discussed by Hirschberg et al. (1996).

Recently, quantitative models of aeroacoustic processes have been proposed. Pelorson et al.'s (1994) model predicting flow separation within the glottis has been discussed in section 3.4. Hirschberg et al. (1996) place this in a more general context. It is shown that viscosity, the friction of the fluid, cannot be neglected; including it predicts not only a pressure drop across the glottis as is observed, but a boundary layer next to the walls. Where the flow separates from the walls, forming a shear layer, the strong gradient across that layer generates vorticity, which causes the edges of the jet to roll up into vortices; the vorticity itself is the source of sound in the jet. Predicting boundary layer behavior precisely is difficult, especially for the complex glottal geometry; Pelorson's model is simplified, but works well, and is crucial in predicting sound generation. The simplification also allows it to be used for speech synthesis.

Hirschberg et al. also note that, although the turbulence of the jet generates noise, sound generation is significantly increased if there is a constriction downstream of the region of jet formation. The false folds result in dipole sources which are more efficient than the jet's quadrupole sources; similarly, teeth and/or lips downstream of a supraglottal constriction can generate efficient dipole sources in stop and fricative production. Sinder's model (Sinder, 1999), discussed in section 3.2, likewise uses a vorticity model to generate sound sources. This has been used as the basis of a synthesizer that generates its own noise sources (Krane et al., 1998).

McGowan and Howe (2007) describe the use of the Green's function, a general transfer function that includes but is not limited to the case of plane-wave propagation of sound, to model the exchange of energy between the hydrodynamic and acoustic modes of motion. With highly simplified geometry they are nevertheless able to explain why the dipole sources that so many studies have shown to be produced by a jet interacting with a downstream solid boundary, whether the glottal jet at the false folds or a supraglottal jet at the teeth, vary in their contribution to the far-field sound, especially at higher frequencies. They note that predicting the hydrodynamic field still must be done either experimentally or by numeric simulations, but their theoretical framework constitutes a model that is conceptually very powerful.

Full continuum simulations of the entire flow field represent another approach to including aerodynamics in a model of the vocal tract. Numerical simulations divide the fluid up into small volume elements, across each of which the conservation laws must hold. The equations of fluid motion are then solved, as well as the interactions of the fluid with the solid boundaries. Two critical decisions are the spatial resolution, i.e., the size of the volume elements, and the time resolution, the time steps for which pressure and velocity distributions will be computed. The spatial resolution determines the fluid structures, such as size of vortices, that can be simulated; the time resolution affects the stability of the solution, and determines the bandwidth for which the results are valid. Thus, simulating turbulence requires shorter time steps than simulating laminar flow (Blazek, 2005).

As a result, early studies (e.g., Thomas, 1986; Iijima et al., 1990; Liljencrants, 1991b) were limited to laminar flow through simplified geometries, simulating,

for instance, pressure–flow relationships in simple glottal models. As computing power has increased, so has the usefulness of numerical simulations. In the mid-1980s Navier-Stokes equations could be solved numerically, allowing viscous flows to be simulated. This led in turn to methods developed for simulating turbulence: Direct Numerical Simulation (DNS), which is computationally the most intensive, Reynolds-averaged Navier-Stokes (RANS), which predicts mean flows only, and Large-Eddy Simulation (LES), which predicts instantaneous flow but of the large-scale motion only. Many other ways of reducing the computational load have been developed, including use of nonuniform grids, so that smaller elements can be used in the boundary layer than in the main flow channel. Ways of dividing a long duct into short sections, and simulating the flow in each in succession, have been developed. When possible, simulation is done in two dimensions, though this does not work well for simulating turbulence. There are also different ways to predict the sound generated. The compressible form of the Navier-Stokes equations (NSE) can be solved, which predicts the sound field directly; however, this is inaccurate for low flow speeds. Alternatively, the incompressible Navier-Stokes equations can be solved, and then an acoustic analogy used on the predicted pressure and velocity fields (Suh & Frankel, 2007).

As examples of the ways these different constraints can be traded off, consider these recent studies. Zhao et al. (2002) used the compressible NSE on an axisymmetric model of the vocal folds. A moving grid was used for the walls, which allowed forced oscillation to be simulated. The model could not predict turbulence, but the results did show vortex formation downstream of the glottis, and predicted that dipole sound sources due to the unsteady motion on the walls of the glottis were dominant. The effect of false folds and subglottal pressure variations could be studied (C. Zhang et al., 2002).

Adachi and Honda (2003) used LES to model fricative sound production. Complex vocal tract shapes were derived from MRI; the airflow in only the most anterior 4 cm of each vocal tract was simulated to make the problem computationally feasible. Even so, approximately 15 million cells were required to simulate turbulence. They were able to generate 10 ms of far-field sound up to 16 kHz, which compared reasonably well to sound produced by mechanical models of the same two vocal tract shapes.

Suh and Frankel (2007) used three-dimensional LES and an unsteady, compressible formulation to study flow through a static glottis, and predict sound generation for convergent and divergent glottis shapes. Their predictions agreed well with experimental results, and they were able to explain the sound generation mechanisms at different frequencies in detail.

There are other examples in the recent literature, but these will suffice to demonstrate on the one hand, the severe constraints of numerical simulation, but on the other hand, that judicious choices can complement experimental results and aid in the development of simpler aeroacoustic models.

In summary, a variety of models exists that incorporate aerodynamics to a greater or lesser degree. Although it is difficult to model such effects as phonation (fluid–solid interaction) or frication (turbulence) because the underlying

phenomena are incompletely understood and resist an analytical solution, the fact that aerodynamics underlies all aspects of speech production makes such efforts important.

NOTES

I would like to thank Dr. Khalil Iskarous for comments on an earlier version of this chapter. Thanks also to Dr. Celia Scully for the raw data incorporated in Figures 2.4 and 2.6, which were obtained as part of an EC SCIENCE award, CEC-SCI*0147C(EDB). NIH grant NIDCD 006705 provided partial support during the preparation of the version for the 2nd edition of the Handbook of Phonetic Sciences.

1 At frequencies below the first cut-on frequency only plane waves propagate, which excite the longitudinal modes. The first cut-on frequency, above which transverse as well as longitudinal modes can propagate, depends on the duct's cross-sectional shape and inversely on its largest cross-dimension. A circular duct 4 cm in diameter has a cut-on frequency of approximately 5 kHz (Kinsler et al., 1982). This is why using only cavity lengths and area ratios to compute formant frequencies works well up to 5 kHz, and less well above that.

APPENDIX: CONSTANTS AND CONVERSION FACTORS

The following values hold for dry air at 37°C. Values for completely saturated air are given in parentheses where available. (From Batchelor, 1967, and Davies, 1991.)

c = speed of sound	= 35,300 cm/s	(35,900)
γ = ratio of specific heats	= 1.400	(1.396)
R = gas constant	= 2.87×10^6 erg/g	(2.977×10^6)
ρ = density	= 1.139×10^{-3} g/cm^3	(1.098×10^{-3})
μ = absolute viscosity	= 1.89×10^{-4} g/cm-s	
$\nu = \mu/\rho$ = kinematic viscosity	= 0.166 cm^2/s	
P_0 = standard atmospheric pressure at sea level	= 760 mm Hg	(760)

The conversion table for units of pressure in Table 2.1 should be interpreted as follows: 1 of the unit chosen from the leftmost column equals x of the unit chosen from the topmost row, where x is the value found at the intersection of the chosen row and column. For example, 1 bar = 10^5 Pa.

The conversion table for units of volume velocity in Table 2.2 should be interpreted as follows: 1 of the unit chosen from the leftmost column equals x of the unit chosen from the topmost row, where x is the value found at the intersection of the chosen row and column. For example, 1 liter/sec = 60.0 liters/min.

Table 2.1 Pressure (force per unit area)

	dyn/cm^2	Pa	bar	atm	$cm\ H_2O$	$in\ H_2O$	$mm\ Hg$	lb/in^2
dyn/cm^2	1	0.1	10^{-6}	9.869×10^{-7}	1.0197×10^{-3}	4.015×10^{-4}	7.501×10^{-4}	1.4503×10^{-5}
Pa	10	1	10^{-5}	9.869×10^{-6}	1.0197×10^{-2}	4.015×10^{-3}	7.501×10^{-3}	1.4503×10^{-4}
bar	10^6	10^5	1	9.869×10^{-1}	1.0197×10^{3}	4.015×10^{2}	7.501×10^{2}	1.4503×10^{1}
atm	1.013×10^6	1.013×10^5	1.013	1	1033.0	406.8	760.0	14.7
cm H_2O	980.71	98.071	9.8071×10^{-4}	9.865×10^{-4}	1	0.3937	0.7355	1.422×10^{-2}
in H_2O	2491.0	249.1	2.491×10^{-3}	2.458×10^{-3}	2.54	1	1.868	3.613×10^{-2}
mm Hg	1.333×10^3	133.3	1.333×10^{-3}	1.316×10^{-3}	1.3597	0.5353	1	1.934×10^{-2}
lb/in^2	6.895×10^4	6.895×10^3	6.895×10^{-2}	6.805×10^{-2}	70.307	27.68	51.71	1

Note: 1 μbar = 1 dyn/cm^2; 1 Nt/m^2 = 1 Pa.; 1 psi = 1 lb/in^2. The threshold of hearing, often used as a reference pressure in computing decibels, is 20 μPa = 2×10^{-5} Pa. Values in this table are derived from Halliday and Resnick (1966).

Table 2.2 Volume velocity (volume flow past a cross-section per unit time)

	cm^3/s	m^3/s	l/s	l/min	ft^3/min	in^3/s
1 cm^3	1	10^{-6}	10^{-3}	6.0×10^{-2}	2.119×10^{-3}	6.102×10^{-2}
1 m^3/s	10^6	1	10^3	6.0×10^4	2.119×10^3	6.102×10^4
1 l/s	1.000×10^3	1.000×10^{-3}	1	60.0	3.531×10^{-2}	61.02
1 l/min	16.67	1.667×10^{-5}	0.0167	1	3.531×10^{-2}	1.017
1 ft^3/min	471.9	4.719×10^{-4}	0.4719	28.32	1	28.80
1 in^3/s	16.39	1.639×10^{-5}	1.639×10^{-2}	0.9834	3.472×10^{-2}	1

Note: cfm = cubic feet per minute = ft^3/min; scfm = standard cubic ft/min, the volumetric flowrate of a gas corrected to standardized conditions of temperature, pressure and relative humidity (but note that there is no universally accepted set of "standard" conditions); 1 US fluid gallon = 4 US fluid quarts = 8 US pints = 128 US fluid ounces = 231 in^3; 1 British imperial gallon = 277.42 in^3. Values in this table are derived from Halliday and Resnick (1966).

REFERENCES

Adachi, S. & Honda, K. (2003) CFD approach to fricative sound sources. In S. Palethorpe & M. Tabain (eds.), *Proceedings of the 6th International Seminar on Speech Production* (pp. 1–6). Sydney: Macquarie University.

Alipour, F., Berry, D. A., & Titze, I. R. (2000) A finite-element model of vocal-fold vibration. *Journal of the Acoustical Society of America*, 108, 3003–3012.

Badin, P. (1989) Acoustics of voiceless fricatives: production theory and data. *Speech Transmission Laboratory – Quarterly Progress and Status Report*, 3, 33–55.

Badin, P. & Fant, G. (1984) Notes on vocal tract computation. *Speech Transmission Laboratory – Quarterly Progress and Status Report*, 2–3, 53–108.

Barney, A., Shadle, C. H., & Davies, P. O. A. L. (1999) Fluid flow in a dynamic mechanical model of the vocal folds and tract, I: Measurements and theory. *Journal of the Acoustical Society of America*, 105, 444–55.

Batchelor, G. K. (1967) *An Introduction to Fluid Dynamics*. Cambridge: Cambridge University Press.

Benade, A. H. (1976) *Fundamentals of Musical Acoustics*. New York: Oxford University Press.

Blazek, J. (2005) *Computational Fluid Dynamics: Principles and Applications*, 2nd edn. Amsterdam: Elsevier Science.

Busnel, R. G. & Classe, A. (1976) *Whistled Languages*. New York: Springer-Verlag.

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

Chanaud, R. C. & Powell, A. (1965) Some experiments concerning the hole and ring tone. *Journal of the Acoustical Society of America*, 37, 902–11.

Conrad, W. A. (1985) Collapsible tube model of the larynx. In I. R. Titze & R. C. Scherer (eds.), *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control* (pp. 328–48). Denver: Denver Center for the Performing Arts.

Cranen, B. & Boves, L. (1985) Pressure measurements during speech production using semiconductor miniature pressure transducers: Impact on models for speech production. *Journal of the Acoustical Society of America*, 77, 1543–51.

Cranen, B. & Boves, L. (1988) On the measurement of glottal flow. *Journal of the Acoustical Society of America*, 84, 888–900.

Davies, P. O. A. L. (1991) Program suite VOAC. Unpublished program documentation, Institute of Sound and Vibration Research, University of Southampton.

Davies, P. O. A. L., McGowan, R. S., & Shadle, C. H. (1993) Practical flow duct acoustics applied to the vocal tract. In I. R. Titze (ed.), *Vocal Fold Physiology: Frontiers in Basic Science* (pp. 93–142). San Diego: Singular Publishing Group, Inc.

Doebelin, E. O. (1983) *Measurement Systems: Application and Design*, 3rd edn. London: McGraw-Hill.

Draper, M. H., Ladefoged, P., & Whitteridge, D. (1959) Respiratory muscles in speech. *Journal of Speech and Hearing Research*, 2, 16–27.

Elder, S. A., Farabee, T. M., & Demetz, F. C. (1982) Mechanisms of flow-excited cavity tones at low Mach number. *Journal of the Acoustical Society of America*, 72, 532–49.

Erath, B. D. & Plesniak, M. W. (2006) The occurrence of the Coanda effect in pulsatile flow through static models of the human vocal folds. *Journal of the Acoustical Society of America*, 120, 1000–11.

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

Flanagan, J. L. (1972) *Speech Analysis Synthesis and Perception*, 2nd edn. Berlin: Springer-Verlag.

Flanagan, J. L. & Ishizaka, K. (1976) Automatic generation of voiceless excitation in a vocal cord-vocal tract speech synthesizer. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24, 163–70.

Flanagan, J. L. & Landgraf, L. L. (1968) Self-oscillating source for vocal-tract synthesizers. *IEEE Transactions on Audio and Electroacoustics*, AU-16, 57–64.

Fritzell, B., Hammarberg, B., Gauffin, J., Karlsson, I., & Sundberg, J. (1986) Breathiness and insufficient vocal fold closure. *Journal of Phonetics*, 14, 549–53.

Gauffin, J. & Sundberg, J. (1989) Spectral correlates of glottal voice source waveform characteristics. *Journal of Speech and Hearing Research*, 32, 556–65.

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

Guérin, B. (1983) Effects of the source–tract interaction using vocal fold models. In I. R. Titze & R. C. Scherer (eds.), *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control* (pp. 482–99). Denver: Denver Center for the Performing Arts.

Gunter, H. (2003) A mechanical model of vocal-fold collision with high spatial and temporal resolution. *Journal of the Acoustical Society of America*, 113, 994–1000.

Halliday, D. & Resnick, R. (1966) *Physics*. New York: John Wiley.

Hammarberg, B., Fritzell, B., & Schiratzki, H. (1984) Teflon injection in 16 patients with paralytic dysphonia: Perceptual and acoustic evaluation. *Journal of Speech and Hearing Disorders*, 49, 72–82.

Heinz, J. M. (1956) Fricative consonants. *MIT Research Laboratory of Electronics Quarterly Report*, Oct–Dec., 57.

Heller, H. H. & Widnall, S. E. (1970) Sound radiation from rigid flow spoilers correlated with fluctuating forces. *Journal of the Acoustical Society of America*, 47, 924–36.

Hertegård, S. & Gauffin, J. (1992) Acoustic properties of the Rothenberg mask. *Speech Transmission Laboratory – Quarterly Progress and Status Report*, 2–3, 9–18.

Hirschberg, A., Pelorson, X., Hofmans, G. C. J., Hassel, R. R. van, & Wijnands, A. P. J. (1996) Starting transient of the flow through an in-vitro model of the vocal folds. In P. J. Davis & N. H. Fletcher (eds.), *Vocal Fold Physiology: Controlling Complexity and Chaos* (pp. 31–46). San Diego: Singular Publishing Group, Inc.

Hixon, T., Goldman, M., & Mead, J. (1973) Kinematics of the chest wall during speech production: Volume displacements of the rib cage, abdomen, and lung. *Journal of Speech and Hearing Research*, 16, 78–115.

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

Hofmans, G. C. F., Groot, G., Ranucci, M., Graziani, G., & Hirschberg, A. (2003) Unsteady flow through in-vitro models of the glottis. *Journal of the Acoustical Society of America*, 113, 1658–75.

Holger, D. K., Wilson, T. A., & Beavers, G. S. (1977) Fluid mechanics of the edgetone. *Journal of the Acoustical Society of America*, 62, 1116–28.

Holmberg, E. B., Hillman, R. E., & Perkell, J. S. (1988) Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *Journal of the Acoustical Society of America*, 84, 511–29.

Howe, M. S. & McGowan, R. S. (2005) Aeroacoustics of [s]. *Proceedings of the Royal Society A*, 461: 2056, 1005–28.

Iijima, H., Miki, N., & Nagai, N. (1990) Finite-element analysis of a vocal cord model with muscle of nonhomogeneous elasticity. *Journal of the Acoustical Society of Japan*, (E)11, 53–6.

Ingard, K. U. & Ising, H. (1967) Acoustic nonlinearity of an orifice. *Journal of the Acoustical Society of America*, 42, 6–17.

Ishizaka, K. & Flanagan, J. L. (1972) Synthesis of voiced sounds from a

two-mass model of the vocal cords. *Bell System Technical Journal*, 51, 1233–68.

Isshiki, N. (1964) Regulating mechanisms of vocal intensity variation. *Journal of Speech and Hearing Research*, 7, 17–29.

Jackson, P. J. B. & Shadle, C. H. (2000) Frication noise modulated by voicing, as revealed by pitch-scaled decomposition. *Journal of the Acoustical Society of America*, 108, 1421–34.

Jackson, P. J. B. & Shadle, C. H. (2001) Pitch-scaled estimation of simultaneous voiced and turbulence-noise components in speech. *IEEE Transactions on Speech and Audio Processing*, 9, 713–26.

Jiang, J. J., Diaz, C. E., & Hanson, D. G. (1998) Finite element modeling of vocal fold vibration in normal phonation and hyperfunctional dysphonia: Implications for the pathogenesis of vocal nodules. *Annals of Otology, Rhinology and Laryngology*, 107, 603–10.

Jiang, J. J. & Titze, I. R. (1994) Measurement of vocal fold intraglottal pressure and impact stress, *Journal of Voice*, 8, 132–44.

Karlsson, I. (1986) Glottal wave forms for normal female speakers. *Journal of Phonetics*, 14, 415–19.

Karlsson, I. (1992) Analysis and Synthesis of Different Voices with an Emphasis on Female Speech (ISRN KTH/ToM/FR–92/3–SE TRITA-TOM 1992:3). Doctoral Dissertation, Royal Institute of Technology, KTH, Department of Speech Communication and Music Acoustics, Stockholm.

Kinsler, L. E., Frey, A. R., Coppens, A. B., & Sanders, J. V. (1982) *Fundamentals of Acoustics*, 3rd edn. New York: John Wiley.

Koizumi, T. & Taniguchi, S. (1990) A novel model of pathological vocal cords and its application to the diagnosis of vocal cord polyp. In *Proceedings of the International Conference on Speech and Language Processing* (pp. 73–6). Kobe: Acoustical Society of Japan.

Krane, M. H. (2005) Aeroacoustic production of low-frequency unvoiced speech sounds. *Journal of the Acoustical Society of America*, 118, 410–27.

Krane, M. H., Sinder, D., & Flanagan, J. (1998) Approximate computational model for sound generation due to unsteady flows in pipes. Proceedings of the 16th International Congress on Acoustics and 135th meeting of the Acoustical Society of America, Seattle, WA; *Journal of the Acoustical Society of America*, 103, 2795.

Ladefoged, P. & Maddieson, I. (1996) *The Sounds of the World's Languages*. Oxford: Blackwell.

Liljencrants, J. (1991a) A translating and rotating mass model of the vocal folds. *Speech Transmission Laboratory – Quarterly Progress and Status Report*, 1, 1–18.

Liljencrants, J. (1991b) Numerical simulations of glottal flow. In J. Gauffin & B. Hammarberg (eds.), *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms* (pp. 99–104). San Diego: Singular Publication Group Inc.

Maeda, S. (1987) On the generation of sound in stop consonants. *Speech Communication Group Working Papers*, Research Laboratory of Electronics, MIT, 5, 1–14.

Massey, B. S. (1984) *Mechanics of Fluids*, 5th edn. Wokingham, UK: Van Nostrand Reinhold.

McGowan, R. S. (1988) An aeroacoustic approach to phonation. *Journal of the Acoustical Society of America*, 83, 696–704.

McGowan, R. S. (1992) Tongue-tip trills and vocal-tract wall compliance. *Journal of the Acoustical Society of America*, 91, 2903–10.

McGowan, R. S. & Howe, M. S. (2007) Compact Green's functions extend the acoustic theory of speech production. *Journal of Phonetics*, 35, 259–70.

Meyer, J. & Gautheron, B. (2006) Whistled speech and whistled languages. In K. Brown (ed.), *Encyclopedia of Language and Linguistics*, 2nd edn. (vol. 13, pp. 573–6). Oxford: Elsevier.

Meyer-Eppler, W. (1953) Zum Erzeugungsmechanismus der Gerauschlaute [On the generating mechanism of noise sounds]. *Zeitschrift für Phonetik*, 7, 196–212.

Mongeau, L., Franchek, N., Coker, C. H., & Kubli, R. A. (1997) Characteristics of a pulsating jet through a small modulated orifice, with application to voice production. *Journal of the Acoustical Society of America*, 102, 1121–33.

Narayanan, S. & Alwan, A. (2000) Noise source models for fricative consonants. *IEEE Transactions on Speech and Audio Processing*, 8, 328–44.

Negus, V. E. (1949) *The Comparative Anatomy and Physiology of the Larynx*. London: Heinemann.

Nelson, P. A. & Morfey, C. L. (1981) Aerodynamic sound production in low speed flow ducts. *Journal of Sound and Vibration*, 79, 263–89.

Ohala, J. J. (1990) Respiratory activity in speech. In W. J. Hardcastle & A. Marchal (eds.), *Speech Production and Speech Modelling* (pp. 23–54). Dordrecht: Kluwer.

Pastel, L. M. P. (1987) Turbulent Noise Sources in Vocal Tract Models. MS Thesis, MIT.

Pelorson, X. (2001) On the meaning and accuracy of the pressure-flow technique to determine constriction areas within the vocal tract. *Speech Communication*, 35, 179–90.

Pelorson, X., Hirschberg, A., Hassel, R. R. van, & Wijnands, A. P. J. (1994) Theoretical and experimental study of quasi-steady flow separation within the glottis during phonation: Application to a modified two-mass model. *Journal of the Acoustical Society of America*, 96, 3416–31.

Pelorson, X., Hirschberg, A., Wijnands, A. P. J., & Bailliet, H. (1995) Description of the flow through in-vitro models of the glottis during phonation. *Acta Acustica*, 3, 191–202.

Pelorson, X., Hofmans, G. C. J., Ranucci, M., & Bosch, R. C. M. (1997) On the fluid mechanics of bilabial plosives. *Speech Communication*, 22, 155–72.

Powell, A. (1961) On the edgetone. *Journal of the Acoustical Society of America*, 33, 395–409.

Powell, A. (1962) Vortex action in edgetones. *Journal of the Acoustical Society of America*, 34, 163–6.

Recasens, D. (1991) On the production characteristics of apicoalveolar taps and trills. *Journal of Phonetics*, 19, 267–80.

Rothenberg, M. T. (1973) A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *Journal of the Acoustical Society of America*, 53, 1632–45.

Rothenberg, M. T. (1981) Acoustic interaction between the glottal source and the vocal tract. In K. N. Stevens & M. Hirano (eds.), *Vocal Fold Physiology* (pp. 305–23). Tokyo: University of Tokyo Press.

Sawashima, M. (1977) Fiberoptic observation of the larynx and other speech organs. In M. Sawashima & F. S. Cooper (eds.), *Dynamic Aspects of Speech Production* (pp. 31–47). Tokyo: University of Tokyo Press.

Scherer, R. C. & Guo, C.-G. (1991) Generalized translaryngeal pressure coefficients for a wide range of laryngeal configurations. In J. Gauffin & B. Hammarberg (eds.), *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms* (pp. 83–90). San Diego: Singular Publishing Group, Inc.

Scherer, R. C. & Titze, I. R. (1983) Pressure–flow relationships in a model of the laryngeal airway with diverging glottis. In D. M. Bless & J. M. Abbs (eds.), *Vocal Fold Physiology: Contemporary Research and Clinical Issues* (pp. 179–93). San Diego: College-Hill Press.

Scherer, R. C., Titze, I. R., & Curtis, J. F. (1983) Pressure–flow relationships in two models of the larynx having rectangular glottal shapes. *Journal of the Acoustical Society of America*, 73, 668–76.

Tao, C. & Jiang, J. J. (2006) Anterior-posterior biphonation in a finite element model of vocal fold vibration. *Journal of the Acoustical Society of America*, 120, 1570–7.

Tao, C., Jiang, J. J., & Zhang, Y. (2006) Simulation of vocal fold impact pressures with a self-oscillating finite-element model. *Journal of the Acoustical Society of America*, 119, 3987–94.

Teager, H. M. (1980) Some observations on oral air flow during phonation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28, 599–601.

Teager, H. M. & Teager, S. M. (1983) Active fluid dynamic voice production models, or there is a unicorn in the garden. In I. R. Titze & R. C. Scherer (eds.), *Vocal Fold Physiology* (pp. 387–401). Denver: The Denver Center for Performing Arts.

Thomas, C. (ed.) (1973) *Taber's Cyclopedic Medical Dictionary*, 12th edn. Philadelphia: F. A. Davis Co.

Thomas, T. J. (1986) A finite element model of fluid flow in the vocal tract. *Computer Speech and Language*, 1, 131–52.

Titze, I. R. (1973) The human vocal cords: A mathematical model, Part I. *Phonetica*, 28, 129–70.

Titze, I. R. (1974) The human vocal cords: A mathematical model, Part II. *Phonetica*, 29, 1–21.

Titze, I. R. (1992) Phonation threshold pressure: A missing link in glottal aerodynamics. *Journal of the Acoustical Society of America*, 91, 2926–35.

Titze, I. R. (2000) *Principles of Voice Production*, 2nd printing. Iowa City, IA: National Center for Voice and Speech.

Titze, I. R. (2006) *The Myoelastic Aerodynamic Theory of Phonation*. Denver: National Center for Voice and Speech.

Titze, I. R. & Talkin, D. T. (1979) A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation. *Journal of the Acoustical Society of America*, 66, 60–74.

Westbury, J. R. (1983) Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America*, 73, 1322–36.

Zhang, C., Zhao, W., Frankel, S. H., & Mongeau, L. (2002) Computational aeroacoustics of phonation, Part II: Effects of flow parameters and ventricular folds. *Journal of the Acoustical Society of America*, 112, 2147–54.

Zhang, Z., Mongeau, L., & Frankel, S. H. (2002) Experimental verification of the quasi-steady approximation for aerodynamic sound generation by pulsating jets in tubes. *Journal of the Acoustical Society of America*, 112, 1652–63.

Zhao, W., Zhang, C., Frankel, S. H., & Mongeau, L. (2002) Computational aeroacoustics of phonation, Part I: Computational methods and sound generation mechanisms. *Journal of the Acoustical Society of America*, 112, 2134–46.

FURTHER READING

Baken, R. J. (1987) *Clinical Measurement of Speech and Voice*. Boston: College-Hill Press.

Sundberg, J. (1987) *The Science of the Singing Voice*. DeKalb, IL: Northern Illinois University Press.

Tennekes, H. & Lumley, J. L. (1972) *A First Course in Turbulence*. Cambridge, MA: MIT Press.

Versteeg, H. K. & Malalasekera, W. (2007) *An Introduction to Computational Fluid Mechanics: The Finite Volume Method*, 2nd edn. Harlow, UK: Pearson Education Ltd., Prentice-Hall.

Wagner, C., Hüttl, T., & Sagaut, P. (eds.) (2007) *Large-Eddy Simulation for Acoustics*. New York: Cambridge University Press.

Scully, C. (1986) Speech production simulated with a functional model of the larynx and the vocal tract. *Journal of Phonetics*, 14, 407–13.

Scully, C. (1990) Articulatory synthesis. In W. J. Hardcastle & A. Marchal (eds.), *Speech Production and Speech Modelling* (pp. 151–86). Dordrecht: Kluwer.

Shadle, C. H. (1983) Experiments on the acoustics of whistling. *The Physics Teacher*, March, 148–54.

Shadle, C. H. (1985) The Acoustics of Fricative Consonants. Ph.D. thesis, MIT Research Laboratory of Electronics, Technology Report 506.

Shadle, C. H. (1990) Articulatory-acoustic relationships in fricative consonants. In W. J. Hardcastle & A. Marchal (eds.), *Speech Production and Speech Modelling* (pp. 187–209). Dordrecht: Kluwer.

Shadle, C. H. (1991) The effect of geometry on source mechanisms of fricative consonants. *Journal of Phonetics*, 19, 409–24.

Shadle, C. H., Barney, A. M., & Davies, P. O. A. L. (1999) Fluid flow in a dynamic mechanical model of the vocal folds and tract, II: Implications for speech production studies. *Journal of the Acoustical Society of America*, 105, 456–66.

Shadle, C. H., Barney, A. M., & Thomas, D. W. (1991) An investigation into the acoustics and aerodynamics of the larynx. In J. Gauffin & B. Hammarberg (eds.), *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms* (pp. 73–82), San Diego: Singular Publishing Group, Inc.

Shadle, C. H. & Scully, C. (1995) An articulatory-acoustic-aerodynamic analysis of [s] in VCV sequences. *Journal of Phonetics*, 23, 53–66.

Shinwari, D., Scherer, R. C., DeWitt, K. J., & Afjeh, A. A. (2003) Flow visualization and pressure distribution in a model of the glottis with a symmetric and oblique divergent angle of 10 degrees. *Journal of the Acoustical Society of America*, 113, 487–97.

Sinder, D. J. (1999) Speech Synthesis Using an Aeroacoustic Fricative Model. Ph.D. thesis, Rutgers State University of New Jersey.

Slifka, J. (2003) Respiratory constraints on speech production: Starting an utterance. *Journal of the Acoustical Society of America*, 114, 3343–53.

Södersten, M. & Lindestad, P.-Å. (1990) Glottal closure and perceived breathiness during phonation in normally speaking subjects. *Journal of Speech and Hearing Research*, 33, 601–11.

Södersten, M., Lindestad, P.-Å., & Hammarberg, B. (1991) Vocal fold closure, perceived breathiness, and acoustic characteristics in normal adult speakers. In J. Gauffin & B. Hammarberg (eds.), *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms* (pp. 217–24), San Diego: Singular Publishing Group, Inc.

Sondhi, M. M. & Schroeter, J. (1987) A hybrid time-frequency domain articulatory speech synthesizer. *IEEE Transactions on Acoustics, Speech, and Signal Processing* (ASSP-35:7, July), 955–67.

Stevens, K. N. (1971) Airflow and turbulence noise for fricative and stop consonants: Static considerations. *Journal of the Acoustical Society of America*, 50, 1180–92.

Stevens, K. N. (1993) Models for the production and acoustics of stop consonants. *Speech Communication*, 13, 367–75.

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

Story, B. H. & Titze, I. R. (1995) Voice simulation with a body-cover model of the vocal folds, *Journal of the Acoustical Society of America*, 97, 1249–60.

Suh, J. & Frankel, S. H. (2007) Numerical simulation of turbulence transition and sound radiation for flow through a rigid glottal model. *Journal of the Acoustical Society of America*, 121, 3728–39.