830

CHAPTER 3

Practical Flow Duct Acoustics Applied to the Vocal Tract

P.O.A.L. DAVIES, Ph.D.
RICHARD S. McGOWAN, Ph.D.
CHRISTINE H. SHADLE, Ph.D.

During phonation, steady flow of air through the vocal tract is converted into pulsed flow, thereby creating sound for a listener. The conversion of the steady flow to pulsed flow, and the subsequent transport of the unsteady air motion in the vocal tract to the atmosphere, determine all the characteristics of the sound: its fundamental frequency and amplitude, its spectral composition, including upper harmonic content and noise, and the way these properties change with time. The mechanical properties and dynamics of the folds are important in determining the quality of sound. Also important are the mechanical properties and dynamics of the unsteady air motion and interacting solid surfaces in the supralaryngeal tract, which determine the way that unsteady motion is transported through the air from the folds to the mouth. This is the subject of sound propagation, or in the terminology of speech scientists, the characteristics of the filter, as opposed to the source. In this paper we primarily consider propagation, and present a realistic computer model of the acoustical characteristics of the vocal tract.

THE ACOUSTIC APPROXIMATION

What are the properties of the air and its motion that help determine whether the motion is acoustic or not? There are the constituent properties, such as density, temperature, pressure, speed of sound, and viscosity, to name some of them. The base values (determined while the air is at rest) of these quantities are used to determine whether perturbations are small or large. If the unsteady variation in pressure is several percent of the pressure of the air at rest, then the perturbation may not be considered to be small, and the acoustic approximation may not apply, depending on the error that can be tolerated. With the acoustic approximation, squares and products of fluctuating quantities are assumed to be negligibly small compared to both the mean and first order fluctuating quantities, and the motion remains isentropic or homentropic[1]. This acoustic approximation then results in linearization of the equations of motion with both time and space.

The acoustic approximation provides a realistic model of the physical processes that exist when a loudspeaker drives the air in a cylindrical tube at low amplitudes. This simple picture, in which all air motion is also acoustic wave motion, can be made progressively more realistic when one adds features that are relevant to sound propagation in the vocal tract. The shape of the tube can be altered; it is widely understood that the complex geometric details of the vocal tract modify the resonances. If the walls were made of a non-rigid material, corresponding more closely to those of the vocal tract, they would react to the presence of sound, producing both attenuation and dispersion. If air flowed through the tube, sound would travel faster downstream than upstream, and these convection effects of the mean flow may shift formant frequencies. Furthermore, the mean flow can generate sound at various stations along the tract by the mechanisms of vortex shedding and flow turbulence (Shadle, 1985, 1990). A further possibility is that sound can be scattered and absorbed by turbulence (Howe, 1984). Having considered the ways in which these phenomena can be characterized in general, we shall then consider features in some depth that directly modify sound propagation. Vortex shedding and flow turbulence are also significant for a complete understanding of sound generation mechanisms, which are briefly considered in the section "A Survey of Phenomena not Accounted for by the Program VOAC."

---

[1] Note that "isentropic" implies that entropy remains the same over a particular volume of air, while "homentropic" means that entropy remains the same over a particular mass of air.

The usual assumptions of classical acoustic theory applied to the vocal tract are as follows:

1. that the acoustic medium is a frictionless, homogeneous (ideal) fluid at rest on the average,
2. that processes associated with the wave motion are isentropic,
3. that the fluctuating pressure amplitudes remain sufficiently small that the linearizing acoustic assumptions are valid,
4. that the wave propagation is wholly axial and directed along the duct axis (x, say),
5. that the tract axis is not curved,
6. that the walls of the tract are acoustically rigid, and
7. that individual sections of the duct have constant cross-sectional area, and thus
8. that the wave fronts remain plane, and normal to the duct axis, and finally
9. that abruptness in the area function is due to the approximate, and not to the original, area function. Therefore transfer across junctions can be treated the same regardless of degree of abruptness, and effects due to abruptness can be neglected (as noted by McGowan, 1987).

Many, though not all, of these assumptions have been relaxed for the present study, as described below.

It is worth mentioning at the outset that the observed acoustic behavior of air in the vocal tract is not that of an ideal fluid; this has a number of consequences. Thus, for (1) above the air is treated as a saturated viscous heat-conducting fluid, with a flow having time-averaged velocity u_0. However, it is assumed that, essentially, u_0 remains uniform over any cross section. The presence of water vapor means that the ratio of the specific heats, γ, the density, ρ, the speed of sound, c, and the characteristic impedance must all be assigned appropriate values; those for saturated air at 37°C are assumed to apply in the tract.

Before considering the rest of the assumptions, we present definitions of some nondimensional parameters based on these physical characteristics of the fluid and the flow which will aid us in describing the implications of keeping or relaxing each of the assumptions.

The influence of viscosity can be characterized by the value of the Reynolds number, Re_D, where

$$Re_D = \frac{vD}{\nu} \qquad (3\text{-}1)$$

where v is a characteristic time- and locally space-averaged velocity, D is a characteristic length scale, and ν is the kinematic viscosity (0.167 cm²/s). The Reynolds number represents the ratio of inertial to viscous forces in the flow, and when D is set equal to the hydraulic diameter[2] of the tract, its value is one factor that characterizes the state of the bulk motion, i.e. whether it remains laminar or turbulent. Another factor is the presence or absence of other inhomogeneities in the flow or on the bounding surfaces. Meyer-Eppler (1953) related the sound output of pinched plastic tubing to the fricatives /s, ʃ, f/ and concluded that Re_D for the vocal tract is around 1700; this value has been used since by Catford (1977) to describe speech production, and by Flanagan and Ishizaka (1976) to synthesize fricatives. Pedley, Schroter, and Sudlow (1970) regarded flow in the bronchial tube as turbulent with Re_D greater than 2,000. For a stressed /h/ at a high mean flow of 1200 cm³/s and a phonated vowel at normal flow of 200 cm³/s their critical values correspond respectively to a tract diameter (or area, if tract cross-section is assumed circular) of 6 cm (28 cm²) and 1 cm (0.8 cm²). These values for bulk flow turbulence seem surprising since it is generally understood that transition to fully developed turbulent flow only exists in relatively long tubes (length/diameter greater than 40, say) when $Re_D > 2500$ or more. Turbulence in the vocal tract will normally be associated with separating shear layers, which are characterized by the form of the local geometry rather than the value of the Reynolds number. It is possible that the problem lies in comparing Reynolds numbers of significantly different geometries, that is, of smooth tubing with a single elliptical constriction and the more complex vocal tract.

When D is instead taken in the x direction (parallel to the tract axis) and set equal to a relevant length scale, such as the distance from area discontinuities, Re_x characterizes the properties of the boundary layer flow adjacent to the walls. The boundary layer is the relatively thin flow region where the streamwise velocity rises rapidly and continuously from zero at the walls (no-slip condition) to the undisturbed value in the main flow outside the layer. The transverse velocity gradient is largest at the walls, decreasing outwards through the layer. Associated with this gradient are corresponding values of the local shear stress. With steady flow along a smooth flat plate the boundary layer flow remains laminar up to values of Re_x of around 5×10^5. This value is reduced by adverse (rising) and increased by favorable (falling) streamwise pressure gradients in the external flow. With laminar flow in the boundary layer, as would normally exist in the vocal tract, the boundary layer thickness is proportional to $1/\sqrt{Re_x}$.

Velocity fluctuations associated with the acoustic waves must also be zero at the walls; the associated fluctuating viscous stresses produce an acoustic boundary layer. The effective thickness of the acoustic boundary layer is characterized by the value of the Stokes number, S_u, defined by

$$S_u = r_0 \sqrt{\frac{\omega}{\nu}} \,, \tag{3-2}$$

where ω is the radian frequency of the sound and r_0 is the radius (or hydraulic radius) of the tract. In this case it represents the width of the acoustic boundary layer relative to that of the tract. The small temperature fluctuations associated with the acoustic pressure fluctuations are also modified in the vicinity of the walls giving an associated thermal boundary layer. The width of this thermal boundary layer in relation to the viscous boundary layer is characterized by the value of the Prandtl number, Pr, where

$$Pr = \frac{\nu}{\kappa} \,, \tag{3-3}$$

which is the ratio of the kinematic viscosity ν (i.e. the viscous diffusivity of momentum) to the thermal diffusivity of heat, $\kappa = k/\rho c_p$, where k = thermal conductivity, ρ = density, and c_p = specific heat at constant pressure of the fluid. For air, the Prandtl number equals 0.7 over a wide temperature range.

Another feature of real fluid flows is that they cannot sustain tensile stresses. Thus the flow separates at sharp corners and expansions, forming vortex sheets, and also from bounding surfaces in the presence of adverse (rising) time-averaged or fluctuating pressure gradients. With steady, unsteady, or pulsatile flow, time-dependent flow separation may occur. The frequency of vortex shedding associated with these flow processes is then characterized by the value of the Strouhal number, St, defined by

$$St = \frac{fD}{v} \tag{3-4}$$

where f is frequency, D is a characteristic length scale, and v is a characteristic time-averaged velocity. This quantity is relevant in deciding whether a flow is unsteady and must be treated as such, or can be analyzed using a quasi-steady approximation, as will be discussed later. Flow separations are also important in relation to the simplifying assumption (2) above, that the process is isentropic, since

---

[2] The hydraulic diameter or radius equals the actual diameter or radius for a tract of circular cross-section. For any cross-sectional shape, the hydraulic radius h, is quantified as $2A/P$ where A = cross sectional area, P = perimeter length.

the entropy fluctuations accompanying flow separation can be acoustically significant.

Assumption (3) above, that the linearizing acoustic assumptions are valid, effectively amounts to the requirement that the speed of sound c in the medium should remain effectively invariant with space or time. When this is not the case, wave propagation is no longer described by linear, or acoustic, equations. Because the speed of sound is proportional to the square root of the absolute temperature, assumption (3) is also equivalent, through conservation of momentum, to the requirement that the ratio of the maximum particle velocity $U = u_0 + |u|$ to the local sound speed remains small, where $|u|$ is the peak amplitude of the velocity fluctuations $u(t)$. This is characterized by the value of the instantaneous Mach number, M_i, where

$$M_i = \frac{U}{c} \; . \tag{3-5}$$

So long as the value of M_i remains small in relation to unity, changes in pressure, with the corresponding changes in temperature, will be negligible and the sound speed c can be regarded as constant with space and time.

The value of the mean flow Mach number M, defined as

$$M = \frac{u_o}{c} \; , \tag{3-6}$$

also has significance, since its value quantifies the speed, defined as $c(1 \pm M)$, at which acoustic disturbances propagate along the vocal tract. Note that since u_o changes as the cross-sectional area varies, M must change, and therefore the phase speed varies along a nonuniform tract. For example, with a high volume flow of 1200 cm^3/s and a highly constricted area of 0.1 cm^2, M attains a value of 1/3 and the ratio of the downstream to upstream speed of propagation is two! At the other extreme, with a volume flow of 200 cm^3/s and a tract area of 4 cm^2, $M = 0.0014$ and its influence on the speed of propagation is correspondingly small. Nevertheless, for consistency the mean flow is included throughout the calculations reported here, while it is assumed that the speed of sound, c, remains constant.

In the current version of the program it is assumed that the linearized acoustic equations (assumption (3) in the list above) remain valid throughout the duct, except when the processes cease to remain isentropic, i.e. when assumption (2) is also relaxed. Curvature of the duct axis may be neglected, thus keeping assumption (5), at least as a first approximation. Tract curvature can readily be included later if this is deemed desirable; this is considered further in the section entitled, "A Survey of Phenomena not Accounted for by the Program VOAC."

Other assumptions in the list are also relaxed: assumption (7), that sections of the duct have constant cross-sectional area, has been relaxed to allow sections in which the area changes linearly and quadratically with distance along the duct. For these additional sections, assumption (8) is relaxed: the wave fronts are plane for constant area, cylindrical for linearly-changing area, and spherical for quadratically-changing area. Assumption (4), that of axial propagation, applies strictly only to sections of constant area, but applies in essence to all sections since cross-modes are not modelled. Assumption (6) is relaxed; the walls can be considered to be rigid or reacting. Finally, end corrections are computed at abrupt area changes, relaxing assumption (9). Some specific analytical problems were solved so that these additional features could be included, but the details are not included here.

The authors are certainly aware that some of the assumptions of classical theory are also relaxed in several existing useful models of the acoustic behavior of the vocal tract during phonation that are in the literature (e.g. Flanagan, 1972; Liljencrants, 1985; Scully, 1990). However, they all appear to be based on the well-documented analogy with electrical networks, while a different and, perhaps, more realistic approach has been followed in the studies reported here. Since, with zero mean flow, the linearized equations describing the inviscid motion of fluids in conduits are similar to those describing electric currents in uniform conductors or waveguides, it is often convenient to use the analogy that this similarity offers, at least as a first approximation. One would expect, with similarly shaped boundaries, that the patterns of motion of both would be similar, but those of fluid flow are often observed to differ profoundly. This is obvious, if one compares the expanding flow of an electric current with the flow of fluid from a tube as illustrated in Figure 3-1.

Helmholtz (1868) published a detailed discussion of the discrepancies between predicted and observed fluid motion, and many further examples have been published since then. He rightly concluded, "... this followed because a fluid cannot withstand tension and remains continuous only so long as the local pressure remains positive.". Alternatively, one might recall that inertial forces are insignificant for the moving electrons constituting an electric current compared with those arising from fluid motion. Thus, as in Figure 3-1, the flow separates at the edge of the tube and the description of the motion must include this discontinuity, where the surface of separation can be treated mathematically as a sheet of vorticity. Such sheets are unstable (Rayleigh, 1896, ch. xxi) and are observed to roll up to form vortices. It is clear that, in realistic models of vocal tract acoustic behavior, due

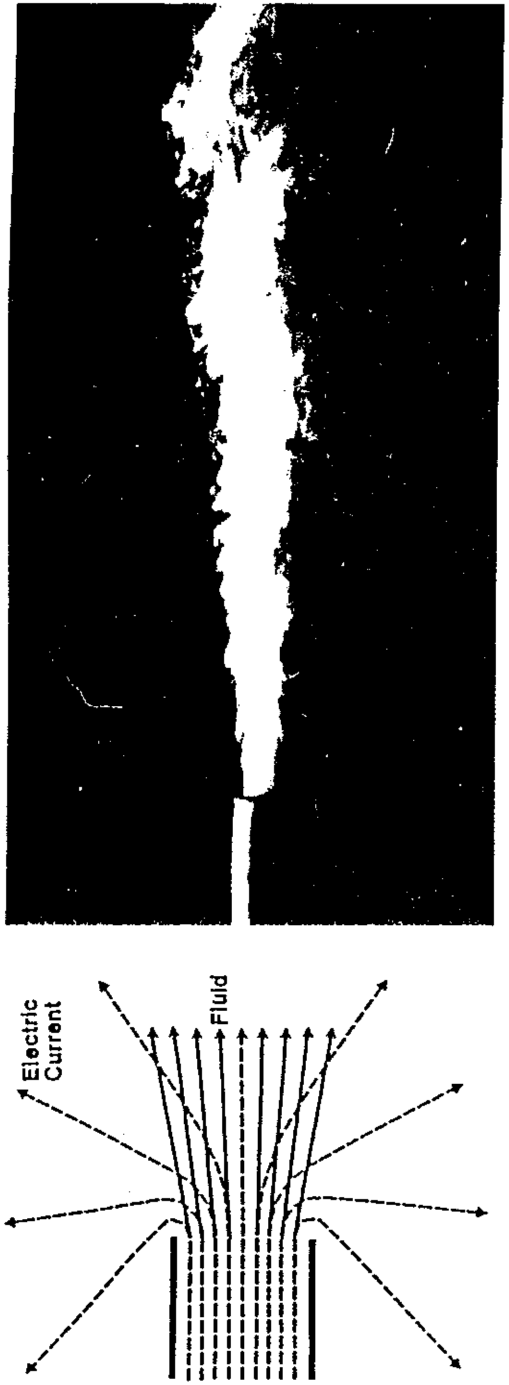

Figure 3-1. Comparison of a fluid jet with the flow of an electric current (from Davies, 1975).

account must be taken of such fluctuating fields of vorticity, when their contributions become significant. This turns out to be the case, for example, at abrupt area discontinuities where the vortex motions modify the wave propagation (Davies, 1988), though their existence might be represented in an electrical analogy as, say, an added reactance. However, this does not appear to have been included in the classic model of such area discontinuities or by, for instance, Kelly and Lochbaum (1962). It is worth repeating that the adoption of electro-acoustic analogies has been avoided in the present algorithms as they are somewhat inflexible in application. A thorough description of the novel features of the program is included in a later section on simulation.

The general approach to our study of the fluid mechanics of the vocal tract is to add complexity to the simple acoustic propagation models, so that the effect of each addition is understood. We believe that this is the most effective way to build a model of propagation in the vocal tract. The following section therefore explains our model of propagation, VOAC, in detail, and the section "Simulations" includes thereafter results of simulations using VOAC. On the other hand, full continuum simulations, which are based on the differential forms of the conservation equations and mostly restricted to the glottal region at present, have a place in studies of the fluid mechanics of the vocal tract (Thomas, 1986; Iijima, Miki, and Nagai, 1990; Liljencrants, 1991). These are discussed further, along with phenomena *not* simulated by VOAC, in the section titled "A Survey of Phenomena not Accounted for by the Program VOAC."

It is imperative when considering such aeroacoustic phenomena to define what constitutes a sound, since not all of the pressure and velocity fluctuations occurring within the vocal tract are necessarily associated with acoustic wave motion. Such motion has by definition a characteristic phase velocity equal to the local speed of sound relative to the medium in the tract. Furthermore, though there is a net transfer of energy by a sound wave, there is no net transfer of mass or momentum. However, when momentum fluctuations being transported by the mean flow reach any area discontinuity, e.g. the lips, they can generate acoustic energy which is then radiated. Thus, in the vocal tract, where there is a net flow of air from the lungs, there is an additional convection of unsteady motion by this flow with the possibility of further sound generation. Such turbulent convection is not included in the simulations here. However, the analysis presented here takes due account of the convective effects on acoustic propagation, so comparison is included in the subsection "Mean Flow Effects" of the consequences of neglecting this factor in the resonances and other acoustic phenomena associated with speech. Generation of sound by the flow is not considered.

The application of non-acoustic aspects of fluid mechanics to speech is already established in regards to vocal fold vibration. For instance, the turbulence formed near the glottis is essential in describing the vibration of the vocal folds (Titze, 1988). However, in the application of fluid mechanics to propagation, only the simplest acoustic theory has been applied. That is, any unsteady air motion near the glottis has been assumed to be acoustic wave motion, and that acoustic wave has been assumed to travel along the tract exclusively by linear, isentropic, one-dimensional, or plane wave propagation. Perhaps other aspects of fluid mechanics should be applied to describe unsteady momentum transport in the vocal tract by the mean flow.

There are practical reasons why there may be concern for the details of propagation when studying phonation. One is the subject of source-tract interaction, whether it is the supralaryngeal air motion affecting the motion of the folds or the flow of air through the glottis. Also, to obtain the source signal it is necessary to remove the effect of the propagation, as much as possible, from the total speech signal. Removing resonance peaks from the spectrum can be done in a relatively straightforward manner. However, there are other aspects of propagation that are more difficult to remove from the source and, in fact, source character can be determined to a large extent by the environment into which it is radiating. For example, by using an electro-acoustic analog of the tract, Rothenberg (1981a) showed that the vocal tract affects the glottal volume velocity pulse skewness, and hence the overall spectral tilt. This source-tract interaction usually is not factored out of the inverse filtered signal, but it could be done with a model that calculated input impedance of the tract. Implications for inverse filtering methods that arise from the model reported in this paper will be discussed in the section titled, "Implications for Inverse Filtering."

It is conceivable that aspects of air motion not included in the acoustic propagation model of the vocal tract can have effects on propagation that need to be factored out, just as in the case of pulse skewness. Thus, here we study the acoustic propagation theory with a view to finding the range of parameters for which it is a valid approximation and finding amendments to it when it is not.

THE VOCAL TRACT ACOUSTIC PREDICTION PROGRAM (VOAC)

BASIC DESCRIPTION OF VOAC

The VOcal tract ACoustic prediction program, VOAC, first calculates the distribution along the axis of the vocal tract of the complex amplitude spectra $p^+(f)$, $p^-(f)$ of the respectively positively and negatively travelling component waves, where the direction from glottis to lips is taken as positive. The calculations, however, start at the lips where the amplitude of the incident wave is assigned the value unity, so that all other amplitudes are expressed relative to this. They then proceed systematically along the tract terminating at the glottis. Although a special program has been written for the vocal tract acoustic predictions, the general approach follows that originally set out in Alfredson and Davies (1971), but includes among others the developments described in Davies (1988) and Davies and Doak (1990a,b).

In the program the axially varying geometry of the vocal tract is represented by a sequence of elements that provide a simplified but close approximation to the observed geometry during phonation. Some individual elements consist of one or more short lengths of tract with constant cross-sectional area in sequence (Types 1, 4, and 5). Of these, Type 1 includes provision for abrupt area changes at the junctions between each of its component units with constant sectional area; Type 4 includes an additional provision for the insertion of sinuses at the junctions; Type 5 is a simple tube. Of the elements that incorporate a changing cross-sectional area, the Type 2 element retains a constant width but has area varying linearly with distance; the Type 3 element has a quadratic change in area with distance. The elements are designed to include *within* them every type of area discontinuity or change; thus adjacent elements have identical areas and hydraulic radii at their adjoining faces. This then allows transmission through each element type to be handled by a separate algorithm, a fact that is used to advantage.

Wave propagation is assumed to be always directed along the axis of the tract, with wave fronts that are plane in the elements with constant area, cylindrical in Type 2 elements, and spherical in Type 3. The radius of curvature of the vocal tract axis remains sufficiently large compared with the transverse dimensions for its influence on wave propagation to remain small, so this has been neglected. (But see the section titled "Curvature of the Duct Axis" for further discussion.) As an initial approximation made in order to speed development of the current program, the wave fronts in Type 2 elements have been assumed to remain plane, but this assumption will be eliminated in due course. Wave propagation along each element is then calculated with the appropriate form of the convected wave equation (Davies, 1988; Davies and Doak, 1990a), which matches the conditions imposed by the local geometry of the tract boundaries. For example, with isentropic plane waves and a steady axial uniform flow with time-averaged velocity u_0 in a tract of constant area, the convected wave equation becomes

$$\frac{D^2p}{Dt^2} + c^2\nabla^2 p = 0\,, \tag{3-7}$$

where p is the fluctuating pressure, c the speed of sound, and the operator D^2/Dt^2 represents the material derivative. In this case D^2/Dt^2 factors to

$$\left(\frac{\partial}{\partial t} + u_o\,\frac{\partial}{\partial x}\right)\left(\frac{\partial}{\partial t} + u_o\,\frac{\partial}{\partial x}\right). \tag{3-8}$$

It is further assumed that squares and products of fluctuating quantities are negligibly small compared to products of mean quantities with both the respective mean and first order fluctuating quantities (the acoustic approximation).

With an ideal inviscid acoustic medium and rigid walls, equation 3-7 is satisfied by

$$\begin{aligned} p^+(t - x/c) &= p^+(0)\exp i(\omega t - k^+ x),\\ u^+(t - x/c) &= u^+(0)\exp i(\omega t - k^+ x), \end{aligned} \tag{3-9}$$

$$\begin{aligned} p^-(t + x/c) &= p^-(0)\exp i(\omega t + k^- x),\\ u^-(t + x/c) &= u^-(0)\exp i(\omega t + k^- x), \end{aligned} \tag{3-10}$$

where $p^+(0)$, $u^+(0)$ correspond respectively to the values of fluctuating component pressure and velocity at $x = 0$, and $k^+ = k/(1 + M)$, $k^- = k/(1 - M)$ where $k = \omega/c$ and M is the Mach number u_o/c. The local values of the fluctuating particle velocity and pressure are given respectively by

$$u(x,t) = \left(\frac{1}{\rho c}\right)\left(p^+(t-x/c) - p^-(t + x/c)\right), \tag{3-11}$$

$$p(x,t) = p^+(t-x/c) + p^-(t + x/c). \tag{3-12}$$

In practice, viscous and thermal influences at the walls, which are a function both of the Stokes number S_n and of the thermal properties of the medium, modify the wave propagation constant. The thermal properties of the medium are represented by the ratio of the specific heats, γ, and the Prandtl number Pr. The wave number $k = \omega/c$ implicit in equations 3-9 and 3-10 is replaced by β for the acoustic calculations by the program, where

$$\beta = \frac{\omega}{c} + \alpha(1-i). \tag{3-13}$$

For $S_n > 10$ and plane wave propagation, the visco-thermal coefficient α can be expressed as

$$\frac{\alpha}{k} = \frac{1}{S_n}\frac{|1 + (\gamma - 1)\sigma|}{\sqrt{2}}\,, \tag{3-14}$$

where $\sigma = \sqrt{(1/Pr)}$. With the thermal properties relevant to the moist air in the vocal tract, $\alpha/k \approx 1/S_n$ (within 4 percent). Recall that $S_n = r_o\sqrt{(\omega/\nu)}$, where r_o is the radius and ν is the kinematic viscosity. Thus the relative influence on wave propagation of this α/k factor diminishes directly with the square root of the frequency. For excitation at 25 Hz in a tract of radius 3 mm, S_n equals 10. Appropriate expressions for the propagation constant with $S_n < 10$ are available (Tijdemann, 1975), but have not been included at present.

Since the walls of the vocal tract are not rigid, the influence of wall vibration on wave propagation has also been included in the program by making a further appropriate modification to the wave propagation constant. For a cylindrical tube with reacting walls, yielding a time-varying duct cross-sectional area $A(t) = \pi(r(t))^2$, the linearized equations for conservation of mass and momentum, with mean flow neglected, become

$$\frac{\partial \rho A}{\partial t} + \rho_o A_o \frac{\partial u}{\partial x} = 0, \tag{3-15}$$

$$\rho_o\frac{\partial u}{\partial t} + \frac{\partial \phi}{\partial x} = 0\,, \tag{3-16}$$

where ρ_o and A_o are respectively the unperturbed or time-averaged values of ρ and A. The wave equation derived from these expressions (Lighthill, 1978) becomes

$$\left(\frac{1}{c}\right)^2\frac{\partial^2 p}{\partial t^2} - \frac{\partial^2 p}{\partial x^2} = 0, \tag{3-17}$$

where the phase speed, c, is now defined by

$$\begin{aligned} c^{-2} &= \frac{1}{A_o}\frac{d(\rho A)}{dp}\bigg|_{p=p_o}\\ &= \frac{d\rho}{dp}\bigg|_{p=p_o} + \frac{\rho_o}{A_o}\frac{dA}{dp}\bigg|_{p=p_o}\\ &= c_a^{-2} + c_w^{-2}\,, \end{aligned} \tag{3-18}$$

where c_a is the adiabatic speed of sound and c_w is the perturbation to the phase speed caused by the wall vibration. Variation with entropy is neglected.

In the linear approximation,

$$\left.\frac{dA}{dp}\right|_{p=p_o} \approx 2\,\pi r_o \left.\frac{dr}{dp}\right|_{p=p_o}. \tag{3-19}$$

Also, assuming a locally reacting wall, with tilde denoting Fourier transform,

$$(-m\omega^2 + iR\omega + K)\,\tilde{r} = \tilde{p}, \tag{3-20}$$

where m = mass/area, R = resistance/area, and K = stiffness/area. The resulting perturbation in phase speed due to wall vibration is then expressed by:

$$\begin{aligned} c_w^{-2} &= \frac{\rho_o}{A_o} \left.\frac{dA}{dp}\right|_{p=p_o} \\ &= \frac{2}{r_o}\,\rho_o\,\frac{[m(\omega_o^2-\omega^2)-iR\omega]}{m^2(\omega_o^2-\omega^2)^2+R^2\omega^2}, \end{aligned} \tag{3-21}$$

where $\omega_o = (K/m)^{1/2}$ is the natural frequency of the wall. The values of m, R and ω_o were obtained from Ishizaka, French, and Flanagan (1975), with $\omega_o = (2\pi)30$ rad/sec. Thus, the wave number modified by wall vibration is expressed to first order as

$$k^* = \frac{\omega}{c}\left\{1 + \frac{\rho_o c^2}{r_o\,[m(\omega_o^2-\omega^2)+iR\omega]}\right\}. \tag{3-22}$$

The effects of both wall vibration and visco-thermal attenuation are taken into account through appropriate changes in the complex wave number. Because the frequencies of interest tend to be greater than the natural frequency of the wall, the wall vibration tends to diminish the real part of the wave number, thus raising the formant frequencies. The visco-thermal damping adds to the real part of the wave number, and therefore tends to lower the formant frequencies. The relative importance of the wall vibration is greatest at the lower frequencies as seen by equation (3-22), and its influence on the wave number decreases more rapidly with frequency than that due to visco-thermal effects.

Turning now to the calculation of wave transfer across area and other axial discontinuities along the vocal tract, one notes first that historically, their influence has been taken into account by organ builders and wind instrument makers who included appropriate end corrections at the discontinuities. A detailed analysis and discussion of their practical application to the end open to atmosphere can be found

in Rayleigh (1896) and other contributions to acoustics in the classical literature. Since the acoustic analysis in the program begins at the lips, one needs to determine both the magnitude and phase of the corresponding pressure reflection coefficient, $p^-(f)/p^+(f)$, which is based on the geometry of the opening and the surrounding region. These spectra are functions of the mean flow and of the Helmholtz number, kr_o, where r_o is the radius of the opening. Experimental results (for example, Sugiyama and Irii, 1991), supported by analytical considerations, suggest that during speech the lips radiate sound effectively as an unflanged opening, at least as far as the near field is concerned, and this will have the dominant influence on wave reflection. The modulus of the reflection coefficient for an unflanged tube with outflow has been determined analytically by Munt (1990), while the phase with zero outflow, expressed as an end correction, has been determined analytically by Levine and Schwinger (1948). Extensive measurements at Southampton with steady outflow up to M = 0.3 (Davies, 1988; Davies, Bento Coelho, and Bhattacharya, 1980) are in close agreement with both analytical predictions. These are incorporated in the program by using a sequence of simplified expressions that match the analytic predictions to better than 0.1 percent.

Wave transfer across the other discontinuities includes abrupt area changes with the optional inclusion of sidebranches (sinuses) at junctions, where the wave motion remains wholly axial. They also include transfers between spherical, cylindrical, and plane wave fronts, where some adjustment to the wave front is necessary. The approach adopted is in many respects similar to that for the open termination at the lips. The modulus of the component waves either side of area changes was determined by solving the integral form of the linearized conservation relations for mass, energy, and momentum over the surfaces of a control volume. This was effectively a plane surface for plane waves and an appropriate volume for the others, as set out in Davies (1988) and Davies and Doak (1990b). In the former case, the control volume (transfer plane) was set at an appropriate distance from the discontinuity to make allowance for the generation of evanescent waves that are required to match the boundary condition of zero velocity over tract surfaces normal to the axis. As before, analytical expressions were derived for such end corrections which were experimentally confirmed and empirically generalized by matching the results with simplified expressions incorporated in the program. The special case of an expanding outflow required appropriate consideration of the accompanying flow separation and entropy generation. Thus the relation between fluctuating pressure and density was modified to $\rho = (p + \delta)/c^2$, and δ was then eliminated between the three conservation equations, to derive the modulus of the transfer.

Having established the component wave amplitude spectra, one can then readily calculate the corresponding fluctuating pressure and particle velocity spectra, reflection coefficient spectra, impedance spectra, acceleration or force spectra, intensity spectra, and so on at any desired position along the tract. Similarly one can calculate the corresponding transfer function spectra relating to any two locations along the duct and in particular between the glottis and lips.

DISCUSSION OF NEW FEATURES

The addition of mean flow terms that multiply perturbation quantities, and the possibility for nonisentropic relations between perturbation pressure and density, other than for visco-thermal attenuation, are novel additions in the speech synthesis field. Also, the inclusion of three conservation equations when appropriate, instead of two, is a novel feature. In the absence of mean flow, conservation of mass and conservation of energy yield the familiar continuity of volume velocity or mass and pressure across junctions. The addition of conservation of axial momentum in the presence of mean flow provides an independent relation between perturbation pressure and flow. This does not lead to an over-determined system of equations, because a nonisentropic relation between pressure and density is allowed.

Another major feature of the program is the description of some physical relationships. This includes the reflection coefficient at the open end as a function of mean flow Mach number (Davies, 1988). Also, at abrupt area changes, as appear in Type 1 and Type 4 elements, due allowance is made for the existence of the evanescent waves necessary to satisfy the boundary conditions at the wall perpendicular to the flow. Evanescent waves are composed of higher order modes that do not propagate and decay exponentially with distance from the discontinuity; they appear only locally in the region of the junction. The net effect on the propagating wave is to produce a phase change in the wave, which has been modelled historically as an end-correction.

The results of running this program with parameters derived from a real vocal tract will be discussed in the next section, but it is possible to mention some of the effects of the novel aspects of this program without recourse to simulation. The net outflow of air can affect the propagation of sound in several ways. Acoustic waves are convected by the flow so that waves travelling toward the mouth have their speed increased by a factor of the mean flow Mach number, M, and waves travelling toward the glottis have their speed decreased by the same factor. One of the results of this is to reduce the distance between the nodes of a standing wave by a factor of $(1 - M^2)$, where M is the mean flow Mach number. However, the mean flow Mach number for voiced speech is usually very small. At a mean volume flow rate of 200 cm^3/sec and a cross-sectional area of 2.0 cm^2, the mean flow Mach number is about 0.003, which is too small to have an effect on formant frequencies.

Another consequence of non-zero mean flow is enhanced loss of acoustic energy at abrupt expansions through vortex shedding (Bechert, 1980; Howe, 1980). Because the mean flow is of the same order of magnitude as the peak flow during phonation, this loss is actually a component of the nonlinear resistance that occurs at high acoustic amplitudes (Ingard and Ising, 1967). Liljencrants (1985) discussed the implementation of such nonlinear losses in a reflection-type analog. To account for vortex shedding, a non-isentropic relation is allowed in VOAC between pressure and density, as discussed above.

The new facility provided for including the influence of wall vibration on acoustic wave number, if so desired, is yet another novel feature. The consequences of its implementation are discussed among the results described in the next section. Further new features concern the calculation of wave propagation along elements of the vocal tract whose area varies continuously either linearly or quadratically with axial displacement. Wave propagation is calculated throughout the tract with the solutions to the appropriate form of the convected wave equation, avoiding the approximations involved when the algorithms are based on electro-acoustic analogs. This has the advantage that the calculated distributions of complex component pressure amplitudes remain continuous between area discontinuities. Thus they can be evaluated at any desired position along the tract, since their values are automatically provided at the junctions between each element which may be inserted at any desired location, excepting conical elements. Furthermore, sinuses can be conveniently included at any appropriate location and are represented by an equivalent sidebranch added at the corresponding discontinuity.

The basic computational output of the program is a transfer function, where the transferred variables are the relative values of the positive-going and negative-going pressure amplitudes. Alternatively, the total transfer function can be calculated as the product of a sequence of scattering transfer matrices, provided that their component elements have been evaluated realistically. This may require a somewhat complex procedure because, unlike electrical networks, acoustic elements do not normally possess reciprocal transfer properties (Davies, 1991).

SIMULATIONS

CONSTRUCTION OF TWO VOCAL TRACTS

It is obviously desirable to compare sound spectra predicted by VOAC with those measured from speech spectra for living subjects. Thus two vocal tract models have been constructed for this purpose. However, a significant parameter required by the program is the hydraulic radius, h_r, used in the program to define the influence of viscosity and heat conduction, among other factors, on sound propagation along the tract. Therefore, it was necessary to find articulatory data that included the shapes of the tract cross-sections from which to compute or at least estimate the distribution of hydraulic radius to accompany the observed area distribution of the tract during phonation. Both were then supplied to the program with the corresponding element lengths to describe the tract geometry.

Two such sources of data were found, namely: Fant (1960), which included the cross-sectional shapes for vowel /i/, and Baer, Gore, Gracco, and Nye (1991), for which the vowel /a/ for subject TB was used. Two different vowel sounds were thus predicted. Both references included acoustic data on the vowels in the form of tabulated formant frequencies. Furthermore, the data for subject TB also included areas of the pyriform sinuses, whose presence may affect the impedance seen from the glottis profoundly.

Fant's data included the area function for a number of vowels, acquired by a "variety of methods," but cross-sectional shapes at different stations along the tract were shown only for the vowel /i/. It was possible from the pictorial information he provided, after enlarging this to "life size" based on the reported tract length and area function, to measure the corresponding perimeters and thus have a sequence of seven estimates of h_r. The distribution of this quantity elsewhere along the tract was then estimated by appropriate interpolation.

With the subject TB, magnetic resonance data were acquired while the vowels /a/, /æ/, /i/, /u/ were sustained. In their article Baer et al. (1991) report area functions and show some of the cross-sectional outlines. Data files were acquired for tract outlines of the axial and coronal sections spaced at 0.5-cm intervals. A program was then written that computed the perimeter length and area enclosed by each section. In the upper pharyngeal region the tract areas were derived along equally spaced radial lines of a circle fit to the radius of curvature of the tract. The resulting projected sections were not easily available, however, and could not be easily recreated since that involved careful location of the axial and coronal sections relative to each other, cross-checking with the original images, and so on, to recreate the precise shapes. Therefore, a quick approximation was made in this upper pharyngeal region: the areas reported in the article were used, and in the absence of a perimeter the hydraulic radius was computed assuming circular cross-sections throughout.

Having arrived at an area function and a corresponding hydraulic radius function for each vowel, the next step was to divide these into the elements used by VOAC. The area functions that resulted are shown in Figure 3-2, contrasted with, respectively, the area function quoted by Fant, and the areas derived directly from the MRI outlines. The differences are due mainly to the relatively small number of elements used: 7 for Fant /i/, 10 for TB /a/. For Fant /i/, note that in the region of the greatest mismatch, between $x = 1$ and 3 cm, the volume of the front cavity is identical although the area functions differ significantly; the same holds for TB /a/ between $x = 3$ and 3.5 cm.

RESULTS

In this section the results of simulations are considered, with an emphasis on showing the effects that can be modelled by the present program, but not by previous programs based on, for example, the Kelly-Lochbaum method. The results will mostly be reported in terms of magnitude or modulus of the dimensionless input impedance of the tract at the glottal plane as a function of frequency. The peaks of these plots correspond to resonance, or formant, frequencies of the vocal tract for the glottal volume velocity source. The dimensionless input impedance is defined by:

$$\frac{Z(f)}{\rho c} = \frac{p^+(f) + p^-(f)}{p^+(f) - p^-(f)}, \qquad (3\text{-}23)$$

where $p^+(f)$ and $p^-(f)$ are the transforms of the fluctuating pressures at the glottis ($x = 0$). The dimensions of $Z(f)$ are the same as for a transfer impedance, that is, pressure divided by velocity, and indeed resonances will be the same for both transfer and input impedances. However, anti-resonances will not be the same, and this should be kept in mind when sinuses are included in TB /a/ or when any other tract with side branches is specified.

Length Corrections

Length corrections were normally applied to tube sections by VOAC to account for phase changes at abrupt area changes, which are caused by evanescent modes (Davies [illegible]). It is interesting to note the size of [illegible]

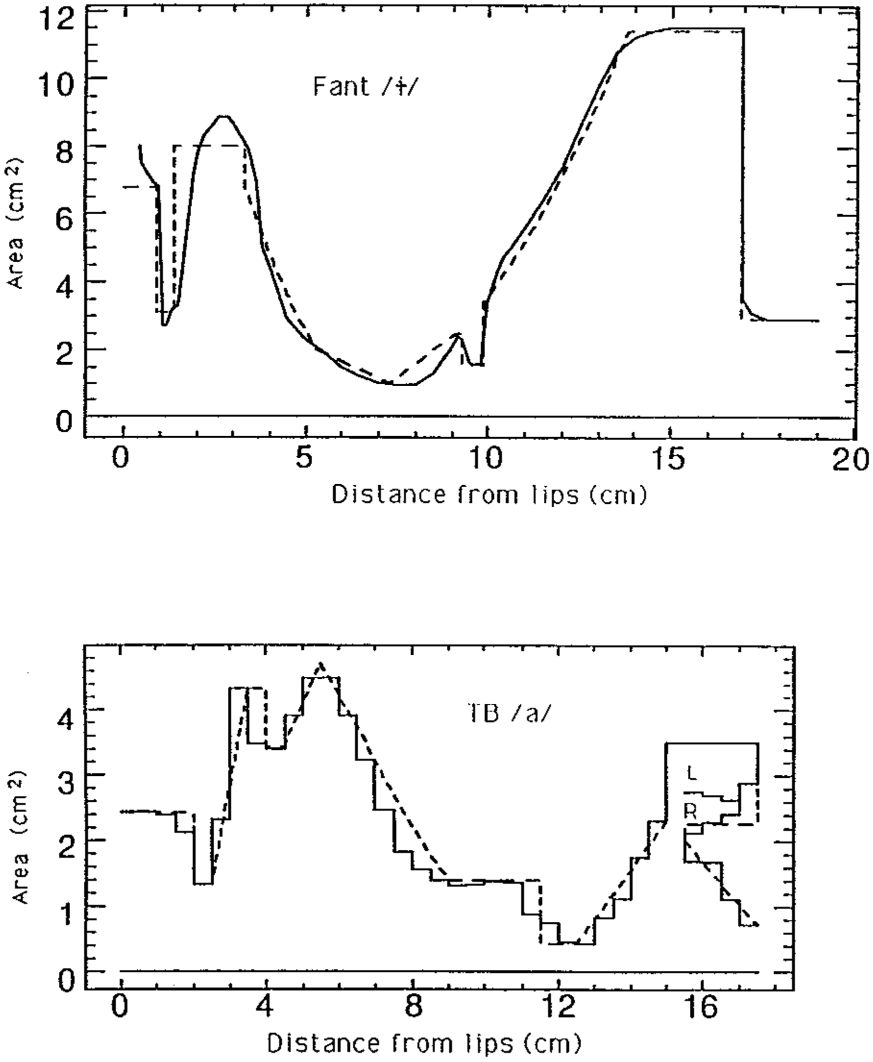

Figure 3-2. Area functions for Fant /i/ (top) and TB /a/ (bottom). Solid lines show area functions given by sources (Fant, 1960; Baer et al., 1991); dashed lines show area functions used by VOAC. In TB /a/ diagram, "L" and "R" indicate left and right pyriform sinus areas respectively.

example occurs for Fant /i/ in the tube section at the glottis. The glottal section had a physical length of 2.10 cm and a corrected length of 2.39 cm, and the section just above the glottis had a physical length of 3.10 cm and a corrected length of 2.81 cm. The length corrections, which are a function of the area ratio between the sections (1:4 here), are on the order of 20% of the physical lengths and so can be considered to be substantial. Note that the inclusion of such corrections does not alter the overall length of the vocal tract, but merely defines the appropriate axial position of each transfer plane for discontinuities.

Comparison of Formant Frequencies; Effects of Wall Impedance, Sinuses

Input impedance spectra were computed for both vocal tracts under different conditions to allow an assessment of the acoustic significance of each condition independently. For each spectrum, formant frequencies were identified by locating the maximum values of the magnitude. These formants were useful for comparison with the acoustic data supplied by Fant and Baer et al., and are presented in Tables 3-1 and 3-2.

Figure 3-3 contrasts rigid and reacting walls for Fant /i/. Inclusion of the reacting walls raises the first formant and increases its bandwidth substantially. Reacting walls also raise the second formant slightly (see Table 3-1). The first formant predicted by VOAC matches the first formant measured by Fant better when the walls are reacting; the predicted second formant is significantly too high in both cases. The BESK and LEA synthesis results reported by Fant were more closely matched to the subject's values than are the VOAC results.

Figures 3-4 and 3-5 show four different cases for TB /a/: rigid and reacting walls are contrasted in each plot, with the vocal tract respectively including, and omitting, the sinuses. In both cases formant frequencies are higher for reacting walls, and this effect is more pronounced for the lower frequency formants. When the sinuses are included, an anti-resonance of high frequency appears as expected, though it is not visible on these plots. Removing the sinuses removes this anti-resonance and also increases the formant frequencies, which matches the direction-of-change predictions made by the Haskins articulatory synthesizer (Mermelstein, 1973; Rubin, Baer and Mermelstein, 1981). Note however that VOAC predictions match formants measured for the subject better than the Haskins synthesizer, which is based on a Kelly-Lochbaum model, and that the most realistic case simulated on VOAC — with sinuses, reacting walls, and flow — matched the subject the best. This raises the possibility that, if data

Table 3-1. *Comparison of formant frequencies (Hz) for Fant /i/*

	VOAC Simulations Flow			Fant Results		
	0	200cm^3/s	1200cm^3/s	Subject	BESK	LEA
Rigid Walls	F1= 265 F2=1620	265 1620	265 1620		296 1517	285 1480
Reacting Walls	F1= 310 F2=1630	320 1630	307 1630	300 1480		

Table 3-2. *Comparison of formant frequencies (Hz) for TB /a/ VOAC simulations.*

	Rigid Walls		Reacting Walls	
	Flow		Flow	
	0	200 cm^3/s	0	200 cm^3/s
With sinuses	F1 = 543 F2 = 980 F3 = 2470	545 980 2470	590 1010 2480	600 1012 2480
Without sinuses	F1 = 570 F2 = 1020 F3 = 2485	570 1020 2485	615 1053 2495	625 1055 2495

	Haskins Synthesis	Subject
With sinuses	F1 = 535 F2 = 1057 F3 = 2430	595 1006 2400
Without sinuses	F1 = 579 F2 = 1091 F3 = 2461	

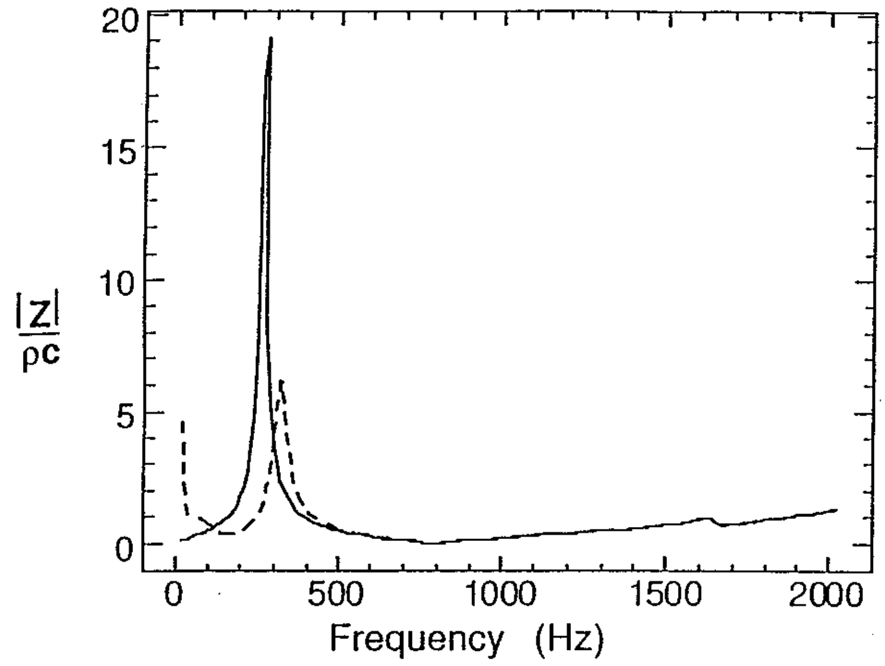

Figure 3-3. Driving-point impedance of Fant/i/, as seen from glottis. Flowrate = 200 cm³/s. Solid line: rigid walls. Dashed line: reacting walls.

were available on the shape of the pyriform sinuses in Fant's subject, the VOAC predictions would be a closer match for that subject as well.

The contrast with and without flow evident in Table 3-2 shows that flow may modify predicted formant frequencies when the walls are reacting. This interaction was tested systematically and is discussed in the next section.

Mean Flow Effects

Direct comparisons of the effect of mean flow could be tested using VOAC. Comparisons between low flow (200 cm³/sec) and high flow conditions (1200 cm³/sec) were made with three different vocal tract configurations. One condition was that of the Fant /i/ vowel described above, and the other three were based on modifications of that vowel: tight constrictions with an area of 0.1 cm² were placed in succession at the lips, behind the teeth, and at the uvula. The effect of the mean flow in this program is through the Mach number parameter, and for a given volume flow, this is maximized by decreasing the area. Thus, at 1200 cm³/sec, the Mach number for flow through the constriction was 0.33, and at 200 cm³/sec the Mach number was one-sixth this amount, or 0.055. In the unconstricted vocal tract for the Fant vowel /i/ the minimum area was 1.0 cm² so that the maximum Mach number attained in the 1200 cm³/sec case is less than 0.033 and in the 200 cm³/sec case it is less than 0.0055. The results are shown in Figures 3-6 to 3-9.

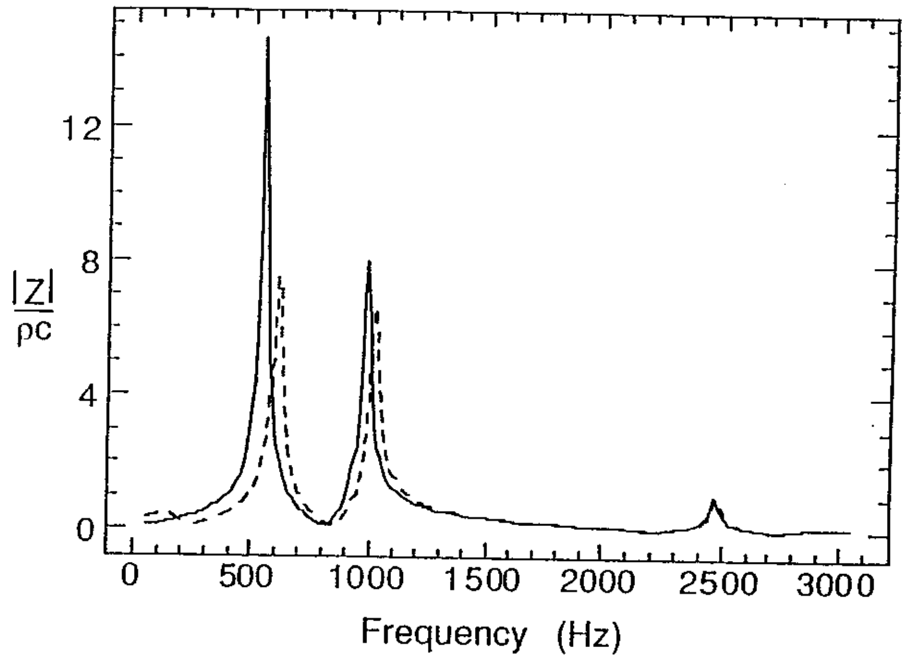

Figure 3-4. Driving-point impedance of TB /a/, as seen from glottis. Flowrate = 200 cm³/s. Solid line: rigid walls. Dashed line: reacting walls. Sinuses are included.

As expected, a six-fold increase in volume velocity had little effect on the formant frequencies, for the relatively unconstricted vocal tract of /i/ (see Table 3-1). While the frequency range from 0 to 2000 Hz is shown in these figures, the range from 2000 to 4000 Hz was also checked to be certain there was no effect on the higher formants. However, for the constricted vocal tracts, a change in volume flow did sometimes have a substantial effect, depending on the location of the constriction. When the constriction is located at the lips or uvula (see Figures 3-7 and 3-9), there is an effect, but not when it is located just behind the teeth (Figure 3-8). The maxima of the input impedance for

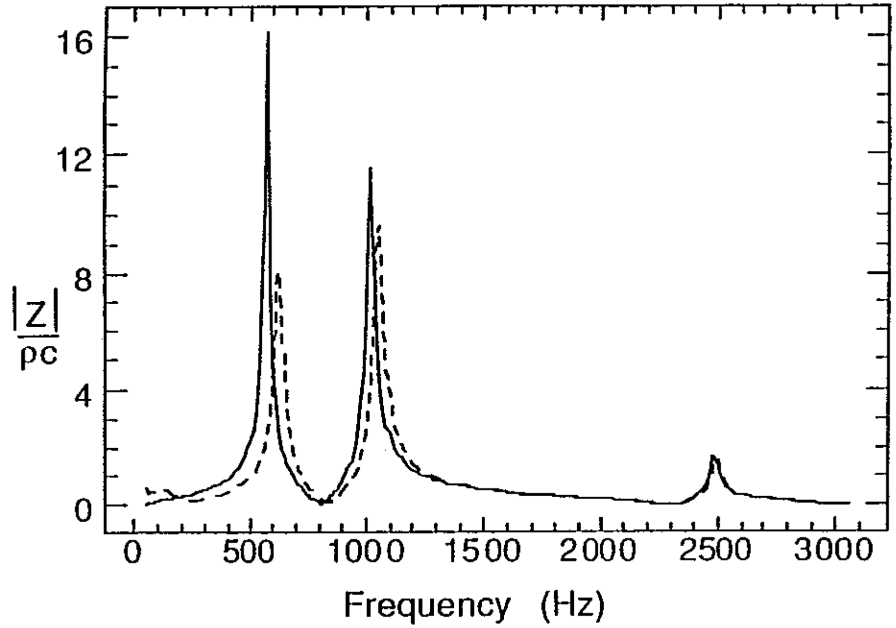

Figure 3-5. Driving-point impedance of TB /a/, as seen from glottis. Flowrate = 200 cm³/s. Solid line: rigid walls. Dashed line: reacting walls. Sinuses are omitted.

the lip constriction shifted from 120 Hz and 680 Hz in the low-flow condition to 110 Hz and 650 Hz respectively in the high flow condition. In the uvular constriction configuration, the maximum at 150 Hz in the low-flow condition shifted to 100 Hz in the high-flow condition[3]. Furthermore, Figures 3-6 to 3-9 show a change in the shape of the input impedance magnitude spectra.

A SURVEY OF PHENOMENA NOT ACCOUNTED FOR IN THE PROGRAM VOAC

The program VOAC just described includes elements of duct air flow and acoustics not normally included in reflection-type or other electro-acoustical analogs. The program's algorithms do include some of the familiar simplifying assumptions, such as linearity of the equations of motion, but attempts a more realistic simulation by giving up others, such as isentropic propagation. However, VOAC is necessarily based

---

[3] All these values are determined in 10 Hz intervals.

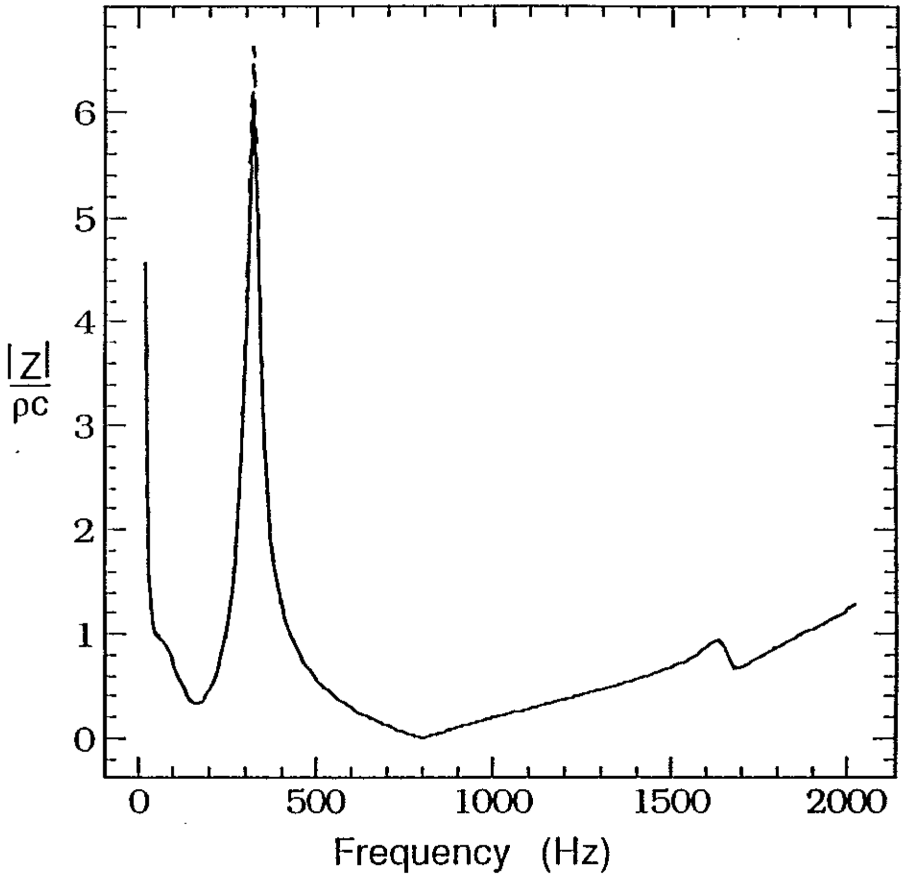

Figure 3-6. Driving-point impedance of Fant/i/, as seen from glottis. Walls are rigid. Solid line: flowrate = 200 cm³/s. Dashed line: flowrate = 1200 cm³/s.

on an approximate simulation of the complete flow field in the vocal tract. Alternative simulations based on the differential forms of the conservation laws are being pursued (Thomas, 1986; Iijima et al., 1990; Alipour and Patel, 1991; Liljencrants, 1991). While these approaches are useful, especially when studying the oscillation of the vocal folds, the approach of going from simple to more complex models also has advantages in studying propagation. Thus, starting with the simplest acoustic models, relevant aspects of the fluid mechanics can be added one at a time to test their influence on the acoustic behavior. More realistic simulations can be used as a tool to develop further understanding, but it is easier to judge their relative importance and understand the effect of various additions to a model if the model is

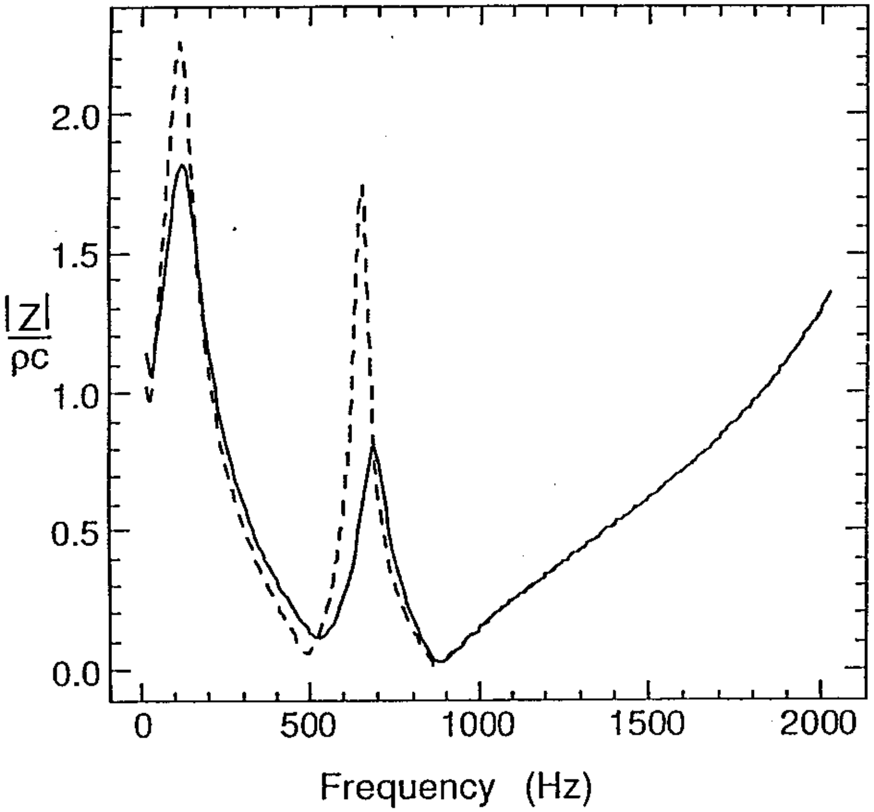

Figure 3-7. Driving-point impedance of Fant/i/, modified to have constriction of 0.1 cm² located at the lips. Walls are rigid. Solid line: flowrate = 200 cm³/s. Dashed line: flowrate = 1200 cm³/s.

relatively simple to begin with. A discussion is presented below of various phenomena that might be included in a program that could yield predictions that are more realistic than the current VOAC.

CURVATURE OF DUCT AXIS

The axis of the vocal tract is in reality curved, while the acoustic calculations reported here assumed that it was straight. The influence of such curvature on acoustic propagation with rectangular tubes depends on the relative sharpness of the bends defined by the ratio, *a*, of its outer to inner radius. Thus, one finds that the tract reported for Fant /i/ follows an arc of some 120 degrees above the tongue with the value of *a* varying from 1.1 to 1.5. The corresponding decrease in phase velocity for the formant frequencies in a curved hard-walled tube compared with a straight one is about 0.5% for *a* = 1.1 to over 5% for *a* = 1.5. This result applies to frequencies below about 20 kHz for the geometry appropriate to the vocal tract, where the value stated is around the lower limit for higher order mode propagation in curved ducts (Rostafinski, 1991; El-Raheb, 1980). Thus tract curvature may go some way towards explaining the relatively high estimates of formant frequency recorded in Table 3-1, compared with observation. However, a realistic inclusion of axis curvature that takes account of the complex changes in area function and shape of cross-section may not be worth the significant effort it would require, particularly considering the

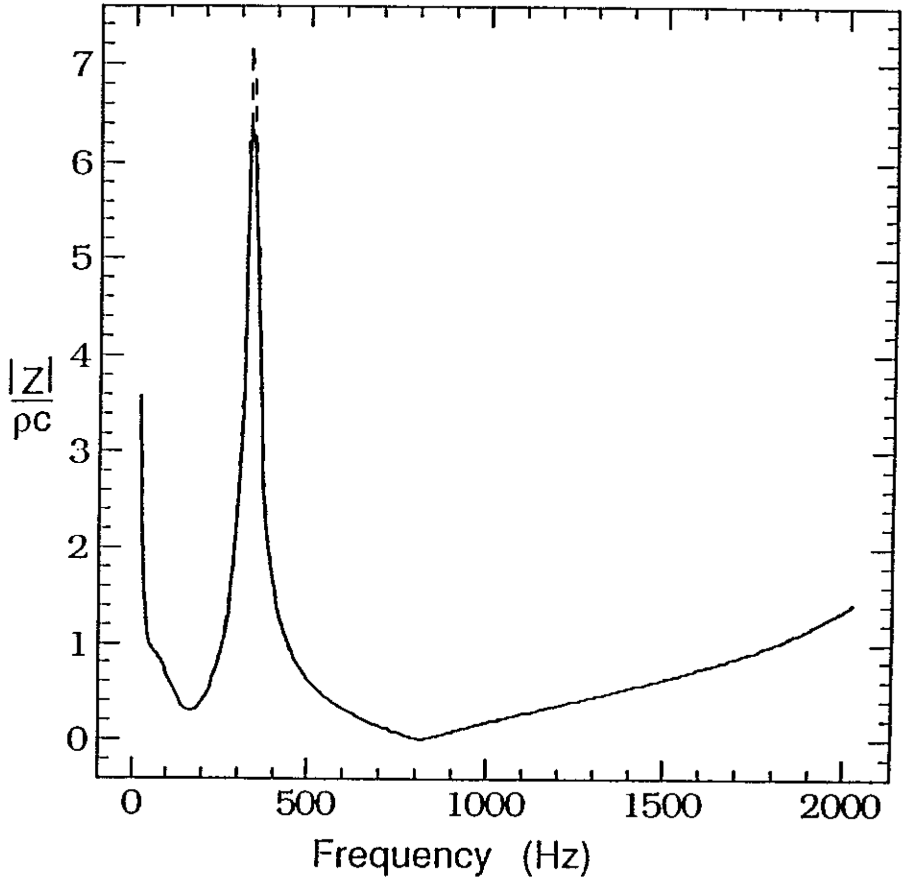

Figure 3-8. Driving-point impedance of Fant/i/, modified to have constriction of 0.1 cm² located behind the teeth. Walls are rigid. Solid line: flowrate = 200 cm³/s. Dashed line: flowrate = 1200 cm³/s.

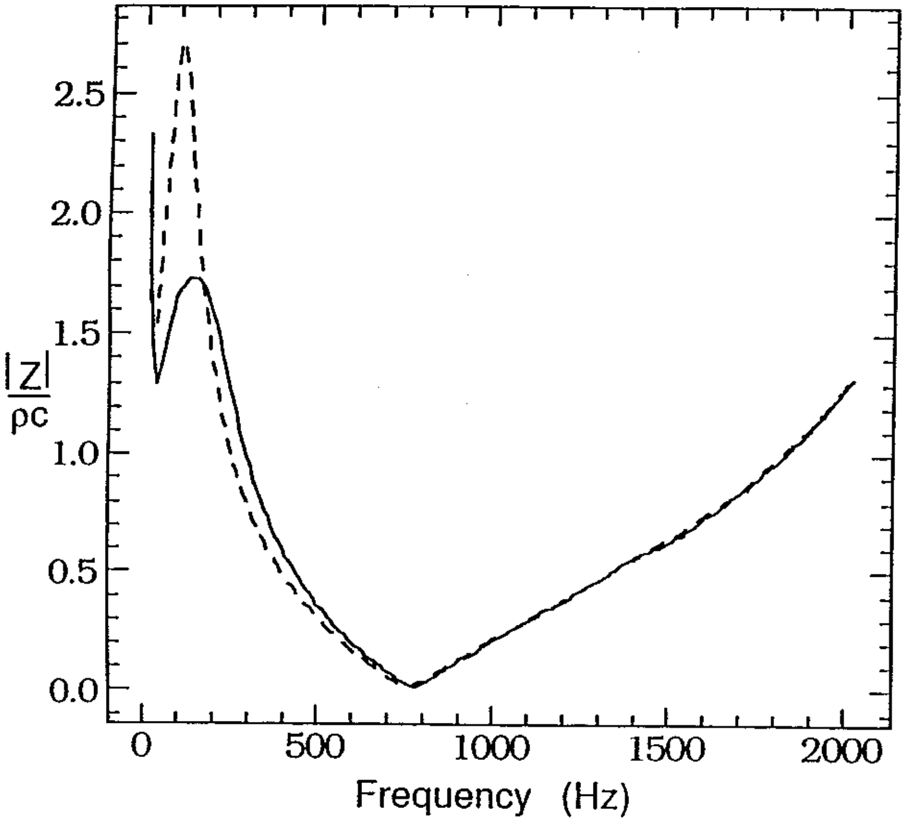

Figure 3-9. Driving-point impedance of Fant/i/, modified to have constriction of 0.1 cm² located at the uvula. Walls are rigid. Solid line: flowrate = 200 cm³/s. Dashed line: flowrate = 1200 cm³/s.

uncertainties that must exist in defining the geometry of a given tract. Uncertainties also exist with the measurement of corresponding formant frequencies.

SOUND ABSORPTION BY TURBULENCE

Turbulence is known to produce sound in the vocal tract, but since this is not strictly to do with propagation, we consider it briefly in the next section under "Future Developments." There are, however, other effects of turbulence on sound that need to be examined. There can be scattering of sound by turbulence when the time scales of the turbulent fluctuations and sound are not comparable (Lighthill, 1953). On the other hand, for comparable turbulence and sound time scales, turbulence can absorb sound as the sound acts irreversibly to stretch the associated vortex lines (Noir and George, 1978). Because the velocity associated with turbulence is of the order of the flow velocity and the velocity associated with sound is the speed of sound, then with flow speeds much less than sonic the wavelength of sound must be much larger than the turbulence scales, if the turbulence and acoustic time scales are to be comparable. This implies that this absorption mechanism will then pertain to sounds below the normal fundamental frequency of voicing.

However, if solid boundaries are present, the absorption of sound by turbulence can be greatly enhanced in a higher frequency range (Ronnenberger and Ahrens, 1977; Howe, 1979, 1984). The important parameter for determining the interaction in the presence of boundaries is the ratio of the acoustic boundary-layer thickness to the viscous sublayer thickness of the turbulent flow. This ratio needs to be large so that the acoustic boundary layer can interact with the turbulence, and this ratio is an increasing function of mean flow velocity and decreasing function of acoustic frequency. The effect can be quantified as a percent increase in the exponential damping coefficient above that observed for visco-thermal boundary-layer damping without mean flow (see equation 3-14). For instance, the exponential damping coefficients are predicted to change about 20% for a 630 Hz excitation from their value at zero flow when the mean flow Mach number is 0.1, in a smooth circular pipe (Ronnenberger and Ahrens, 1977). The effect becomes more pronounced as the frequency decreases, so that the damping coefficient could be as much as 80% more than it would be with zero mean flow at 150 Hz. However, to obtain the necessary relatively high mean flow Mach number in the vocal tract with turbulent flow it must be highly constricted, as well as having a high volume flow rate. Thus, these conditions can only be expected to be attained at restricted locations within the vocal tract. Normally mean flow Mach numbers of less than 0.01 are more likely, where the effect just described would hardly be noticeable, even at 150 Hz. It is probably safe to ignore extra attenuation of the acoustic wave caused by sound-turbulence interaction in the vocal tract and at its boundaries.

OTHER ATTENUATION AND MODAL PROPAGATION

Other attenuation mechanisms have been studied in regard to wind instruments (Keefe, 1983; Keefe and Benade, 1983). These include bulk viscous losses in highly curved sections and acoustic streaming off of sharp edges. These are difficult to quantify and will be left for future consideration.

A necessary component of turbulence is the accompanying rotational, or vortex, motion. To model vorticity explicitly requires fluctuations in all three spatial dimensions. A further limitation of the essentially one-dimensional modelling adopted for the current program is that propagating modes of higher order cannot be included in the predictions. The fact that such higher-order modes cannot thus be included explicitly means that high frequency propagation may not be modelled correctly above about 8kHz.

VORTICITY WAVES

There are other phenomena related to the vorticity generated at junctions and elsewhere that can contribute to the overall acoustic behavior of the vocal tract. Davies (1983) reports that the shear layer associated with a steady jet flow exiting into an expansion where an acoustic standing wave is present can also exhibit wave-like behavior. That is, along with the acoustic standing wave, convecting vortical disturbances in the shear layer also contribute to the observed pressure field at the same frequency as the sound. However, their phase velocity is the mean speed of the shear layer and thus an order of magnitude or so less than the local speed of sound. A situation corresponding to this may occur during voiced frication. While this type of interaction may transfer energy both to and from the acoustic wave within the vocal tract, the net effect could be an increased fluctuating volume velocity at the mouth. Teager and Teager (1983a, 1983b) discussed the glottal jet, which is necessarily a region of high vorticity, and the possibility of its interaction with other modes of fluid movement.

There is also literature reporting the existence of waves of vorticity within tubes in the presence of an oscillating boundary (Stephanoff, Pedley, Lawrence, and Secomb, 1983; Pedley and Stephanoff, 1985[4]; Sobey, 1985; Shadle, Barney, and Thomas, 1991; Barney, Shadle, and Thomas, 1991). Observations cited in the two studies SP were made in a channel with a rectangular cross-section (10 cm x 1 cm) while a piston drove a thick rubber membrane sinusoidally at the side of a channel as fluid flowed through the channel. The relevant parameters were a Reynolds number, a Strouhal number, both based on channel width (1 cm) and mean flow velocity, and the ratio of piston amplitude to channel width as an amplitude parameter. The Reynolds numbers were between 360 and 1260, the Strouhal numbers ranged up to 0.077, and the amplitude parameter was on the order of 0.5 (Pedley and Stephanoff, 1985). By way of comparison, in a circular duct having a typical vocal tract diameter of 2 cm, a mean particle velocity of 60 cm/s (typical for phonation), and a frequency of oscillation of 150 Hz, the Reynolds number would be 800 and the Strouhal number would be 5. Since Re is in the range studied by SP and St is much higher, this means that any unsteady behavior found in SP's studies may possibly be present in the vocal tract and thus should be considered.

---

[4] These two papers are subsequently referred to as SP.

SP reported that for $St > 0.005$ the flow was not quasi-steady, but a wave of vorticity was seen travelling down the channel. The wave consisted of vorticity of alternating orientation bending the streamlines of the main flow into a wavy pattern, first toward one wall and then toward the other. Theoretically, this phenomenon is described by a linearized Korteweg-de-Vries equation, and could conceivably account for the Coanda effect cited by Teager and Teager (1983a).

Similar experiments using models of vocal-tract dimensions support the view that vortices are formed and do persist downstream of an oscillating boundary. Shadle et al. (1991) observed vortex formation downstream of a single moving shutter oscillating against a fixed shutter. This could possibly have been ascribed to the cross-channel asymmetry, which was a feature of the SP studies as well. However, a later model with two moving shutters using hot-wire measurements (Barney et al., 1991) also showed evidence of vortex formation near the glottis. Just how these vortices are affected by structures such as the pyriform sinuses and the false folds remains to be investigated. It is, however, clear from Barney et al. that flow is not at all uniform across the tract near the glottis (1 cm and 4 cm downstream), and only gradually becomes uniform.

There is a further important aspect to this work, not directly pertinent to propagation in the vocal tract. Even at small Strouhal numbers it was found in the SP studies that the quasi-steady approximation is not valid. Flow separation can be very sensitive to acceleration effects, even at low Strouhal numbers, so that the flow is not normally quasi-steady when separation is involved (Sobey, 1980, 1983). This has important consequences in determining the stresses on the surfaces of vibrating objects in the vocal tract, and could provide a mechanism for single-degree-of-freedom vibration, as in falsetto mode.

ONE-DIMENSIONAL NONLINEAR PROPAGATION

A further assumption that has been incorporated in the program is that the one-dimensional propagation is linear, and it is not known how important this assumption may be with respect to the sound that the listener hears. The effect of nonlinearity in one-dimensional propagation is expected to be greatest when the fundamental frequency is close to the first formant and the amplitude of voicing is large (McGowan, 1991). In the case of nonlinearity, energy is transferred from the lower harmonics of the voice to the higher harmonics, while the formant frequencies should show only small perturbations. Thus, nonlinear propagation should affect the overall spectral tilt of the speech signal, and an inverse filtering procedure which removes only spectral peaks will attribute all the spectral tilt to the voice source. A way to test the importance of nonlinearity in one-dimensional propagation is by numerical simulation. The appropriate equations have been derived for a straight tube with circular cross-section by Chester (1964). These equations contain both the effects of nonlinearity and visco-thermal boundary-layer losses and computational solutions of these equations in an infinite tube have been discussed recently by Sugimoto (1991). It is an easy task to extend Chester's equations to one-dimensional propagation in variable area tubes.

The next step in the computational study of the mechanics of air motion in the vocal tract might be to write wave propagation algorithms based on the equations derived for one-dimensional, nonlinear propagation with losses. These algorithms would be based on time-marching techniques, rather than the frequency-domain techniques used in the present program. The question then arises whether reflection coefficients at abrupt area changes used in the present program are still valid, or need new development. Based on the short review here, this approach may return new information on the actual filtering function of the vocal tract without resorting to a full numerical simulation of the fluid mechanics in the vocal tract.

IMPLICATIONS FOR INVERSE FILTERING

Inverse filtering has been used for some time now as a way of gathering information about glottal behavior from externally measurable signals. Inverse filtering with a version of VOAC could produce even more accurate glottal waveforms because of the aspects of flow that have been included in its algorithms. However, at this point, in light of the above discussions, it is worth taking a closer look at the basic theory underlying inverse filtering.

In any of the inverse filtering methods used to date, an externally-measured signal (or signals) is filtered to derive an equivalent plane-wave glottal source. The hypothesized form of the source is clearly simplified; studies cited in the earlier section, "Vorticity Waves," show clearly that flow near the glottis that is unsteady because of the action of the vocal folds is neither uniform across the tract nor localized at the exit plane of the glottis. But whether these simplifications matter might depend on the particular application.

In some studies the inverse-filtered waveform is used for resynthesis (e.g. Karlsson, 1989), or as a point of comparison across subjects or phonation types (e.g. Sundberg and Gauffin, 1979; Holmberg, Hillman, and Perkell, 1988; Gobl and Karlsson, 1991). As long as the parameters of the inverse filter are retained, that filter can be used to reconstruct the original measured waveform from the inverse-filtered waveform. No information has been lost, and thus, in a sense, it does not matter how realistic the hypothetical source function is.

However, if one wishes to use the inverse-filtered waveform to deduce the mechanism that produced the external signal, several difficulties arise. First, the same external signal and filter function will always produce the same inverse-filtered waveform, and yet it is possible that different mechanical systems, such as a siren, a piston in a tube, and vocal folds, each of which would produce different local flow patterns, could produce the same inverse-filtered waveform. Such non-unique source mechanisms pose one problem. In addition, any mechanism affecting the radiated sound and not explicitly modelled in the inverse filter will be effectively lumped with the source. Thus any sound generation downstream of the glottis due to vorticity waves or turbulence will not be identified separately; similarly, any nonlinearities in propagation or losses due to convection will be attributed to the "glottal" source mechanism.

These problems have in a sense already been recognized. Some studies have therefore focussed on deriving the necessarily complex relationship between the simplified inverse-filtered waveform and the complicated vocal fold and air motion by empirical means, for instance, by comparing the inverse-filtered waveform with simultaneously recorded electroglottograph (Fant and Sonesson, 1962), laryngograph (Rothenberg, 1981b), or photoglottograph waveforms (Ananthapadmanabha and Gauffin, 1983). Unless, and until, we can model the production mechanism more realistically, these empirical studies remain the only way of bridging the mechanics-to-sound gap at the glottis.

A final problem for even these empirical studies, however, concerns the role of convection and DC flow. The Rothenberg mask is used to derive volume velocity at the mouth from the measured trans-mask pressure drop, over a range of 0 to 1000 Hz or more. It is referred to as the glottal volume velocity, and since the Rothenberg mask was introduced (Rothenberg, 1973), the DC value has been presumed to be the mean airflow emerging from the glottis, giving information regarding whether the glottis closed completely or not. However, note that if a pulse of air at the glottis generates sound, the sound will travel to the lips at the speed of sound (actually slightly faster since the sound is travelling downstream), while the air itself will flow at the average convection speed. The difference in these two velocities results in a significant difference in travel time through the tract: sound travelling at 35,900 cm/s takes 0.5 ms to travel 17 cm, whereas air travelling at a particle velocity of 100 cm/s (based on an average volume velocity of 200 cm^3/s and average cross-sectional area of 2 cm^2) takes 170 ms to travel 17 cm. This means in practice that the DC and higher-frequency components of the flow measured simultaneously at a Rothenberg mask were generated at the glottis at times differing by 170 ms or more[5]. For a vowel steady-state, this difference may not matter; for onset, offset, and other transient phenomena, it is clearly crucial.

It may seem a simple matter to apply delays to all acoustic frequencies, but the amount of the delay needed would depend on a time history of flow velocities throughout the tract, including that of the desired output, the glottal velocity. Further, if the mean flow of air has led to generation of sound downstream of the glottis, that component of the propagated sound will have taken longer to travel from the glottis than a component that was generated at the glottis. It remains to be established how significant such intermediate sound generation is for vowels.

In summary, a consideration of inverse filtering in light of the flow mechanisms discussed in this chapter has revealed several unresolved issues concerning the technique and the interpretation of the inverse-filtered waveform. Many of these issues may not be significant, depending on the particular application; however, the role of convection in DC flow measurement seems likely to be crucial to all studies using the Rothenberg mask. A more thorough analysis is planned.

CONCLUSION

This paper has considered the aeroacoustics of the vocal tract in some detail to develop a realistic model of propagation. Predictions of formant frequencies were compared with measurements for two cases. Future suggested developments including both additional features and new areas of study conclude the chapter.

CONCERNING THE PROGRAM

The program VOAC includes features of real fluid behavior that are necessarily omitted when electro-acoustic analogues are adopted for modelling the acoustic behavior of the vocal tract (Kelly and Lochbaum, 1962; Flanagan, 1972; Liljencrants, 1985; Scully, 1990). These additions include mean flow, non-isentropic processes associated with vortex generation, and, when appropriate, axial mean velocity gradients. Other additions include more realistic boundary conditions such as generation of evanescent waves at axial area discontinuities, flow-dependent wave reflections at the lips, and the inclusion of sinuses. The program has been written so that the influence on wave propagation of mean flow, wall vibration, and viscothermal losses may each be evaluated independently.

The calculation of the continuous axial distribution of the complex component wave amplitude spectra predicts the acoustic conditions existing throughout the vocal tract. The program has been written to provide other optional outputs, such as attenuation spectra, pressure and particle velocity modulus and phase spectra, similar reflection coefficient and impedance spectra, and also particle acceleration or pressure gradient spectra at any desired and specified location along the tract. It is also possible to add other additional outputs if required, such as transmission coefficients or transmission matrix spectra. Thus the program is very flexible.

FURTHER DEVELOPMENTS OF VOAC

There are two immediate improvements that can be made to the algorithms in VOAC. More realistic modelling of cylindrical wavefront propagation with the accompanying transfers at junctions between elements is currently in hand. The influence of axial curvature of the vocal tract could also be incorporated (Rostafinski, 1991), at least to a first approximation. Although its influence on acoustic propagation may remain insignificant in many instances, its relative importance should be demonstrated by appropriate adjustment to the wave number.

There are some other aspects of the flow of real fluids that have not been included in VOAC. However, examination of the literature suggests that only a few will have any effect on the sound heard by a listener. One-dimensional nonlinear propagation, in particular, can be attacked in the same general spirit as in the program exhibited here: that is, make judicious additions to the existing approximations to discover the significance of the relevant aspects of the dynamics of real fluid motion that have been neglected. Nonlinearities should be examined because it has been observed that the time-varying particle velocity can be of the same order as the mean particle velocity, which did show some effect on the propagation. In such cases, if products of mean-flow terms and time-varying terms are retained, then the products of time-varying terms should also be retained for consistency.

---

[5] These predictions are borne out by experimental results, to be presented in a forthcoming Ph.D. thesis by A.M. Barney, Department of Electronics and Computer Science, University of Southamptom.

There are others, such as the influence of vorticity on propagation, that may respond to a full three-dimensional, high Reynolds number numerical study of the fluid mechanics of the vocal tract and that remain to be fully investigated.

CONCERNING THE SIMULATIONS

Acoustic input impedance spectra were calculated for two vocal tracts for which appropriate geometric information was available. The results and comparisons of the derived formant frequencies are summarized in Tables 3-1 and 3-2, which demonstrate in general a successful match, particularly for the vowel /a/ of subject TB. Reactive walls, non-zero flow, and omission of sinuses all independently act to raise formant frequencies. It was seen, however, that flow has a greater effect when the walls are reacting; other interactions remain to be investigated.

It is important to note the following: 1) the tract geometry has been simplified. It is possible that using more elements, or tailoring the elements for a more perfect fit, will change the predictions significantly. 2) Wall parameters for the case of reacting walls are only poorly known. We need more data on mass, stiffness, and damping. 3) Termination at the lips might be more accurately modelled, particularly at the higher Helmholtz numbers. 4) Duct curvature could be included in future.

Experience indicates that more valid data on vocal tract geometry are necessary in order to provide reliable acoustic predictions. MRI gives good resolution in three dimensions, but the associated experimental environment raises further difficulties. These include the unnatural nature of the speaking task, the hostile environment for the acoustic measurements, and the limitation to static configurations of the vocal tract.

Confidence in the comparisons between measurements *in vivo* and predictions of the model needs strengthening. Ideally, geometric, acoustic, and volume velocity data should be determined simultaneously, and presented with error bounds. Failing this, it would be helpful if the acoustic data included formant bandwidths as well as frequencies, and similar data defining spectral shape. A time history of volume velocity at the mouth for the subject would also be helpful.

FUTURE DEVELOPMENTS

The program described in this paper is very flexible and represents a different approach to modelling vocal tract acoustics, and thus offers a range of possibilities for future studies. Obviously, it is important to extend our comparisons to other phonemes and subjects for which appropriate data are available. It would also be possible to investigate the acoustic consequences of vocal tract movements under more realistic flow conditions. The facility for calculating forces can be applied to investigations of forced motions of the tract boundaries, viz. the vocal folds. This could then lead to the study of new aspects of source-tract interaction.

A useful addition to the program would be sound generation by the flow. This would include periodically shed vortex motions, and turbulence, both of which are concerned in voiced fricatives. This opens new possibilities for a unified approach to source and propagation modelling, rather than an ad hoc separation of the two.

ACKNOWLEDGMENTS

We wish to thank Pat Nye and Tom Baer for their help in supplying the MRI data. Preparation of this chapter was supported by Scientific and Engineering Research Council of Great Britain grant GR/G 44260 to Christine H. Shadle, and by NIH grants DC-00121, DC-00865, and HD-01994 to Haskins Laboratories.

DISCUSSION

Dr. Scherer: Could the turbulence created at glottal exit during breathy phonation act to attenuate the AC component of voicing?

Dr. McGowan: No, I do not believe that this is possible because of the Mach number of the mean flow. This is explained in the subsection titled, "Sound Absorption by Turbulence."

Dr. Herzel: Can you estimate the effect of vorticity generation and jet instabilities on vocal fold vibration? This feedback might be of interest to understand normal phonation as well as irregularities.

Dr. McGowan: In the current models of vocal fold vibration (Titze, Ishizaka, etc.), the nonlinear resistance caused by the formation of vorticity just upstream and downstream of the glottis is an essential part of the feedback mechanism allowing energy flow from the air to the folds. (This is for modal vibration). However, it is unknown how the modulation of tract vorticity by supralaryngeal wave motion affects the vibration of the folds.

Dr. Titze: A few years ago Dr. Teager tried to convince us that the acoustic waves in the vocal tract are confined to the boundaries of the emerging jet rather than the actual physical boundaries of the tract. Would you care to comment on that?

Dr. Davies: The waves always occupy the whole cross section of the tract. The portion of the wave front where mean flow exists is of course influenced by the mean flow and will have a local phase velocity of $c(1+M)$ for the positively and $c(1-M)$ for the negatively travelling waves, while the corresponding phase velocity in the surrounding region where the mean velocity is zero will be c. However the difference between $1 \pm M$ and unity is so small (for the vocal tract), the waves remain effectively plane.

Dr. Fujimura: Regarding the singing formant: is it created by a passive resonance, or by an active amplifier involving an energy source due to vortices in the larynx?

Dr. Davies: I can only speculate for this particular case since I don't have all the relevant parametric data at my fingertips! However experience in analogous situations (not related to the vocal tract) would suggest that it is almost certainly an "active amplifier," particularly when it represents the result of "directed training" by the singer who learns how to extract the required filter frequency spectral energy from the mean flow, involving resonances excited by vortices. Both the resonators and their excitation by vortices represent complex parts of the acoustic "amplifier" which are both physically necessary for the phenomena to exist. Thus the vortex train provides the energy transfer from mean flow while the resonator determines the frequency and amplitude of the oscillations. Therefore, I come down strongly on the side of an active amplifier mechanism, not a passive resonance.

Dr. Berke: We have had much difficulty in empirically measuring volume velocity. Inverse filtering is a far field measurement and it is difficult to understand its relationships to the true flow and gives no information on turbulence and its acoustic amplification or damping.

Dr. Shadle: It is unfortunately a basic fact of life that if you have one signal that you wish to decompose into source(s) and filter, you must either have or estimate other information about the system. For inverse filtering of normal phonation, the filter characteristic is estimated from the vowel quality (and from the signal itself), and the source is assumed to be localized at the other end of that filter, resulting in a unique decomposition. If another source such as turbulence is operating, one must also know/estimate where that source is and the degree of coherence of the two sources. This problem is of course compounded by the fact that the transducers used for inverse filtering are not sensitive above 3 kHz, where most of the noise is likely to be!

To do an accurate decomposition, more measurements are needed in different places along the tract. This is one reason I do so much work with mechanical models.

Dr. Fujimura: We are discussing aspects of speech signals that cannot be captured by standard theory, which we all are biased toward. Inverse filtering and spectrographic displays can mislead us in these specific aspects of signal characteristics. We certainly need physical experiments.

Dr. Berke: Can you suggest means of visualizing turbulence in the vocal tract?

Dr. Shadle: There are many ways of visualizing turbulence, but few of these can be used in the vocal tract. One of the simpler visualization methods involves seeding the flow with particles — e.g. smoke — and photographing the flow downstream of the seeding point. This could be done in the vocal tract provided that illumination and a viewpoint for the camera could be supplied by optical fiber. Laser anemometry detects particle motion by using the particle's Doppler shift, and thus provides more information than simply particle position. However, it involves much more elaborate technology and is expensive. For both these methods, it must be noted that the **streaklines** followed by particles in the flow do not necessarily equal **streamlines** of the flow itself when the flow rate is high or the flow is unsteady.

Laser interferometry visualizes density fluctuations. The equipment must be very precisely positioned for results to be meaningful; also, in the vocal tract it appears unlikely that density changes would be large enough to be visible.

Dr. Scherer: As the vocal folds vibrate during phonation, is there sufficient time for vortices to be created at glottal exit, and is the supraglottal geometry sufficient to allow significant acoustic generation from the vortices?

Dr. Davies: If the shear layer is thin enough then there is no reason why vortices spaced a cm or so apart (corresponding to a fundamental frequency range of 100-300 Hz) should not form. With regard to your second question, it seems possible that the supraglottal geometry could result in significant acoustic generation, but this should be tested experimentally.

Dr. Scherer: How would you now interpret the glottal volume velocity from inverse filtering? Does it have connection to reality? Does the three-dimensional particle velocity sufficiently integrate downstream close to the glottis to give validity to the concept of the volume velocity waveform?

Dr. Shadle: Since the inverse-filtered waveform is a linear transformation of the signal recorded at the mouth, and that recorded signal includes the part that constitutes radiated sound, the inverse-filtered waveform is definitely "connected to reality": it has a close and reversible relationship to the sound pressure near the lips, which is measurable. However, it does not closely correspond to any parameter measurable near the glottis. The velocity profile is highly nonuniform not only at the glottis but for more than one-quarter the length of the entire vocal tract, based on hot-wire measurements in a model of the folds and tract (see Barney et al, 1991). All of the studies done relating the inverse-filtered waveform to vocal fold motion and other measurable glottal parameters are in fact teasing out what the relationship is between the hypothesized, idealized source parameters and measurable glottal parameters by the only method currently available: experiment. However, a more realistic inverse filter should lead to a more realistic predicted source, and could mean that the relationship to measurable glottal parameters is simpler. This is obviously an important direction to pursue.

PRACTICAL FLOW DUCT ACOUSTICS APPLIED TO THE VOCAL TRACT: RESPONSE

RONALD C. SCHERER, Ph.D.

The paper and presentations by Drs. Davies, McGowan, and Shadle emphasized the application of Dr. Davies' computer model to speech synthesis by the incorporation of the effects of mean flow rate, evanescent modes at abrupt area changes, non-isentropic relations between pressure and density (which account for certain aspects of flow separation), and wall displacement. Mean flows, wall vibration, and visco-thermal damping have been shown to affect formant frequencies. The approach and discussions reinforce the concept of acoustic propagation as a subset of fluid mechanics for vocal tract effects. The model may be the most physically based articulatory-type speech synthesis model to date.

This work, and subsequent discussions with Drs. Davies, McGowan, and Shadle, on aeroacoustic aspects, prompt several practical questions and comments that might be worthwhile to discuss.

VORTICES IN THE LARYNX

Questions have arisen concerning the involvement of jets and vortices in creating sound in the vocal tract, especially near the larynx. This is a worthy area of concern. The idea of additional sources of acoustic energy that may help improve intelligibility or voice quality and intensity of voice is of keen interest in both the clinic and performance settings.

A review article by I. Namer and M. Otugen (1988) may have some application here, although it is mentioned primarily to provoke discussion rather than to prove anything. A glottal (let us say) turbulent plane jet (of Reynolds numbers expected for phonation) would be a free shear flow driven by the momentum introduced at the glottal exit. At exit the jet begins to entrain the ambient air. Momentum is exchanged with more entrainment as the jet propagates. The length of the potential core, wherein the centerline velocity equals the exit velocity, is approximately 5 glottal widths, which would be on the order of 1 cm. Beyond the potential core, the centerline velocity decays while turbulence intensity grows. Regarding initial jet growth, immediately downstream of the glottal exit the unstable laminar shear layers break down to form vortices. These induce mixing by enwrapping the ambient air. Vortex formation occurs on both sides of the plane jet.

At the end of the potential core, there is found a dominant frequency which corresponds to a Strouhal number *(fD/u)* between 0.15 and 0.42, which appears to be independent of Reynolds number. Extending this to the glottis, it is noted that taking a characteristic volume velocity of 300 cc/s, and a characteristic glottal diameter of 0.15 cm, the corresponding vortex frequency is 2000 Hz for $St = 0.15$, 5600 Hz for $St = 0.42$, and the intermediate frequency of 3000 Hz for $St = 0.225$. That is, the voice enhancement feature so desired by performance (as well as clinically) appears well within these calculations from experiments made with plane jets. These frequencies would be present in the flow at a location essentially above the false vocal folds within the laryngeal vestibule. This discussion assumes that the false folds are sufficiently placed laterally that the jet is undisturbed, and that the laryngeal confinement in the larynx tube (the tissue boundaries of the laryngeal airway) allows the jet process to happen in the way described. There also appears evidence that externally produced sound near the dominant vortex frequency may force the vortices to grow.

Given that the discussion so far is accurate and there may be the possibility of dominant frequency formation in the velocity fluctuations at the end of the potential core, we might look to the laryngeal region for resonance of similar frequency. A quick estimate of the resonance frequency of the bilateral sinus of Morgagni between the true and false folds suggests that this region has a high resonance over 15,000 Hz. Thus, a jet-edge-resonator oscillation is unlikely here (it is noted that the sinus of Morgagni may qualify for self-sustained oscillation due to inherent instability of the finite thickness shear layer, resulting in a calculated frequency of 4000-6000 Hz; ref. D. Rockwell, 1977). However, if the larynx region, including the vestibule, which opens quickly to the larger pharyngeal region, is modeled as a closed-open tube of about 3 cm, a resonance frequency is obtained between 2500 and 3000 Hz, the frequency area described as the singer's formant region. Thus, there may be a match of a vortex source, even though weak, with an intimate resonance. This reasoning appears in line with Sundberg's earlier discussions.

It is noted that the pulsatile nature of glottal flow means that the jet flow from the glottis varies constantly, even to no flow conditions during glottal closure. That is, the glottal airflow is pulsed with the creation of variable jet bursts. This means that the jet flow is never sustained, and if acoustically helpful vortices can be created quickly enough, they are trains modulated by the fundamental frequency with intra-train aspect variations. The pulsatile nature makes the problem of vorticity contributions more complicated.

GLOTTAL SOUND SOURCE

There has been some question about the glottal sound source in relation to the inverse filtered airflow exiting the mouth as is performed with a Glottal Enterprises system. That is, inverse filtering the output oral flow results in a volume velocity signal attributed to the glottal end of the vocal tract, assuming the inverse filtering deals with the vocal tract resonances. [Inverse filtering a microphone signal at the mouth output essentially results in the glottal volume flow derivative.] The question is whether the volume velocity signal obtained with inverse filtering is a valid representation of the flow at the larynx.

At any instant the glottal flow is an air jet made up of a distribution over the glottal opening of particle air velocities that vary widely in value. The sum (integral) of the particle velocities at glottal exit (i.e., over the glottal exit area) most likely sufficiently equals the instantaneous volume velocity obtained by inverse filtering. At glottal exit it certainly is not the case that at any instant there is a uniform velocity profile across the tract as would be assumed under some linear acoustic concept. One might ask if the glottal jet spreads quickly so that the velocity profile does appear somewhat uniform after, say, 0.5 cm distance. This is most likely impossible at the Reynolds numbers expected in phonation. The jet probably travels at least one cm prior to any significant spreading. The upper surfaces of the false vocal folds slope up to perhaps encourage reattachment of the flow in the vestibule, however.

The shape of the glottal volume velocity signal has been of interest because of the amplitude, skewing, and open quotient characteristics. The amplitude of the signal would refer to the maximum volume flow for the cycle, and the inverse filtered waveform most likely accurately depicts this. The open quotient refers more to a kinematic aspect, that is, minimal (or zero) flow points rather than to the volume velocity per se, so to the extent that flow "corners" are sharp and resonance elimination is adequate, the inverse filtered signal should be helpful in this measure. Waveform skewing has been related to reactance of the vocal tract by Rothenberg, Titze, and others, and the question I presume is whether this aspect should have been part of the inverse filtering (vocal tract rendering) or retained in the glottal volume flow. This issue of what characterizes the vocal tract and what would characterize the glottal flow was well emphasized by Drs. Davies, McGowan, and Shadle. It is central to the problem of how to separate the influences of glottal, vocal tract acoustic, and aeroacoustic sound sources, a primary issue of the paper by Davies, McGowan, and Shadle.

Acoustically, the glottis is considered monopole source-like, for which the derivative of the volume flow (or particle flow) is the sound. That is, acoustic disturbances are created and travel at the speed of sound according to the time derivative of the flow. Naively, I would assume that the particle velocity distribution across the glottis varies with time, and therefore there must be a distribution of the time derivatives of velocity across the glottis creating a distribution of sound intensities. Apparently the acoustic disturbances within a very short distance sum together to create effectively a single acoustic wave front, and the intensity of this wave front would strongly correspond to the differentiated flow at the corresponding time in the phonatory cycle.

From a practical viewpoint, I would want to have the glottis function so that it can create, with little respiratory effort, a glottal particle velocity distribution with large velocity values near the time of peak volume flow (and near the time of peak glottal area), followed by a moment just before glottal closure of very large glottal resistance and almost no flow, so that the greatest particle velocity change over a short period of time could take place, generating the largest possible sound intensity. From a strictly glottal resistance point of view, this would be achieved by having the glottal flow peak occur with an optimum diffuser condition as the shape of the glottis, whereby there is greatest pressure recovery through the glottis, and relatively least resistance and greatest flow. And then near glottal closure, there should be a highly resistive situation of nearly parallel and very close vocal folds. How aeroacoustic interactions might facilitate this optimization of flow and glottal function needs to be seen.

Although the glottal flow may be considered a monopole source, if the glottal flow jet impinges onto the underside of the false vocal folds, a dipole source of sound may be created due to this interaction, its intensity depending upon the velocity of the flow. Again, this would be a source of sound that would be difficult to relate by measurement to that location in the vocal tract, rather than to the glottal flow itself.

Other potential dipole sources such as a jet impinging onto the epiglottis would suffer the same fate, most likely.

PATHOLOGY AND PERFORMANCE

Models of voice and speech need to account for anatomical abnormalities, normal and abnormal innervation conditions, variable structural instabilities and unsteadiness, and resonance enhancements that are special to pathology and performance. Many abnormal conditions might appear merely as perturbations of normal conditions, except, for purposes here, at high levels of pressure or flow, when special aeroacoustic considerations may be needed. Also, the vortex and nonlinearity issues are obviously of great importance for the creation of intense sounds, and perhaps for the creation of resonance enhancement in both the studio and the clinic. They certainly should be issues in "efficiency" concepts. From a vocal tract modelling point of view, the issue of "voice placement" and its acoustic and aeroacoustic concomitants are important and worthy of immediate attention.

SUMMARY

The synthesis model discussed by Drs. Davies, McGowan, and Shadle is highly promising, and its flexibility and physical orientation is reassuring for future development. Their discussions of the aeroacoustic aspects of phonation and the vocal tract are highly enlightening. It would be fortuitous if their modelling could be used to enhance the accurate deduction of sound sources and their locations within the vocal tract with the intent of helping the diagnosis, rehabilitation, and training of those with voice concerns, for it is an area virtually untouched empirically or theoretically.

For the future, we need computer models that are physiologically and physically complete, including the effects of aeroacoustics. This may require new unsteady numerical techniques to handle the complexity from basic principles, including three-dimensionality and a variety of flow regimes and types of sound sources that are simultaneously present. Accompanying this development, we need empirical modeling that also looks at the subtle aspects of physiology and physics of voice and speech, in order to understand, for example, the aeroacoustic aspects. Consideration should be given to an important point made by Dr. Davies, that it is important to model as complete a system as possible, in order to differentiate the salient aspects and not waste time studying parts that turn out to be of negligible import.

DISCUSSION OF RESPONSE

Dr. Davies: One should be careful when applying the observations of vortex shedding with a steady flow to a time varying flow.

As an extreme example consider the vortex generation at an open tube when a steady flow is suddenly switched on. The result is the development of one large starting vortex over a significant time followed by the gradual development towards the steady flow situation. This generally ignores the influence of developing image vortices resulting from the presence of the tract boundaries. These further complicate and reinforce the developing rotational motions.

Dr. Scherer: Thank you for this caution. We need to examine examples such as the one you give, including those which are even closer mechanically to the variable orifice — variable flow glottis.

REFERENCES

Alfredson, R.J. and Davies, P.O.A.L. (1971). Performance of exhaust silencer components. *Journal of Sound and Vibration, 15,* 175-196.

Alipour, F. and Patel, V.C. (1991). A two-dimensional model of laryngeal flow. *Journal of the Acoustical Society of America, 121,* 1978 (abstract).

Ananthapadmanabha, T. and Gauffin, J. (1983). Some results on the acoustic and aerodynamic factors in phonation. In I.R. Titze and R.C. Scherer (eds.), *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control* (pp 402-413), Denver, CO: Denver Center for the Performing Arts.

Baer, T., Gore, J.C., Gracco, L.C. and Nye, P.W. (1991). Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels. *Journal of the Acoustical Society of America, 90:2:1,* 799-828.

Barney, A.M., Shadle, C.H. and Thomas, D.W. (1991). Airflow measurement in a dynamic mechanical model of the vocal folds. In A. Pickering (ed.), *1991 Research Journal, Department of Electronics and Computer Science* (pp 155-158), Southampton, U.K.: Univ. of Southampton.

Bechert, D.W. (1980). Sound absorption caused by vorticity shedding, demonstrated with a jet flow. *Journal of Sound and Vibration, 70,* 389-405.

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

Chester, W. (1964). Resonant oscillations in closed tubes. *Journal of Fluid Mechanics, 18,* 44-65.

Davies, P.O.A.L. (1975). *Order in disorder: studies in turbulence.* An Inaugural Lecture, pub. by University of Southampton.

Davies, P.O.A.L. (1983). Flow-acoustic coupling in ducts. *Journal of Sound and Vibration, 77,* 191-209.

Davies, P.O.A.L. (1988). Practical flow duct acoustics. *Journal of Sound and Vibration, 124,* 91-115.

Davies, P.O.A.L. (1991). Transmission matrix representation of exhaust system acoustic characteristics. *Journal of Sound and Vibration, 131,* 333-338.

Davies, P.O.A.L., Bento Coelho, J.L. and Bhattacharya, (1980). Reflection coefficients for an unflanged pipe with flow. *Journal of Sound and Vibration, 72,* 543-546.

Davies, P.O.A.L. and Doak, P.E. (1990a). Spherical wave propagation in a conical pipe with mean flow. *Journal of Sound and Vibration, 37,* 343-346.

Davies, P.O.A.L. and Doak, P.E. (1990b). Wave transfer to and from conical diffusers with mean flow. *Journal of Sound and Vibration, 138,* 345-350.

El-Raheb, M. (1980). Acoustic propagation in rigid three-dimensional waveguides. *Journal of the Acoustical Society of America, 67,* 1924-1930.

Fant, G. (1960). *Acoustic theory of speech production.* The Hague: Mouton and Co.

Fant, G. and Sonesson, B. (1962). Indirect studies of glottal cycles by synchronous inverse filtering and photo-electrical glottography. *Speech Transmission Laboratory — Quarterly Progress and Status Report, 4,* 1-3.

Flanagan, J. L. (1972). *Speech analysis synthesis and perception.* New York: Springer-Verlag.

Flanagan, J.L. and Ishizaka, K. (1976). Automatic generation of voiceless excitation in a vocal cord-vocal tract speech synthesizer. *IEEE Transactions on Acoustics, Speech and Signal Processing, 24,* 163-170.

Gobl, C. and Karlsson, I. (1991). Male and female voice source dynamics. In J. Gauffin and B. Hammerberg (eds.), *Vocal Fold Physiology* (pp 121-128). San Diego, CA: Singular Publishing Group.

Helmholtz, H. von (1868). On discontinuous movements of fluids. *Philosophical Magazine, 36,* 337-346.

Holmberg, E.B., Hillman, R.E. and 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,* 511-529.

Howe, M.S. (1979). The interaction of sound with low Mach number wall turbulence, with application to sound propagation in turbulent pipe flow. *Journal of Fluid Mechanics, 94,* 729-744.

Howe, M. S. (1980). The dissipation of sound at an edge. *Journal of Sound and Vibration, 70,* 407-411.

Howe, M. S. (1984). On the absorption of sound by turbulence and other hydrodynamic flows. *IMA Journal of Applied Mathematics, 32,* 187-209.

Iijima, H., Miki, N. and 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,* 53-56.

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

Ishizaka, K., French, J.C. and Flanagan, J.L. (1975). Direct determination of vocal tract wall impedance. *IEEE Transactions on Acoustics, Speech and Signal Processing, 23,* 370-373.

Karlsson, I. (1989). A female voice for a text-to-speech system. *Proceedings of European Conference on Speech Communications and Technology,* Paris, 349-352.

Keefe, D.H. (1983). Acoustic streaming, dimensional analysis of nonlinearities, and tone hole mutual interactions in woodwinds. *Journal of the Acoustical Society of America, 73,* 1804-1820.

Keefe, D.H. and Benade, A.H. (1983). Wave propagation in strongly curved ducts. *Journal of the Acoustical Society of America, 74,* 320-332.

Kelly, J. L. and Lochbaum, C. C. (1962). Speech synthesis. *Proceedings of the 4th International Congress of Acoustics,* Copenhagen, Denmark, paper C42, 1-4.

Levine, H. and Schwinger, (1948). On the radiation of sound from an unflanged circular pipe. *Physical Review, 73,* 383-406.

Lighthill, M. J. (1953). On the energy scattered from the interaction of turbulence with sound or shock waves. *Proceedings of the Cambridge Philosophical Society, 49,* 531-551.

Lighthill, M. J. (1978). *Waves in Fluids.* Cambridge, England: Cambridge University Press.

Liljencrants, J. (1985). *Speech synthesis with a reflection-type line analog.* Ph.D. Dissertation, Royal Institute of Technology, Stockholm, Sweden.

Liljencrants, J. (1991). Numerical simulations of glottal flow. In J. Gauffin and B. Hammerberg (eds.), *Vocal Fold Physiology* (pp 99-104). San Diego: Singular Publishing Group.

McGowan, R.S. (1987). Articulatory synthesis: Numerical solution of a hyperbolic differential equation. *Haskins Laboratories Status Report on Speech Research, SR-89/90,* 69-80.

McGowan, R.S. (1991). Nonlinearities for one-dimensional propagation in the vocal tract. *Journal of Phonetics, 19,* 425-432.

Mermelstein, P. (1973). Articulatory model for the study of speech production. *Journal of the Acoustical Society of America, 53,* 1070-1082.

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

Munt, R.M. (1990). Acoustic transmission properties of a jet pipe with subsonic jet-flow. *Journal of Sound and Vibration, 142,* 413-436.

Namer, I. and Otugen, M. (1988). Velocity measurements in a plane turbulent air jet at moderate Reynolds numbers. *Experiments in Fluids, 6,* 387-399.

Noir, D.T. and George, A.R. (1978). The absorption of sound by homogeneous turbulence. *Journal of Fluid Mechanics, 86,* 593-608.

Pedley, T.J., Schroter, R.C., and Sudlow, M.F. (1970). The prediction of pressure drop and variation of resistance within the human bronchial airways. *Respiratory Physiology, 9,* 387-405.

Pedley, T. J. and Stephanoff, K.D. (1985). Generation of vorticity waves in a channel. *Journal of Fluid Mechanics, 160,* 337-367.

Rayleigh, J.W.S. (1896). *The theory of sound.* 2nd ed., London: Macmillan.

Rockwell, D. (1977). Prediction of oscillation frequencies for unstable flow past cavities, *Journal of Fluids Engineering, Trans. ASME 99,* 294-300.

Ronnenberger, D. and Ahrens, C.D. (1977). Wall shear stress caused by small amplitude perturbations of turbulent boundary-layer flow: an experimental investigation. *Journal of Fluid Mechanics, 83,* 433-464.

Rostafinski, W. (1991). *Propagation of sound waves in curved ducts.* NASA Reference Publication 1248.

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,* 1632-1645.

Rothenberg, M.T. (1981a). Acoustic interaction between the glottal source and the vocal tract. In K.N. Stevens and M. Hirano (eds.), *Vocal Fold Physiology* (pp 305-323), Tokyo: University of Tokyo Press.

Rothenberg, M.T. (1981b). Some relations between glottal airflow and vocal fold contact area. In C.L. Ludlow and M.O. Hart (eds.), *Proceedings of the Conference on the Assessment of Vocal Pathology.* ASHA Reports, Monograph 11.

Rubin, P., Baer, T., and Mermelstein, P. (1981). An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America, 70,* 321-328.

Scully, C. (1990). Articulatory synthesis. In W.J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modelling* (pp 151-186), Netherlands: Kluwer Academic Press.

Shadle, C. H. (1985). *The acoustics of fricative consonants.* Ph.D. Dissertation, Dept. of Elect. Eng. and Comp. Science, Mass. Inst. of Tech. Released as RLE Technical Report no. 506.

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

Shadle, C.H., Barney, A. and Thomas, D.W. (1991). An investigation into the acoustics and aerodynamics of the larynx. In J. Gauffin and B. Hammerberg (eds.), *Vocal Fold Physiology* (pp 73-82). San Diego, CA: Singular Publishing Group.

Sobey, I.J. (1980). On flow through furrowed channels. Part I: Calculated flow patterns. *Journal of Fluid Mechanics, 96,* 1-26.

Sobey, I.J. (1983). The occurrence of separation in oscillatory flow. *Journal of Fluid Mechanics, 134,* 263-287.

Sobey, I.J. (1985). Observation of waves during oscillatory channel flow. *Journal of Fluid Mechanics, 151,* 395-426.

Stephanoff, K.D., Pedley, T.J., Lawrence, C.J. and Secomb, T.W. (1983). Fluid flow along a channel with an asymmetric oscillating constriction. *Nature, 305,* 692-695.

Sugimoto, N. (1991). Burger's equation with a fractional derivative: Hereditary effects on nonlinear acoustic waves. *Journal of Fluid Mechanics, 225,* 631-653.

Sugiyama, K. and Irii, H. (1991). Comparison of sound pressure radiation from a prolate spheroid and the human mouth. *Acustica, 73,* 271-276.

Sundberg, J. and Gauffin, J. (1979). Waveform and spectrum of the glottal voice source. In S. Öhman and B. Lindblom (eds.), *Frontiers of Speech Communication* (pp 301-322), London: Academic Press.

Teager, H.M. and Teager, S.M. (1983a). The effects of separated air flow on vocalization. In D.M. Bless and J.H. Abbs (eds.), *Vocal Fold Physiology* (pp 124-141), San Diego: College-Hill Press.

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

Thomas, T. J. (1986). A finite element model of fluid flow in the vocal tract. *Computer Speech and Language, 1,* 131-152.

Tijdemann, H. (1975). On the propagation of sound waves in cylindrical tubes. *Journal of Sound and Vibration, 39,* 1-33.

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

Vocal Fold Physiology Series
Organizing Committee:

Ingo R. Titze, Ph.D.
Minoru Hirano, M.D.
Osamu Fujimura, Ph.D.
Robert Thayer Sataloff, M.D., D.M.A.

Primary Sponsor

The Voice Foundation

Vocal Fold Physiology

Frontiers in Basic Science

Edited by

Ingo R. Titze, Ph.D.
Professor, Department of Speech Pathology and Audiology
Director, National Center for Voice and Speech
The University of Iowa
Director of Research, The Recording and Research Center
The Denver Center for the Performing Arts

SINGULAR PUBLISHING GROUP, INC.
San Diego, California