Table 2.5. *Glides and semi-vowels*

Place	
Palatal	/j/ you
Labial	/w/ we
	(no final form)
Palatal	/r/ read
Alveolar	/l/ let

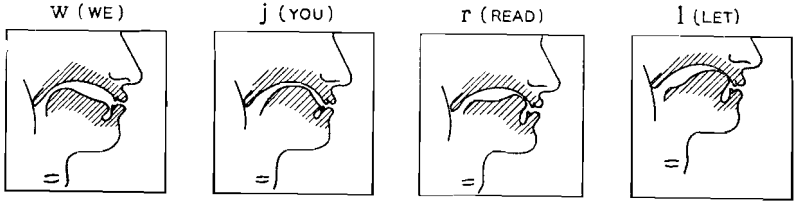


Fig. 2.9. Vocal tract configurations for the beginning positions of the glides and semivowels. (After POTTER, KOPP and GREEN)

2.225. Combination Sounds: Diphthongs and Affricates. Some of the preceding vowel or consonant elements can be combined to form basic sounds whose phonetic values depend upon vocal tract motion. An appropriate pair of vowels, so combined, form a diphthong. The diphthong is vowel-like in nature, but is characterized by change from one vowel position to another. For example, if the vocal tract is changed from the /e/ position to the /ɪ/ position, the diphthong /eɪ/ as in *say* is formed. Other GA diphthongs are /ɪu/ as in *new*, /ɔɪ/ as in *boy*; /aʊ/ as in *out*, /aɪ/ as in *I*, and /oʊ/ as in *go*.

As vowel combinations form the diphthongs, stop-fricative combinations likewise create the two GA affricates. These are the /tʃ/ as in *chew* and the /dʒ/ as in *jar*.

2.3. Quantitative Description of Speech

The preceding discussion has described the production of speech in a completely qualitative way. It has outlined the mechanism of the voice and the means for producing an audible code which, within a given language, consists of distinctive sounds. However, for any transmission system to benefit from prior knowledge of the information source, this knowledge must be cast into a tractable analytical form that can be employed in the design of signal processing operations. Detailed inquiry into the physical principles underlying the speech-producing mechanism is therefore indicated.

The following chapter will consider the characteristics of the vocal system in a quantitative fashion. It will treat the physics of the vocal and nasal tracts in some depth and will set forth certain acoustical properties of the vocal excitations. The primary objective—as stated earlier—is to describe the acoustic speech signal in terms of the physical parameters of the system that produced it. Because of physiological and linguistic constraints, such a description carries important implications for analysis-synthesis telephony.

III. Acoustical Properties of the Vocal System

The collection of olfactory, respiratory and digestive apparatus which man uses for speaking is a relatively complex sound-producing system. Its operation has been described qualitatively in the preceding chapter. In this chapter we would like to consider in more detail the acoustical principles underlying speech production. The treatment is not intended to be exhaustive. Rather it is intended to circumscribe the problems of vocal tract analysis and to set forth certain fundamental relations for speech production. In addition, it aims to outline techniques and method for acoustic analysis of the vocal mechanism and to indicate their practical applications. Specialized treatments of a number of these points can be found elsewhere[1].

3.1. The Vocal Tract as an Acoustic System

The operations described qualitatively in the previous chapter can be crudely represented as in Fig. 3.1. The lungs and associated respiratory muscles are the vocal power supply. For voiced sounds, the expelled air causes the vocal cords to vibrate as a relaxation oscillator, and the air stream is modulated into discrete puffs or pulses. Unvoiced sounds are generated either by passing the air stream through a constriction in the tract, or by making a complete closure, building up pressure behind the closure and abruptly releasing it. In the first case, turbulent flow and incoherent sound are produced. In the second, a brief transient excitation occurs. The physical configuration of the vocal tract is highly variable and is dictated by the positions of the articulators; that is, the jaw, tongue, lips and velum. The latter controls the degree of coupling to the nasal tract.

[1] For this purpose G. FANT, *Acoustic Theory of Speech Production*, is highly recommended. Besides presenting the acoustical bases for vocal analysis, this volume contains a wealth of data on vocal configurations and their calculated frequency responses. An earlier but still relevant treatise is T. CHIBA and M. KAJIYAMA, *The Vowel: Its Nature and Structure*. Another excellent and more recent analysis of vowel articulation is G. UNGEHEUER, *Elemente einer akustischen Theorie der Vokalartikulation*.

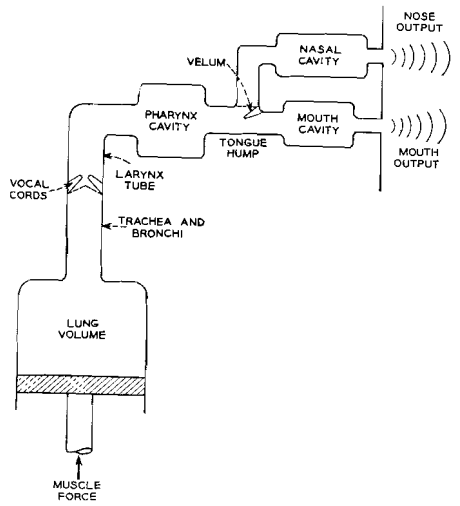


Fig. 3.1. Schematic diagram of functional components of the vocal tract

In general, several major regions figure prominently in speech production. They are: (a) the relatively long cavity formed at the lower back of the throat in the pharynx region; (b) the narrow passage at the place where the tongue is humped; (c) the variable constriction of the velum and the nasal cavity; (d) the relatively large, forward oral cavity; (e) the radiating ports formed by the mouth and nostrils.

Voiced sounds are always excited at the same point in the tract, namely at the vocal cords. Radiation of voiced sounds can take place either from the mouth or nose, or from both. Unvoiced excitation is applied to the acoustic system at the point where turbulent flow or pressure release occurs. This point may range from an anterior position [such as the labio-dental excitation for /f/] to a posterior position [such as the palatal excitation for /k/]. Unvoiced sounds are normally radiated from the mouth. All sounds generated by the vocal apparatus are characterized by properties of the source of excitation and the acoustic transmission system. To examine these properties, let us first establish some elementary relations for the transmission system, then consider the sound sources, and finally treat the combined operation of sources and system.

The length of the vocal tract (about 17 cm in man) is fully comparable to the wavelength of sound in air at audible frequencies. It is therefore not possible to obtain a precise analysis of the tract operation from a lumped-constant approximation of the major acoustic components. Wave motion in the system must be considered for frequencies

above several hundred cps. The vocal and nasal tracts constitute lossy tubes of non-uniform cross-sectional area. Wave motion in such tubes is difficult to describe, even for lossless propagation. In fact, exact solutions to the wave equation are available only for two nonuniform geometries, namely for conical and hyperbolic area variations (MORSE). And then only the conical geometry leads to a one-parameter wave.

So long as the greatest cross dimension of the tract is appreciably less than a wavelength (this is usually so for frequencies below about 4000 cps), and so long as the tube does not flare too rapidly (producing internal wave reflections), the acoustic system can be approximated by a one-dimensional wave equation. Such an equation assumes cophasic wave fronts across the cross-section and is sometimes called the Webster equation (WEBSTER). Its form is

$$\frac{1}{A(x)}\frac{\partial}{\partial x}\left[A(x)\frac{\partial p}{\partial x}\right]=\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}, \tag{3.1}$$

where $A(x)$ is the cross-sectional area normal to the longitudinal dimension, p is the sound pressure (a function of t and x) and c is the sound velocity. In general this equation can only be integrated numerically, and it does not include loss. At least three investigations, however, have made use of this formulation for studying vowel production (CHIBA and KAJIYAMA; UNGEHEUER; HEINZ, 1962a, b).

A more tractable approach to the analysis problem (both computationally and conceptually) is to impose a further degree of approximation upon the nonuniform tube. The pipe may be represented in terms of incremental contiguous sections of right circular geometry. The approximation may, for example, be in terms of cylinders, cones, exponential or hyperbolic horns. Although quantizing the area function introduces error, its effect can be made small if the lengths of the approximating sections are kept short compared to a wavelength at the highest frequency of interest. The uniform cylindrical section is particularly easy to treat and will be the one used for the present discussion.

3.2. Equivalent Circuit for the Lossy Cylindrical Pipe

Consider the length dx of lossy cylindrical pipe of area A shown in Fig. 3.2a. Assume plane wave transmission so that the sound pressure and volume velocity are spatially dependent only upon x. Because of its mass, the air in the pipe exhibits an inertance which opposes acceleration. Because of its compressibility the volume of air exhibits a compliance. Assuming that the tube is smooth and hard-walled, energy losses can occur at the wall through viscous friction and heat conduction. Viscous losses are proportional to the square of the particle velocity, and heat conduction losses are proportional to the square of the sound pressure.

The characteristics of sound propagation in such a tube are easily described by drawing upon elementary electrical theory and some well-known results for one-dimensional waves on transmission lines. Consider sound pressure analogous to the voltage and volume velocity analogous to the current in an electrical line. Sound pressure and volume velocity for plane wave propagation in the uniform tube satisfy the same wave equation as do voltage and current on a uniform transmission line. A dx length of lossy electrical line is illustrated in Fig. 3.2b. To develop the analogy let us write the relations for the electrical line. The per-unit-length inductance, capacitance, series resistance and shunt conductance are L, C, R, and G respectively. Assuming sinusoidal time dependence

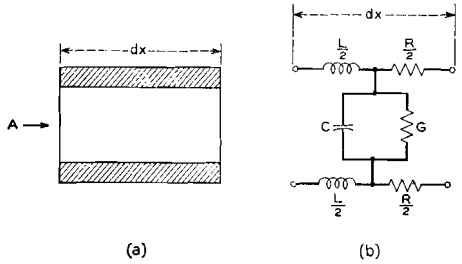


Fig. 3.2a and b. Incremental length of lossy cylindrical pipe. (a) acoustic representation; (b) electrical equivalent for a one-dimensional wave

for voltage and current, ($Ie^{j\omega t}$ and $Ee^{j\omega t}$), the differential current loss and voltage drop across the dx length of line are

$$dI=-Eydx \quad \text{and} \quad dE=-Izdx, \tag{3.2}$$

where $y=(G+j\omega C)$ and $z=(R+j\omega L)$.

The voltage and current therefore satisfy

$$\frac{d^2E}{dx^2}-zyE=0 \quad \text{and} \quad \frac{d^2I}{dx^2}-zyI=0, \tag{3.3}$$

the solutions for which are

$$\begin{aligned} E&=A_1e^{\gamma x}+B_1e^{-\gamma x}\\ I&=A_2e^{\gamma x}+B_2e^{-\gamma x}, \end{aligned} \tag{3.4}$$

where $\gamma=\sqrt{zy}=(\alpha+j\beta)$ is the propagation constant, and the A's and B's are integration constants determined by terminal conditions.

For a piece of line l in length, with sending-end voltage and current E_1 and I_1, the receiving-end voltage and current E_2 and I_2 are given by

$$\begin{aligned} E_2&=E_1\cosh\gamma l-I_1Z_0\sinh\gamma l\\ I_2&=I_1\cosh\gamma l-E_1Y_0\sinh\gamma l, \end{aligned} \tag{3.5}$$

where $Z_0=\sqrt{z/y}$ and $Y_0=\sqrt{y/z}$ are the characteristic impedance and admittance of the line. Eq. (3.5) can be rearranged to make evident the impedance parameters for the equivalent four-pole network

$$\begin{aligned} E_1&=Z_0I_1\coth\gamma l-Z_0I_2\operatorname{csch}\gamma l\\ E_2&=Z_0I_1\operatorname{csch}\gamma l-Z_0I_2\coth\gamma l. \end{aligned} \tag{3.6}$$

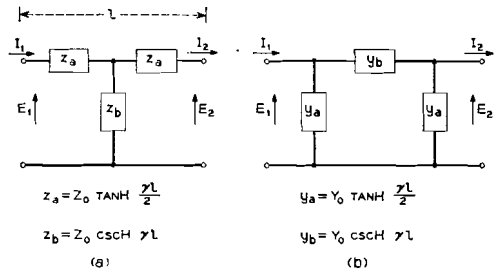


Fig. 3.3a and b. Equivalent four-pole networks for a length l of uniform transmission line. (a) T-section; (b) π-section

The equivalent T-network for the l length of line is therefore as shown in Fig. 3.3a. Similarly, a different arrangement makes salient the admittance parameters for the four-pole network.

$$\begin{aligned} I_1&=Y_0E_1\coth\gamma l-Y_0E_2\operatorname{csch}\gamma l\\ I_2&=Y_0E_1\operatorname{csch}\gamma l-Y_0E_2\coth\gamma l. \end{aligned} \tag{3.7}$$

The equivalent π-network is shown in Fig. 3.3b.

One recalls also from conventional circuit theory the lossless case corresponds to $\gamma=\sqrt{zy}=j\beta=j\omega\sqrt{LC}$, and $Z_0=\sqrt{L/C}$. The hyperbolic functions then reduce to circular functions which are purely reactive. Notice, too, for *small loss* conditions, (that is, $R\ll\omega L$ and $G\ll\omega C$) the attenuation and phase constants are approximately

$$\begin{aligned} \alpha&\cong\frac{R}{2}\sqrt{C/L}+\frac{G}{2}\sqrt{L/C}\\ \beta&\cong\omega\sqrt{LC}. \end{aligned} \tag{3.8}$$

Having recalled the relations for the uniform, lossy electrical line, we want to interpret plane wave propagation in a uniform, lossy pipe in analogous terms. If sound pressure, p, is considered analogous to voltage and acoustic volume velocity, U, analogous to current, the lossy, one-dimensional, sinusoidal sound propagation is described by the same equations as given in (3.3). The propagation constant is complex (that is, the velocity of propagation is in effect complex) and the wave attenuates as it travels. In a smooth hard-walled tube the viscous and heat conduction losses can be represented, in effect, by an I^2R loss and an E^2G loss, respectively. The inertance of the air mass is analogous to the electrical inductance, and the compliance of the air volume is analogous to the electrical capacity. We can draw these parallels quantitatively[1].

3.21. The Acoustic "L"

The mass of air contained in the dx length of pipe in Fig. 3.2a is $\rho A\,dx$, where ρ is the air density. The differential pressure drop in accelerating this mass is by NEWTON'S law:

$$dp=\rho\,dx\,\frac{du}{dt}=\rho\frac{dx}{A}\cdot\frac{dU(x,t)}{dt},$$

where u is particle velocity and U is volume velocity.

For $U(x,t)=U(x)\,e^{j\omega t}$

$$dp=j\omega\rho\frac{dx}{A}U$$

and (3.9)

$$\frac{dp}{dx}=j\omega L_a U,$$

where $L_a=\rho/A$ is the acoustic inertance *per unit length.*

3.22. The Acoustic "R"

The acoustic R represents a power loss proportional to U^2 and is the power dissipated in viscous friction at the tube wall (INGÅRD). The easiest way to put in evidence this equivalent surface resistance is to consider the situation shown in Fig. 3.4. Imagine that the tube wall is a plane surface, large in extent, and moving sinusoidally in the x-direction with velocity $u(t)=u_m e^{j\omega t}$. The air particles proximate to the wall experience a force owing to the viscosity, μ, of the medium. The power expended per unit area in dragging the air with the plate is the loss to be determined.

[1] The reader who is not interested in these details may omit the following four sections and find the results summarized in Eq. (3.33) of Section 3.25.

Consider a layer of air dy thick and of unit area normal to the y axis. The net force on the layer is

$$\mu\left[\left(\frac{\partial u}{\partial y}\right)_{y+dy}-\left(\frac{\partial u}{\partial y}\right)_{y}\right]=\rho\,dy\,\frac{\partial u}{\partial t},$$

where u is the particle velocity in the x-direction. The diffusion equation specifying the air particle velocity as a function of the distance above the wall is then

$$\frac{\partial^2 u}{\partial y^2}=\frac{\rho}{\mu}\frac{\partial u}{\partial t}. \tag{3.10}$$

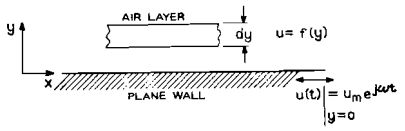


Fig. 3.4. Relations illustrating viscous loss at the wall of a smooth tube

For harmonic time dependence this gives

$$\frac{d^2u}{dy^2}=j\frac{\omega\rho}{\mu}u=k_v^2\,u, \tag{3.11}$$

where $k_v=(1+j)\sqrt{\omega\rho/2\mu}$, and the velocity distribution is

$$u=u_m e^{-k_v y}=u_m e^{-\sqrt{\omega\rho/2\mu}\,y}e^{-j\sqrt{\omega\rho/2\mu}\,y}. \tag{3.12}$$

The distance required for the particle velocity to diminish to $1/e$ of its value at the driven wall is often called the boundary-layer thickness and is $\delta_v=\sqrt{2\mu/\omega\rho}$. In air at a frequency of 100 cps, for example, $\delta_v\cong 0.2$ mm.

The viscous drag, per unit area, on the plane wall is

$$F=-\mu\left(\frac{\partial u}{\partial y}\right)_{y=0}=\mu\,k_v\,u_m,$$

or (3.13)

$$F=u_m(1+j)\sqrt{\omega\mu\rho/2}.$$

Notice that this force has a real part and a positive reactive part. The latter acts to increase the apparent acoustic L. The average power dissipated per unit surface area in this drag is

$$\bar{P}=\tfrac{1}{2}|F|\,u_m\cos\vartheta=\tfrac{1}{2}u_m^2 R_s, \tag{3.14}$$

where $R_s=\sqrt{\omega\rho\mu/2}$ is the per-unit-area surface resistance and ϑ is the phase angle between F and u, namely, 45°. For a length l of the acoustic tube, the inner surface area is $S\cdot l$, where S is the circumference. Therefore, the average power dissipated *per unit length* of the tube is $\bar{P}\cdot S=\frac{1}{2}u_m^2\cdot S\cdot R_s$, or in terms of the acoustic volume velocity

$$\bar{P}\cdot S=\tfrac{1}{2}U_m^2 R_a,$$

where

$$R_a=\frac{S}{A^2}\sqrt{\omega\rho\mu/2}, \tag{3.15}$$

and A is the cross-sectional area of the tube. R_a is then the per-unit-length acoustic resistance for the analogy shown in Fig. 3.2.

As previously mentioned, the reactive part of the viscous drag contributes to the acoustic inductance per unit length. In fact, for the same area and surface relations applied above, the acoustic inductance obtained in the foregoing section should be increased by the factor $\frac{A^2}{S}\sqrt{\mu\rho/2\omega}$, or

$$L_a=\frac{\rho}{A}\left(1+\frac{S}{A}\sqrt{\frac{\mu}{2\rho\omega}}\right). \tag{3.16}$$

Thus, the viscous boundary layer increases the apparent acoustic inductance by effectively diminishing the cross-sectional area. For vocal tract analysis, however, the viscous boundary layer is usually so thin that the second term in (3.16) is negligible. For example, for a circular cross-section of 9 cm², the second term at a frequency of 500 cps is about (0.006) ρ/A.

3.23. The Acoustic "*C*"

The analogous acoustic capacitance, or compliance, arises from the compressibility of the volume of air contained in the dx length of tube shown in Fig. 3.2a. Most of the elemental air volume $A\,dx$ experiences compressions and expansions which follow the adiabatic gas law

$$PV^{\eta}=\text{constant},$$

where P and V are the total pressure and volume of the gas, and η is the adiabatic constant[1]. Differentiating with respect to time gives

$$\frac{1}{P}\frac{dP}{dt}=-\frac{\eta}{V}\frac{dV}{dt}.$$

[1] η is the ratio of specific heat at constant pressure to that at constant volume. For air at normal conditions, $\eta=c_p/c_v=1.4$.

The diminution of the original air volume, owing to compression caused by an increase in pressure, must equal the volume current into the compliance; that is,

$$U=-\frac{dV}{dt},$$

and

$$\frac{1}{P}\frac{dP}{dt}=\frac{\eta U}{V}.$$

For sinusoidal time dependence $P=P_0+pe^{j\omega t}$, where P_0 is the quiescent pressure and is large compared with p. The volume flow into the compliance of the $A\,dx$ volume is therefore approximately

$$U=j\omega\frac{V}{P_0\eta}\cdot p=j\omega\frac{A\,dx}{P_0\eta}\cdot p. \tag{3.17}$$

From wave considerations $P_0\eta$ can be shown to equal ρc^2. The volume velocity into the per-unit-length compliance can therefore be written as

$$U=j\omega\cdot C_a\cdot p,$$

where

$$C_a=\frac{A}{P_0\eta}=\frac{A}{\rho c^2} \tag{3.18}$$

is the per-unit-length acoustic compliance.

3.24. The Acoustic "*G*"

The analogous shunt conductance provides a power loss proportional to the square of the local sound pressure. Such a loss arises from heat conduction at the walls of the tube. The per-unit-length conductance can be deduced in a manner similar to that for the viscous loss. As before, it is easier to treat a simpler situation and extend the result to the vocal tube.

Consider a highly conductive plane wall of large extent, such as shown in Fig. 3.5. The air above the boundary is essentially at constant pressure and has a coefficient of heat conduction λ and a specific heat c_p. Suppose the wall is given an oscillating temperature $T|_{y=0}=T_m e^{j\omega t}$. The vertical temperature distribution produced in the air is described by the diffusion equation (HILDEBRAND)

$$\frac{\partial^2 T}{\partial y^2}=\frac{c_p\rho}{\lambda}\frac{\partial T}{\partial t},$$

or

$$\frac{\partial^2 T}{\partial y^2}=j\omega\frac{c_p\rho}{\lambda}T. \tag{3.19}$$

The solution is $T=T_m e^{-k_h y}$, where

$$k_h=(1+j)\sqrt{\frac{\omega c_p \rho}{2\lambda}}, \tag{3.20}$$

which is the same form as the velocity distribution due to viscosity. In a similar fashion, the boundary layer depth for temperature is $\delta_h=\sqrt{2\lambda/\omega c_p \rho}$, and $k_h=(1+j)/\delta_h$.

Now consider more nearly the situation for the sound wave. Imagine an acoustic pressure wave moving parallel to the conducting boundary,

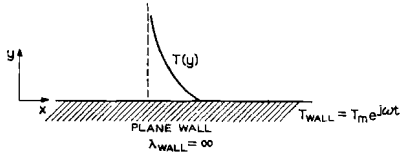


Fig. 3.5. Relations illustrating heat conduction at the wall of a tube

that is, in the x-direction. We wish to determine the temperature distribution above the wall produced by the sound wave. The conducting wall is assumed to be maintained at some quiescent temperature and permitted no variation, that is, $\lambda_{\text{wall}}=\infty$. If the sound wavelength is long compared to the boundary extent under consideration, the harmonic pressure variation above the wall may be considered as $P=P_0+p$, where P_0 is the quiescent atmospheric pressure and $p=p_m e^{j\omega t}$ is the pressure variation. (That is, the spatial variation of p with x is assumed small.) The gas laws prescribe

$$PV^{\eta}=\text{constant} \quad \text{and} \quad PV=RT \quad \text{(for unit mass)}.$$

Taking differentials gives

$$\frac{dV}{V}=-\frac{1}{\eta}\frac{dP}{P} \quad \text{and} \quad \frac{dP}{P}+\frac{dV}{V}=\frac{dT}{T}. \tag{3.21}$$

Combining the equations yields

$$\frac{dP}{P}\left(1-\frac{1}{\eta}\right)=\frac{dT}{T}, \tag{3.22}$$

where

$$dP=p=p_m e^{j\omega t}$$

$$dT=\tau=\tau_m e^{j\omega t},$$

so from (3.22)

$$\tau_m=\frac{T_0}{P_0}\left(\frac{\eta-1}{\eta}\right)p_m. \tag{3.23}$$

At the wall, $y=0$ and $\tau(0)=0$ (because $\lambda_{\text{wall}}=\infty$). Far from the wall (i.e., for y large), $|\tau(y)|=\tau_m$ as given in (3.23). Using the result of (3.20), the temperature distribution can be constructed as

$$\tau(y,t)=[1-e^{-k_h y}]\tau_m e^{j\omega t},$$

or

$$\tau(y,t)=\frac{P_0}{T_0}\left(\frac{\eta-1}{\eta}\right)[1-e^{-k_h y}]p_m e^{j\omega t}. \tag{3.24}$$

Now consider the power dissipation at the wall corresponding to this situation. A long wavelength sound has been assumed so that the acoustic pressure variations above the boundary can be considered $p=p_m e^{j\omega t}$, and the spatial dependence of pressure neglected. Because of the temperature distribution above the boundary, however, the particle velocity will be nonuniform, and will have a component in the y-direction. The average power flow per unit surface area into the boundary is $\overline{p u_{y_0}}^t$, where u_{y_0} is the velocity component in the y direction at the boundary. To examine this quantity, u_y is needed.

Conservation of mass in the y-direction requires

$$\rho\frac{\partial u_y}{\partial y}=-\frac{\partial \rho}{\partial t}. \tag{3.25}$$

Also, for a constant mass of gas $d\rho/\rho=-dV/V$ which with the second equation in (3.21) requires

$$\frac{dP}{P}-\frac{d\rho}{\rho}=\frac{dT}{T}. \tag{3.26}$$

Therefore,

$$\frac{\partial u_y}{\partial y}=\left(\frac{1}{T_0}\frac{\partial \tau}{\partial t}-\frac{1}{P_0}\frac{\partial p}{\partial t}\right), \tag{3.27}$$

and

$$u_y=\int\frac{\partial u_y}{\partial y}\cdot dy$$

$$u_y=\frac{j\omega p}{P_0}\left\{\frac{\eta-1}{\eta}\left(y+\frac{e^{-k_y y}}{ky}\right)-y\right\}. \tag{3.28}$$

And,

$$u_{y_0}=p\frac{\omega}{c}\frac{\eta-1}{\rho c}\frac{j}{1+j}\delta_h. \tag{3.29}$$

The equivalent energy flow into the wall is therefore

$$W_h=\overline{p u_{y_0}}^t=\frac{\omega}{c}\frac{\eta-1}{\rho c}\delta_h\frac{1}{\sqrt{2}}\frac{1}{T}\int_0^T P_m^2\cos\left(\omega t+\frac{\pi}{4}\right)\cos\omega t\cdot dt$$

$$W_h=\frac{1}{4}\frac{\omega}{c}\frac{\eta-1}{\rho c}\delta_h p_m^2=\frac{1}{2}G_\alpha p_m^2, \tag{3.30}$$

where G_α is an equivalent conductance per unit wall area and is equal

$$G_\alpha=\frac{1}{2}\frac{\omega}{c}\frac{\eta-1}{\rho c}\sqrt{\frac{2\lambda}{\omega c_p\rho}}. \tag{3.31}$$

The equivalent conductance per unit length of tube owing to heat conduction is therefore

$$G_\alpha=S\frac{\eta-1}{\rho c^2}\sqrt{\frac{\lambda\omega}{2c_p\rho}}, \tag{3.32}$$

where S is the tube circumference.

To reiterate, both the heat conduction loss G_a and the viscous loss R_a are applicable to a smooth, rigid tube. The vocal tract is neither, so that in practice these losses might be expected to be somewhat higher. In addition, the mechanical impedance of the yielding wall includes a mass reactance and a conductance which contribute to the shunt element of the equivalent circuit. The effect of the wall reactance upon the tuning of the vocal resonances is generally small, particularly for open articulations. The contribution of wall conductance to tract damping is more important. Both of these effects are estimated in a later section.

3.25. Summary of the Analogous Acoustic Elements

The per-unit-length analogous constants of the uniform pipe can be summarized.

$$L_a=\frac{\rho}{A}, \qquad C_a=\frac{A}{\rho c^2},$$

$$R_a=\frac{S}{A^2}\sqrt{\frac{\omega\rho\mu}{2}}, \qquad G_a=S\frac{\eta-1}{\rho c^2}\sqrt{\frac{\lambda\omega}{2c_p\rho}}, \tag{3.33}$$

where A is tube area, S is tube circumference, ρ is air density, c is sound velocity, μ is viscosity coefficient, λ is coefficient of heat conduction,

η is the adiabatic constant, and c_p is the specific heat of air at constant pressure[1].

Having set down these quantities, it is possible to approximate the nonuniform vocal tract with as many right circular tube sections as desired. The transmission characteristics can be determined either from calculations on equivalent network sections such as shown in Fig. 3.3, or from electrical circuit simulations of the elements. When the approximation involves more than three or four network loops, manual computation becomes prohibitive. Computer techniques can then be used to good advantage.

A further level of approximation can be made for the equivalent networks in Fig. 3.3. For a given length of tube, the hyperbolic elements may be approximated by the first terms of their series expansions, namely,

$$\tanh x=x-\frac{x^3}{3}+\frac{2x^5}{15}\cdots,$$

and

$$\sinh x=x+\frac{x^3}{3!}+\frac{x^5}{5!}\cdots,$$

so that

$$z_a=Z_0\tanh\frac{\gamma l}{2}\cong\tfrac{1}{2}(R_a+j\omega L_a)\,l$$

and

$$\frac{1}{z_b}=\frac{1}{Z_0}\sinh\gamma l\cong(G_a+j\omega C_a)\,l. \tag{3.34}$$

The error incurred in making this approximation is a function of the elemental length l and the frequency, and is

$$\left(1-\frac{x}{\tanh x}\right) \text{ and } \left(1-\frac{x}{\sinh x}\right),$$

respectively. In constructing electrical analogs of the vocal tract it has been customary to use this approximation while keeping l sufficiently small. We shall return to this point later in the chapter.

We will presently apply the results of this section to some simplified analyses of the vocal tract. Before doing so, however, it is desirable to establish several fundamental relations for sound radiation from the mouth and for certain characteristics of the sources of vocal excitation.

[1] $\rho=1.14\times10^{-3}$ gm/cm^3 (moist air at body temperature, 37° C).
$c=3.5\times10^4$ cm/sec (moist air at body temperature, 37° C).
$\mu=1.86\times10^{-4}$ dyne-sec/cm^2 (20° C, 0.76 m. Hg).
$\lambda=0.055\times10^{-3}$ cal/cm-sec-deg (0° C).
$c_p=0.24$ cal/gm-degree (0° C, 1 atmos.).
$\eta=1.4$.

3.3. The Radiation Load at the Mouth and Nostrils

At frequencies where the transverse dimensions of the tract are small compared with a wavelength, the radiating area of the mouth or nose can be assumed to have a velocity distribution that is approximately uniform and cophasic. It can therefore be considered a vibrating surface, all parts of which move in phase. The radiating element is set in a baffle that is the head. To a rough approximation, the baffle is spherical and about 9 cm in radius for a man.

MORSE has derived the radiation load on a vibrating piston set in a spherical baffle and shows it to be a function of frequency and the relative sizes of the piston and sphere. The analytical expression for the load is involved and cannot be expressed in closed form. A limiting condition, however, is the case where the radius of the piston becomes small compared with that of the sphere. The radiation load then approaches that of a piston in an infinite, plane baffle. The latter is well known and can be expressed in closed form. In terms of the normalized acoustic impedance

$$z = Z_A \cdot \frac{A}{\rho c} = \frac{p}{U} \cdot \frac{A}{\rho c}$$

(that is, per-unit-free-space impedance), it is

$$z_p = \left[1 - \frac{J_1(2ka)}{ka}\right] + j\left[\frac{K_1(2ka)}{2(ka)^2}\right], \tag{3.35}$$

where $k = \omega/c$, a is the piston radius, A the piston area, $J_1(x)$ the first order Bessel function, and $K_1(x)$ a related Bessel function given by the series

$$K_1(x) = \frac{2}{\pi}\left[\frac{x^3}{3} - \frac{x^5}{3^2 \cdot 5} + \frac{x^7}{3^2 \cdot 5^2 \cdot 7} \cdots\right].$$

For small values of ka, the first terms of the Bessel functions are the most significant, and the normalized radiation impedance is approximately

$$z_p \cong \frac{(ka)^2}{2} + j\frac{8(ka)}{3\pi}; \quad ka \ll 1. \tag{3.36}$$

This impedance is a resistance proportional to ω^2 in series with an inductance of normalized value $8a/3\pi c$. The parallel circuit equivalent is a resistance of $128/9\pi^2$ in parallel with an inductance of $8a/3\pi c$.

By way of comparison, the normalized acoustic load on a vibrating sphere is also well known and is

$$z_s = \frac{jka}{1 + jka}, \tag{3.37}$$

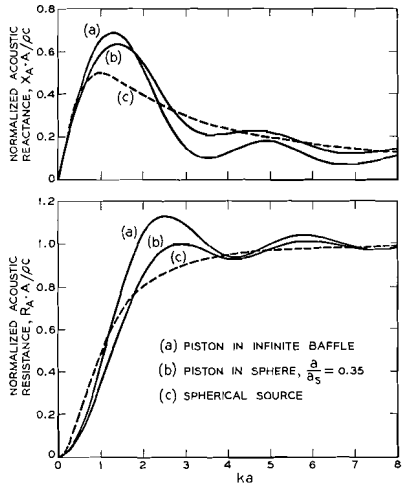


Fig. 3.6. Normalized acoustic radiation resistance and reactance for (a) circular piston in an infinite baffle; (b) circular piston in a spherical baffle whose radius is approximately three times that of the piston; (c) pulsating sphere. The radius of the radiator, whether circular or spherical, is a

where a is the radius of the sphere. Note that this is the parallel combination of a unit resistance and an a/c inductance. Again, for small ka,

$$z_s \cong (ka)^2 + j(ka); \quad ka \ll 1. \tag{3.38}$$

Using MORSE's results for the spherical baffle, a comparison of the real and imaginary parts of the radiation impedances for the piston-in-sphere, piston-in-wall, and pulsating sphere is made in Fig. 3.6. For the former, a piston-to-sphere radius ratio of $a/a_s = 0.35$ is illustrated. The piston-in-wall curves correspond to $a/a_s = 0$. For $ka < 1$, one notices that the reactive loads are very nearly the same for all three radiators. The real part for the spherical source is about twice that for the pistons.

These relations can be interpreted in terms of mouth dimensions. Consider typical extreme values of mouth area (smallest and largest) for vowel production. A man articulating a rounded vowel such as /u/ produces a mouth opening on the order of 0.9 cm^2. For an open vowel such as /a/ an area of 5.0 cm^2 is representative. The radii of circular pistons with these areas are 0.5 cm and 1.3 cm, respectively. For frequencies less than about 5000 cps, these radii place ka less than unity. If the head is approximated as a sphere of 9 cm radius, the ratios of piston-to-sphere radii for the extreme areas are 0.06 and 0.1, respectively. For these dimensions and frequencies, therefore, the radiation load on the mouth is not badly approximated by considering it to be the load on

a piston in an infinite wall. The approximation is even better for the nostrils whose radiating area is smaller. For higher frequencies and large mouth areas, the load is more precisely estimated from the piston-in-sphere relations. Notice, too, that approximating the normalized mouth-radiation load as that of a pulsating sphere leads to a radiation resistance that is about twice too high.

3.4. Spreading of Sound about the Head

In making acoustic analyses of the vocal tract one usually determines the volume current delivered to the radiation load at the mouth or nostrils. At these points the sound energy is radiated and spreads spatially. The sound is then received by the ear or by a microphone at some fixed point in space. It consequently is desirable to know the nature of the transmission from the mouth to the given point.

The preceding approximations for the radiation impedances do not necessarily imply how the sound spreads about the head. It is possible for changes in the baffling of a source to make large changes in the spatial distribution of sound and yet produce relatively small changes in the radiation load. For example, the piston-in-wall and piston-in-sphere were previously shown to be comparable assumptions for the radiation load. Sound radiated by the former is of course confined to the half-space, while that from the latter spreads spherically. The lobe structures are also spatially different.

One might expect that for frequencies where the wavelength is long compared with the head diameter, the head will not greatly influence the field. The spatial spreading of sound should be much like that produced by a simple spherical source of strength equal to the mouth volume velocity. At high frequencies, however, the diffraction about the head might be expected to influence the field.

A spherical source, pulsating sinusoidally, produces a particle velocity and sound pressure at r distance from its center equal respectively to

$$u(r)=\frac{a u_0}{r}\,\frac{jka}{1+jka}\,\frac{1+jkr}{jkr}\,e^{-jk(r-a)},$$

and

$$p(r)=\frac{\rho c a u_0}{r}\,\frac{jka}{1+jka}\,e^{-jk(r-a)}, \tag{3.39}$$

where a is the radius, u_0 is the velocity magnitude of the surface, and $k=\omega/c$. [Note the third factor in $u(r)$ accounts for the "bass-boost" that is obtained by talking close to a velocity microphone, a favorite artifice of nightclub singers.] If $ka\ll 1$, the source is a so-called simple

(point) source, and the sound pressure is

$$p(r)=\frac{j\omega\rho U_0}{4\pi r}\,e^{-jkr}, \tag{3.40}$$

where $U_0=4\pi a^2 u_0$ is the source strength or volume velocity. The simple source therefore produces a sound pressure that has spherical symmetry and an amplitude that is proportional to $1/r$ and to ω.

MORSE has derived the pressure distribution in the far field of a small vibrating piston set in a spherical baffle. Assuming that the mouth and

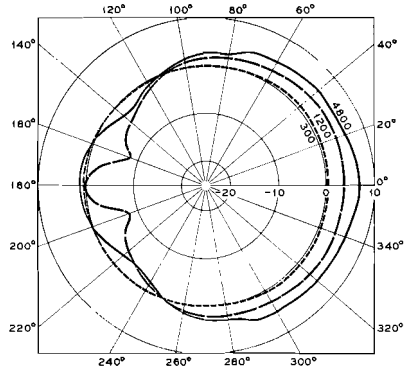


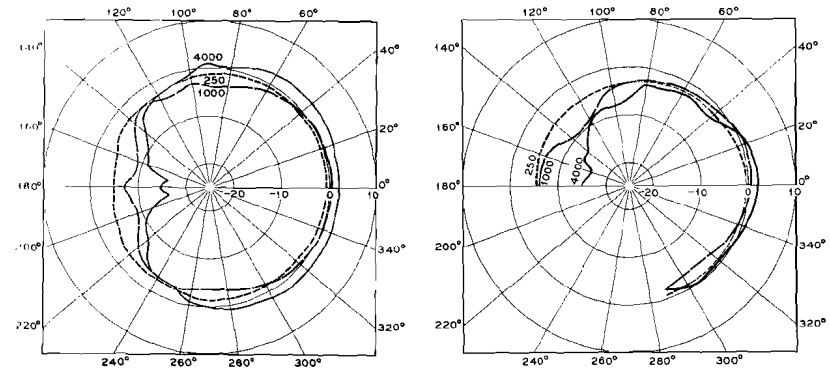
Fig. 3.7. Spatial distributions of sound pressure for a small piston in a sphere of 9 cm radius. Pressure is expressed in db relative to that produced by a simple spherical source of equal strength

head are approximately this configuration, with a 9 cm radius roughly appropriate for the sphere, the radiation pattern can be expressed relative to that which would be produced by a simple source of equal strength located at the same position. When this is done, the result is shown in Fig. 3.7. If the pressure field were identical to that of a simple spherical source, all the curves would fall on the zero db line of the polar plot. The patterns of Fig. 3.7 are symmetrical about the axis of the mouth (piston) which lies at zero degrees. One notices that on the mouth axis the high frequencies are emphasized slightly more than the +6 db/oct variation produced by the simple source (by about another +2 db/oct for frequencies greater than 300 cps). Also some lobing occurs, particularly at the rear of the "head".

The question can be raised as to how realistic is the spherical approximation of the real head. At least one series of measurements has been carried out to get a partial answer and to estimate spreading of sound about an average life-sized head (FLANAGAN, 1960a). A sound

Fig. 3.8. Life-size mannequin for measuring the relation between the mouth volume velocity and the sound pressure at an external point. The transducer is mounted in the mannequin's head

transducer was fitted into the head of the adult mannequin shown in Fig. 3.8. The transducer was calibrated to produce a known acoustic volume velocity at the lips of the dummy, and the amplitude and phase of the external pressure field were measured with a microphone. When the amplitudes are expressed relative to the levels which would be produced by a simple source of equal strength located at the mouth, the results for the horizontal and vertical planes through the mouth are shown in Fig. 3.9.



(a) (b)

Fig. 3.9 a and b. Distribution of sound pressure about the head, relative to the distribution for a simple source; (a) horizontal distribution for the mannequin; (b) vertical distribution for the mannequin

One notices that for frequencies up to 4000 cps, the pressures within vertical and horizontal angles of about ± 60 degrees, centered on the mouth axis, differ from the simple source levels by no more than ± 3 db. Simultaneous phase measurements show that within this same solid angle, centered on the mouth axis, the phase is within approximately 30 degrees of that for the simple source. Within these limits, then, the function relating the volume velocity through the mouth to the sound pressure in front of the mouth can be approximated as the simple source function of Eq. (3.40). Notice that $p(r)/U_0 \sim \omega$, and the relation has a spectral zero at zero frequency.

3.5. The Source for Voiced Sounds

3.51. Glottal Excitation

The nature of the vocal tract excitation for voiced sounds has been indicated qualitatively in Figs. 2.1 through 2.4. It is possible to be more quantitative about this mechanism and to estimate some of the acoustical properties of the glottal sound source. (The glottis, as pointed out earlier, is the orifice between the vocal cords.) Such estimates are based mainly upon a knowledge of the subglottal pressure, the glottal dimensions, and the time function of glottal area.

The principal physiological components of concern are illustrated schematically in Fig. 3.10. The diagram represents a front view of the subglottal system. The dimensions are roughly appropriate for an adult male (JUDSON and WEAVER). In terms of an electrical network, this system might be thought analogous to the circuit shown in Fig. 3.11.

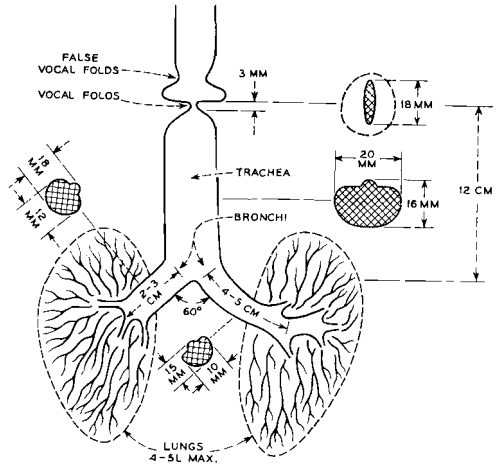


Fig. 3.10. Schematic diagram of the human subglottal system

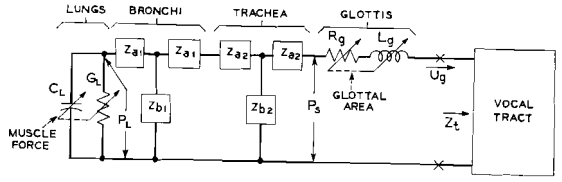


Fig. 3.11. An equivalent circuit for the subglottal system

A charge of air is drawn into the lungs and stored in their acoustic capacity C_L. The lungs are spongy tissues and exhibit an acoustic loss represented by the conductance G_L. The loss is a function of the state of inflation. The muscles of the rib cage apply force to the lungs, raise the lung pressure P_L, and cause air to be expelled—via the bronchi and trachea—through the relatively small vocal cord orifice. (Recall Fig. 3.1.) Because of their mass and elastic characteristics, the cords are set vibrating by the local pressure variations in the glottis. The quasi-periodic opening and closing of the cords varies the series impedance $(R_g+j\omega L_g)$ and modulates the air stream. The air passing into the vocal tract is therefore in the form of discrete puffs or pulses. As air is expelled, the rib-cage muscles contract and tend to maintain a constant lung pressure for a constant vocal effort. The lung capacity is therefore reduced so that the ratio of air charge to capacity remains roughly constant.

The bronchial and tracheal tubes—shown as equivalent T-sections in Fig. 3.11—are relatively large so that the pressure drop across them is small[1]. The subglottal pressure P_s and the lung pressure P_L are therefore nearly the same. The variable-area glottal orifice is the time-varying impedance across which most of the subglottic pressure is expended. The subglottal potential is effectively converted into kinetic energy in the form of the glottal volume velocity pulses, U_g.

For frequencies less than a couple of thousand cps, the main component of the glottal impedance is the resistive term. For many purposes in vocal tract analysis, it is convenient to have a small-signal (ac) equivalent circuit of the glottal resistance; that is, a Thevenin equivalent of the circuit to the left of the X's in Fig. 3.11. Toward deducing such an equivalent, let us consider the nature of the time-varying glottal impedance and some typical characteristics of glottal area and volume flow.

3.52. Glottal Impedance

To make an initial estimate of the glottal impedance, assume first that the ratio of the glottal inertance to resistance is small compared to the period of area variation (that is, the L_g/R_g time constant is small compared with the fundamental period, T). We will show presently the conditions under which this assumption is tenable. For such a case, the glottal volume flow may be considered as a series of consecutively established steady states, and relations for steady flow through an orifice can be used to estimate the glottal resistance.

Flow through the vocal cord orifice in Fig. 3.10 can be approximated as steady, incompressible flow through the circular orifice shown in Fig. 3.12. The subglottal and supraglottal pressures are P_1 and P_2, respectively. The particle velocity in the port is u, the orifice area is A and its depth (thickness) is d. If the cross-sectional areas of the adjacent tubes are much larger than A, variations in P_1 and P_2 caused by the flow

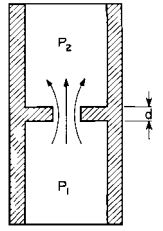


Fig. 3.12. Simple orifice approximation to the human glottis

[1] The branching bronchi are represented as a single tube having a cross-sectional area equal to the sum of the areas of the branches.

are small, and the pressures can be assumed sensibly constant. Also, if the dimensions of the orifice are small compared with the wavelength of an acoustic disturbance, and if the mean flow is much smaller than the speed of sound, an acoustic disturbance is known essentially instantaneously throughout the vicinity of the orifice, and incompressibility is a valid assumption. Further, let it be assumed that the velocity distribution over the port is uniform and that there is no viscous dissipation.

Under these conditions, the kinetic energy per-unit-volume possessed by the air in the orifice is developed by the pressure difference (P_1-P_2) and is

$$(P_1-P_2)=\frac{\rho u^2}{2}. \tag{3.41}$$

The particle velocity is therefore

$$u=\left[\frac{2(P_1-P_2)}{\rho}\right]^{\frac{1}{2}}. \tag{3.42}$$

We can define an orifice resistance, R_g^*, as the ratio of pressure drop to volume flow

$$R_g^*=\frac{\rho u}{2A}=\frac{\rho U}{2A^2}, \tag{3.43}$$

where $U=u\cdot A$ is the volume velocity. In practice, P_2 is essentially atmospheric pressure, so that $(P_1-P_2)=P_s$, the excess subglottal pressure, and

$$R_g^*=\frac{(2\rho P_s)^{\frac{1}{2}}}{2A}. \tag{3.44}$$

In situations more nearly analogous to glottal operation, the assumptions of uniform velocity distribution across the orifice and negligible viscous losses are not good. The velocity profile is generally not uniform, and the streamlines are not straight and parallel. There is a contraction of the jet a short distance downstream where the distribution is uniform and the streamlines become parallel (vena contracta). The effect is to reduce the effective area of the orifice and to increase R_g^*. Also, the pressure-to-kinetic energy conversion is never accomplished without viscous loss, and the particle velocity is actually somewhat less than that given in (3.42). In fact, if the area and flow velocity are sufficiently small, the discharge is actually governed by viscous laws. This can certainly obtain in the glottis where the area of opening can go to zero. Therefore, an expression for orifice resistance—valid also for small velocities and areas—might, as a first approximation, be a linear combination of kinetic and viscous terms

$$R_g=R_v+k\left(\frac{\rho U}{2A^2}\right), \tag{3.45}$$

where R_v is a viscous resistance and k is a real constant. For steady laminar flow, R_v is proportional to the coefficient of viscosity and the length of the conducting passage, and is inversely proportional to a function of area.

To find approximations of the form (3.45), WEGEL and VAN DEN BERG *et al.* have made steady-flow measurements on models of the human larynx. Both investigations give empirical formulas which agree in order of magnitude. VAN DEN BERG'S data are somewhat more extensive and were made on plaster casts of a normal larynx. The glottis was idealized as a rectangular slit as shown in Fig. 3.13. The length, l, of the slit was maintained constant at 18 mm, and its depth, d, was

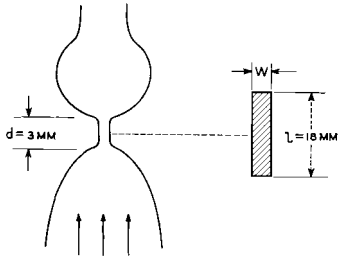


Fig. 3.13. Model of the human glottis. (After VAN DEN BERG *et al.*)

maintained at 3 mm. Changes in area were made by changing the width, w. Measurements on the model show the resistance to be approximately

$$R_g=\frac{P_s}{U}=\frac{12\mu d}{lw^3}+0.875\,\frac{\rho U}{2(lw)^2}, \tag{3.46}$$

where μ is the coefficient of viscosity. According to VAN DEN BERG, (3.46) holds within ten per cent for $0.1\leqq w\leqq 2.0$ mm, for $P_s\leqq 64$ cm H_2O at small w, and for $U\leqq 2000$ cc/sec at large w. As (3.46) implies, values of P_s and A specify the volume flow, U.

The glottal area is $A=lw$ so that the viscous (first) term of (3.46) is proportional to A^{-3}. The kinetic (second) term is proportional to uA^{-1} or, to the extent that u can be estimated from (3.42), it is approximately proportional to $P_s^{\frac{1}{2}}A^{-1}$. Whether the viscous or kinetic term predominates depends upon both A and P_s. They become approximately equal when $(\rho P_s)^{\frac{1}{2}}A^2=19.3\,\mu dl^2$. For typical values of vocal P_s, this equality occurs for glottal areas which generally are just a fraction (usually less than $\frac{1}{5}$) of the maximum area. In other words, over most of the open cycle of the vocal cords the glottal resistance is determined by the second term in (3.46).

As pointed out previously, (3.46) is strictly valid only for steady flow conditions. A relevant question is to what extent might (3.46) be applied in computing the glottal flow as a function of time when $A(t)$ and P_s are known. The question is equivalent to inquiring into the influence of the inertance of the glottal air plug. Because the pressure drop across the bronchi and trachea is small, and because P_s is maintained sensibly constant over the duration of several pitch periods by the low-impedance lung reservoir[1], the circuit of Fig. 3.11 can, for the present purpose, be simplified to that shown in Fig. 3.14. Furthermore, it is possible to show that at most frequencies the driving point impedance of the vocal

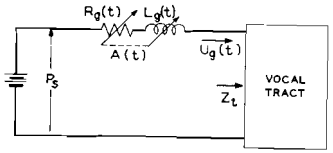


Fig. 3.14. Simplified circuit for the glottal source

tract, Z_t, is small compared with the glottal impedance. If the idealization $Z_t=0$ is made, then $U_g(t)$ satisfies

$$U_g(t)\,R_g(t)+\frac{d}{dt}\left[L_g(t)\,U_g(t)\right]=P_s, \tag{3.47}$$

where Eq. (3.46) can be taken as the approximation to $R_g(t)$ and, neglecting end corrections, $L_g(t)=\rho d/A(t)$.

Because R_g is a flow-dependent quantity, Eq. (3.47) is a nonlinear, first-order differential equation with nonconstant coefficients. For an arbitrary $A(t)$, it is not easily integrated. However, a simplification in the area function provides some insight into the glottal flow. Consider that $A(t)$ is a step function so that

$$A(t)=A_0; \quad t\geqq 0$$
$$=0; \quad t<0, \quad \text{and} \quad U_g(0)=0.$$

Then dL_g/dt is zero for $t>0$, and the circuit acts as a flow-dependent resistance in series with a constant inductance. A step of voltage (P_s) is applied at $t=0$. The behavior of the circuit is therefore described by

$$\frac{dU_g}{dt}=\frac{1}{L_g}(P_s-R_g\,U_g). \tag{3.48}$$

[1] VAN DEN BERG *et al.* estimate the variation to be less than five per cent of the mean subglottal pressure. P_s was measured by catheters inserted in the trachea and esophagus.

At $t=0$, $U_g(0)=0$ and

$$\left.\frac{dU_g}{dt}\right|_{t=0}=\frac{P_s}{L_g},$$

so that initially

$$U_g(t)\cong\frac{P_s}{L_g}\,t \quad \text{(for positive } t \text{ near zero)}.$$

Similarly, at $t=\infty$, $dU_g/dt=0$ and $U_g(\infty)=P_s/R_g$. The value of $U_g(\infty)$ is the steady-flow value which is conditioned solely by R_g. In this case U_g is the solution of $P_s-U_g R_g=0$, and is the positive root of a second-degree polynominal in U_g.

A time constant of a sort can be estimated from these asymptotic values of the flow build-up. Assume that the build-up continues at the initial rate, P_s/L_g, until the steady-state value $U_g(\infty)$ is achieved. The time, T, necessary to achieve the build-up is then

$$U_g(t)=\frac{P_s}{L_g}\,T=U_g(\infty)=\frac{P_s}{R_g},$$

or

$$T=\frac{L_g}{R_g}. \tag{3.49}$$

Since R_g is a sum of viscous and kinetic terms R_v and R_k, respectively, the time constant $L_g/(R_v+R_k)$ is smaller than the smaller of L_g/R_v and L_g/R_k. If the step function of area were small, R_v would dominate and the L_g/R_v time constant, which is proportional to A^2, would be more nearly appropriate. If the area step were large, the L_g/R_k constant would apply. In this case, and to the extent that R_v might be neglected [i.e., to the extent that R_g might be approximated as $R_k=0.875(2\rho P_s)^{\frac{1}{2}}/2A$], the L_g/R_k constant is proportional to $P_s^{-\frac{1}{2}}$ and is independent of A.

On the basis of these assumptions, a plot of the factors L_g/R_v and L_g/R_k is given in Fig. 3.15. Two values of P_s are shown for L_g/R_k, namely 4 cm H_2O and 16 cm H_2O. The first is about the minimum (liminal) intensity at which an adult male can utter a vowel. The latter corresponds to a fairly loud, usually high-pitched utterance. The value of L_g/R_g is therefore less than the solid curves of Fig. 3.15.

The curves of Fig. 3.15 show the greatest value of the time constant (i.e., for liminal subglottic pressure) to be of the order of a quarter millisecond. This time might be considered negligible compared with a fundamental vocal cord period an order of magnitude greater, that is, 2.5 msec. The latter corresponds to a fundamental vocal frequency of 400 cps which is above the average pitch range for a man's voice. To a first order approximation, therefore, the waveform of glottal volume velocity can be estimated from P_s and $A(t)$ simply by applying (3.46).

Notice also from the preceding results that for $L_g/R_g \cong 0.25$ msec (i.e., $P_s \cong 4$ cm H_2O) the inductive reactance becomes comparable to the resistance for frequencies between 600 and 700 cps. For $P_s = 16$ cm H_2O, the critical frequency is about doubled, to around 1300 cps. This suggests that for frequencies generally greater than about 1000 to 2000 cps, the glottal impedance may exhibit a significant frequency-proportional term, and the spectrum of the glottal volume flow may reflect the influence of this factor.

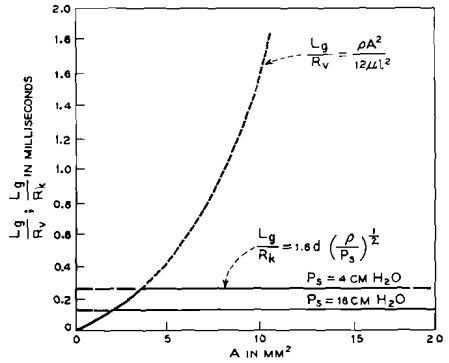


Fig. 3.15. Ratios of glottal inertance (L_g) to viscous and kinetic resistance (R_v, R_k) as a function of glottal area (A)

If the effects of inertance are neglected, a rough estimate of the glottal volume velocity can be made from the resistance expression (3.46). Assuming constant subglottal pressure, the corresponding volume velocity is seen to be proportional to A^3 at small glottal areas and to A at larger areas. Typical volume velocity waves deduced in this manner for a man are shown in Fig. 3.16 (FLANAGAN, 1958). The area waves are measured from high speed motion pictures of the glottis (see Fig. 2.3 in Chapter 2), and the subglottal pressure is estimated from the sound intensity and direct tracheal pressure measurements. The first condition is for the vowel /æ/ uttered at the lowest intensity and pitch possible. The second is for the same sound at a louder intensity and the same pitch. In the first case the glottis never completely closes. This is characteristic of weak, voiced utterances. Note that the viscous term in R_g operates to sharpen the leading and trailing edges of the velocity wave. This effect acts to increase the amplitude of the high-frequency components in the glottal spectrum.

The spectrum of the glottal volume flow is generally irregular and is characterized by numerous minima, or spectral zeros. For example,

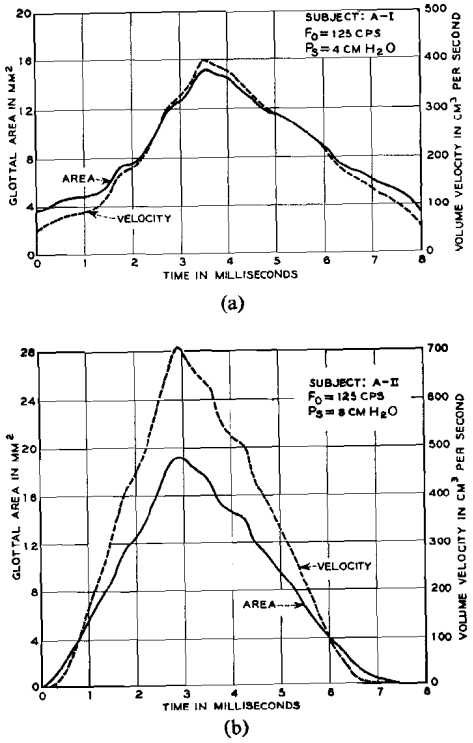


Fig. 3.16a and b. Glottal area and computed volume velocity waves for single vocal periods. F_0 is the fundamental frequency: P_s is the subglottal pressure. The subject is an adult male phonating /æ/. (After FLANAGAN, 1958)

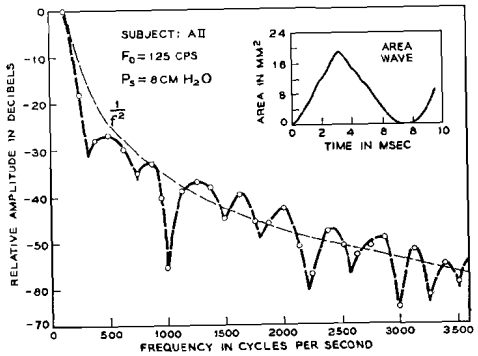


Fig. 3.17. Calculated amplitude spectrum for the glottal area wave AII shown in Fig. 3.16. (After FLANAGAN, 1961b)

if the wave in Fig. 3.16b were idealized as a symmetrical triangle, its spectrum would be of the form $(\sin x/x)^2$ with double-order spectral zeros occurring for $\omega = 4n\pi/\tau_0$, where n is an integer and τ_0 is the open time of the glottis. If the actual area wave of Fig. 3.16b is treated as periodic with period 1/125 sec, and its Fourier spectrum computed (most conveniently on a digital computer), the result is shown in Fig. 3.17 (FLANAGAN, 1961b). The slight asymmetry of the area wave causes the spectral zeros to lie at complex frequencies, so that the spectral minima are neither equally spaced nor as pronounced as for the symmetrical triangle.

3.53. Small-Signal Equivalent Source for the Glottis

Considering only the resistance R_g, given in Eq. (3.46), it is possible to approximate an ac or small-signal equivalent source for the glottal source. Such a specification essentially permits the source impedance to be represented by a time-invariant quantity and is useful in performing vocal tract calculations. The Thevenin (or Norton) equivalent generator for the glottis can be obtained in the same manner that the ac equivalent circuit for an electronic amplifier is derived. According to (3.46)

$$U_g(t) = f(P_s, A).$$

The glottal volume velocity, area and subglottic pressure are unipolar time functions. Each has a varying component superposed upon a mean value. That is,

$$U_g(t) = U_{g0} + U'(t)$$
$$A(t) = A_0 + A'(t)$$
$$P_s(t) = P_{s0} + P'_s(t).$$

Expanding $U_g(t)$ as a Taylor series about (P_{s0}, A_0) and taking first terms gives

$$U_g(P_s, A) = U_g(P_{s0}, A_0) + \left.\frac{\partial U_g}{\partial P_s}\right|_{P_{s0}, A_0} (P_s - P_{s0}) + \left.\frac{\partial U_g}{\partial A}\right|_{P_{s0}, A_0} (A - A_0) + \cdots,$$
$$= U_{g0} + U'_g(t),$$

and

$$U'_g(t) = \left.\frac{\partial U_g}{\partial P_s}\right|_{P_{s0}, A_0} P'_s + \left.\frac{\partial U_g}{\partial A}\right|_{P_{s0}, A_0} A'(t). \tag{3.50}$$

One can interpret (3.50) as an ac volume velocity (current) source of value $\partial U_g/\partial A|_{P_{s0}, A_0} A'(t)$ with an inherent conductance $\partial U_g/\partial P_s|_{P_{s0}, A_0}$. The source delivers the ac volume current $U'_g(t)$ to its terminals. The source configuration is illustrated in Fig. 3.18. The instantaneous polarity of $P'_s(t)$ is reckoned as the pressure beneath the glottis relative to that above.

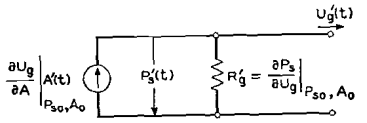


Fig. 3.18. Small-signal equivalent circuit for the glottal source. (After FLANAGAN, 1958)

The partials in (3.50) can be evaluated from (3.46). Let

$$R'_g = \left.\frac{\partial P_s}{\partial U_g}\right|_{P_{s0}, A_0}.$$

Then

$$\frac{\partial P_s}{\partial U_g} = R_g + U_g \frac{\partial R_g}{\partial U_g},$$

and

$$R'_g = (R_v + 2R_k)_{P_{s0}, A_0}. \tag{3.51}$$

The magnitude of the equivalent velocity source is simply

$$\left.\frac{\partial U_g}{\partial A}\right|_{P_{s0}, A_0} A'(t) = \left[u + A\frac{\partial u}{\partial A}\right]_{P_{s0}, A_0} A'(t).$$

Neglecting the viscous component of the resistance, Eq. (3.42) may be used to approximate u, in which case $\partial u/\partial A = 0$ and

$$\left.\frac{\partial U_g}{\partial A}\right|_{P_{s0}, A_0} \cong \left(\frac{2P_{s0}}{\rho}\right)^{\frac{1}{2}} A'(t). \tag{3.52}$$

The approximations (3.51) and (3.52) therefore suggest that the ac resistance of the glottal source is equal the viscous (first) term of (3.46) plus twice the kinetic (second) term, and that the ac volume current source has a waveform similar to the time-varying component of $A(t)$. To consider a typical value of R'_g, take $P_{s0} = 10$ cm H_2O and $A_0 = 5$ mm^2. For these commonly encountered values R'_g is computed to be approximately 100 cgs acoustic ohms. This source impedance can be compared with typical values of the acoustic impedance looking into the vocal tract (i.e., the tract driving point impedance). Such a comparison affords an insight into whether the glottal source acts more nearly as a constant current (velocity) generator or a voltage (pressure) source.

The driving point impedance of the tract is highly dependent upon vocal configuration, but it can be easily estimated for the unconstricted shape. Consider the tract as a uniform pipe, 17 cm long and open at the far end. Assuming no nasal coupling, the tract is terminated only by the mouth radiation impedance. The situation is illustrated in Fig. 3.19.

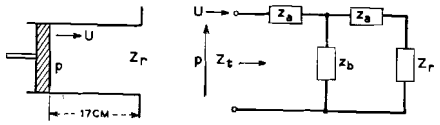


Fig. 3.19. Simplified representation of the impedance looking into the vocal tract at the glottis

Using the transmission line relations developed earlier in the chapter, the impedance Z_t looking into the straight pipe is

$$Z_t = Z_0 \frac{Z_r \cosh \gamma l + Z_0 \sinh \gamma l}{Z_0 \cosh \gamma l + Z_r \sinh \gamma l}, \tag{3.53}$$

where $l = 17$ cm, and the other quantities have been previously defined. If for a rough estimate the pipe is considered lossless, $\gamma = j\beta$ and (3.53) can be written in circular functions

$$Z_t = Z_0 \frac{Z_r \cos \beta l + j Z_0 \sin \beta l}{Z_0 \cos \beta l + j Z_r \sin \beta l}, \tag{3.54}$$

where $Z_0 = \rho c/A$, $\beta = \omega/c$. The maxima of Z_t will occur at frequencies where $l = (2n+1)\,\lambda/4$, so that $\beta l = (2n+1)\,\pi/2$ and $\cos \beta l = 0$. The maxima of Z_t for the lossless pipe are therefore

$$Z_{t_{max}} = Z_0^2/Z_r, \tag{3.55}$$

and the pipe acts as a quarter-wave transformer. The minima, on the other hand, are $Z_{t_{min}} = Z_r$ and the pipe acts as a half-wave transformer.

To estimate $Z_{t_{max}}$, we can use the radiation impedance for the piston in the infinite baffle, developed earlier in the chapter [see Eq. (3.36)].

$$Z_r = z_p \frac{\rho c}{A} = \frac{\rho c}{A}\left[\frac{(ka)^2}{2} + j\frac{8}{3\pi}(ka)\right], \tag{3.56}$$

where

$$a = \sqrt{A/\pi}, \quad \text{and} \quad ka \ll 1.$$

As a reasonable area for the unconstricted tract, take $A = 5\,\text{cm}^2$. The first quarter-wave resonance for the 17 cm long pipe occurs at a frequency of about 500 cps. At this frequency

$$Z_r|_{500\,\text{cps}} = (0.18 + j0.81), \quad \text{and} \quad Z_{t_{max}}|_{500\,\text{cps}} = \frac{(\rho c/A)^2}{Z_r} = 86/\underline{-77^\circ}$$

cgs acoustic ohms. This driving point impedance is comparable in size to the ac equivalent resistance of the glottal source just determined.

As frequency increases, the magnitude of Z_r increases, and the load reflected to the glottis at the quarter-wave resonances becomes smaller.

At the second resonance, for example, $Z_r|_{1500\,\text{cps}} = (1.63 + j\,2.44)$ and $Z_{t_{max}}|_{1500\,\text{cps}} = 24/\underline{-56^\circ}$ cgs acoustic ohms. The reflected impedance continues to diminish with frequency until at very high frequencies $Z_t = Z_0 = 8.4$ cgs acoustic ohms. Note, too, that at the half-wave resonances of the tract, i.e., $l = n\lambda/2$, the sine terms in (3.54) are zero and $Z_t = Z_r$.

The input impedance of the tract is greatest therefore at the frequency of the first quarter-wave resonance (which corresponds to the first formant). At and in the vicinity of this frequency, the driving point impedance (neglecting all losses except radiation) is comparable to the ac resistance of the glottal source. At all other frequencies it is less. For the unconstricted pipe the reflected impedance maxima are capacitive because the radiation load is inductive. To a first approximation, then, the glottal source appears as a constant volume velocity (current) source except at frequencies proximate to the first formant. As previously discussed, the equivalent vocal cord source sends an ac current equal to $u \cdot A'(t)$ into Z_t in parallel with R_g'. So long as constrictions do not become small, changes in the tract configuration generally do not greatly influence the operation of the vocal cords. At and near the frequency of the first formant, however, some interaction of source and tract might be expected, and in fact does occur. Pitch-synchronous variations in the tuning and the damping of the first formant – owing to significant tract-source interaction – can be observed experimentally[1].

3.6. The Source for Noise and Transient Excitation of the Tract

Our present knowledge of the mechanism and properties of noise and transient excitation of the vocal tract is considerably less than our understanding of voiced excitation. Not least among the reasons are the difficulties connected with direct measurement of the tract configuration, the size of constrictions, the spectral properties and inherent impedance of the source, and its spatial distribution. Noise excitation is generated by the air stream at a constriction. The resulting rotational flow and eddies produce a sound pressure which is largely random. The sound /ʃ/, for example, is produced by forcing air through the narrow constriction between the tongue and the roof of the mouth. Turbulent flow can also be generated by directing an air jet across an obstacle or sharp edge. The upper teeth serve this purpose in the production of

[1] The acoustic mechanism of vocal-cord vibration and the interactions between source and system are discussed in more detail later. An acoustic oscillator model of the cords is derived in Chapter VI and a computer simulation of the model is described.

dental fricatives such as /f/. One fricative consonant, /h/, is produced by turbulent flow generated at the glottis. The excitation mechanism is similar to that for the front-excited fricatives except the nonvibrating vocal cords create the constriction.

Stop consonants are produced by making a complete closure at an appropriate point (labial, dental or palatal), building up a pressure behind the occlusion, and sharply releasing the pressure by an abrupt opening of the constriction. This excitation is therefore similar to exciting an electrical network with a step function of voltage. The stop explosion is frequently followed by a fricative excitation. This latter element of the stop is similar to a brief fricative continuant of the same articulation.

Because it is spatially distributed, the location of the noise source in the tract is difficult to fix precisely. Generally it can be located at the constriction for a short closure, and just anterior to a longer constriction. In terms of a network representation, the noise source and its inherent impedance can be represented as the series elements in Fig. 3.20. P_s is the sound pressure generated by the turbulent flow and Z_s is the inherent impedance of the source. The series connection of the source can be qualitatively justified by noting that a shunt connection of a low-impedance pressure source would alter the mode structure of the vocal network. Furthermore, experimentally measured mode patterns for consonants appear to correspond to the series connection of the exciting source (FANT, 1960).

Although the spectral characteristics and inherent impedance of the noise source are not well known, estimates of these quantities can be made from a knowledge of the sound output and the tract configuration, and from measurements on tube models (HEINZ, 1958). Data obtained in this manner suggest that the spectrum is relatively flat in the mid-audio frequency range and that the source impedance is largely resistive. In fact, the relations for orifice resistance developed in the previous section appear to give reasonable estimates for the inherent impedance.

Voiced fricative sounds, such as /v/, are produced by simultaneous operation of the glottal and turbulent sources. Because the vibrating vocal cords cause a pulsive flow of air, the turbulent sound generated at the constriction is modulated by the glottal puffs. The turbulent sound is therefore generated as pitch-synchronous bursts of noise.

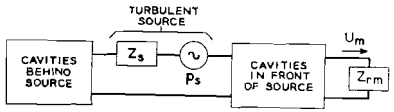


Fig. 3.20. Equivalent circuit for noise excitation of the vocal tract

It is possible to be a little more quantitative about several aspects of fricative and stop excitation. For example, MEYER-EPPLER (1953) has carried out measurements on fricative generation in constricted plastic tube models of the vocal tract. He has related these measurements to human production of the fricative consonants /f, s, ʃ/. For these vocal geometries a critical Reynold's number, R_{ec}, apparently exists below which negligible turbulent sound is produced. MEYER-EPPLER found that the magnitude of the noise sound pressure P_r—measured at a distance r from the mouth of either the model or the human—is approximately described by

$$P_r = K(R_e^2 - R_{ec}^2), \tag{3.57}$$

where K is a constant, R_e is the dimensionless Reynold's number $R_e = uw\rho/\mu$ and, as before, u is the particle velocity, ρ the air density, μ the coefficient of viscosity and w the effective width of the passage.

We recall from the earlier discussion [Eq. (3.41)] that for turbulent flow at a constriction the pressure drop across the orifice is approximately $P_d = \rho u^2/2 = \rho U^2/2A^2$. Therefore, $R_e^2 = 2\rho(w/\mu)^2 P_d$ and (3.57) can be written

$$P_r = (K_1 w^2 P_d - K_2); \quad P_r \geqq 0, \tag{3.58}$$

where K_1 and K_2 are constants. This result indicates that, above some threshold value, the fricative sound pressure in front of the mouth is proportional to the pressure drop at the constriction (essentially the excess pressure behind the occlusion) and to the square of the effective width of the passage.

By way of illustrating typical flow velocities associated with consonant production, a constriction area of 0.2 cm^2 and an excess pressure of 10 cm H_2O are not unusual for a fricative like /s/. The particle velocity corresponding to this pressure is $u = (2P_d/\rho)^{\frac{1}{2}} \cong 4100$ cm/sec[1] and the volume flow is $U \cong 820\ cm^3/sec$.

If the constricted vocal passage is progressively opened and the width increased, a constant excess pressure can be maintained behind the constriction only at the expense of increased air flow. The flow must be proportional to the constriction area. The power associated with the flow is essentially $P_d U$ and hence also increases. Since the driving power is derived from the expiratory muscles, their power capabilities determine the maximum flow that can be produced for a given P_d. At some value of constriction area, a further increase in area, and consequently in w, is offset by a diminution of the P_d that can be maintained. The product $w^2 P_d$ in (3.58) then begins to decrease and so does the intensity of the fricative sound.

[1] Note this velocity is in excess of 0.1 Mach!

Voiceless stop consonants contrast with fricatives in that they are more transient. For strongly articulated stops, the glottis is held open so that the subglottal system contributes to the already substantial volume behind the closure (V_B). The respiratory muscles apply a force sufficient to build up the pressure, but do not contract appreciably to force air out during the stop release. The air flow during the initial part of the stop release is mainly turbulent, with laminar streaming obtaining as the flow decays. In voiced stops in word-initial position (for example /d, g/), voicing usually commences following the release, but often (for example, in /b/) can be initiated before the release.

In very crude terms, stop production can be considered analogous to the circuit of Fig. 3.21. The capacitor C_B is the compliance ($V_B/\rho c^2$)

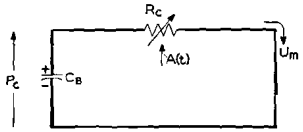


Fig. 3.21. Approximate vocal relations for stop consonant production

of the cavities back of the closure and is charged to the excess pressure P_c. The resistance R_c is that of the constriction and is, according to the previous discussion [Eq. (3.43)], approximately $R_c=\rho U_m/2A^2$. Suppose the constriction area is changed from zero as a step function, that is,

$$\begin{aligned} A(t)&=0; \quad t<0 \\ &=A; \quad t\geqq 0. \end{aligned}$$

The mouth volume current then satisfies

$$U_m R_c+\frac{1}{C_B}\int_0^t U_m\,dt=P_c,$$

or

$$\frac{\rho U_m^2}{2A^2}+\frac{1}{C_B}\int_0^t U_m\,dt=P_c, \quad \text{for } U_m>0,$$

and the solution for positive values of U_m is

$$U_m(t)=\left(\frac{2P_c}{\rho}\right)^{\frac{1}{2}} A\left[1-\frac{At}{C_B(\rho 2P_c)^{\frac{1}{2}}}\right]. \tag{3.59}$$

According to (3.59) the flow diminishes linearly with time during the initial phases of the stop release. At the indicated rate, the time to deplete the air charge would be

$$t_1=\frac{C_B(\rho 2P_c)^{\frac{1}{2}}}{A}. \tag{3.60}$$

As the flow velocity becomes small, however, the tendency is toward laminar streaming, and the resistance becomes less velocity dependent [see first term in Eq. (3.46)]. The flow decay then becomes more nearly exponential[1].

[1] This can be seen exactly by letting R_c include a constant (viscous) term as well as a flow-dependent term. Although the differential equation is somewhat more complicated, the variables separate, and the solution can be written in terms of U_m and $\ln U_m$.

Let

$$R_c=r_v A^{-3}(t)+r_k A^{-2}(t)|U_m|,$$

where r_v and r_k are constants involving air density and viscosity [as described in Eq. (3.46)]. If the constriction area is changed stepwise from zero to A at time zero, the resulting flow will again be unipolar and now will satisfy

$$(r_k/A^2)\,U_m^2+(r_v/A^3)\,U_m+1/C_B\int_0^t U_m\,dt=P_C.$$

The variables in this equation are separable and the solution can be obtained by differentiating both sides with respect to time. This yields

$$\frac{r_v}{A^3}\left(\frac{dU_m}{dt}\right)+2\frac{r_k}{A^2}U_m\frac{dU_m}{dt}+\frac{U_m}{C_B}=0$$

and

$$\frac{r_v C_B}{A^3}\left(\frac{dU_m}{U_m}\right)+2\frac{r_k C_B}{A^2}\,dU_m=-dt.$$

Integrating termwise gives

$$\frac{r_v C_B}{A^3}\ln U_m\Big]_0^t+2\frac{r_k C_B}{A^2}U_m\Big]_0^t=-t.$$

At $t=0$, $U_m=U_0$, where U_0 is the positive real root of the quadratic

$$\left(\frac{r_k}{A^2}\right)U_0^2+\frac{r_v}{A^3}U_0-P_C=0.$$

Then

$$\ln\left(\frac{U_m}{U_0}\right)+\frac{2r_k A}{r_v}(U_m-U_0)+\frac{tA^3}{r_v C_B}=0.$$

Note

for A large: $U_m\approx\left[U_0-\left(\frac{A^2}{2r_k C_B}\right)t\right]$

for A small: $U_m\approx U_0 e^{-\left(\frac{A^3}{r_v C_B}\right)t}$.

It also follows that

$$\begin{aligned}\frac{dU_m}{dt}&=\frac{-U_m}{\frac{r_v C_B}{A^3}+\frac{2r_k C_B}{A^2}U_m}\\ &\approx\frac{-A^2}{2r_k C_B}, \text{ for large } A\\ &\approx\frac{-U_m A^3}{r_v C_B}, \text{ for small } A.\end{aligned}$$

To fix some typical values, consider the production of a voiceless stop such as /t/. According to FANT (1960), realistic parameters for articulation of this sound are $P_c = 6$ cm H_2O, $V_B = \rho c^2 C_B = 4$ liters (including lungs) and $A = 0.1$ cm^2. Assuming the area changes abruptly, substitution of these values into (3.59) and (3.60) gives $U_m(0) = 320$ cm^3/sec and $t_1 = 130$ msec. The particle velocity at the beginning of the linear decay is $u_m(0) = 3200$ cm/sec. After 50 msec it has fallen to the value 1300 cm/sec which is about the lower limit suggested by MEYER-EPPLER for noise generation. As FANT points out, the amount of air consumed during this time is quite small, on the order of 10 cm^3.

Both STEVENS (1956) and FANT (1960) emphasize the importance of the open glottis in the production of a strong stop consonant. A closed glottis reduces V_B to something less than 100 cm^3, and the excess pressure which can be produced behind the constriction is typically on the order of 3 cm H_2O. For such conditions is it difficult to produce flows sufficient for noise generation. The turbulent noise produced during the stop release is essentially a secondary effect of the excitation. The primary excitation is the impact of the suddenly applied pressure upon the vocal system. As mentioned earlier, this excitation for an abrupt area change is analogous to a step function of voltage applied to an electrical circuit. Such a source is characterized by a spectrum which is proportional to $1/\omega$, or diminishes in amplitude at -6 db/oct.

3.7. Some Characteristics of Vocal Tract Transmission

Some of the fundamental relations developed in the foregoing sections can now be used to put in evidence certain properties of vocal transmission. These characteristics are easiest demonstrated analytically by highly simplifying the tract geometry. Calculations on detailed approximations are more conveniently done with computers. Although our examples generally will be oversimplified, the extensions to more exact descriptions will in most cases be obvious.

As a first step, consider the transmission from glottis to mouth for nonnasal sounds. Further, as an ultimate simplification, consider that the tract is uniform in cross section over its whole length l, is terminated in a radiation load whose magnitude is negligible compared with the characteristic impedance of the tract, and is driven at the glottis from a volume-velocity source whose internal impedance is large compared to the tract input impedance. The simple diagram in Fig. 3.22 represents this situation. The transmission function relating the mouth and glottal volume currents is then

$$\frac{U_m}{U_g} = \frac{z_b}{z_b + z_a} = \frac{1}{\cosh \gamma l}. \tag{3.61}$$

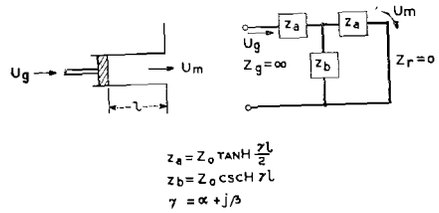


Fig. 3.22. Relation between glottal and mouth volume currents for the unconstricted tract. The glottal impedance is assumed infinite and the radiation impedance is zero

The normal modes (poles) of the transmission are the values of γl which make the denominator zero. These resonances produce spectral variations in the sound radiated from the mouth. They are

$$\begin{aligned} &\cosh \gamma l = 0 \\ &\gamma l = \pm j(2n+1)\frac{\pi}{2}, \quad n = 0, 1, 2, \ldots . \end{aligned} \tag{3.62}$$

The poles therefore occur at complex values of frequency. Letting $j\omega = \sigma + j\omega = s$, the complex frequency, and recalling from (3.8) that $\gamma = \alpha + j\beta$ and $\beta \cong \omega/c$ for small losses, the complex pole frequencies may be approximated as

$$s_n \cong -\alpha c \pm j\frac{(2n+1)\pi c}{2l}, \quad n = 0, 1, 2, \ldots^1. \tag{3.63}$$

The transmission (3.61) can be represented in factored form in terms of the roots of the denominator, namely

$$H(s) = \frac{U_m(s)}{U_g(s)} = \prod_n \frac{s_n s_n^*}{(s - s_n)(s - s_n^*)}, \tag{3.64}$$

where s_n^* is the complex conjugate of s_n, and the numerator is set to satisfy the condition

$$\left.\frac{U_m(j\omega)}{U_g(j\omega)}\right|_{j\omega = 0} = \frac{1}{\cosh \alpha l} \cong 1,$$

for small α. The transmission is therefore characterized by an infinite number of complex conjugate poles[2]. The manifestations of these normal modes as spectral peaks in the output sound are called *formants*. The

[1] Actually α is an implicit function of ω [see Eq. (3.33)]. However, since its frequency dependence is relatively small, and since usually $\sigma_n \ll \omega_n$, the approximation (3.63) is a convenient one.

[2] Rigorous justification of the form (3.64) has its basis in function theory (TITCHMARSH; AHLFORS). See Chapter VI, Sec. 6.22 for further discussion of this point.

transmission (3.64) exhibits no zeros at finite frequencies. Maxima occur in

$$|H(j\omega)| \quad \text{for } \omega = \pm(2n+1)\frac{\pi}{2}\frac{c}{l},$$

and the resonances have half-power cps bandwidths approximately equal to $\Delta f = \sigma/\pi = \alpha c/\pi$. For an adult male vocal tract, approximately 17 cm in length, the unconstricted resonant frequencies therefore fall at about $f_1 = 500$ cps, $f_2 = 1\,500$ cps, $f_3 = 2\,500$ cps, and continue in $c/2l$ increments.

In the present illustration the only losses taken into account are the classical heat conduction and viscous losses discussed earlier. A calculation of formant bandwidth on this basis alone will consequently be abnormally low. It is nevertheless instructive to note this contribution to the formant damping. Recall from Eq. (3.8) that for small losses

$$\alpha \cong \frac{R_a}{2}\sqrt{\frac{C_a}{L_a}} + \frac{G_a}{2}\sqrt{\frac{L_a}{C_a}},$$

where R_a, G_a, L_a and C_a have been given previously in Section (3.25). At the first-formant frequency for the unconstricted tract (i.e., 500 cps), and assuming a circular cross-section with typical area 5 cm^2, α is computed to be approximately 5.2×10^{-4}, giving a first-formant bandwidth $\Delta f_1 = 6$ cps. At the second formant frequency (i.e., 1 500 cps) the same computation gives $\Delta f_2 = 10$ cps. The losses increase as $f^{\frac{1}{2}}$, and at the third formant (2 500 cps) give $\Delta f_3 = 13$ cps.

It is also apparent from (3.64) that $H(s)$ is a minimum phase function (that is, it has all of its zeros, namely none, in the left half of the s-plane) so that its amplitude and phase responses are uniquely linked (that is, they are Hilbert transforms). Further, the function is completely specified by the s_n's, so that the frequency *and* amplitude of a formant peak in $|H(j\omega)|$ are uniquely described by the pole frequencies. In particular if the formant damping can be considered known and constant, then the amplitudes of the resonant peaks of $|H(j\omega)|$ are implicit in the imaginary parts of the formant frequencies $\omega_1, \omega_2, \ldots$, (FANT, 1956; FLANAGAN, 1957c). In fact, it follows from (3.61) that

$$\begin{aligned} |H(j\omega)|_{\omega=\omega_n} &= \frac{1}{|\cosh(\alpha + j\beta)l|_{\omega=\omega_n}} \\ &= \frac{1}{|j\sinh\alpha l|} \\ &\cong \frac{1}{\alpha l}, \end{aligned} \tag{3.65}$$

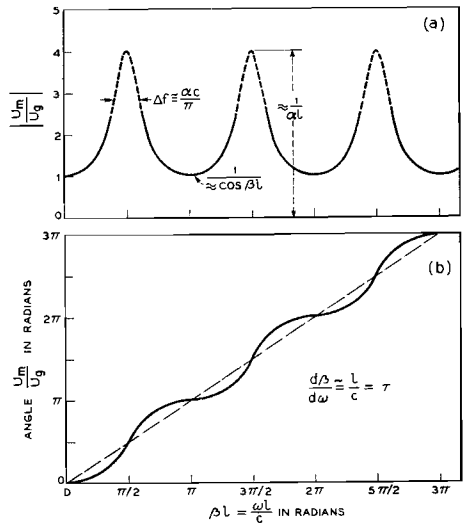


Fig. 3.23. Magnitude and phase of the glottis-to-mouth transmission for the vocal tract approximation shown in Fig. 3.22

where $\beta = \omega/c$ and $\omega_n = (2n+1)\,\pi c/2l$. Notice, too, that the phase angle of $H(j\omega)$ advances π radians in passing a formant frequency ω_n so the amplitude and phase response of $H(j\omega)$ appear as in Fig. 3.23. In the same connection, note that for the completely lossless case

$$H(j\omega) = \frac{1}{\cos\dfrac{\omega l}{c}}.$$

3.71. Effect of Radiation Load upon Mode Pattern

If the radiation load on the open end of the tube is taken into account, the equivalent circuit for the tube becomes that shown in Fig. 3.24. Here A_t is the cross-sectional area of the tract and A_m is the radiating area of the mouth with equivalent radius a_m. The thickness of the mouth constriction is assumed negligible, the glottal impedance is high, and

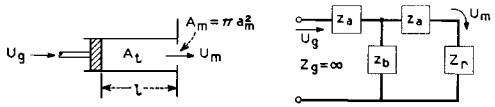


Fig. 3.24. Equivalent circuit for the unconstricted vocal tract taking into account the radiation load. The glottal impedance is assumed infinite

cross dimensions are small compared with a wavelength. The transmission from glottis to mouth is therefore

$$\frac{U_m}{U_g}=\frac{1}{\cosh\gamma l+\frac{Z_r}{Z_0}\sinh\gamma l},$$

or, more conveniently,

$$\frac{U_m}{U_g}=\frac{\cosh\gamma_r l}{\cosh(\gamma+\gamma_r)l}, \tag{3.66}$$

where $\gamma_r l=\tanh^{-1}Z_r/Z_0$. Note that for $Z_r\ll Z_0$, $\cosh\gamma_r l\cong 1$ and for low loss $Z_0\cong\rho c/A_t$.

By the transformation (3.66), the radiation impedance is carried into the propagation constant, so that

$$(\gamma+\gamma_r)=\left[\alpha+j\beta+\frac{1}{l}\tan^{-1}\frac{Z_r}{Z_0}\right]$$
$$=(\alpha+j\beta+\alpha_r+j\beta_r)=(\alpha'+j\beta')=\gamma'.$$

If the radiation load is taken as that on a piston in a wall [see Eq. (3.36) in Sec. 3.3] then

$$Z_r\cong\frac{\rho c}{A_m}\left[\frac{(ka)^2}{2}+j\frac{8ka}{3\pi}\right],\quad ka\ll 1, \tag{3.67}$$

where a equals the mouth radius a_m. Expanding $\tanh^{-1}Z_r/Z_0$ as a series and taking only the first term (i.e., assuming $Z_r\ll Z_0$) gives

$$\gamma_r\cong\frac{1}{l}\frac{A_t}{A_m}\left[\frac{(ka)^2}{2}+j\frac{8ka}{3\pi}\right]$$
$$=\alpha_r+j\beta_r. \tag{3.68}$$

For low loss $\beta\cong\omega/c=k$, so that

$$(\alpha'+j\beta')=\left[\alpha+\frac{A_t}{A_m}\frac{(\beta a)^2}{2l}\right]+j\beta\left[1+\frac{A_t}{A_m}\frac{8a}{3\pi l}\right]. \tag{3.69}$$

Again the poles of (3.66) occur for

$$e^{2\gamma' l}+1=0$$

or

$$\gamma'=\pm j\frac{(2n+1)\pi}{2l},\quad n=0,1,2,\ldots. \tag{3.70}$$

Letting $j\omega\to s=(\sigma+j\omega)$, and remembering that in general $\sigma_n\ll\omega_n$, the poles are approximately

$$s_{nr}\cong\frac{1}{1+\frac{A_t}{A_m}\frac{8a}{3\pi l}}\left[-\left(\alpha c+\frac{A_t\omega^2}{2\pi lc}\right)\pm j\frac{(2n+1)\pi c}{2l}\right], \tag{3.71}$$
$$n=0,1,2,\ldots\quad(Z_r\ll Z_0).$$

The general effect of the radiation, therefore, is to decrease the magnitude of the imaginary parts of the pole frequencies and to make their real parts more negative.

For the special case $A_m=A_t$, the modes are

$$s_{nr}\cong\left(\frac{3\pi l}{3\pi l+8a}\right)\left[-\left(\alpha c+\frac{a^2\omega^2}{2lc}\right)\pm j\frac{(2n+1)\pi c}{2l}\right]. \tag{3.72}$$

Using the values of the example in the previous section, $A_t=5\text{ cm}^2$, $l=17$ cm, the spectral resonances (formants) are lowered in frequency by the multiplying factor $3\pi l/(3\pi l+8a)=0.94$. The original 500 cps first formant is lowered to 470 cps, and the 1500 cps second formant is lowered to 1410 cps. The first formant bandwidth is increased to about $\Delta f_1\cong 0.94(6+4)=9$ cps, and the second formant bandwidth to about $\Delta f_2\cong 0.94(10+32)=40$ cps. The same computation for the third formant gives $\Delta f_3\cong 100$ cps. The latter figures begin to be representative of formant bandwidths measured on real vocal tracts with the glottis closed (HOUSE and STEVENS, 1958; DUNN, 1961; VAN DEN BERG, 1955). The contributions of the radiation, viscous and heat losses to Δf_1 are seen to be relatively small. Glottal loss and cavity wall vibration generally are more important contributors to the first formant damping.

As (3.72) indicates, the contribution of the radiation resistance to the formant damping increases as the square of frequency, while the classical heat conduction and viscous loss cause α to grow as $\omega^{\frac{1}{2}}$. The radiation reactance is inertive and causes the formant frequencies to be lowered. For $A_m=A_t$, Eq. (3.71) shows that the radiation reactance has the same effect as lengthening the vocal tract by an amount $(8a/3\pi)$.

3.72. Effect of Glottal Impedance upon Mode Pattern

The effect of the equivalent glottal impedance can be considered in much the same manner as the radiation load. To keep the illustration simple, again assume the radiation load to be negligible compared with the characteristic impedance of the uniform tract, but take the glottal impedance as finite. This situation is depicted by Fig. 3.25. Similar to

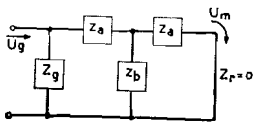


Fig. 3.25. Equivalent circuit for the unconstricted vocal tract assuming the glottal impedance to be finite and the radiation impedance to be zero

the previous instance, the volume velocity transmission function can be put in the form

$$\frac{U_m}{U_g}=\frac{1}{\frac{z_a}{Z_g}\left(\frac{Z_g}{z_b}+\frac{z_a}{z_b}+1\right)+1+\frac{z_a}{Z_g}}$$
$$=\frac{1}{\cosh\gamma l+\frac{Z_0}{Z_g}\sinh\gamma l} \qquad (3.73)$$
$$=\frac{\cosh\gamma_g l}{\cosh(\gamma+\gamma_g)l},$$

where $\gamma_g l=\tanh^{-1}Z_0/Z_g$, and the glottal impedance is transformed into the propagation constant. Again taking the first term of the series expansion for $\tanh^{-1}Z_0/Z_g$ (i.e., assumming $Z_g\gg Z_0$) gives

$$(\gamma+\gamma_g)\cong\left(\alpha+j\beta+\frac{1}{l}\frac{Z_0}{Z_g}\right).$$

The equivalent glottal impedance may be approximated as $Z_g=(R_g'+j\omega L_g)$, where R_g' is the ac equivalent resistance determined previously in Eq. (3.51), and L_g is the effective inductance of the glottal port. The zeros of the denominator of (3.73) are the poles of the transmission, and an argument similar to that used in the preceding section for low losses ($Z_0\cong\rho c/A_t$, $\beta\cong\omega/c$) leads to

$$s_{ng}\cong\frac{1}{1-\left(\frac{L_g Z_0 c}{l|Z_g|^2}\right)}\left\{-\left(\alpha c+\frac{R_g' Z_0 c}{l|Z_g|^2}\right)\pm j\frac{(2n+1)\pi c}{2l}\right\}. \qquad (3.74)$$

According to (3.74), the effect of the finite glottal impedance is to increase the damping of the formant resonances (owing to the glottal loss R_g') and to increase the formant frequencies by the factor multiplying the bracketed term (owing to the glottal inductance). A sample calculation of the effect can be made. As typical values, take a subglottic pressure (P_s) of 8 cm H_2O, a mean glottal area (A_0) of 5 mm^2, a glottal orifice thickness (d) of 3 mm, a vocal tract area (A_t) of 5 cm^2 and

a tract length (l) of 17 cm. For these conditions the glottal resistance, computed according to Eq. (3.51), is $R_g'\cong 91$ cgs acoustic ohms. The glottal inductance is $L_g=\sigma d/A_0=6.8\times10^{-3}$ cgs units. At about the frequency of the first formant, that is, $\omega\cong\pi c/2l=2\pi$ (500 cps), the multiplying factor has a value $1/(1-0.014)$, so that the first formant resonance is increased from its value for the infinite glottal impedance condition by about 1.4%. The effect of the glottal inductance upon formant tuning is greatest for the lowest formant because $|Z_g|$ increases with frequency. The same computation for the second formant ($\approx$1 500 cps) shows the multiplying factor to be $1/(1-0.010)$. One notices also that the effect of the multiplying term is to shorten the apparent length of the tract to

$$\left(l-\frac{L_g Z_0 c}{|Z_g|^2}\right).$$

The resonant bandwidth for the first formant is computed to be

$$\Delta f_1=\frac{1}{(1-0.014)}[6\text{ cps}+56\text{ cps}]=63\text{ cps},$$

which is reasonably representative of first formant bandwidths measured in real speech. The contribution of the glottal loss R_g' to formant damping is greatest for the lowest formant. It diminishes with increasing frequency because $|Z_g|$ grows with frequency. At the second formant frequency, the same calculation gives $\Delta f_2=(1/1-0.010)$ (10 cps+40 cps)=51 cps. One recalls, too, that the heat conduction and viscous losses (which specify α) increase as $\omega^{\frac{1}{2}}$, while the radiation loss increases as ω^2 (for $ka\ll1$). The lower-formant damping is therefore influenced more by glottal loss, and the higher-formant damping is influenced more by radiation loss.

In this same connection, one is reminded that the glottal resistance and inductance (used here as equivalent constant quantities) are actually time varying. There is consequently a pitch-synchronous modulation of the pole frequencies s_{ng} given in (3.74). That is, as the vocal cords open, the damping and resonant frequency of a formant increase, so that with each glottal period the pole frequency traverses a small locus in the complex-frequency plane. This pitch-synchronous change in formant damping and tuning can often be observed experimentally, particularly in inverse filtering of formants. It is most pronounced for the first formant.

3.73. Effect of Cavity Wall Vibration

The previous discussion has assumed the walls of the vocal tract to be smooth and rigid. The dissipative elements of concern are then the radiation resistance, the glottal resistance, and the viscous and heat

conduction losses at the cavity walls. The human vocal tract is of course not hard-walled, and its surface impedance is not infinite. The yielding walls can consequently contribute to the energy loss in the tract and can influence the mode tuning. We would like to estimate this effect.

The finite impedance of the tract wall constitutes an additional shunt path in the equivalent "T" (or π) section for the pipe (see Fig. 3.3). Because the flesh surrounding the tract is relatively massive and exhibits viscous loss, the additional shunt admittance for the frequency range of interest (i.e., speech frequencies) can be approximated as a per-unit-length reciprocal inductance or inertance ($\Gamma_w = 1/L_w$) and a per-unit-length conductance ($G_w = 1/R_w$) in parallel[1]. The modified equivalent "T" section is shown in Fig. 3.26.

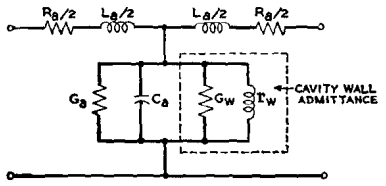


Fig. 3.26. Representation of wall impedance in the equivalent T-section for a length l of uniform pipe

Let us note the effect of the additional shunt admittance upon the propagation constant for the tube. As before, the basic assumption is that a plane wave is propagating in the pipe and that the sound pressure at any cross section is uniform and cophasic. Recall that

$$\gamma = \alpha + j\beta = \sqrt{yz},$$

where y and z are the per-unit-length shunt admittance and series impedance, respectively. The latter quantities are now

$$z = (R_a + j\omega L_a)$$

$$y = (G_a + G_w) + j\left(\omega C_a - \frac{\Gamma_w}{\omega}\right). \tag{3.75}$$

Again, most conditions of interest will be relatively small-loss situations for which

$$R_a \ll \omega L_a$$

[1] For describing the behavior at very low frequencies, a compliance element must also be considered.

and

$$(G_a + G_w) \ll \left(\omega C_a - \frac{\Gamma_w}{\omega}\right).$$

Also, in general, the susceptance of the air volume will exceed that of the walls and $\omega C_a \gg \Gamma_w/\omega$. Following the earlier discussion [see Eq. (3.8)] the attenuation constant for this situation can be approximated by

$$\alpha \cong \frac{1}{2} R_a \sqrt{\frac{C_a}{L_a}} + \frac{1}{2}(G_a + G_w)\sqrt{\frac{L_a}{C_a}}. \tag{3.76}$$

In a like manner, the phase constant is given approximately by

$$\beta \cong \omega \sqrt{L_a\left(C_a - \frac{\Gamma_w}{\omega^2}\right)} = \frac{\omega}{c'}. \tag{3.77}$$

The effective sound velocity c' in a pipe with "massive" walls—that is, with negative susceptance—is therefore faster than for free space. The pipe appears shorter and the resonant frequencies are shifted upward. The effect is greatest for the lower frequencies. The same result can be obtained more elegantly in terms of specific wall admittance by writing the wave equation for the cylindrical pipe, noting the radial symmetry and fitting the boundary impedance conditions at the walls (MORSE). In addition to the plane-wave solution, the latter formulation also gives the higher cylindrical modes.

Results (3.76) and (3.77) therefore show that vibration of the cavity wall contributes an additive component to the attenuation constant, and when the wall is predominantly mass-reactive, its effect is to diminish the phase constant or increase the speed of sound propagation. Following the previous technique [see Eq. (3.63)], the natural modes for a uniform tube of this sort are given by

$$s_{nw} = \left[-\alpha c' \pm j\frac{(2n+1)\pi c'}{2l}\right] \tag{3.78}$$

$$= (\sigma_{nw} + j\omega_{nw}); \quad n = 0, 1, 2, \ldots.$$

To calculate the shunting effect of the walls in the real vocal tract, it is necessary to have some knowledge of the mechanical impedance of the cavity walls. Such measurements are obviously difficult and apparently have not been made. An order-of-magnitude estimate can be made, however, by using mechanical impedance values obtained for other surfaces of the body. At best, such measurements are variable, and the impedance can change appreciably with place. The data do, however, permit us to make some very rough calculations.

One set of measurements (FRANKE) has been made for chest, thigh and stomach tissues, and these have been applied previously to estimate the wall effect (HOUSE and STEVENS, 1958). For frequencies above about 100 cps, the fleshy areas exhibit resistive and mass reactive components. The specific impedances fall roughly in the range 4000–7000 dyne-sec/cm^3. A typical measurement on the stomach surface gives a specific impedance that is approximately

$$z_s=(r_s+jx_s)=(r_s+j\omega l_s)$$
$$=(6500+j\omega\, 0.4), \tag{3.79}$$

for $(2\pi\cdot 200)\leqq\omega\leqq(2\pi\cdot 1000)$.

This specific series impedance can be put in terms of equivalent parallel resistance and inductance by

$$r_p=\frac{r_s^2+x_s^2}{r_s} \quad \text{and} \quad jx_p=j\frac{r_s^2+x_s^2}{x_s}.$$

These specific values (per-unit-area) can be put in terms of per-unit-length of tube by dividing by S, the inner circumference, to give

$$R_w=\frac{r_s^2+x_s^2}{r_s S} \quad \text{and} \quad jX_w=j\frac{r_s^2+x_s^2}{x_s S}.$$

Therefore,

$$G_w=\frac{r_s S}{r_s^2+x_s^2} \quad \text{and} \quad -j\frac{\Gamma_w}{\omega}=-j\frac{\omega l_s S}{r_s^2+x_s^2},$$

where,

$$\Gamma_w=\frac{\omega^2 l_s S}{r_s^2+x_s^2}. \tag{3.80}$$

Assuming the vocal tract to be unconstricted and to have a uniform cross-sectional area of 5 cm^2 (i.e., $S=7.9$ cm), we can compute the effect of the wall admittance upon the propagation constant, the formant bandwidth and formant frequency. According to (3.76) and (3.77), the wall's contribution to α and β is

$$\alpha_w\cong\frac{G_w}{2}\sqrt{\frac{L_a}{C_a}},$$

and

$$\beta_w\cong\omega\sqrt{L_a\left(C_a-\frac{l_s S}{r_s^2+x_s^2}\right)}$$
$$\cong\frac{\omega}{c}\left[1-\frac{\rho c^2 l_s}{a(r_s^2+x_s^2)}\right], \tag{3.81}$$

where the radius of the tube is $a=\sqrt{A/\pi}$, and the bracketed expression is the first two terms in the binomial expansion of the radical.

Substituting the measured values of r_s and l_s and computing α_w, β_w and formant bandwidths at approximately the first three formant frequencies gives[1]

Frequency	α_w	β_w	$\Delta f_w=\frac{\alpha_w c'}{\pi}$
500 cps	4.7×10^{-3}	$\frac{\omega}{c}(1-0.011)$	50 cps
1500 cps	3.6×10^{-3}	$\frac{\omega}{c}(1-0.008)$	40 cps
2500 cps	2.5×10^{-3}	$\frac{\omega}{c}(1-0.006)$	30 cps

[1] Using $c=3.5\times10^4$ cm/sec and $\rho=1.14\times10^{-3}$ gm/cm^3.

The contribution of wall loss to the formant bandwidth is therefore greatest at the lowest formant frequency and diminishes with increasing formant frequency. These computed values, however, when combined with the previous loss contributions actually seem somewhat large. They suggest that the walls of the vocal tract are more rigid than the stomach tissue from which the mechanical impedance estimates were made.

The increase in formant tuning, occasioned by the mass reactance of the cavity walls, is seen to be rather slight. It is of the order of one per cent for the lower formants and, like the damping, diminishes with increasing frequency.

3.74. Two-Tube Approximation of the Vocal Tract

The previous sections utilized a uniform-tube approximation of the vocal tract to put in evidence certain properties. The uniform tube, which displays modes equally spaced in frequency, comes close to a realistic vocal configuration only for the unconstricted schwa sound /ə/. Better insight into the interaction of vocal cavities can be gained by complicating the approximation one step further; namely, by approximating the tract as two uniform, cascaded tubes of different cross section. To keep the discussion tractable and focused mainly upon the transmission properties of the tubes, we again assume the glottal impedance to be high compared with the input impedance of the tract, and the radiation load to be negligible compared with the impedance level at the mouth. This situation is represented in Fig. 3.27.

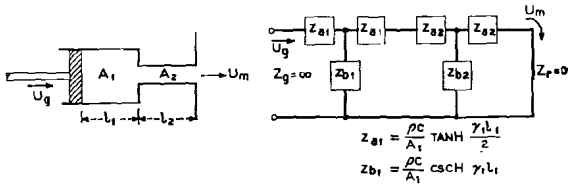


Fig. 3.27. Two-tube approximation to the vocal tract. The glottal impedance is assumed infinite and the radiation impedance zero

For the circuit shown in Fig. 3.27, the mouth-to-glottis volume current ratio is

$$\frac{U_m}{U_g}=\frac{1}{\left(1+\frac{z_{a2}}{z_{b2}}\right)\left(1+\frac{z_{a1}}{z_{b1}}+\frac{z_{a2}}{z_{b1}}\right)+\frac{z_{a2}}{z_{b1}}},$$

which reduces to

$$\frac{U_m}{U_g}=\frac{1}{(\cosh\gamma_1 l_1)(\cosh\gamma_2 l_2)\left(1+\frac{A_1}{A_2}\tanh\gamma_1 l_1\tanh\gamma_2 l_2\right)}. \quad (3.82)$$

The poles of (3.82) occur for

$$\frac{A_1}{A_2}\tanh\gamma_2 l_2=-\coth\gamma_1 l_1. \quad (3.83)$$

If the tubes are lossless, the hyperbolic functions reduce to circular functions and all impedances are pure reactances. The normal modes then satisfy

$$\frac{A_1}{A_2}\tan\beta l_2=\cot\beta l_1. \quad (3.84)$$

Because the vocal tract is relatively low loss, Eq. (3.84) provides a simple means for examining the mode pattern of the two-tube approximation. For example, consider the approximations shown in Fig. 3.28 to the articulatory configurations for four different vowels. The reactance functions of (3.84) are plotted for each case, and the pole frequencies are indicated.

One notices that the high front vowel /i/ exhibits the most disparate first and second formants, while the low back vowel /a/ gives rise to the most proximate first and second formants. The neutral vowel /ə/, corresponding to the unconstricted tract, yields formants uniformly spaced 1000 cps apart. The reactance plots also show that increasing the area ratio (A_1/A_2) of the back-to-front cavities results in a decrease of the first formant frequency. On the classical $F1$ vs $F2$ plot, the first two

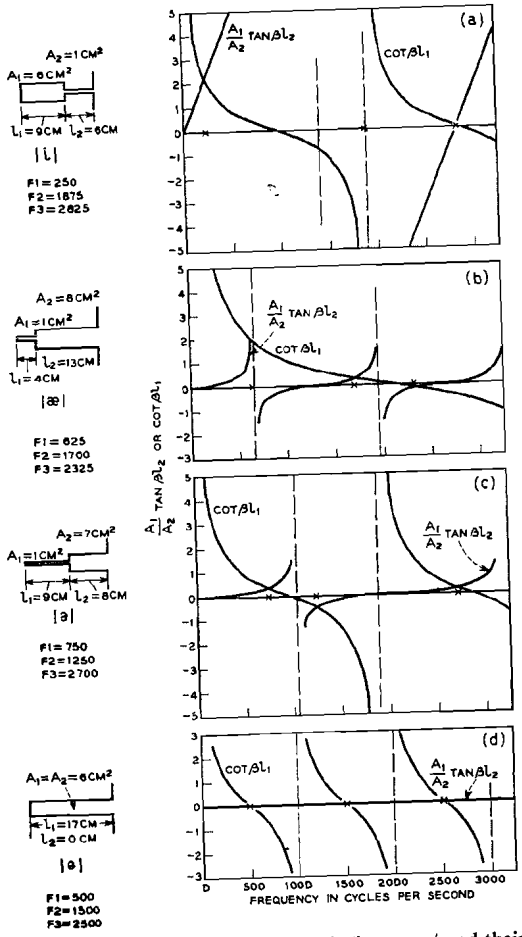


Fig. 3.28a-d. Two-tube approximations to the vowels /i, æ, a, ə/ and their undamped mode (formant) patterns

modes for the four approximations fall as shown in Fig. 3.29. The unconstricted /ə/ sound occupies the central position. For comparison, formant data for four vowels—as spoken by adult males—are also plotted (PETERSON and BARNEY)[1]. The lower left corner of the classical

[1] Most of the vocal tract dimensions used to illustrate acoustic relations in this chapter are appropriate to adult males. Women and children have smaller vocal apparatus. Since the frequencies of the resonant modes are inversely related to the tract length, the vowel formants for women and children are higher than for the men. According to CHIBA and KAJIYAMA, the young adult female vocal tract is 0.87 as long as the young adult male. The female formants, therefore, should be about 15% higher than those of the male. This situation is also reflected in the measurements of PETERSON and BARNEY.

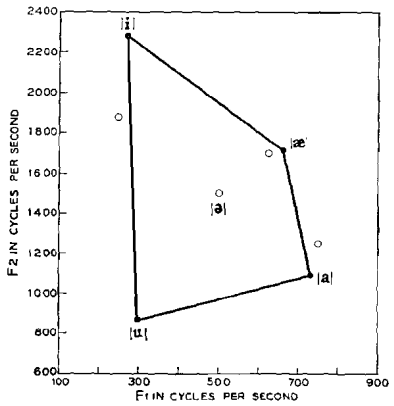


Fig. 3.29. First formant ($F1$) versus second formant ($F2$) for several vowels. Solid points are averages from PETERSON and BARNEY'S data for real speech uttered by adult males. Circles are for the two-tube approximation to the vowels shown in Fig. 3.28

vowel plot, the area appropriate to the vowel /u/, has been indicated for completeness. Because of lip rounding, however, the vowel /u/ cannot be approximated in terms of only two tubes.

Eq. (3.84) also makes salient an aspect of compensatory articulation. The mode pattern for $l_1=a$, $l_2=b$, is exactly the same as for $l_1=b$, $l_2=a$. In other words, so long as the area ratio for the back and front cavities is maintained the same, their lengths may be interchanged without altering the formant frquencies. This is exactly true for the idealized lossless tubes, and is approximately so for practical values of loss. This interchangeability is one freedom available to the ventriloquist. It is also clear from (3.84) that if $l_1=2l_2$, the infinite values of $\cot \beta l_1$ and $\tan \beta l_2$ are coincident (at $\beta l_2=\pi/2$) and indicate the second mode. The second formant frequency can therefore be maintained constant by keeping the tube lengths in the ratio of 2:1. The same constancy applies to the third formant if the length ratio is maintained at 3:2.

3.75. Excitation by Source Forward in Tract

As pointed out earlier, fricative sounds (except for /h/) are excited by a series pressure source applied at a point forward in the tract. It is pertinent to consider the mouth volume velocity which such an excitation produces.

A previous section showed that for glottal excitation the maxima of glottis-to-mouth transmission occurred at the natural (pole) frequencies of the vocal system, and the transmission exhibited no zeros. If excitation

is applied at some other point in the system, without altering the network, the normal modes of the response remain the same. The transmission can, however, exhibit zeros. For the series excitation these zeros must occur at frequencies where the impedance looking back from the source (toward the glottis) is infinite.

By way of illustration let us retain the simple two-tube model used previously. Because the turbulent source for voiceless sound is spatially distributed, its exact point of application is difficult to fix. Generally it can be thought to be applied either at or just forward of the point of greatest constriction. The former seems to be more nearly the case for sounds like /ʃ, f, p, k/; the latter for /s, t/. Consider first the case where the source is forward of the constriction. The two-tube circuit is shown in Fig. 3.30. The back cavity is shown closed, and the impedance of the glottis and larynx tube is considered to be high (compared

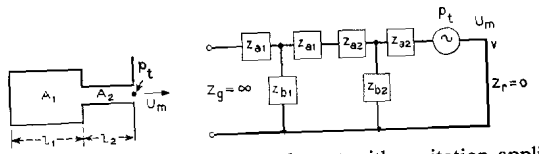


Fig. 3.30. Two-tube approximation to the vocal tract with excitation applied forward of the constriction

to the impedance level of the back cavity) even though the glottis may be open. The radiation impedance is again considered small compared with the impedance level at the mouth, and the inherent impedance of the source *per se* is considered small.

The complex frequency (LAPLACE) transform of the transmission (U_m/p_t) can be written in the form

$$\frac{U_m(s)}{p_t(s)}=H(s)\,G(s), \tag{3.85}$$

where $H(s)$ is a given in (3.64) and contains all the poles of the system, and $G(s)$ is a function which includes all the zeros and constants appropriate to nonglottal excitation. In this particular case, U_m/p_t is simply the driving point admittance at the lips. It is

$$\frac{U_m}{p_t}=\frac{(z_{b2}+z_{b1}+z_{a1}+z_{a2})}{z_{a2}(z_{b2}+z_{b1}+z_{a1}+z_{a2})+z_{b2}(z_{b1}+z_{a1}+z_{a2})},$$

which can be put into the form

$$\frac{U_m}{p_t}=\frac{\frac{1}{Z_{01}}\sinh\gamma_1 l_1 \sinh\gamma_2 l_2\left(\coth\gamma_2 l_2+\frac{A_2}{A_1}\coth\gamma_1 l_1\right)}{\cosh\gamma_1 l_1\cosh\gamma_2 l_2\left[1+\frac{A_1}{A_2}\tanh\gamma_1 l_1\tanh\gamma_2 l_2\right]}. \tag{3.86}$$

The zeros of transmission occur at frequencies which make the numerator zero, and therefore satisfy

$$\coth \gamma_2 l_2 = -\frac{A_2}{A_1} \coth \gamma_1 l_1$$

or

$$\tanh \gamma_1 l_1 = -\frac{A_2}{A_1} \tanh \gamma_2 l_2 ,$$

which for lossless conditions reduces to

$$\tan \beta l_1 = -\frac{A_2}{A_1} \tan \beta l_2 . \qquad (3.87)$$

As an example, let us use (3.87) and (3.84) to determine the (lossless) zeros and poles of U_m/p_t for an articulatory shape crudely representative of /s/. Take

$$A_1 = 7\,\text{cm}^2, \qquad A_2 = 0.2\,\text{cm}^2$$
$$l_1 = 12.5\,\text{cm}, \qquad l_2 = 2.5\,\text{cm} .$$

The pertinent reactance functions are plotted in Fig. 3.31, and the poles and zeros so determined are listed.

The lower poles and zeros lie relatively close and essentially nullify one another. The first significant uncompensated zero lies in the vicinity of 3400 cps, with the first uncompensated pole in the neighborhood of

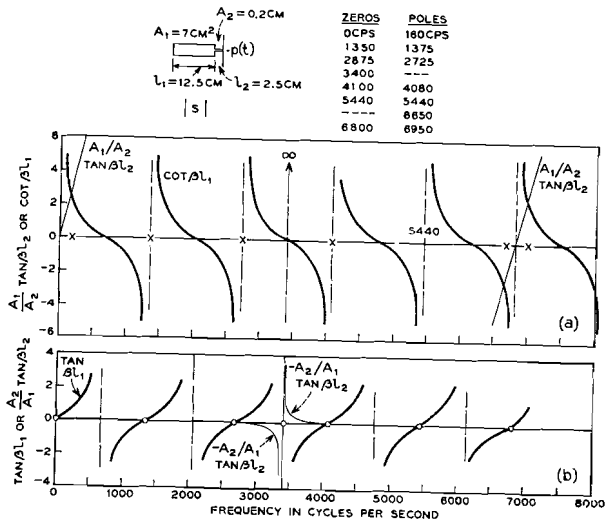


Fig. 3.31a and b. Two-tube approximation to the fricative /s/. The undamped pole-zero locations are obtained from the reactance plots

6650 cps. These two features, as well as the near-cancelling pole-zero pairs, can often be seen in the spectra of real /s/ sounds. For example, Fig. 3.32 shows two measurements of the natural speech fricative /s/ (HUGHES and HALLE). For this speaker, the peak in the vicinity of 6000–7000 cps would appear to correspond with the uncompensated pole, the dip in the vicinity of 3000 cps with the zero. The peak and valley alternations at the lower frequencies reflect roughly the effect of

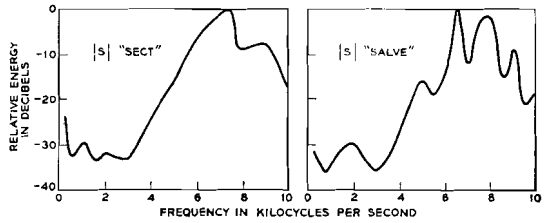


Fig. 3.32. Measured spectra for the fricative /s/ in real speech. (After HUGHES and HALLE)

pole-zero pairs such as indicated in the reactance diagrams. The measured spectra presumably include the transformation from mouth volume current to pressure at a fixed point in space, as described in Eq. (3.40). The spectra therefore include a zero at zero frequency owing to the radiation.

To further examine the influence of source position upon the transmission, suppose the turbulent source is applied more nearly at the junction between the two tubes rather than at the outlet. This situation is crudely representative of sounds like /f/, /k/ or possibly /ʃ/. In /f/, for example, the turbulent flow is produced at the constriction formed by the upper teeth and lower lip. The cavities behind the teeth are large, and the lips forward of the constriction form a short, small-area tube. The circuit for such an arrangement is shown in Fig. 3.33. The transmission from source to mouth is

$$\frac{U_m}{p_t} = \frac{z_{b2}}{z_{b2}(z_{a1}+z_{a2}+z_{b1})+z_{a2}(z_{b2}+z_{a1}+z_{a2}+z_{b1})}$$

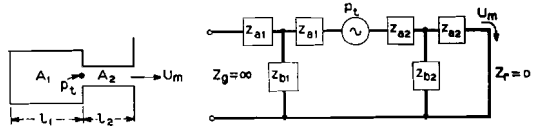


Fig. 3.33. Two-tube approximation to the vocal tract with the source of excitation applied at the tube junction

or

$$\frac{U_m}{p_t}=\frac{\frac{1}{Z_{01}}\sinh\gamma_1 l_1}{\cosh\gamma_1 l_1\cosh\gamma_2 l_2\left[1+\frac{A_1}{A_2}\tanh\gamma_1 l_1\tanh\gamma_2 l_2\right]}. \quad (3.88)$$

The system poles are the same as before, but the zeros now occur at

$$\frac{1}{Z_{01}}\sinh\gamma_1 l_1=0,$$

or

$$s_m=\left(-\alpha_1 c\pm j\frac{m\pi c}{l_1}\right); \quad m=0, 1, 2, \ldots. \quad (3.89)$$

Again for the lossless case, the zeros occur for $\sin\beta l_1=0$, or for frequencies

$$f_m=m\frac{c}{2l_1}\text{ cps} \quad (m=0, 1, 2, \ldots),$$

where the length of the back cavity is an integral number of half wavelengths. The zeros therefore occur in complex-conjugate pairs except for $m=0$. The real-axis zero arises from the impedance of the back cavity volume at zero frequency. Specifically, for the lossless situation at low frequencies, the numerator of (3.88) approaches

$$\lim_{\omega\to 0}\frac{1}{Z_{01}}\sin\beta l_1\cong\frac{\omega l_1}{Z_{01}c}=\frac{A_1 l_1}{\rho c^2}\omega=\omega C_1, \quad \text{where } C_1=\frac{V_1}{\rho c^2}$$

is the acoustic compliance of the back cavity.

The result (3.89) makes clear the reason that a labio-dental fricative such as /f/ exhibits a relatively uniform spectrum (devoid of large maxima and minima) over most of the audible frequency range. A crude approximation to the articulatory configuration for /f/ might be obtained if the parameters of Fig. 3.33 are taken as follows: $A_1=7\text{ cm}^2$, $A_2=0.1\text{ cm}^2$, $l_1=14$ cm, $l_2=1$ cm. As before the poles occur for $\cot\beta l_1=A_1/A_2\tan\beta l_2$. Because of the large value of A_1/A_2 and the small value of l_2, the poles occur very nearly at the frequencies which make $\cot\beta l_1$ infinite; namely

$$f_n\cong n\frac{c}{2l_1}, \quad n=0, 1, 2, \ldots.$$

(The first infinite value of $\tan\beta l_2$ occurs at the frequency $c/4l_2$, in the vicinity of 8 500 cps.) The zeros, according to (3.89), occur precisely at the frequencies

$$f_m=m\frac{c}{2l_1}, \quad m=0, 1, 2, \ldots,$$

so that each pole is very nearly cancelled by a zero. The transmission U_m/P_t is therefore relatively constant until frequencies are reached where the value of $A_1/A_2\tan\beta l_2$ has its second zero. This relative flatness is generally exhibited in the measured spectra of real /f/ sounds such as shown in Fig. 3.34 (HUGHES and HALLE).

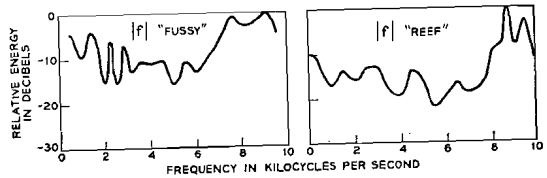


Fig. 3.34. Measured spectra for the fricative /f/ in real speech. (After HUGHES and HALLE)

3.76. Effects of the Nasal Tract

This highly simplified and approximate discussion of vocal transmission has so far neglected the properties of the nasal tract. The nasal tract is called into play for the production of nasal consonants and for nasalizing certain sounds primarily radiated from the mouth. Both of these classes of sounds are voiced. For the nasal consonants, an oral closure is made, the velum is opened and the sound is radiated chiefly from the nostrils. The blocked oral cavity acts as a side branch resonator. In producing a nasalized vowel, on the other hand, coupling to the nasal tract is introduced by opening the velum while the major radiation of sound continues from the mouth. Some radiation, usually lower in intensity, takes place from the nostrils.

The functioning of the combined vocal and nasal tracts is difficult to treat analytically. The coupled cavities represent a relatively complex system. Precise calculation of their interactions can best be done by analog or digital computer simulation. Nevertheless, it is possible to illustrate computationally certain gross features of the system by making simplifying approximations. More specifically, suppose the pharynx cavity, mouth cavity and nasal cavity are each approximated as uniform tubes. The equivalent network is shown in Fig. 3.35.

Notice that, in general, the parallel branching of the system at the velum causes zeros of nasal output at frequencies where the driving point impedance (Z_m) of the mouth cavity is zero, and vice versa. At such frequencies, one branch traps all the velar volume flow. In particular for nasal consonants, /m, n, ŋ/, $Z_{rm}=\infty$ and $U_m=0$. Zeros then occur in the nasal output at frequencies for which $Z_m=0$ for the closed oral cavity. Nasal consonants and nasalized vowels are generally characterized by resonances which appear somewhat broader, or more

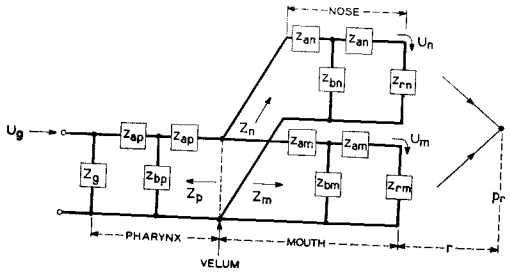


Fig. 3.35. An equivalent circuit for the combined vocal and nasal tracts. The pharynx, mouth and nasal cavities are assumed to be uniform tubes

highly damped, than those for vowels. Additional loss is contributed by the nasal tract which over a part of its length is partitioned longitudinally. Its inner surface is convoluted, and the cavity exhibits a relatively large ratio of surface area to cross-sectional area. Viscous and heat conduction losses are therefore commensurately larger.

Following the approach used earlier, and with the purpose of indicating the origin of the poles and zeros of a nasal consonant, let us make a crude, simple approximation to the vocal configuration for /m/. Such an approximation is illustrated in Fig. 3.36. The poles of the nasal output will be determined by the combined pharynx, mouth and nasal cavities, while the side-branch resonator—formed by the closed oral cavity—will introduce zeros wherever its input impedance is zero. Considering the system to be lossless, the radiation load to be negligible, and the glottal impedance to be high, the easiest way to estimate the pole frequencies is to find the frequencies where the velar admittance (at the point where the three cavities join) is zero. This requires

$$\sum_{k=p,\,m,\,n} Y_k = 0 = \frac{1}{Z_{0m}} \tan \beta l_m + \frac{1}{Z_{0p}} \tan \beta l_p - \frac{1}{Z_{0n}} \cot \beta l_n \qquad (3.90)$$
$$= A_m \tan \beta l_m + A_p \tan \beta l_p - A_n \cot \beta l_n .$$

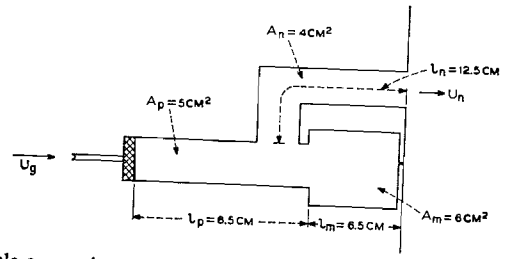


Fig. 3.36. A simple approximation to the vocal configuration for the nasal consonant /m/

The zeros of transmission occur for

$$Z_m = 0 = \frac{\rho c}{A_m} \cot \beta l_m$$

or

$$\beta l_m = (2n+1)\frac{\pi}{2}, \quad n = 0, 1, 2, \ldots$$

or

$$f = (2n+1)\frac{c}{4 l_m}. \qquad (3.91)$$

The mode pattern determined by relations (3.90) and (3.91) is shown in Fig. 3.37. One sees that the first pole of the coupled systems is fairly low, owing to the substantial length of the pharynx and nasal tract and the mouth volume. A pole and zero, additional to the poles of the pure vowel articulation, are introduced in the region of 1000 cps. This mode pattern is roughly representative of all the nasal consonants in that the pharynx and nasal tract have roughly the same shape for all. The first zero falls at approximately 1300 cps in the present example. For the consonants /n and ŋ/, the oral cavity is progressively shorter, and the zero would be expected to move somewhat higher in frquency. By way of comparison, the measured spectrum of a real /m/ is shown in Fig. 3.38 (FANT, 1960). In this measured spectrum, the nasal zero appears to be reflected by the relatively broad spectral minimum near 1200 cps. The larger damping and appreciable diminution of spectral amplitude at the higher frequencies is characteristic of the nasal consonants.

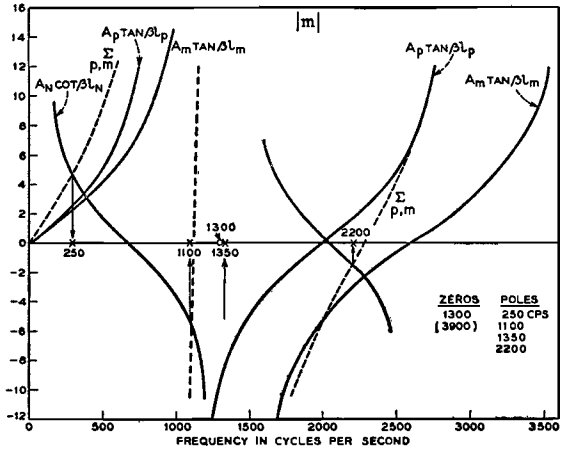


Fig. 3.37. Reactance functions and undamped mode pattern for the articulatory approximation to /m/ shown in Fig. 3.36

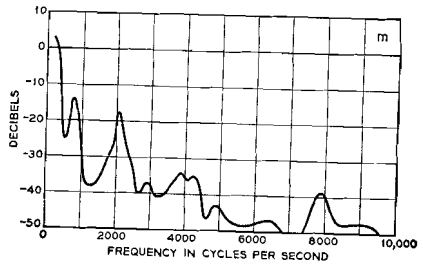


Fig. 3.38. Measured spectrum for the nasal consonant /m/ in real speech. (After FANT, 1960)

3.77. Four-Tube, Three-Parameter Approximation of Vowel Production

To illustrate fundamental relations, the preceding sections have dealt with very simple approximations to the vocal system. Clearly these crude representations are not adequate to describe the gamut of articulatory configurations employed in a language. The approximations can obviously be made better by quantizing the vocal system into more and shorter tube sections. For vowel production in particular, one generally can identify four main features in the tract geometry. These are the back pharynx cavity, the tongue hump constriction, the forward mouth cavity and the lip constriction (see Fig. 3.1). Approximation of these features by four abutting tubes gives a description of vocal transmission substantially more precise than the two-tube approximation. The first several normal modes of the four-tube model are reasonably good approximations to the lower formants of real vowels. Such a four-tube model is illustrated in Fig. 3.39a (adapted from FANT, 1960).

If the glottal impedance is taken as large and the radiation load small, the glottal-to-mouth transmission is

$$\frac{U_m}{U_g}=\frac{1}{\prod_{n=1}^{4}(\cosh\gamma_n l_n)(a\,b+c\,d)},$$

where

$$\begin{aligned}
a&=\left(1+\frac{A_1}{A_2}\tanh\gamma_1 l_1\tanh\gamma_2 l_2\right)\\
b&=\left(1+\frac{A_3}{A_4}\tanh\gamma_3 l_3\tanh\gamma_4 l_4\right)\\
c&=\frac{A_2}{A_3}\left(\tanh\gamma_3 l_3+\frac{A_3}{A_4}\tanh\gamma_4 l_4\right)\\
d&=\frac{A_1}{A_2}(\tanh\gamma_1 l_1+\tanh\gamma_2 l_2).
\end{aligned}\qquad(3.92)$$

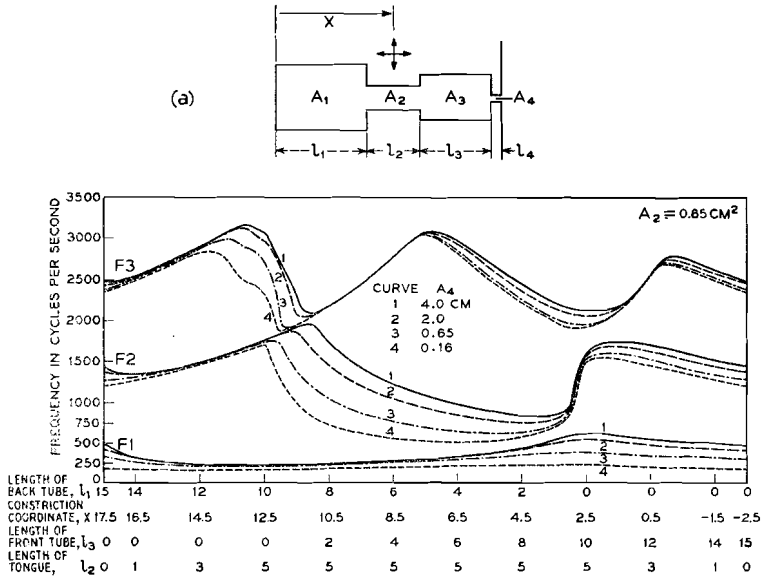


Fig. 3.39a and b. Nomogram for the first three undamped modes ($F1$, $F2$, $F3$) of a four-tube approximation to the vocal tract. (Data adapted from FANT, 1960.) The parameter is the mouth area, A_4. Curves 1, 2, 3 and 4 represent mouth areas of 4, 2, 0.65 and 0.16 cm^2, respectively. Constant quantities are $A_1=A_3=8$ cm^2, $l_4=1$ cm and $A_2=0.65$ cm^2. Abscissa lengths are in cm

One notices that if $l_3=l_4=0$, Eq. (3.92) reduces to the two-tube relations given by Eq. (3.82).

To demonstrate how the first several normal modes of such a cavity arrangement depend upon configuration, FANT (1960) has worked out detailed nomograms for several combinations of A's and l's. One of these is particularly relevant and essentially depicts the scheme followed by DUNN (1950) in his development of an electrical vocal tract analog. It is reproduced in adapted form in Fig. 3.39b. The constraints are as follows: $l_2+l_2+l_3=15$ cm; $l_4=1$ cm; $A_1=A_3=8$ cm^2; $A_2=0.65$ cm^2; and $l_2=5$ cm, provided tube 2 is terminated by cavities on both sides. The parameters are the distance from the glottis to the center of the tongue constriction, x, and the mouth area, A_4. For very large and very small values of x, l_3 and l_1 are zero, respectively, and the length l_2 is varied to satisfy the total length condition. The variation of the first three normal modes for a range of values of the parameters and for one value of the tongue constriction ($A_2=0.65$ cm^2) are shown in Fig. 3.39b.

These data show that a shift of the tongue constriction from a back ($x\approx 3$ cm) to a front position ($x\approx 9$ cm) is generally associated with a

transition from high $F1$-low $F2$ to low $F1$-high $F2$. (This general tendency was also evident in the two-tube models discussed in Section 3.74.) Increasing the lip rounding, that is decreasing A_4 (as well as increasing l_4), generally reduces the frequencies of all formants. Although not shown here, decreasing the tongue constriction reduces the frequency variations of the formants with place of constriction. In terms of absolute cps, the variations in $F1$ are generally smaller than those of the higher formants. Perceptually, however, the percentage change in formant frequency is more nearly the important quantity. This point will be discussed further in Chapter VII.

Owing to the substantial coupling between the connecting tubes, a particular formant cannot be strictly associated with a particular resonance of a particular vocal cavity. The normal mode pattern is a characteristic of the whole coupled system. Numerous efforts have been made in the literature to relate specific formants to specific vocal cavities, but this can be done exactly only when the constrictions are so small in size that the cavities are, in effect, uncoupled. In instances where the coupling is small, it is possible to loosely associate a given formant with a particular resonator. The treachery of the association, however, can be simply illustrated. If a forward motion of the tongue hump causes a resonant frequency to rise—for example, $F2$ for $3<x<9$ cm in Fig. 3.39—the suggestion is that the resonance is mainly influenced by a cavity of diminishing length, in this case the mouth cavity. On the other hand, the same resonance might be caused to rise in frequency by a tongue retraction and a consequent shortening of the pharynx cavity—for example, $F2$ for $16>x>13$ cm. It is therefore clear that a given formant may be principally dependent upon different cavities at different times. It can change its cavity-mode affiliation with changes in vocal configuration. In fact, its dependence upon the mode of vibration of a particular cavity may vary.

The four-tube approximation to vowel production implies that vowel articulation might be grossly described in terms of three parameters, namely, the distance from the glottis to the tongue-hump constriction, x; the size of the tongue constriction, A_2; and a measure of lip rounding such as the area-to-length ratio for the lip tube, A_4/l_4. This basis notion has long been used qualitatively by phoneticians to describe vowel production. It has been cast into quantitative frameworks by DUNN (1950), STEVENS and HOUSE (1955), FANT (1960) and COKER (1968), in connection with work on models of the vocal mechanism.

As pointed out earlier, DUNN has used the scheme much as represented in Fig. 3.39, that is, with constant-area tubes approximating the tract adjacent to the constriction. STEVENS and HOUSE and FANT have extended the scheme by specifying constraints on the taper of the vocal

tract in the vicinity of the constriction. STEVENS and HOUSE use a parabolic function for the area variation, and FANT uses a section of a catenoidal horn (i.e., a hyperbolic area variation). Both use fixed dimensions for the larynx tube and the lower pharynx. In perceptual experiments with synthetic vowels, STEVENS and HOUSE find that a reasonably unique relation exists between the allowed values of x, A_2 and A_4/l_4 and the first three vowel formants. Although these three parameters provide an adequate description of most nonnasal, nonretroflex, vowel articulations, it is clear that they are not generally sufficient for describing consonant and nasal configurations.

Later work by COKER has aimed at a more detailed and physiologically meaningful description of the vocal area function. COKER's articulatory model is specified by seven, relatively-orthogonal parameters: the $x-y$ position coordinates of the tongue body; the degree and the place of the tongue tip constriction; the mouth area; the lip protrusion; and the degree of velar (nasal) coupling. Each parameter has an associated time constant representative of its vocal feature. This articulatory model has been used as the synthesis element in an automatic system for converting printed text into synthetic speech (COKER, UMEDA and BROMAN)[1].

3.78. Multitube Approximations and Electrical Analogs of the Vocal Tract

As the number of elemental tubes used to approximate the vocal shape becomes large, the computational complexities increase. One generally resorts to analog or digital aids in solving the network when the number of approximating sections exceeds about four. In early work analog electrical circuitry has proven a useful tool for simulating both vocal and nasal tracts. It has been used extensively by DUNN (1950); STEVENS, FANT and KASOWSKI; FANT (1960); STEVENS and HOUSE (1955, 1956); and ROSEN. The idea is first to approximate the linear properties of the vocal mechanism by a sufficiently large number of tube sections and then to approximate, in terms of lumped-constant electrical elements, the hyperbolic impedances of the equivalent T or π networks shown in Fig. 3.3. At low frequencies the lumped-constant circuit behaves as a distributed transmission line and simulates the one-dimensional acoustic wave propagation in the vocal tract. The number of approximating tube sections used, the approximation of the hyperbolic elements, and the effect of cross modes in the actual vocal tract determine the highest frequency for which the electrical transmission line is an adequate analog.

[1] See further discussion of this system in Chapters V and VI.

As shown previously, the elements of the T-section equivalent of the cylindrical tube are

$$z_a = Z_0 \tanh \frac{\gamma l}{2} \quad \text{and} \quad z_b = Z_0 \operatorname{csch} \gamma l .$$

Taking first-order approximations to these quantities gives

$$\begin{aligned} z_a &\cong Z_0 \left(\frac{\gamma l}{2}\right) \quad \text{and} \quad z_b \cong Z_0 \left(\frac{1}{\gamma l}\right) \\ z_a &\cong Z_0 \tfrac{1}{2}(\alpha + j\beta) l \qquad z_b \cong Z_0 \frac{1}{(\alpha + j\beta) l} . \end{aligned} \tag{3.93}$$

From the relations developed earlier, $Z_0 = [(R+j\omega L)/(G+j\omega C)]^{\frac{1}{2}}$ and $\gamma = [(R+j\omega L)(G+j\omega C)]^{\frac{1}{2}}$, where R, G, L and C have been given in terms of per-unit-length acoustical quantities in Eq. (3.33). The T-elements are therefore approximately

$$z_a = \tfrac{1}{2}(R + j\omega L) l \quad \text{and} \quad z_b = \frac{1}{(G + j\omega C) l} .$$

In general, the acoustical quantities R_a, L_a, G_a and C_a [in Eq. (3.33)] will not correspond to practical electrical values. It is usually convenient to scale the acoustical and electrical impedance levels so that

$$Z_{0e} = k Z_{0a}$$

or

$$\left[\frac{R_e + j\omega L_e}{G_e + j\omega C_e}\right]^{\frac{1}{2}} = \left[\frac{k R_a + j\omega k L_a}{\frac{G_a}{k} + \frac{j\omega C_a}{k}}\right]^{\frac{1}{2}} . \tag{3.94}$$

By way of indicating the size of a practical scale constant k, consider the low-loss situation where

$$Z_{0e} = \sqrt{\frac{L_e}{C_e}} = k Z_{0a} = k \sqrt{\frac{L_a}{C_a}} = k \left(\frac{\rho c}{A}\right), \tag{3.95}$$

where A is the cross-sectional area of the acoustic tube. A practical value for Z_{0e} is 600 electrical ohms, and a typical value of A is 8 cm^2. Therefore $k = 600/5.3 = 113$, and the mks impedances of the per-unit-length electrical elements are scaled up by 113 times the cgs impedances of the per-unit-length acoustic elements.

Note, too, that $\beta l \cong \omega l/c = \omega l_e \sqrt{L_e C_e} = \omega l_a \sqrt{L_a C_a}$. Since the velocity of sound and the air density in a given length of tube are constant, maintaining the $L_e C_e$ product constant in the electrical line is equivalent to maintaining constant velocity of sound propagation in the simulated pipe. Similarly, changes in the pipe area A are represented by proportional changes in the C_e/L_e ratio.

The electrical simulation is of course applicable to both vocal and nasal tracts. Choice of the elemental cylinder length l, the electrical scale constant k, and a knowledge of the cross-sectional area A along the tract are the only parameters needed to determine the lossless elements of the transmission line. An estimate of tract circumference along its length is needed to compute the viscous and heat conduction losses (R and G). The radiation loads at the mouth and nostrils are obtained by applying the electrical scale constant to the acoustic radiation impedances obtained earlier in the chapter. It is likewise possible to apply these techniques to the subglottal system and to incorporate it into the electrical simulation. At least four designs of electrical vocal tracts have been developed for studying vocal transmission and for synthesizing speech (DUNN, 1950; STEVENS, FANT and KASOWSKI; FANT, 1960; ROSEN). At least one design has been described for the subglottal system (VAN DEN BERG, 1960).

The digital computer is also an exceedingly effective tool for analyzing multi-tube approximations to the vocal tract. Its ability to carry out complex calculations at high speed makes the solution of 20 or 30-section approximations to the tract almost elementary. At least two computer programs for calculating transfer functions and normal modes for multitube approximations have been used (FANT, 1960; MATHEWS and WALKER).

Another approach has been to represent the cylindrical sections in terms of the reflection coefficients at their junctions (KELLY and LOCHBAUM; MERMELSTEIN; STRONG). This simulation also produces a response which, after digital-to-analog conversion, represents the speech waveform. It therefore can be used effectively as a synthesizer.

In another study of speech synthesis a computer program has been derived that is the difference equation equivalent of the multi-section, bilateral transmission line (FLANAGAN and LANDGRAF). This formulation allows computation of instantaneous pressure and velocity along the transmission line, including the sound pressure radiated from the mouth. When supplied a time-varying area function representative of realistic articulation, its calculated output represents samples of the synthesized speech waveform. Both analog and digital representations of the vocal system will be considered further in a later discussion on speech synthesis.

3.8. Fundamentals of Speech and Hearing in Analysis-Synthesis Telephony

The preceding sections have set forth certain basic acoustic principles for the vocal mechanism. Not only do these relations concisely describe the physical behavior of the source of speech signals, but they imply a

good deal about efficient communication. They suggest possibilities for coding speech information in forms other than merely the transduced pressure wave. The normal mode and excitation relations, for example, indicate a schema on which an analysis-synthesis transmission system might be based. The same can be said for describing the vocal tract by articulatory parameters. Both results reflect constraints peculiar to the speech-producing mechanism.

As yet, however, the properties of hearing and the constraints exhibited by the ear have not entered the discussion. The next chapter proposes to establish certain fundamental properties of the mechanism of hearing – so far as they are known. The exposition will follow a pattern similar to that of the present chapter. The results of both fundamental discussions will then be useful in subsequent consideration of speech analysis and speech synthesis.

IV. The Ear and Hearing

The ultimate recipient of information in a speech communication link usually is man. His perceptual abilities dictate the precision with which speech data must be processed and transmitted. These abilities essentially prescribe fidelity criteria for reception and, in effect, determine the channel capacity necessary for the transmission of voice messages. It consequently is pertinent to inquire into the fundamental mechanism of hearing and to attempt to establish capabilities and limitations of human perception.

As suggested earlier, speech information – originating from a speaker, traversing a transmission medium and arriving at a listener – might be considered at a number of stages of coding. On the transmitter side, the stages might include the acoustic wave, the muscular forces manipulating the vocal mechanism, or the physical shape and excitation of the tract. On the receiver side, the information might be considered in terms of the acoustic-mechanical motions of the hearing transducer, or in terms of the electrical pulses transmitted to the brain over the auditory nerve. Characteristics of one or more of these codings might have application in practicable transmission systems.

The previous chapter set forth fundamental relations between the acoustics and the physiology of the vocal mechanism. We will subsequently have occasion to apply the results to analysis-synthesis telephony. In the present chapter we wish to establish similar relations for the ear. Later we will utilize these in discussions of auditory discrimination and speech perception.

4.1. Mechanism of the Ear

The acousto-mechanical operation of the peripheral ear has been put on a rather firm base. This knowledge is due primarily to the brilliant experiments carried out by G. VON BÉKÉSY, and for which he was awarded the Nobel Prize in 1961. In contrast, present knowledge is relatively incomplete about inner-ear processes for converting mechanical motion into neural activity. Still less is known about the transmission of neural information to the brain and the ultimate mechanism of perception.

Despite these difficulties, it is possible to quantify certain aspects of perception without knowing in detail what is going on inside the "black box". Subjective behavior, in response to prescribed auditory stimuli, can of course be observed and measured, and such data are useful guideposts in the design of speech communication systems. In some instances the correlations between perceptual behavior and the physiological operation of the peripheral ear can be placed in clear evidence. The present discussion aims to indicate current understanding of auditory physiology and psychoacoustic behavior, and to illustrate the extent to which the two can be brought into harmony.

The primary acoustic transducer of the human is shown schematically in Fig. 4.1. The acousto-mechanical components of the organ are conventionally divided according to three regions, namely, the outer ear, the middle ear, and the inner ear.

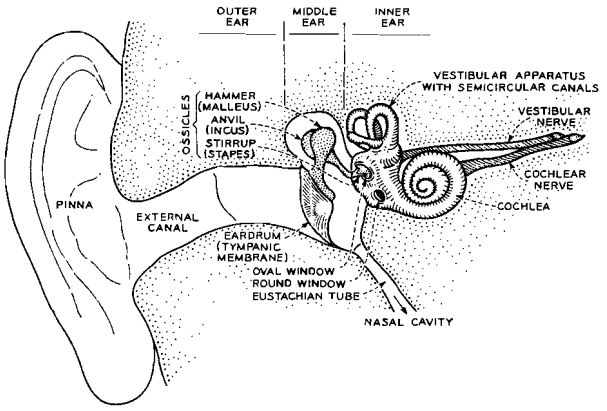


Fig. 4.1. Schematic diagram of the human ear showing outer, middle and inner regions. The drawing is not to scale. For illustrative purposes the inner and middle ear structures are shown enlarged