Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

Vocal tract wall effects, losses, and resonance bandwidths

Fant, G.

journal: STL-QPSR
volume: 13
number: 2-3
year: 1972
pages: 028-052

C. VOCAL TRACT WALL EFFECTS, LOSSES, AND RESONANCE BANDWIDTHS

G. Fant

Abstract

The finite impedance of the vocal tract cavity walls accounts for a shift in formant frequencies and additional losses compared to the idealized hard wall conditions. These effects are treated by: 1) low frequency approximation, 2) single tube tract and lumped element representation of the wall shunt, 3) generalized distributed element treatment of mass loading and losses. The superposition of sound radiated from the walls and from the mouth is calculated for case 2) above. Data on resonance bandwidths are reviewed and simplified formulas are derived for estimating formant bandwidth contributions from: a) radiation, b) classical friction and heat conduction losses, c) dissipation within the cavity walls. The relative importance of these factors in various classes of vocal tract configurations and with respect to different resonances is discussed. Formulas have also been derived for predicting resonance bandwidths from the set of resonance frequencies.

1. Introduction. The closed tract resonance and divers' speech

The effect of the vocal tract wall impedance on the tuning of vocal resonances was first discussed by van den Berg (1953) and by van den Berg (1955 B). When the mouth and the glottis are shut, the closed tract resonance is of the order of 150-200 Hz which is typical of F_1 for voiced stops. This resonance is determined by the lumped inductance of the cavity walls and the capacitance of the enclosed volume. With a finite mouth opening the low frequency equivalent network, see section 2, contains the wall inductance in parallel with the mouth opening inductance and the first formant frequency may be calculated from:

$$F_1 = (F_{1i}^2 + F_W^2)^{\frac{1}{2}} \qquad (1)$$

where F_{1i} is the calculated F_1 without the wall mass shunt and F_W the closed tract resonance. Thus with F_W = 150 Hz and F_{1i} = 300 Hz, F_1 = 335 Hz (Fant, 1960).

A measurement of F_W from spectrograms of voiced occlusions is rather inaccurate. One direct method available is through the vocal tract sine wave response as adopted by Fujimura and Lindqvist (1971). This study which provides us with an extensive material of vocal tract resonances and bandwidths under closed glottis conditions included a few samples of voiced occlusions of /b/ in contact with [ʉ] and [u]. Frequencies of 189 and 204 Hz and bandwidths of 73 respectively 62 Hz were reported.

Speech under hyperbaric conditions provides an interesting technique for estimation of F_W. We shall first derive an analytical expression of this resonance

$$F_W = \frac{1}{2\pi}(C_T L_W)^{-\frac{1}{2}} \qquad (2)$$

where

$$C_T = \frac{V}{\rho c^2} \qquad (3)$$

is the capacitance of the enclosed air and

$$L_W = \frac{\rho_w d}{S \ell} \qquad (4)$$

is the lumped wall inductance. Here

V = vocal tract volume
ρ = density of air
ρ_w = density of walls
S = perimeter of walls, average of the tract
ℓ = vocal tract length
d = wall thickness
c = velocity of sound in the gas

Thus,

$$F_W = \frac{1}{2\pi} c(\rho)^{\frac{1}{2}} \left(\frac{S\ell}{V d \rho_w}\right)^{\frac{1}{2}} \qquad (5)$$

In compressed air c is almost independent of pressure whilst the density ρ is proportional to atmospheric pressure, P. Thus, with F_{W0} denoting the tract resonance under normal pressure conditions

$$F_W = F_{W0}(P)^{\frac{1}{2}} \qquad (6)$$

Breathing a different gas than air the sound velocity c and the density ρ will differ from that of reference conditions c_o and ρ_o. Since for an ideal gas

$$c = \left(\frac{\gamma P}{\rho}\right)^{\frac{1}{2}} \qquad (7)$$

where γ is the ratio of specific heats at constant pressure and volume we may combine Eq. (5) and Eq. (7) to provide a more general expression than Eq. (6)

$$F_W = F_{W0}\left(\frac{\gamma}{\gamma_o} P\right)^{\frac{1}{2}} \qquad (8)$$

At high concentrations of helium in the gas the factor $(\gamma/\gamma_o)^{\frac{1}{2}}$ is close to 1.1. It is interesting to note that F_W is almost independent of gas mixtures and varies as a function of P only whilst other vocal tract resonances are transposed by the factor c/c_o which is of the order of 2-3 in commonly used helium-oxygen gas mixtures for deep-sea diving.

The transposition of F_W with P increases the accuracy in estimation of F_{W0} when F_W can be measured from spectrograms and P is known. From the study of Fant and Sonesson (1964) and that of Fant and Lindqvist (1968) values of F_W = 150-180 Hz and 180-200 Hz, respectively, were reported. These measures were derived from voiced occlusions using Eq. (6) and Eq. (8) and from the observed F_1 of front vowels with additional reference to Eq. (1).

From several independent studies we may thus conclude that the vocal tract closed resonance of male subjects is of the order of 150-200 Hz. The associated bandwidth is of the order of 75-100 Hz, Fant and Sonesson (1964), Fujimura and Lindqvist (1971).

Divers can operate down to a depth of 1200 feet or 40 atmosphere pressures which raises F_W by the factor $(40)^{1/3} = 6.3$. Assuming F_W = 175 Hz, F_W = 1100 Hz, and since $F_{1i} < F_W$ all vowels attain approximately the same F_1 equal to that of the voiced stop voice bar as implied by Eq. (1). The means of restoring the natural range of F_1 values by electronic "unscramblers" are therefore much limited. A complete phoneme recognition and re-synthesis would be needed. Even at the moderate depth of 180 feet the contrast in F_1 between voiced stops, voiced fricatives, glides, and low F_1 vowels is lost and the speech sounds nasal as a result of the average increase in F_1 at fairly constant F_2.

2. Lumped element representation. Superposition effects

Before treating the more general case of distributed mass loading we shall study some simplified networks. Fig. I-C-1a would be a valid approximation for the F1 range of vowels produced with a narrowing in the mouth, e.g. narrow front vowels. L_M is the inductance of the mouth opening with end correction included and L_W the lumped mass load of the walls. An additional large capacitance in series with L_W representing vocal wall elasticity could be added for the sub-audio frequency range. The mouth opening resistance R_M and the wall resistance R_W are treated as small compared to ωL_M and ωL_W. Solving for the poles $\sigma_1 \pm j\omega_1$ of the transfer of volume velocity from the glottis to the mouth opening I_0/I_g we find the frequency and bandwidth

$$\left\{\begin{aligned} F_1 &= \frac{\omega_1}{2\pi} = \frac{1}{2\pi}\left(\frac{1}{C_T}\cdot\frac{L_M+L_W}{L_M L_W}\right)^{1/2} \\ B_1 &= \frac{-\sigma_1}{\pi} = \frac{1}{2\pi}\left(\frac{R_M}{L_M}\cdot\frac{L_W}{(L_W+L_M)} + \frac{R_W}{L_W}\cdot\frac{L_M}{L_M+L_W}\right) \end{aligned}\right. \tag{9}$$

Denoting

$$\left.\begin{aligned} F_W &= \frac{1}{2\pi}(C_T L_W)^{-1/2} \\ F_i &= \frac{1}{2\pi}(C_T L_M)^{-1/2} \end{aligned}\right\} \tag{10}$$

We find $F_1^2 = F_W^2 + F_i^2$ (11)

as in Eq. (1). The bandwidth of the closed tract resonance is

$$B_W = \frac{1}{2\pi}\cdot\frac{R_W}{L_W} \tag{12}$$

and the bandwidth due to the mouth opening resistance when neglecting the R_W L_W branch is

$$B_M = \frac{1}{29}\cdot\frac{R_M}{L_M} \tag{13}$$

Accordingly

$$B = B_W\left(\frac{F_W}{F_1}\right)^2 + B_M\left(\frac{F_i}{F_1}\right)^2 \tag{14}$$

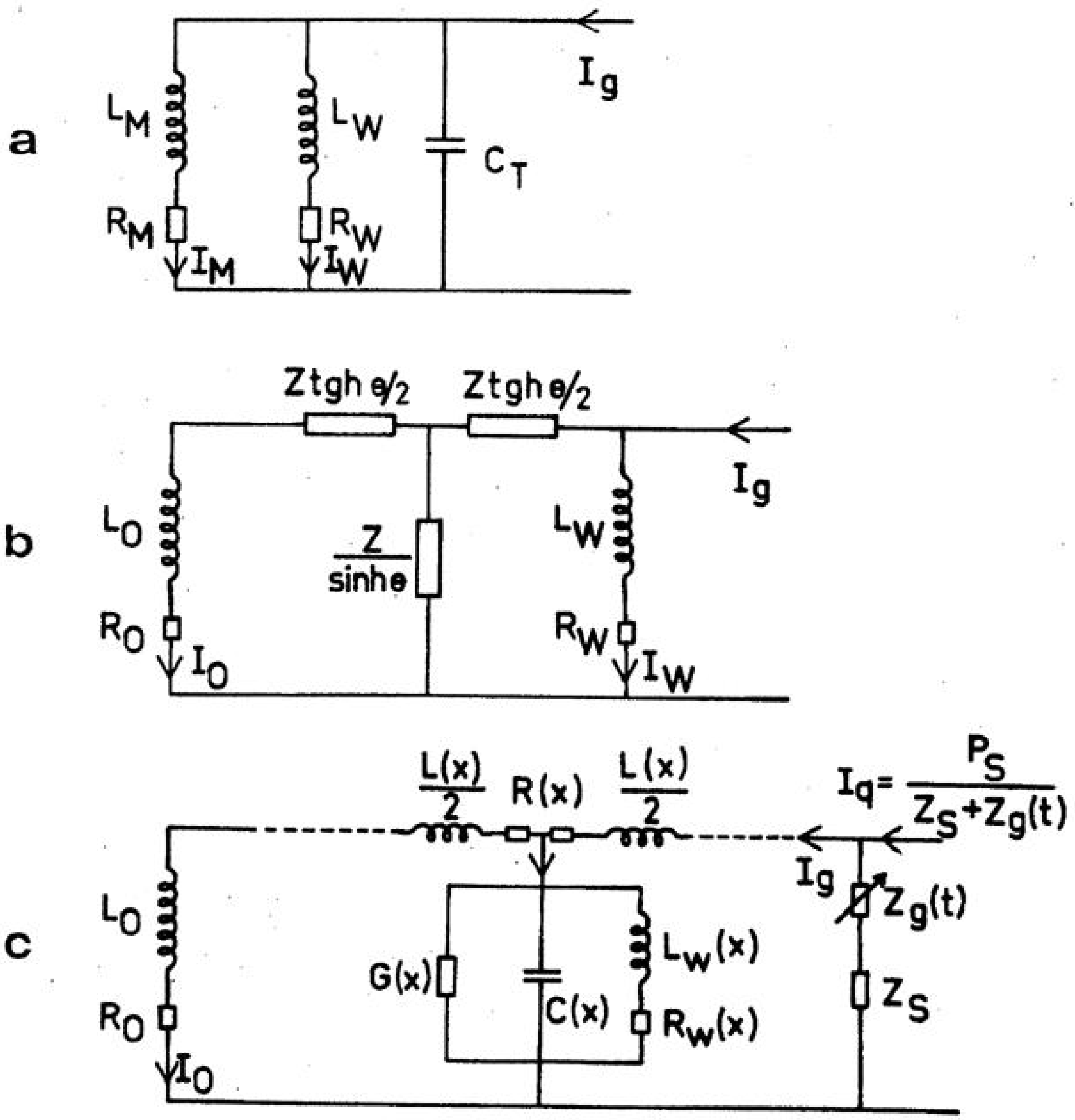

Fig. I-C-1. a Lumped element low frequency representation of the vocal tract with mouth opening and cavity wall shunt.

b Line analog of a tube with lumped element wall shunt at the driving end.

c Distributed element representation of the vocal tract with wall impedance included.

Shunting the capacitance C_T with a conductance $G = 1/R_T$ the bandwidth is increased by the term

$$B_G = \frac{1}{2\pi R_T C_T} \tag{15}$$

A small part of R_M is the radiation resistance

$$R_0 = \frac{\rho \omega^2}{4\pi c} \cdot K_S(\omega) \tag{16}$$

where the $K_S(\omega)$ factor attains the limiting value 1 for zero frequency and 2 for an infinite baffle. $K_s(\omega) = 1.4$ is a representative mean for speaking conditions, Fant (1960). Except for changes in $K_S(\omega)$ the radiation resistance is independent of the area of the radiating surface. This would also be true of the radiation from the walls of the vocal tract. At low frequencies the radiation resistance component of R_W is accordingly equal to R_0.

We are now in a position to calculate the superposition of sound radiated from the mouth opening and from the walls.

$$P = \frac{\rho \omega I_M K_{TM}(\omega)}{4\pi \ell_M} + \frac{\rho \omega I_W K_{TW}(\omega)}{4\pi \ell_W} \tag{17}$$

Assuming equal baffle effects $K_T(\omega)$ and equal distances ℓ_M and ℓ_W from the two radiating surfaces.

$$P = \frac{\rho \omega}{4\pi \ell}(I_M + I_W)K_T(\omega) \tag{18}$$

The combined output has actually a simpler transform than each of the currents I_M and I_W since these have a real pole at $(R_M+R_W)/(L_M+L_W)$ and zeros at R_W/L_W re. R_M/L_M which cancel in the summation.

As an intermediate stage of sophistication we shall now treat the vocal tract as a homogeneous tube and the wall impedance as a mass element lumped at the input end at the throat, as in Fig. I-C-1b. The actual distribution of the mass loading is not known and probably varies with the articulation as will be discussed later on, but we have reasons to assume, from anatomical considerations and measurements of vibrational amplitude on the surface of the head and throat, that the posterior placement of the lumped load is more representative than an anterior location within the tract.

The sum of the currents I_W and I_0 is found to be

$$I_0 + I_W = I_q \frac{1 + \frac{Z}{R_W + sL_W} \sinh \theta}{\frac{Z}{R_W + sL_W} \sinh \theta + \cosh \theta} \qquad (19)$$

where $Z = \rho c/A$ and $\theta = j\omega \ell/c$.

Neglecting losses we find

$$I_0 + I_W = I_q \frac{1 + \frac{\sin x}{x} x_W^2}{\frac{\sin x}{x} x_W^2 + \cosh \theta} \qquad (20)$$

Here $x = \frac{\omega \ell}{c}$

$$x_W = \frac{\omega_W \ell}{c} = \frac{2\pi F_W \ell}{c}$$

and L_W has been substituted from the relation

$$(2\pi F_W)^2 = \omega_W^2 = \frac{\rho c^2}{L_W \cdot A \cdot \ell} \qquad (21)$$

where F_W is an equivalent closed tract resonance frequency and ℓ the total length of the tract. A numerical estimate of the constant x_W is now possible by reference to $F_1 = c/4\ell$ of the open single tube

$$x_W = \frac{2\pi F_W}{4F_1} = \frac{2\pi \cdot 180}{4 \cdot 500} = 0.56 \qquad (22)$$

In the general case of speaking under hyperbaric conditions at a pressure of P ata and with reference to Eq. (6) we may rewrite Eq. (20) as

$$I_0 + I_W = I_q \frac{1 + 0.32 \cdot P \frac{\sin x}{x}}{\cosh \theta'} \qquad (23)$$

The numerator provides a sinx/x shaped correction of the single tube response as shown in Fig. I-C-2 for normal air pressures. The

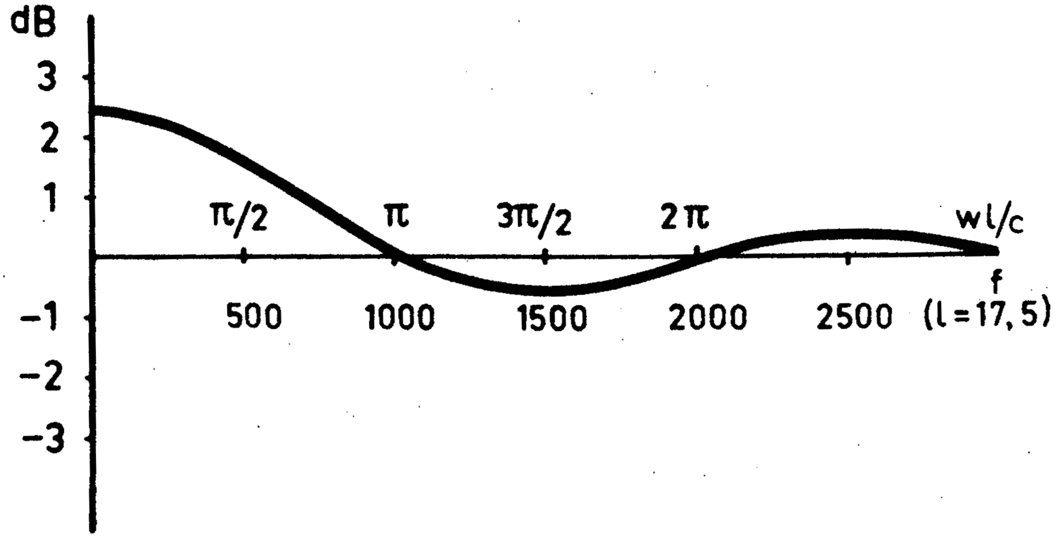

Fig. I-C-2. Response correction when adding sound radiated from the walls near the driving end of a neutral tract to the sound radiated from the mouthopening.

essential feature of this correction is a low frequency boost of 2.4 dB. The minimum of 0.6 dB at 1500 Hz is of no importance. However, at a pressure of P = 15 ata there would appear a 15 dB low frequency boost and a pronounced zero at 1500 Hz, which would affect the F2 level. By solving for the zeros of the denominator of Eq. (19) we find an F_1 = 560 Hz, i.e. a 12 % increase due to the mass shunt at the driving end.

3. Distributed losses and wall impedance

The generalized equivalent network of Fig. I-C-1c contains the distributed series elements $sL(x) + z(x)$ and parallel elements $sC(x) + y(x)$, where

$$\begin{cases} z(s,x) = R(x) \\ y(s,x) = G(x) + [sL_w(x) + R_w(x)]^{-1} \end{cases} \quad (24)$$

per unit length at a coordinate x along the tract. The radiation and glottal source terminations are included in the figure. $L_w(x)$ and $R_w(x)$ are the distributed inductance and series resistance of the cavity walls. R_0 is the radiation resistance, Eq. (16) and L_0 is the radiation inductance which for simplicity of calculations may be included as an end correction adding approximately

$$\Delta \ell_0 = 0.8(A_0/\pi)^{1/2} \quad (25)$$

to the vocal tract length at the lips. As long as ℓ_0 is shorter than wavelengths to be considered the error in calculations of resonance frequencies is very small. The glottal termination is set to infinity load for calculation of the closed glottis response.

The calculation of the response I_0/I_g may now proceed according to the standard matrix techniques. It would, however, be smart to be able to start out with the loss-less and infinite wall impedance case and derive both the wall effects and the damping of resonances by adding correction terms.

The technique of indirect frequency transformations developed by Laurent (1964) offers one such possibility. Multiply all impedances of the original net by a factor ψ such that inductances L(x) become connected with series elements z(s,x) and capacitances C(x) become paralleled by branches y(s,x) as implied by Eq. (24). As a result of the transformation:

$$\left.\begin{array}{ll} \psi s' \cdot L(x) = sL(x) + z(s,x) & \text{(a)} \\ \dfrac{s'C(x)}{\psi} = sC(x) + y(s,x) & \text{(b)} \end{array}\right\} \quad (24)$$

Eliminating the impedance function ψ Eq. (24b) is rewritten as

$$s'^2 = [s+y(s,x)/C(x)][s+y(s,x)/L(x)] = [s+\alpha_1)(s+\alpha_2)] \quad (25)$$

A property, e.g. a pole of a transfer function $H_i(s') = I_o/I_g$, in the original net is accordingly transposed to a frequency s in the transformed net

$$H_i(s') = H(s) \quad (26)$$

A similar transformation was used by Sondhi and Gopinath (1971).

Consider a pair of conjugate poles $s' = \pm j\omega_i$. The associated complex frequencies in the s domain of the transformed network shall satisfy Eq. (25). Providing

$$\left.\begin{array}{ll} \alpha_R = \dfrac{R(x)}{L(x)} & \text{(a)} \\ \alpha_G = \dfrac{G(x)}{C(x)} & \text{(b)} \\ \alpha_W = \dfrac{R_w(x)}{L_w(x)} & \text{(c)} \\ \omega_w^2 = \dfrac{1}{L_w(x)C(x)} = \dfrac{1}{L_W C_T} = (2\pi F_W)^2 & \text{(d)} \end{array}\right\} \quad (27)$$

are all independent of x and A(x).

$$-\omega_i^2 = s^2 + \alpha_R s + \alpha_G B + \alpha_G \alpha_R + \frac{\omega_W^2 (s+\alpha_R)}{s+\alpha_W} \quad (28)$$

$$s^3 + s^2(\alpha_R + \alpha_G + \alpha_W) + s(\omega_i^2 + \omega_W^2 + \alpha_G\alpha_R + \alpha_W\alpha_R + \alpha_W\alpha_G) + \\ + \alpha_W \cdot \omega_i^2 + \alpha_R \cdot \omega_W^2 + \alpha_W\alpha_R\alpha_G = 0 \tag{29}$$

Eq. (29) is solved by assuming one real and one conjugate complex root. Identity with

$$(s+\alpha_1)(s^2 + 2\alpha s + \omega_o^2) = 0 \tag{30}$$

$$\left.\begin{aligned} &\alpha_1 + 2\alpha = \alpha_R + \alpha_G + \alpha_W && \text{(a)} \\ &\omega_o^2 + 2\alpha_1\alpha = \omega_i^2 + \omega_W^2 + \alpha_W\alpha_R + \alpha_G(\alpha_R + \alpha_W) && \text{(b)} \\ &\alpha_1\omega_o^2 = \alpha_W\omega_i^2 + \alpha_R\omega_W^2 + \alpha_W\alpha_R\alpha_G && \text{(c)} \end{aligned}\right\} \tag{31}$$

The last term of Eq. (31c) can be neglected

$$\alpha_1 = \frac{\alpha_W \cdot \omega_i^2 + \alpha_R\omega_W^2}{\omega_o^2} \tag{32}$$

which is the real pole. From Eq. (31a)

$$2\alpha = \alpha_W\left(1 - \frac{\omega_i}{\omega_o^2}\right) + \alpha_R\left(1 - \frac{\omega_W^2}{\omega_o^2}\right) + \alpha_G \tag{33}$$

Furthermore Eq. (30b) is approximated by

$$\omega_o^2 = \omega_i^2 + \omega_W^2 \tag{34}$$

and we may write

$$2\alpha = \alpha_W \cdot \frac{\omega_W^2}{\omega_o^2} + \alpha_R\frac{\omega_i^2}{\omega_o^2} + \alpha_G \tag{35}$$

The frequency and bandwidth of the resonance is thus

$$\begin{cases} F = \frac{\omega_o}{2\pi} = \frac{1}{2\pi}\sqrt{\omega_i^2 + \omega_W^2} = \sqrt{F_i^2 + F_W^2} & (36) \\ B = \frac{\alpha}{\pi} = \frac{1}{2\pi}\left[\frac{\omega_W^2}{\omega_o^2}\ \alpha_W + \frac{\omega_i^2}{\omega_o^2}\cdot \alpha_R + \alpha_G\right] = \end{cases}$$

$$= B_W\left(\frac{F_W}{F}\right)^2 + B_R\left(\frac{F_i}{F}\right)^2 + B_G \qquad (37)$$

These expressions have formal identity with Eqs. (1), (11), and (14) but are more general, being valid for any resonance. Eq. (36) requires

$$\left. \begin{array}{l} L_w(x)C(x) = \text{const} \\ \text{or} \\ \frac{S(x)}{A(x)\cdot d(x)} = \text{const} \end{array} \right\} \qquad (38)$$

as seen from Eq. (5). The perimeters S(x) should accordingly be proportional to the product of the area A(x) and the thickness d(x) of the walls.

It is not unreasonable to assume that the equivalent thickness of the walls increases when the tract is narrowed which would compensate for the increasing perimeter to area ratio. However, an alternative hypothesis is that the total lumped wall mass and the total enclosed air volume in the vocal tract are independent of articulation which would ensure a low-frequency validity of Eq. (36). The α-terms of Eq. (27) are probably not independent of A(x) which limits the validity of the bandwidth derivation from Eq. (37).

An alternative approach to derive Eq. (36) under conditions of uniform mass-loading and constant $L_w(x)C(x) = \omega_W^{-2}$ is to introduce a frequency dependent velocity of sound

$$c_e(\omega) = c[1-\omega_W^2/\omega^2]^{-1/2} \qquad (39)$$

which follows from the substitution

$$sC_e(x) = sC(x) + \frac{1}{sL_W(x)} \qquad (40)$$

and the definitions

$$\left.\begin{aligned} c &= [L(x)C(x)]^{-1/2} \\ c_e &= [L(x)C_e(x)]^{-1/2} \end{aligned}\right\} \qquad (41)$$

Any resonance frequency must be proportional to the velocity of sound as can be seen directly from the Webster's wave equation in pressure P(x) along the tract

$$\frac{1}{A(x)} \cdot \frac{d}{dx}\left[A(x)\frac{dP(x)}{dx}\right] - \left(\frac{s}{c}\right)^2 P(x) = 0 \qquad (42)$$

$$\frac{F}{F_i} = \frac{c_e}{c} = [1 - F_W^2/F^2]^{-1/2} \qquad (43)$$

$$F^2 = F_i^2 + F_W^2$$

as In Eq. (36) and subject to the constraints of Eq. (38) as earlier discussed by Fant and Sonesson (1964). A derivation of wall effects via a correction in the velocity of sound has also been made by Flanagan (1965), but the numerical values of frequency shifts he reports are not representative. For the open neutral tube with F_W = 180 Hz we find F_1 = 530 Hz or a 6 % increase which is one half that found when all the mass-loading was lumped at the glottal end, Eq. (20). The effect on F_2 is 11 Hz only.

Although the validity of Eqs. (1), (6), and (11) has been experimentally verified in an average basis in the studies of Fant and Sonesson (1964) and Fant and Lindqvist (1968) we still lack data on the particular mass distribution along the tract, how it varies with particular articulations.

From these studies of speech under hyperbaric conditions it is found that the derived numerical value of the closed tract resonance F_W does not vary much with the particular vowel.

4. Bandwidth data

The direct vocal tract sweep-frequency technique of injecting a gliding frequency sinusoidal vibration externally at the neck and recording the response from a microphone just outside the lips was first introduced by Fant (1961). A frequency bandwidth plot from this study is shown in Fig. I-C-3. Bandwidths are typically of the order of 30-70 Hz at frequencies below 2000 Hz and increase at a high rate, approximately proportional to frequency square above 2000 Hz. These magnitudes compare well with those of House and Stevens (1958) and Dunn (1961) and with the very detailed measurements of Fujimura and Lindqvist (1964, A, B and 1971).

These authors have specifically drawn attention to the inverse bandwidth frequency relation for frequencies below 500 Hz, as shown in Fig. I-C-4, which is taken from their 1971-article. For male voices we may accordingly extrapolate a limiting value of B_1 = 85 Hz at a closed tract F_W of 175 Hz. According to Eq. (14) and (37) we would expect the bandwidth contribution due to the wall dissipation to be 85/4 = 21 Hz at F_1 = 350 Hz and about just as much from other sources to account for the observed B_1. In general but with important exceptions the wall losses dominate the low F_1 range, fricational losses the F_2 range, and radiation loss the F_3 range, as will be discussed in more detail in the following sections.

One general trend displayed by Fig. I-C-4 is that the first formant bandwidth of females is about 25 % higher than for males. Fujimura and Lindqvist (1971) ascribe it to thinner cavity walls of the females. An alternative explanation offered by Eq. (38) would be that the smaller

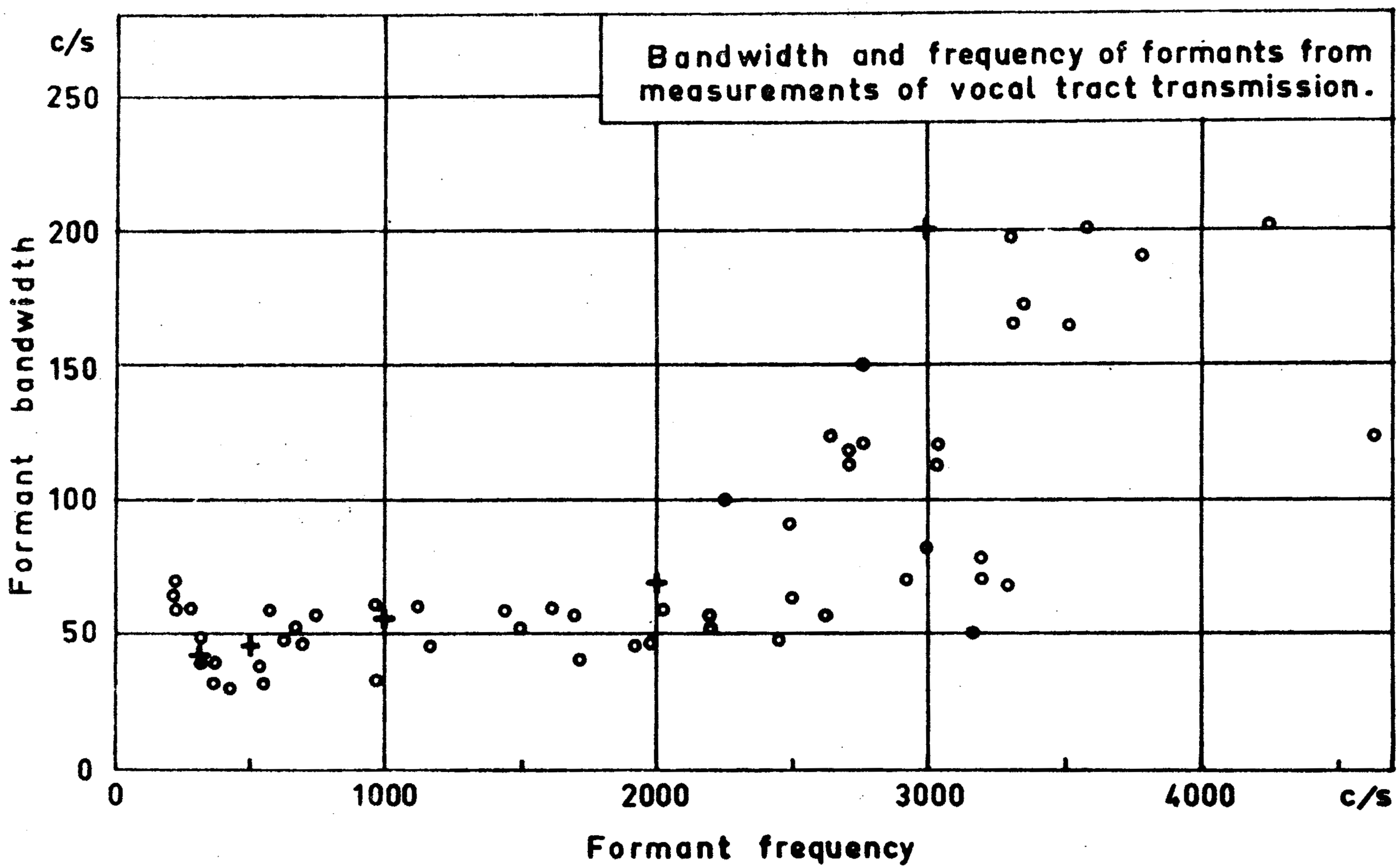

Fig. I-C-3. Vocal tract resonance bandwidth versus frequency. After Fant (1961).

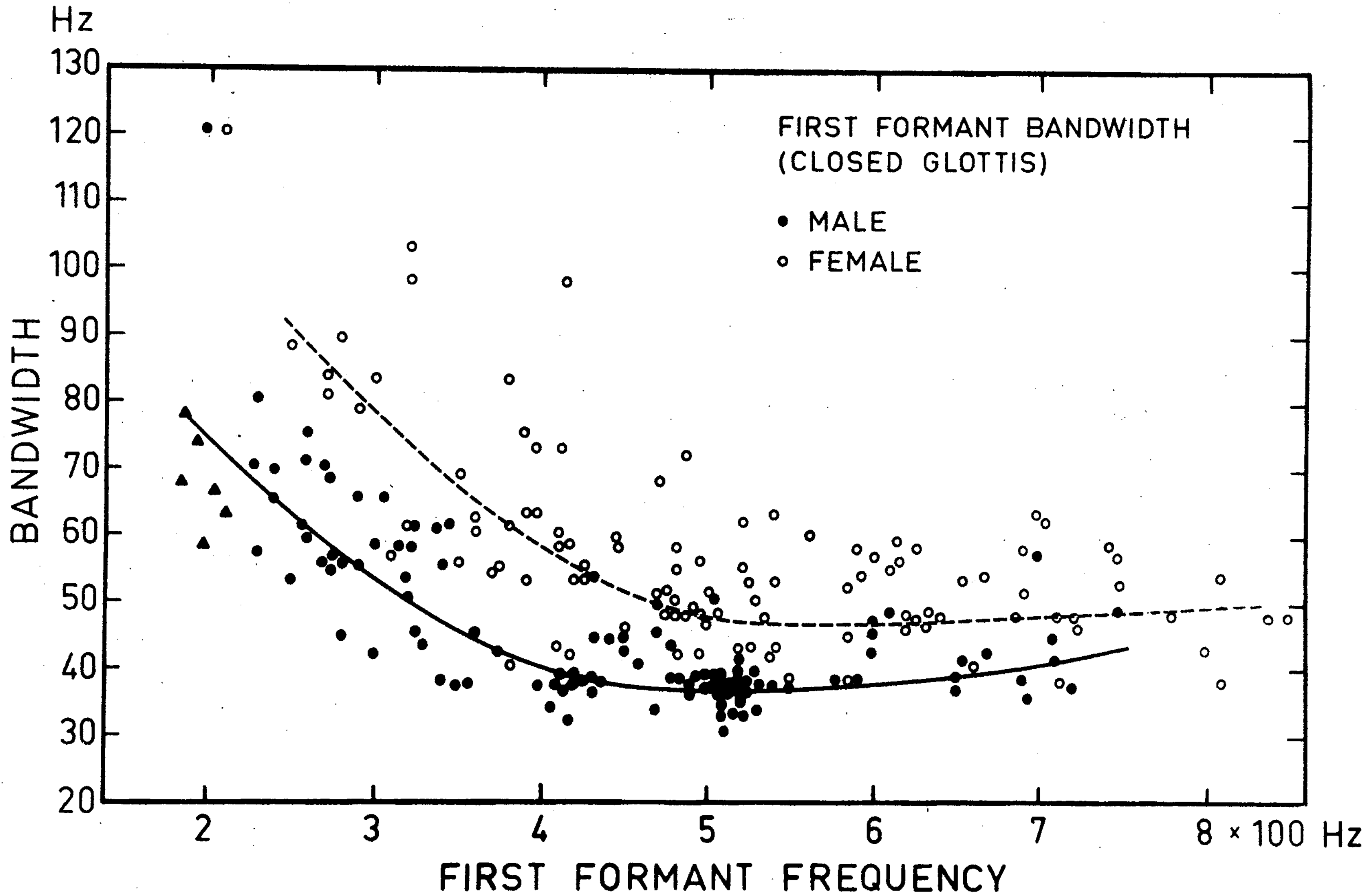

Fig. I-C-4. Bandwidth versus frequency of vocal resonances in the F1 domain for males and females. Fujimura-Lindqvist(1971).

female cavities have a smaller A(x)/S(x) ratio as a consequence of the smaller cross-section dimensions.

For the sake of bringing out the relations between specific categories of vowels and bandwidth patterns I have replotted in Fig. I-C-5 and Fig. I-C-6 some of Fujimura's and Lindqvist's original data on bandwidth and resonant frequencies of Swedish vowels. These have been ordered in minimal phonetic steps and represent the average of two male speakers and of two female speakers respectively.

Among general trends valid for both categories of speakers is that B_3 is substantially higher than B_1 and B_2 in unrounded front and neutral vowels and in open back vowels. In the extreme rounded back and front vowels, [u] and [ʉ] B_3 is nearly always smaller than B_1 and B_2. The bandwidth B_1 has a somewhat greater range of variation than B_2 and about the same average value.

The male and female data are in substantial agreement and their average difference is not so pronounced as would have been expected from Fig. I-C-4. Incidently, as seen from Figs. I-C-5 and I-C-6 the difference in the F-patterns is also rather small for this particular subset of subjects. The B_1 and B_2 variations dependency with phonetic category are essentially the same in the male and the female group. Thus B_1 is greater than B_2 in almost all vowels of maximally low F_1, i.e. in [u], [i], [y], and female [ʉ]. The crossover points of B_1=B_2 occur at approximately the same vowels, i.e. at [ɑ], [e], [ʉ] for the males and [ɑ], [I], and [ø] for the females. The F-pattern-B-pattern correlation will be taken up in the final section.

5. Bandwidth calculations from articulatory models

Under the conditions of closed glottis resonance bandwidths are the sum of terms B_W derived from cavity wall vibrational losses, B_R from friction, and B_G from heat-conduction losses at the interior surfaces of the tract and radiation losses, B_0

$$B = B_W + B_R + B_G + B_0 \qquad (44)$$

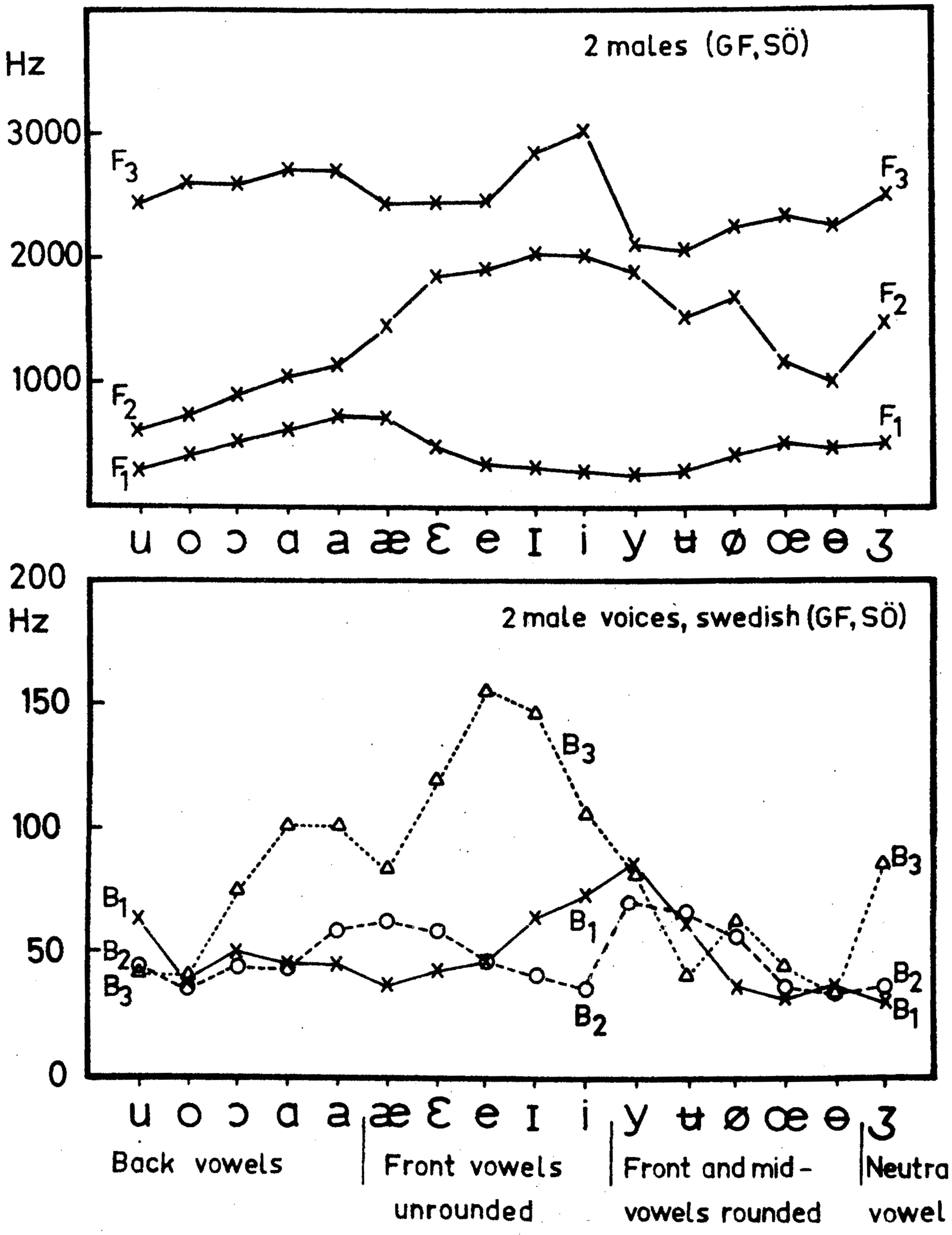

Fig. I-C-5. Resonance frequencies and bandwidths of male vowels in a phonetic sequence to bring out systematic relations.

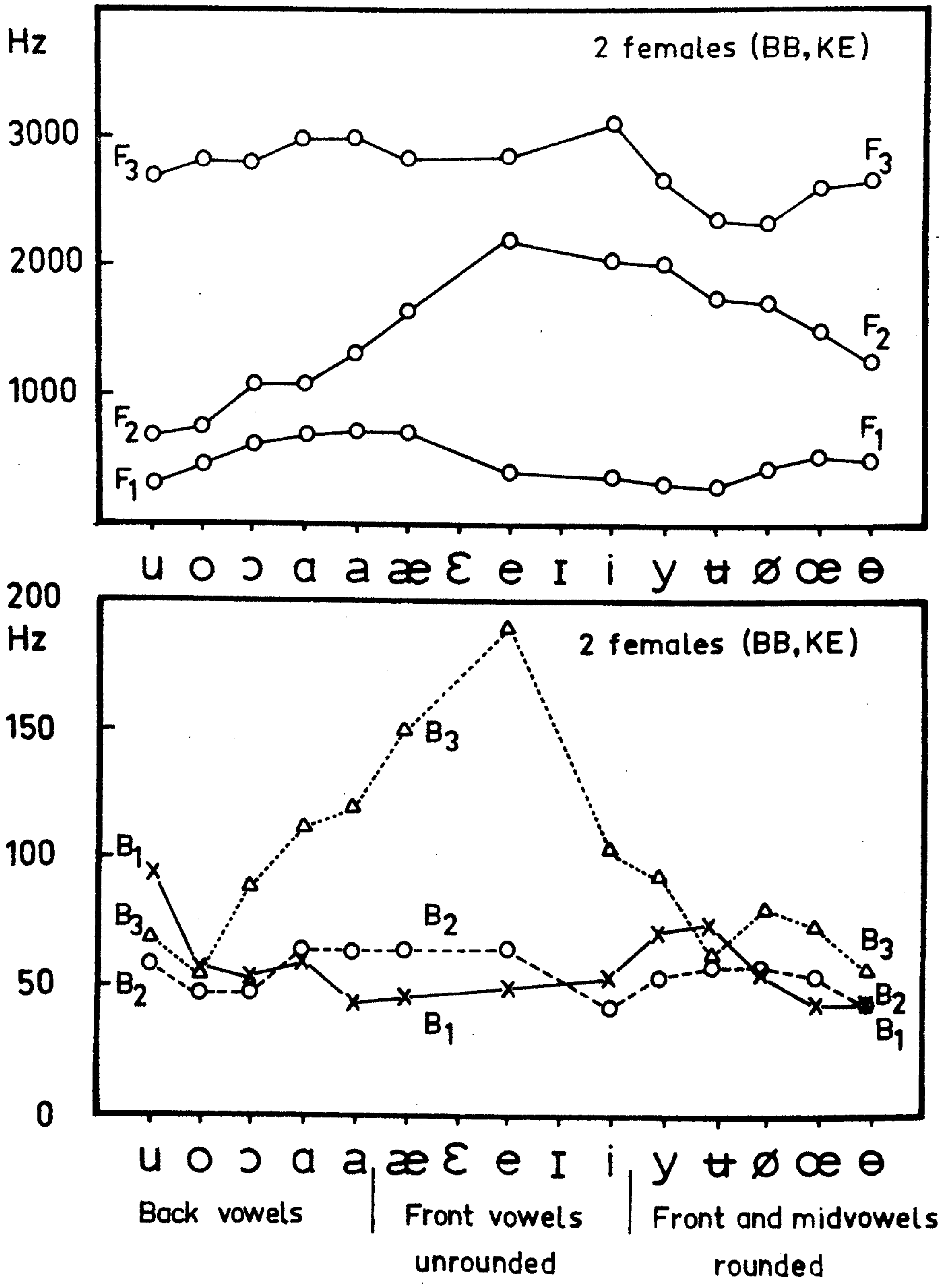

Fig. I-C-6. Resonance frequencies and bandwidths of female vowels arranged in a phonetic sequence to bring out systematic relations. The two subjects selected differ from the male group less than the average for females.

Wall and surface losses. According to Eq. (37) and the Fujimura and Lindqvist (1971) experimental data quoted in the previous section the wall losses account for

$$B_W = B_{W0}\left(\frac{F_W}{F}\right)^2 = B_{W0}(175/F)^2 \qquad (45)$$

where B_{W0} is of the order of 85-100 Hz when normalized to F_W=175 Hz.

The internal surface losses have been treated extensively by Fant (1960, pp. 32-33, 136, and 303-311). Frictional losses derive from expressions of the type R/2L, see Eqs.(13), (35), and (37). With

$$R(x) = \frac{S(x)\,\ell\,(2\mu\,\omega\,\rho)^{1/2}}{2A(x)^2} \qquad (46)$$

and

$$L(x) = \frac{\rho\,\ell}{A(x)} \qquad (47)$$

where ℓ is the length, S(x) the perimeter, A(x) the area, and μ the viscosity coefficient of a unit section of the tract, we obtain

$$B_R = (\mu F/\rho A)^{1/2} \cdot S_A \qquad (48)$$

where

$$S_A = \frac{S}{2\pi(A/\pi)^{\frac{1}{2}}} \qquad (49)$$

is a shape factor which is unity for a circular area and 2 for an elliptical cross-sectional area of width to height ratio equal to 9 which could be representative for narrow constrictions in the vocal tract. Apart from this shape factor it is probable that the friction is greater than in hard walled ideal structures. A value of S_A = 2 has accordingly been adopted for all calculations.

It should be noted that Eq. (48) holds for the distributed case, Eq.(37), as well as for the Helmholtz resonator, Eq. (13). As discussed by Fant (1960, pp. 304-305) the frictional losses in a resonator neck may alternatively be expressed by reference to the volume of the resonator and the length of the neck in which case frequency is eliminated from the expression or by elimination of neck area A in which case frequency enters with negative exponent. Thus with numerical values inserted the following normalized expressions were derived.

$$B_R = 13(f/1000A)^{\frac{1}{2}} S_A = 9.7(100/A\ell_e V)^{\frac{1}{4}} S_A = $$
$$= 7.3(1000/f)^{\frac{1}{2}}(100/V\ell_e)^{\frac{1}{2}} S_A \text{ (Hz)} \qquad (50)$$

The third expression draws attention to the fact that if the resonance frequency is tuned by variations in neck area A alone, the bandwidth contribution will display an $f^{-1/2}$ behavior. This is the typical case of narrowing the lips or raising the tongue against the palate. Accordingly, this negative exponent adds to the cavity wall induced bandwidth increase towards low frequencies, Eq. (45).

For other modes of a tube than the fundamental and for open tubes in general the $f^{1/2}$ dependency is more representative. In this case we should add the head conduction losses

$$\alpha_G = 0.45\,\alpha_R \qquad (51)$$

and the normalized numerical expression for the bandwidth contribution is

$$B_{RG} = B_R + B_G = 18.5(f/1000A)^{\frac{1}{2}} \text{ Hz} \qquad (52)$$

as shown by Fant (1960, p. 308).

To what extent do these simplified models apply to arbitrary shaped vocal tracts? We are in a position to make such an evaluation by reference to exact calculation from distributed element representations and complex matrix calculation techniques, as outlined by Fant (1960, pp. 37-41) and reported on p. 136 in this reference.* In Table I-C-I these are compared with those which could be calculated as a mean of $R(x)/2L(x)$, Eq. (37) and regarding the area not as an arbitrary constant of the same value for all sounds but calculated to satisfy the condition

$$\frac{1}{A_e} = \frac{1}{\ell}\int_0^{\ell} \frac{dx}{A(x)} \qquad (53)$$

i.e. spatial mean over the particular area function. This value is then inserted into Eq. (52) above for use in calculations of all modes.

---

* These calculations were made in 1954 with the first Swedish electronic computer Besk.

Table I-C-I. Exact and approximate calculations of bandwidths B_1, B_2, and B_3 of vowels

	F_1	F_2	F_3	A_e	A_o
u	231	615	2375	1.75	0.65
ɑ	616	1072	2470	2.1	5
ε	432	1959	2722	4.1	5
i	222	2244	3140	3.8	2
Ǝ	500	1500	2500	7	7

from Fant (1960)

Vowel	Walls B_W			Surfaces B_{RG} "Exact"			Surfaces B_{RG} Approx.			Radiation B_0 "Exact"			Radiation B_0 Approx.		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
u	57	8	0	15	16	40	9	20	43	0	0	1	0	0	9
ɑ	8	2	0	17	20	33	20	26	39	4	13	35	5	15	72
ε	15	0	0	11	19	27	12	25	29	3	28	85	3	40	81
i	62	0	0	14	22	36	9	28	33	0	2	190	0	24	45
Ǝ	12.5	1.3	0	10	17	22	10	17	22	5	39	110	5	39	110

Total bandwidth

Vowel	"Exact"			Approx.			Similar vowel, Fujimura-Lindqvist		
	1	2	3	1	2	3	1	2	3
u	72	24	41	63	29	52	63	43	42
ɑ	29	35	68	34	44	112	45	42	100
ε	29	47	112	31	65	110	41	58	119
i	76	24	226	67	53	78	72	35	105
Ǝ	27	57	132	27	57	132	29	37	87

The cavity wall bandwidth term Eq. (45) has been included in the tabulation with B_{W0} = 100 Hz. The $(F_i/F)^2$ factor of Eq. (37) to be applied to B_R has been omitted since the calculations were made without regard to the mass-loading of the walls.

As seen from Table I-C-I the approximation by using Eqs. (52) and (53) instead of the complete vocal tract complex matrix is very good. As expected the relative greatest departures occur at very low F_1 where the mean A_e is no longer representative and the approximate values are of the order of 10 Hz compared to the 15 Hz of the exact calculations. In B_2 and B_3 the mean error is 4.5 Hz.

Radiation. Radiation losses have also been included in the tabulations. These are more critically dependent on the particular vocal tract shape. The following expression was derived by Fant (1960) for a single tube

$$B_0 = 29(f/1500)^2(A/8)(17.6/\ell_e)K_s(f) \qquad (54)$$

where ℓ_e is the length of the tube and $K_s(f)$ a frequency dependent factor which we will set to 1.4 as an average value intermediate between the point source $K_s(f) = 1$ and the infinite baffle $K_s(f) = 2$ extreme which apparently provides too large measures. Eq. (54) would apply for a resonance which is a quarter wavelength or higher mode of a front cavity in a constricted vocal tract, e.g. F3 of /i/, but it would apply very badly to standing wave resonances in semi-closed systems as, for instance, is the case of the second formant of a vowel [i]. It can be noted that the calculated second formant radiation bandwidth of [i], Table I-C-I is 2 Hz only whilst radiation adds 190 Hz to B_3. In general for a semi-closed system the radiation damping is very small and independent of frequency, as shown by Fant (1960), Eq. A.36-21. For a two-tube model

$$B_0 = 90(A_1/\ell_{1e})^2(10/\ell_2 A_2)K_s(f) \qquad (55)$$

The radiation bandwidth term decreases with the square of the A/ℓ ratio of the front part, which has immediate implication for the low radiation damping of liprounded sounds. Apparently the particular cavity resonance dependency is the main factor for determining the extent to which radiation losses enter. The ICA paper of Ichikawa and Nakata (1971) demonstrates these effects. As the point of maximum constriction in a three-parameter model is moved from the glottis to the lips the radiation component of B_3 shows great undulations.

Since the radiation damping of an open tube is inversely proportional to its length, see Eq. (54), an additional source of complication enters. Resonances of a short front cavity become more highly damped than a higher mode of a longer tube at the same frequency.

With all this variability in mind it may seem meaningless to attempt anything like a good estimate of the radiation damping without taking into account the specific shape of the vocal tract. However, for the present study we have adopted Eq. (54) with $\ell_e = 17.6$ cm as in the standard neutral vowel. Furthermore we have chosen to define the radiation area A_0 as the mean of the two smallest areas in the three most anterior sections of the area function quantized in $\Delta x = 0.5$ cm steps. These ℓ_e and A_0 conventions have contributed to underestimate front cavity damping but are necessary in order to avoid excessive values of radiation damping of interior standing wave modes, as F2 of [i].

As may be seen from the tabulation the main error in the approximate radiation loss calculation lies in B_2 and B_3 of [i] and in B3 of [u] and [ɑ]. The approximation is good for the open front vowel [ɛ].

When surface, wall, and radiation terms are summed the calculated data and the experimental data from the Fujimura-Lindqvist study may be compared. With a few exceptions the overall agreement is good and the approximate derivations provide just as good a match as the exact expressions. On the average the approximate B_2 values come out too high and the "exact" B_2 values too low. The low calculated B_2 of [u] indicates that either the frictional losses or the wall losses occasionally play a more important role than anticipated. It must be remembered however, that the spread in human data is rather large and that the vowels of the two studies are not quite the same.

There are good reasons to question the validity of Eq. (45) for calculating the cavity wall damping of the second formant of [u]. As shown by Fant (1960, p. 116) a double Helmholtz resonator model applied well to the articulatory configuration of [u]. It was also stated on p. 121 that the second formant is to a large extent tuned by the back cavity and tongue constriction. The back cavity volume of the [u] is 31 cm^3 whereas the overall tract volume is close to 100 cm^3. Now under the assumption that the mass load of the walls is largely confined to the lower part of the neck just above the glottis, as suggested by Fujimura and Lindqvist (1971), and furthermore that we approach a limit of completely closed tongue hump passage the closed tract resonance of the back cavity becomes $175(100/31)^{1/2} = 315$ Hz. The bandwidth of this resonance is entirely determined by the $R_w(x)/L_w(x)$ as in the undivided tract, i.e. $B_{W0} = 100$ Hz. With a finite coupling to the mouth through the tongue hump of $\ell_2/A_2 = 4.2$ the F_{2i} is calculated to be 495 Hz and when the mass-loading is added $F_2 = (F_{2i}^2 + F_W^2)^{1/2} = 590$ Hz. The bandwidth due to wall losses is then $B_{W0}(F_W/F_2)^2 = 28$ Hz which is about three times or more precisely the ratio of total volume to back cavity volume greater than anticipated by the standard formula, see Table I-C-I. An additional 28-8=20 Hz should be added to B_2 of [u] which happens to be 19 Hz short of the experimental value. This analysis should be significant at least qualitatively.

As a general recommendation for gaining a substantial economy in bandwidth calculations I suggest that the surface losses, i.e. the friction plus heat conduction component, be calculated from the approximate expressions whilst radiation losses if possible should be taken care of by exact calculations. With a more profound insight in the distribution of vocal tract mass-loading it will eventually be possible to gain more exact expressions for the B_W component.

6. Prediction of bandwidths from resonance frequencies

In the previous section vocal tract bandwidths were calculated from a knowledge of their particular frequencies and two vocal tract area measures, one representing the mean of $A(x)^{-1}$, the other the radiating area A_o at the front part of the mouth. In this section we shall attempt to construct empirical formulas for predicting bandwidths from the F-pattern, without reference to articulatory data. Each of B_1, B_2, and B_3 is to be expressed as a function of both F_1, F_2, F_3, and F_4. To what extent is this possible? The search for suitable formulas was particlly aided by the theory of damping mechanisms outlined in the previous sections. The following expression was arrived at for B_1

$$B_1 = 15(500/F_1)^2 + 20(F_1/500)^{1/2} + 5(F_1/500)^2 \text{ (Hz)} \tag{56}$$

The three terms can be identified with the effects of cavity wall losses, classical surface losses, and radiation losses respectively. The expressions for B_2 and B_3 are less directly interpretable.

$$B_2 = 22 + 16(F_1/500)^2 + 12000/(F_3-F_2) \tag{57}$$

$$B_3 = 25(F_1/500)^2 + 4(F_2/500)^2 + 10\ F_3(F_{4a}-F_3) \tag{58}$$

B_2 has a F_1 proportional component which accounts for the increase of radiation loss at greater degrees of articulatory opening. The third terms in B_2 represents radiation loss of a constricted tract under the condition of an F_2 close to F_3 which accounts for energy transfer inspite of the narrowing. The relation of B_2 to F_2 is by no means simple. When both F_2 and F_3 are high and not close as in the vowel [i] B_2 can be as small as 30 Hz and lower than in the vowel [u], where F_2 is at the extreme low end.

The B_3 formula contains an F_1 proportional term representing the degree of articulatory opening, an F_2^2 component, and in addition a term where the degree of proximity of F_3 to the average F_4 of the speaker enters. Lip rounded sounds as [ʉ] and [ø] have larger distances between F_4 and F_3 and on the same time small B_3.

Measured formant frequencies and bandwidths from the male and female groups and bandwidths predicted from Eqs. (56-58) are contained in Table I-C-2 and Table I-C-3 respectively together with decibel values of prediction errors. The only difference made in the female calculations is that the first term of B_1 has been increased by 30 %, i.e. by a 14 % higher equivalent F_W and that F_{4a} is set to 3700 instead of 3400 Hz for the male group.

The bandwidth error represents relative peak levels in the spectrum envelope, $20 \log_{10}(B_m/B_p)$, where B_m is measured and B_p predicted bandwidths. The average value of the error is close to 1 dB except for B_3 of the male group where the mean error in estimates is 2 dB. This accuracy should be sufficient for any practical purpose and is remarkably high in view of the fact that each measured value is the mean of two subjects only with one or two determinations per sound. The

Table I- C-II. Male bandwidths

	Measurements			Measurements			Predictions			Error		
Vowel	F_1 Hz	F_2 Hz	F_3 Hz	B_1 Hz	B_2 Hz	B_3 Hz	B_1 Hz	B_2 Hz	B_3 Hz	B_1 dB	B_2 dB	B_3 dB
u	315	605	2450	63	43	42	57	35	39	+1	+2	+0.5
o	410	690	2600	38	34	40	43	39	54	-1	-1	-2.5
ɔ	505	900	2950	50	44	75	40	43	71	+2	0	+0.5
ɑ	600	1030	2700	45	42	100	40	52	86	+1	-2	+1
a	710	1150	2700	43	58	100	42	62	104	0	-0.5	-0.5
æ	700	1430	2450	37	61	89	42	65	106	-1	-0.5	-1.5
ε	450	1825	2425	41	58	119	42	73	96	0	-2	+2
e	325	1900	2450	43	43	155	56	51	92	-1.5	-0.5	+4.5
I	290	2025	2875	63	40	145	63	41	120	0	0	+1.5
i	230	2000	3000	72	35	105	86	37	129	-1.5	0	-2
y	245	1875	2075	87	75	80	84	86	78	+0.5	-1	0
ʉ	265	1480	2060	61	65	40	68	46	56	-1	+3	-3
ø	400	1650	2250	37	57	61	45	52	77	-2	+1	-2
œ	510	1150	2350	32	36	43	40	47	68	-2	-2	-4
θ	420	1000	2250	38	36	34	36	42	52	+0.5	-1.5	-3.5
ɜ	520	1500	2500	29	37	87	39	51	87	-2.5	-2.5	0

Table I-C-III. Female bandwidths

	Measurements			Measurements			Predictions			Error		
Vowel	F_1 Hz	F_2 Hz	F_3 Hz	B_1 Hz	B_2 Hz	B_3 Hz	B_1 Hz	B_2 Hz	B_3 Hz	B_1 Hz	B_2 Hz	B_3 Hz
u	365	690	2700	96	60	70	56	36	75	+4.5	+4.5	-0.5
o	445	780	2850	60	48	67	48	41	64	+2	+1	0
ɔ	610	1100	2800	56	49	90	43	53	87	+2	-0.5	0
ɑ	700	1100	3000	60	66	114	45	59	111	+2.5	+1	0
a	760	1360	3000	44	66	120	46	64	130	-0.5	0	-0.5
æ	730	1660	2850	48	67	150	44	66	111	+0.5	0	+2.5
e	395	2200	2850	50	67	190	52	51	128	-0.5	+2.5	+3.5
i	345	2060	3100	55	41	105	61	42	132	-1	0	-2
y	295	2000	2650	71	53	92	73	47	98	0	+1	-0.5
ʉ	305	1770	2380	77	59	62	70	48	77	+1	+1.5	-2
ø	410	1740	2350	54	59	80	51	53	84	+0.5	+1	-0.5
œ	525	1490	2600	44	54	75	44	51	87	0	+0.5	-1
θ	490	1240	2650	46	46	58	45	47	74	0	0	-2

positive error in the [u] domain, i.e. the underestimated B_1 and B_2, is especially apparent in the female data. It could be that the cavity wall losses become relatively more prominent under these conditions as outlined in the previous section or that frictional losses become excessive. The former explanation seems probable.

7. Final discussion

(1) The general theory of predicting formant levels and spectrum envelopes from formant frequency proposed by Fant (1956) and (1960) requires a predictability of formant bandwidths as an intermediate step. Eqs.(56-58) allow a prediction of bandwidths of the closed glottis vocal tract from F_1, F_2, F_3, F_4 with an average accuracy of 1 dB.

(2) In real speech the finite glottis impedance must be taken into account. As discussed by Fant (1960) any finite glottis impedance conjectured must be interpreted as a time average over a fundamental glottal period, whilst the instantaneous value of the bandwidth or damping during glottal closure would correspond to the ideal minimum. The extent to which the average glottal dissipation affects the energy level of a formant in speech is dependent on the extent to which the corresponding resonance stores energy in the pharynx cavity. Accordingly B_1 of any front vowel, or B_2 of [i] and [u], and B_3 of many vowels except [i] would have to be corrected with varying degrees that ultimately depend on the degree of opening and pressure drop at the glottis. A weak breathy voice would account for high damping and a strong effective voice a very small damping through glottis. A flow dependent increase of the resistance is also to be expected in real speech under conditions of extreme articulatory narrowing, Fant (1960). There remains much work to be done to quantify these effects.

(3) To simplify vocal tract computations of the F-pattern (F_1, F_2, F_3, F_4) and the associated B-pattern (B_1, B_2, B_3, B_4) it is recommended that the surface losses, i.e. friction and heat conduction, are estimated from overall vocal tract parameters as an alternative to the more elaborate complex matrix calculations. Radiation losses should be taken into account by the exact procedure. The effect of the wall mass load and associated losses can be handled by average formulas as a correction of the ideal hard walled tract data. Further work is needed to investigate more precisely the distribution of the wall load. A lumped

element representation at the glottal end of the line might provide improved accuracy at a suitable level of calculation cost.

(4) The superposition of sound from the external surface of the vocal tract walls and sound radiated from the mouth has been calculated for a simple tube resonator mass loaded at the glottal end. It is found that the wall output adds a low frequency level increase in the transfer function. In hyperbaric speech at pressures of the order of 15 ata or more the superposition may give rize to pronounced spectral zeros and a large low frequency boost.

References

van den BERG, Jw. (1953): Physica van de stemvorming met toepassingen, thesis, University of Groningen ('s-Gravenhage).

van den BERG, Jw. (1955A): "Transmission of the Vocal Cavities", J. Acoust.Soc.Am. 27, pp. 161-168.

van den BERG, Jw. (1955B): "Calculations on a Model of the Vocal Tract for Vowel /i/ (Meat) and on the Larynx", J.Acoust.Soc.Am. 27, pp. 332-338.

DUNN, H.K. (1961): "Methods of Measuring Vowel Formant Bandwidths", J.Acoust.Soc.Am. 33, pp. 1737-1746.

FANT, G. (1956): "On the Predictability of Formant Levels and Spectrum Envelopes from Formant Frequencies", pp. 109-120 in For Roman Jakobson ('s-Gravenhage).

FANT, G. (1960): Acoustic Theory of Speech Production ('s-Gravenhage) (2nd edition 1970).

FANT, G. (1961): "The Acoustics of Speech", pp. 188-201 in Proc. of the 3rd ICA Stuttgart, Vol. I (Amsterdam).

FANT, G. (1962): "Formant Bandwidth Data", STL-QPSR 1/1962, pp. 1-2.

FANT, G. and LINDQVIST, J. (1968): "Pressure and Gas Mixture Effects on Diver's Speech", STL-QPSR 1/1968, pp. 7-17.

FANT, G. and SONESSON, B. (1964): "Speech at High Ambient Air-Pressure", STL-QPSR 2/1964, pp. 9-21.

FLANAGAN, J.L. (1965): Speech Analysis Synthesis and Perception (New York).

FUJIMURA, O. and LINDQVIST, J. (1964A): "The Sinewave Response of the Vocal Tract", STL-QPSR 1/1964, pp. 5-10.

FUJIMURA, O. and LINDQVIST, J. (1964B): "Experiments on Vocal Tract Transfer", STL-QPSR 3/1964, pp. 1-7.

FUJIMURA, O. and LINDQVIST, J. (1965): "Sweep-Tone Measurements of the Vocal Tract Characteristics", STL-QPSR 1/1965, pp. 1-3.

FUJIMURA, O. and LINDQVIST, J. (1971): "Sweep-Tone Measurements of Vocal-Tract Characteristics", J.Acoust.Soc.Am. 49, pp. 541-558.

HOUSE, A.S. and STEVENS, K.N. (1958): "Estimation of Formant Band Widths from Measurements of Transient Response of the Vocal Tract", J. Speech and Hearing Research 1, pp. 309-315.

ICHIKAWA, A. and NAKATA, K. (1971): "Vocal Tract Resonances with Losses", paper 23 C 9, pp. 161-164 in Proc. of the 7th ICA Budapest, Vol. 3 (Budapest).

LAURENT, T. (1964): Frequency Filter Methods (New York).

SONDHI, M.M. and GOPINATH, B. (1971): "Determination of the Shape of a Lossy Vocal Tract", paper 23 C 10, pp. 165-167 in Proc. of the 7th ICA Budapest, Vol. 3 (Budapest).