Models for Approximating Basilar Membrane Displacement—Part II. Effects of Middle-Ear Transmission and Some Relations between Subjective and Physiological Behavior

By JAMES L. FLANAGAN

(Manuscript received December 26, 1961)

This report presents the second half of results of a study on the peripheral ear. There are two objectives: (1) to derive computational models for approximating the mechanical displacement of the basilar membrane when the sound pressure at the eardrum is known, and (2) to demonstrate certain relations between subjective behavior measured experimentally and physiological behavior calculated from the models. The report describes a rational function approximation of middle-ear transmission. This result, in combination with previously derived models for the inner ear, permits an analytical approximation of basilar membrane displacements in both apical and basal regions. Because the models are rational functions, they can, if desired, be simulated by lumped-constant electrical networks. Their computational tractability also permits straightforward approximations to temporal and spatial derivatives of displacement. Relations between computed membrane displacement and subjective behavior are illustrated for several psychoacoustic phenomena, namely pitch perception, binaural lateralization, binaural time-intensity trade, threshold discrimination, and pure-tone masking. The extent to which some of these phenomena can be correlated with, identified in, and predicted by the mechanical operation of the peripheral ear appears to be substantial.

Part I of this report[1] described three analytical models for approximating the displacement of the basilar membrane when the human ear is stimulated by sound. These models were valid for points lying roughly in the apical half of the membrane, that is, for frequencies less than about

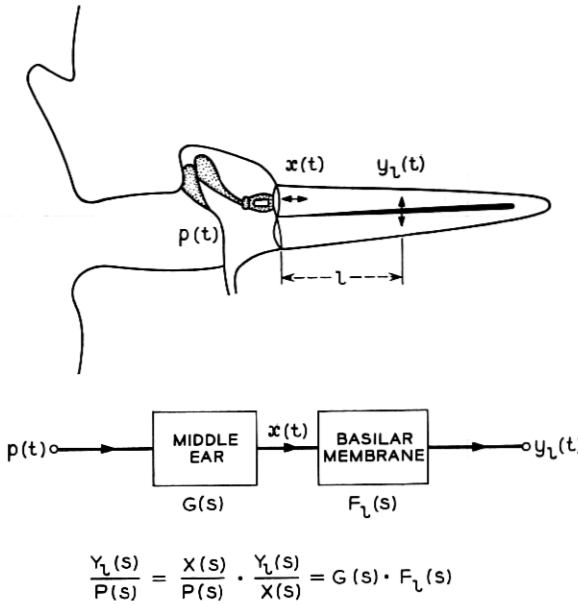

Fig. 1 — Schematic diagram of peripheral ear and functional relations between acousto-mechanical quantities.

1000 cps. Over this frequency range the elastic effects of the middle ear predominate, and the displacement of the stapes footplate is essentially proportional to, and in phase with, the sound pressure at the eardrum. At higher frequencies the mass and viscous properties of the middle ear become important, and the displacement transmission to the stapes is no longer constant with frequency. Applicability of the previously derived models to this range of frequencies depends upon being able to account for middle-ear transmission. This report describes an effort to derive a computational model for middle-ear transmission and to examine its relationship with the models for membrane displacement. Subsequent to this, an attempt is made to relate the mechanical operation of the ear, as described by the models, to several facets of subjective auditory behavior.

I. EFFECTS OF MIDDLE-EAR TRANSMISSION UPON MEMBRANE DISPLACEMENT*

The physiological functions to be considered are illustrated schematically in Fig. 1. $p(t)$ represents the sound pressure at the eardrum as a

* The material in this section was presented orally before the 60th meeting of the Acoustical Society of America, San Francisco, California, October, 1960. The abstract appears in J. Acoust. Soc. Am. **32,** 1960, p. 1494.

function of time; $x(t)$ is the equivalent linear displacement of the stapes footplate;* and $y_l(t)$ is the displacement of the basilar membrane (cochlea shown uncoiled) at a distance l from the stapes. In terms of frequency-domain (Laplace) transforms, the middle-ear transmission is represented by $G(s)$ and the stapes-to-membrane transmission by $F_l(s)$.

In deducing approximations to these functions, the peripheral ear is assumed both to be mechanically linear over the range of interest and to constitute a passive system. A passive system is stable by definition. It has no normal modes whose amplitudes increase indefinitely with time. The functions $G(s)$ and $F_l(s)$ can therefore be approximated by rational functions of frequency whose coefficients are real and whose poles and zeros are either real or occur in complex conjugates. The functions can have no poles with positive real parts and only simple poles with zero real parts.

The earlier paper[1] essentially treated functional approximations to $F_l(s)$ (that is, middle-ear transmission was assumed constant with frequency, or, in the present notation, $G(s) = 1$). Two of the previously derived models will be useful in the present discussion. They are the first and third which, according to the notation used earlier, were called $F_1(s)$ and $F_3(s)$. For convenience they are reproduced here and are:

$$F_1(s) = c_1\beta_l^{4+r}\left(\frac{s+\epsilon}{s+\gamma}\right)\left[\frac{1}{(s+\alpha_l)^2+\beta_l^2}\right]^2 e^{-s\frac{3\pi}{4\beta_l}}, \tag{1}$$

and

$$F_3(s) = c_3\beta_l^{4+r}\frac{\left[s^2+2\alpha_l s+\left(\alpha_l^2-\frac{\beta_l^2}{3}\right)\right]}{[(s+\alpha_l)^2+\beta_l^2]^3} e^{-s\frac{3\pi}{4\beta_l}}, \tag{2}$$

where

$s = (\sigma + j\omega)$ is the complex frequency,

β_l is the radian frequency to which the point l distance from the stapes responds maximally,

β_l^{4+r} is a factor which matches the physiologically measured variations in peak amplitude of displacement with resonant frequency β_l†,

$e^{-s\frac{3\pi}{4\beta_l}}$ is a delay factor of $3\pi/4\beta_l$ seconds which brings the

* In man, the stirrup does not move longitudinally as a planar piston but usually exhibits some rotational motion. $x(t)$ is taken here as the volume displacement of the footplate divided by its area.

† The present form of this factor is applicable only to the frequency range below 1000 cps. Here, as previously discussed,[1] the value of $r = 0.8$. A minor modification will be made in this factor presently to make it appropriate for higher frequencies.

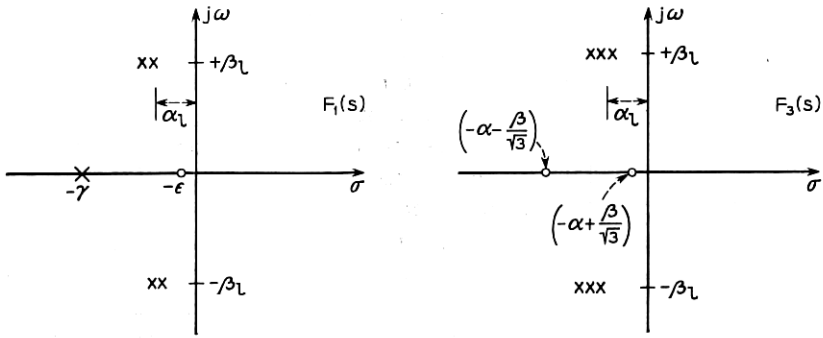

Fig. 2 — Pole zero diagrams for two functional approximations of $F(s)$.

phase delay of the model into line with the phase measured on the human ear. This factor is principally transit delay from stapes to point l on the membrane*.

The membrane characteristics are therefore approximated in terms of the poles and zeros of these two functions. Because the resonant properties of the membrane are nearly constant Q in character, the real and imaginary parts of the pole frequencies are related by a constant factor, i.e., $\beta_l = k\alpha_l$. For the present models, the best fits to the experimental data are obtained for the following choice of parameters:

$$\text{For } F_1(s): \frac{\epsilon}{\beta_l} = 0.1 \text{ to } 0.0\dagger, \qquad \text{For } F_3(s): \frac{\beta_l}{\alpha_l} = 1.7$$

$$\frac{\gamma}{\beta_l} = 1.0 \tag{3}$$

$$\frac{\beta_l}{\alpha_l} = 2.0.$$

Therefore, to within a multiplicative constant, the imaginary part of the pole frequency β_l completely describes the model. The pole-zero diagrams for the two models are shown in Fig. 2.

The real frequency responses of the models are evidenced by letting $s = j\omega$. If frequency is normalized in terms of $\zeta = \omega/\beta_l$, then the rela-

* At low frequencies the phase of the model departs somewhat from the experimental data. See the discussion of this point in Ref. 1 and also in J. L. Flanagan and C. M. Bird, Minimum Phase Responses for the Basilar Membrane, J. Acoust. Soc. Am. **34,** 1962, p. 114.

† See earlier comments about fitting phase response.

tive phase and amplitude responses of $F_1(j\zeta)$ and $F_3(j\zeta)$ are shown in Fig. 3 for the parameters stated in (3).

The inverse Laplace transforms of (1) and (2) are the displacement responses of the membrane to an impulse of displacement by the stapes. These representations will also be useful in the present discussion. If the mathematics is carried out the inverse transforms are found to be:

$$f_1(t) = c_1\beta_l^{1+r}\{[0.033 + 0.360\beta_l(t - T)]e^{-\frac{\beta_l(t-T)}{2}} \sin \beta_l(t - T)$$
$$+ [0.575 - 0.320\beta_l(t - T)]e^{-\frac{\beta_l(t-T)}{2}} \cos\beta_l(t - T) - 0.575e^{-\beta_l(t-T)}\} \quad (4)$$
$$\text{for} \quad t \geqq T; \quad \epsilon/\beta_l = 0.1$$
$$= 0; \qquad \text{for} \quad t < T,$$

and

$$f_3(t) = \frac{c_3\beta_l^{1+r}}{6}[\beta_l(t - T)]^2 e^{-\beta_l(t-T)/1.7} \sin \beta_l(t - T) \quad \text{for} \quad t \geqq T \quad (5)$$
$$= 0 \qquad \text{for} \quad t < T,$$

where the delay $T = 3\pi/4\beta_l$, as previously stated. In the earlier paper, the simplicity of $f_3(t)$ was the main reason that $F_3(s)$ was considered as an approximation to the experimental frequency domain data. A plot

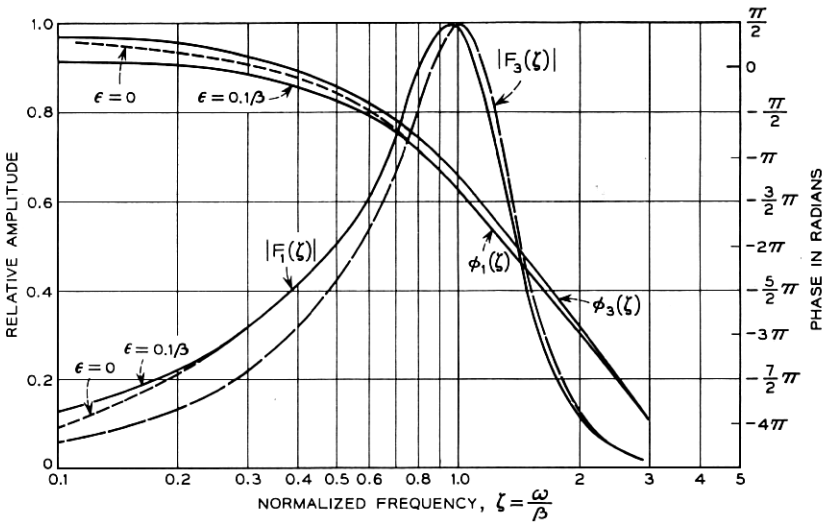

Fig. 3 — Amplitude and phase responses for two $F(s)$ models.

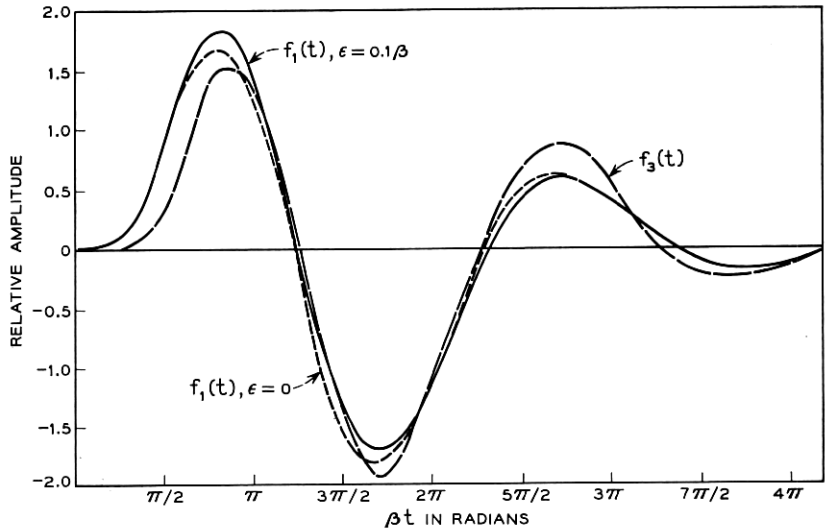

Fig. 4 — Impulse responses for the membrane models. These responses are the inverse transforms of the frequency data in Fig. 3 and represent membrane displacement caused by an impulse of stapes displacement.

of the responses (4) and (5), on a relative amplitude scale and with delay equalized, is shown in Fig. 4. The absolute time origins for the two traces are to the left of the relative origin by 1.9 and 1.5 radians respectively.

As indicated above, the factor β_l^{4+r} in (1) and (2) has a form appropriate to the frequency range below 1000 cps. If the membrane models are to be used at higher frequencies, this factor should be modified according to data given by Bekesy[2] on the peak membrane displacement as a function of frequency (see Fig. 4 in Ref. 1). For constant sinusoidal displacement of the stapes, Bekesy's data indicate that the peak membrane displacement increases at about 5 db/octave up to around 1000 cps, and then tends to flatten off and become roughly constant (at least up to about 2000 cps).

This amplitude variation can be accounted for by altering the multiplicative amplitude factor to $\beta_l^{4+r}(2\pi \cdot 1000/\beta_l + 2\pi \cdot 1000)^r$. The modification causes the peak response (of the curve shown in Fig. 3) to rise at about 5 db/octave below 1000 cps, and to flatten off above this frequency. At low β_l frequencies the amplitude factor is the same as before if the constant c_1 is readjusted by multiplying it by 2^r. With this minor modification, then, the functional approximations to $F_l(s)$ are appropriate for use at frequencies higher than 1000 cps.

1.1 *A Model for Middle-Ear Transmission*

To account for middle-ear transmission one would like an analytical specification of the stapes displacement produced by a given sound pressure at the eardrum for all frequencies of interest. Quantitative physioacoustical data on the operation of the middle ear are very sparse. The data which are available are due largely to Bekesy[3] and to Zwislocki.[4] By considering the topology of the mechanical circuit and the values of elastic, mass, and viscous constants measured in physiological preparations, it is possible to deduce information about the middle-ear transmission. Zwislocki used this approach to develop an analog electrical circuit for the middle ear in which voltage is analogous to pressure, and current to volume velocity. The circuit includes ten components representing the acousto-mechanical elements of the middle ear. Seven of the elements are energy storage elements.

Using the constants suggested by Zwislocki, we measured the transfer characteristics of the middle-ear circuit when terminated in an impedance analogous to the input mechanical impedance of the cochlea. For a constant pressure at the eardrum, the amplitude and phase responses of the stapes displacement are shown by the curves in Fig. 5.*

Although the characteristic equation corresponding to Zwislocki's analog circuit is of seventh degree, the stapes displacement can be analytically approximated reasonably well by a function of third degree. (As discussed in the earlier paper, the criterion of fit is again taken as an intuitive one.) Such an approximating function is of the form:

$$G(s) = \frac{c_0}{(s + a)[(s + a)^2 + b^2]}, \tag{6}$$

where c_0 is a positive real constant. [When combined with $F_t(s)$, the multiplying constants are chosen to yield proper absolute membrane displacement. For convenience, therefore, consider $c_0 = a(a^2 + b^2)$ so

* After the present work was carried out, an excellent paper by A. Møller (Network Model of the Middle Ear, J. Acoust. Soc. Am. **33,** 1961, p. 168) appeared in which analogous electrical circuits for the middle ear are deduced on the basis of input impedance measurements at the drum and the middle-ear topology. For a comparison with Zwislocki's data (which we had already used), we constructed several of Møller's circuits and measured their transfer characteristics. Although their frequency responses differ in fine detail, the results of Zwislocki and Møller agree in the gross aspects of the transmission characteristics. As do Bekesy's earlier results, both sets of the latter data suggest some uncertainty and variability in the middle-ear transmission, particularly in regard to the frequency at which the transmission begins to diminish appreciably. Apparently the function differs among individuals. One of the main objectives of the present paper, however, is to demonstrate a computational technique which has been found useful in explaining certain auditory phenomena. Whenever physiological data are improved and extended, the results can easily be incorporated into the analytical technique presented here.

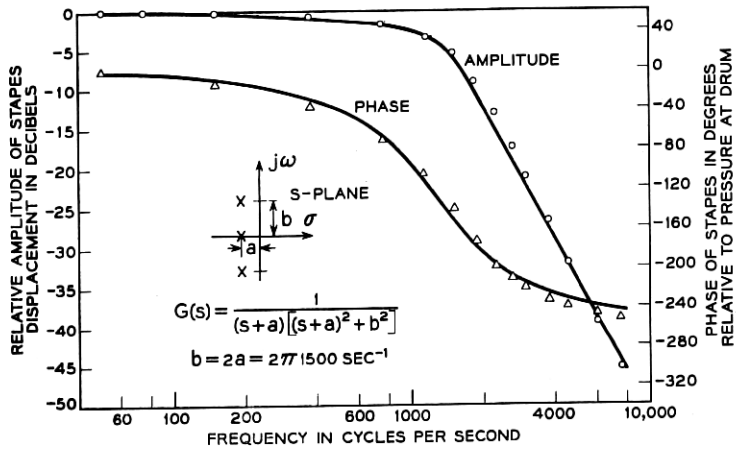

Fig. 5 — Functional approximation of middle ear transmission. The solid curves are from Zwislocki[4] and the plotted points are amplitude and phase values of $G(s)$.

that the low-frequency transmission of $G(s)$ is unity.] When the pole frequencies of $G(s)$ are related according to

$$b = 2a = 2\pi(1500) \text{ rad/sec}, \tag{7}$$

the fit to Zwislocki's data is shown by the plotted points in Fig. 5.

The inverse transform of (6) is the displacement response of the stapes to an impulse of pressure at the eardrum. It is obtained easily and will be useful in the subsequent discussion. Let

$$G(s) = G_1(s)G_2(s),$$

where

$$G_1(s) = \frac{c_0}{s + a}; \qquad G_2(s) = \frac{1}{(s + a)^2 + b^2}. \tag{8}$$

The inverses of the parts are:

$$g_1(t) = c_0 e^{-at}; \qquad g_2(t) = \frac{e^{-at}}{b} \sin bt. \tag{9}$$

The inverse of $G(s)$ is then

$$g(t) = \int_0^t g_1(\tau)g_2(t - \tau)\, d\tau,$$

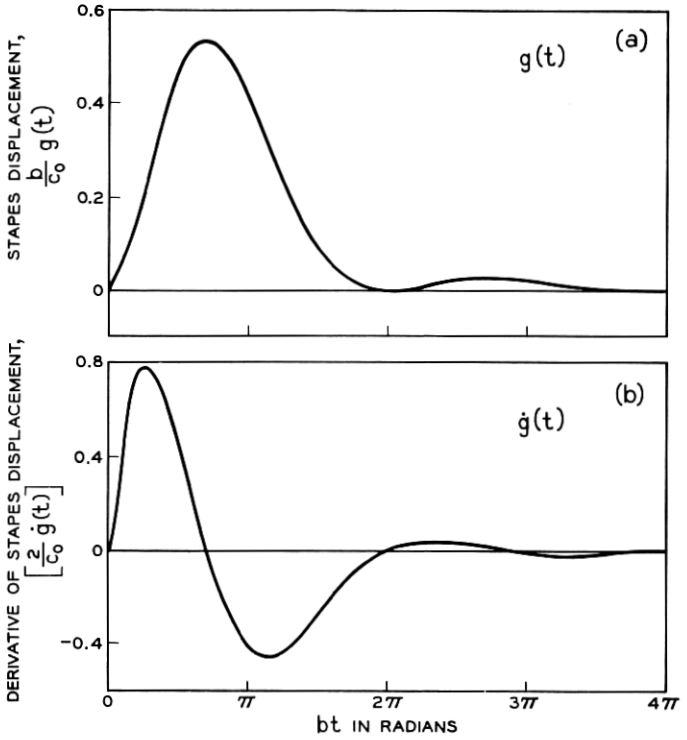

Fig. 6 — Displacement response of the stapes and its time derivative to an impulse of pressure at the eardrum.

or

$$g(t) = c_0 \frac{e^{-at}}{b} (1 - \cos bt) = \frac{c_0 e^{-bt/2}}{b} (1 - \cos bt). \tag{10}$$

Also for use at a future point in the discussion we note that the time derivative of the stapes displacement is:

$$\dot{g}(t) = \frac{c_0 e^{-bt/2}}{2} (2 \sin bt + \cos bt - 1). \tag{11}$$

Plots of $g(t)$ and $\dot{g}(t)$ are shown in Fig. 6.

1.2 *Combined Response of Middle Ear and Basilar Membrane*

The combined response of the models for the middle ear and basilar membrane is simply:

$$\begin{aligned} H_l(s) &= G(s)F_l(s) \\ h_l(t) &= g(t) * f_l(t). \end{aligned} \tag{12}$$

To simplify the computations and to illustrate both results, the response for model $F_1(s)$ will be computed in the frequency domain and that for $F_3(s)$ will be computed in the time domain.

1.2.1 *Inverse Transform of* $H_1(s)$

Disregarding for the moment the constant delay and amplitude terms, which can be resupplied at the end if needed, the problem of transforming $H_1(s) = G(s)F_1(s)$ amounts to computing the inverse of:

$$H_1'(s) = \frac{1}{s+a} \cdot \frac{1}{(s+a)^2 + b^2} \cdot \frac{s+\epsilon}{s+\gamma} \cdot \frac{1}{[(s+\alpha)^2 + \beta^2]^2}. \tag{13}$$

Expand $H_1'(s)$ as partial fractions:

$$H_1'(s) = \frac{A}{s+a} + \frac{Bs+c}{(s+a)^2+b^2} + \frac{D}{s+\gamma} + \frac{E(s)}{[(s+\alpha)^2+\beta^2]^2} \tag{14}$$

where

$$A = \frac{1}{b^2} \cdot \frac{\epsilon - a}{\gamma - a} \cdot \frac{1}{[(\alpha - a)^2 + \beta^2]^2}, \quad \gamma \neq a$$

$$B = 2 \text{ Re } B'$$

$$B' = \frac{1}{2b^2} \cdot \frac{\epsilon - a + jb}{\gamma - a + jb} \cdot \frac{1}{[(\alpha - a)^2 + \beta^2 - b^2 + j2b(\alpha - a)]^2}$$

$$C = [a(2 \text{ Re } B') - b(2 \text{ Im } B')]$$

$$D = \frac{1}{a - \gamma} \cdot \frac{1}{(a-\gamma)^2 + b^2} \cdot \frac{\epsilon - \gamma}{[(\alpha - \gamma)^2 + \beta^2]^2}$$

$$E(s) = (a_0 + a_1 s + a_2 s^2 + a_3 s^3).$$

On the basis of the previous findings the problem is particularized to the conditions:

$$\begin{aligned} \beta &= 2\alpha \qquad \text{Also let:} \quad \eta = \beta/b \\ b &= 2a \qquad\qquad\quad \gamma \neq a \\ \gamma &= \beta \qquad\qquad\quad \beta \neq b. \\ \epsilon &= 0 \end{aligned} \tag{15}$$

If the arithmetic is followed through, the constants are found to be:

$$A = -\frac{1}{b^6(2\eta - 1)(1.25\eta^2 - 0.50\eta + 0.25)^2}$$

$$B' = \frac{(0.50 - j1.00)}{2b^6(\eta - 0.50 + j1.00)[(1.25\eta^2 - 0.50\eta - 0.75) + j(\eta - 1.00)]^2}$$

$$B = 2 \text{ Re } B'$$

$$C = b(\text{Re } B' - 2 \text{ Im } B')$$

$$D = \frac{1}{b^6} \frac{1}{(\eta - 0.50)(\eta^2 - \eta + 1.25)(1.25)^2\eta^3},$$

and the coefficients of $E(s)$ are found to be:

$$\begin{aligned}
a_0 &= -\eta^3 b^2(3.12\eta bA + 1.25\eta C + 1.56bD) \\
a_1 &= -b^2\left[A(3.50\eta^2 - \eta + 0.25) + B(3.50\eta^2 - 2.00\eta - 0.25)\right. \\
&\quad \left. + D(2.5\eta^2) + C\left(\frac{2\eta - 1}{b}\right)\right] \qquad (16) \\
a_2 &= -[Ab(2\eta - 0.50) + Bb(2\eta - 1) + Db\eta + C] \\
a_3 &= -(A + B + D).
\end{aligned}$$

Although somewhat involved numerically, the inverse transform of $H_1'(s)$ can now proceed termwise as indicated in (14). The basic procedure from this point has already been indicated in the appendix of the earlier paper. When the details of this instance are carried through, the result is:

$$\begin{aligned}
h_1'(t) &= A\, e^{-bt/2} + B\, e^{-bt/2}(\cos bt - 0.50 \sin bt) \\
&\quad + \frac{C}{b}(e^{-bt/2} \sin bt) + D\, e^{-\eta bt} + (e^{-\eta bt/2} \sin \eta bt) \\
&\quad \cdot\left\{\frac{1}{2\eta^3 b^3}\left[a_0 - a_1\frac{\eta b}{2} + a_2(1.25\eta^2 b^2) - a_3(1.63\eta^3 b^3)\right]\right\} \\
&\quad + (\eta bt\, e^{-\eta bt/2} \sin \eta bt)\left\{\frac{1}{2\eta^2 b^2}[a_1 - a_2\eta b - a_3(0.25\eta^2 b^2)]\right\} \qquad (17) \\
&\quad + a_3\, e^{-\eta bt/2} \cos \eta bt + (\eta bt\, e^{-\eta bt/2} \cos \eta bt)\left\{\frac{1}{2\eta^3 b^3}\left[-a_0 + a_1\frac{\eta b}{2}\right.\right. \\
&\quad \left.\left. + a_2(0.75\eta^2 b^2) - a_3(1.38\eta^3 b^3)\right]\right\}; \quad \text{for} \quad t \geqq 0.
\end{aligned}$$

$h_1(t)$ is obtained from $h_1'(t)$ by resupplying the constant delay $T =$

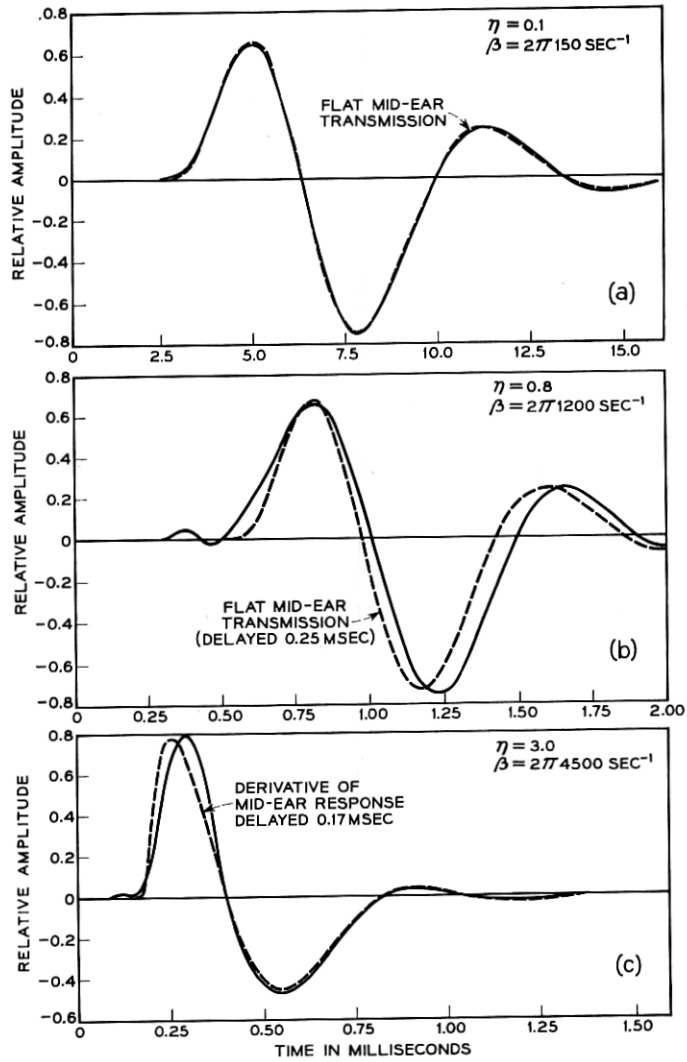

Fig. 7 — Displacement responses of apical, middle and basal points on the membrane to an impulse of pressure at the eardrum. These are computed from the inverse transform of $[G(s)F_1(s)]$.

$3\pi/4\beta$ and the multiplicative amplitude constants; that is, by letting $t = (t - T)$ and by multiplying $h'(t - T)$ by

$$[c_0 c_1 \beta_l^{4+r}(2\pi \cdot 1000/\beta_l + 2\pi \cdot 1000)^r], \text{ where } r = 0.8.$$

The form of the impulse response is thus seen to depend upon the

parameter $\eta = \beta/b$. Values of $\eta < 1.0$ refer to (apical) membrane points whose frequency of maximal response is less than the critical frequency of the middle ear. For these points the middle-ear transmission is essentially constant with frequency, and the membrane displacement is very nearly indicated by $f_1(t)$ in (4). On the other hand, values of $\eta > 1.0$ refer to (basal) points which respond maximally at frequencies greater than the critical frequency of the middle ear. For these points the middle-ear transmission is highly dependent upon frequency and would be expected to influence strongly the membrane displacement. To illustrate this point, (17) has been evaluated for $\eta = 0.1$, 0.8, and 3.0. The result, with the delay resupplied, is shown in Fig. 7.

For an impulse of pressure delivered to the eardrum, the three solid curves represent the membrane displacements at points which respond maximally to frequencies of 150, 1200, and 4500 cps. Each of the plots also includes a dashed curve. In Figs. 7(a) and 7(b), the dashed curve is the membrane displacement computed by assuming the middle-ear transmission to be flat with zero phase. [This is simply the response $\mathfrak{L}^{-1}F_1(s)$.] In Fig. 7(c) the dashed curve is the time derivative of the stapes displacement, $\dot{g}(t)$, taken from Fig. 6. The suggestion is that in the basal region, the form of the membrane displacement is very similar to the derivative of the stapes displacement. This apparently is the case, and this point will be considered again presently.

1.2.2 *Inverse Transform for* $H_3(s)$

If, as in the previous section, delay and scale constants are temporarily disregarded, the inverse transform for $[G(s)F_3(s)]$ is given by the time-domain convolution:

$$h_3'(t) = \left[(\beta t)^2 e^{-\frac{\beta t}{1.7}} \sin \beta t\right] * \left[\frac{e^{-\frac{bt}{2}}}{b}(1 - \cos bt)\right];$$

or

$$h_3'(t) = \int_0^t [(\beta\tau)^2 e^{-\frac{\beta\tau}{1.7}} \sin \beta\tau] \left\{\frac{e^{-\frac{b(t-\tau)}{2}}}{b}[1 - \cos b(t-\tau)]\right\} d\tau, \qquad (18)$$

for $t \geqq 0$.

When this integration is carried through, the result is:

$$h_3'(t) = \left(\frac{\eta}{b}\right)^2 [\mathrm{Im}(P) - \tfrac{1}{2}\,\mathrm{Im}(Q) - \tfrac{1}{2}\,\mathrm{Im}(R)]; \quad t \geqq 0 \qquad (19)$$

where

$$(P) = \frac{1}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j\eta\right]}\left[\left[e^{-\frac{\eta bt}{1.7}}\{\cos \eta bt + j \sin \eta bt\}\right.\right.$$
$$\cdot\left\{(bt)^2 - \frac{2bt}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j\eta\right]} + \frac{2}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j\eta\right]^2}\right\}$$
$$\left.\left. - \frac{2\, e^{-bt/2}}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j\eta\right]^2}\right]\right]$$

$$(Q) = \frac{1}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j(\eta - 1)\right]}\left[\left[e^{-\frac{\eta bt}{1.7}}\{\cos \eta bt + j \sin \eta bt\}\right.\right.$$
$$\cdot\left\{(bt)^2 - \frac{2bt}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j(\eta - 1)\right]} + \frac{2}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j(\eta - 1)\right]^2}\right\}$$
$$\left.\left. - \frac{2\, e^{-bt/2}(\cos bt + j \sin bt)}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j(\eta - 1)\right]^2}\right]\right]$$

$$(R) = \frac{1}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j(\eta + 1)\right]}\left[\left[e^{-\eta bt/1.7}\{\cos \eta bt + j \sin \eta bt\}\right.\right.$$
$$\cdot\left\{(bt)^2 - \frac{2bt}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j(\eta + 1)\right]} + \frac{2}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j(\eta + 1)\right]^2}\right\}$$
$$\left.\left. - \frac{2\, e^{-bt/2}(\cos bt - j \sin bt)}{\left[\left(\frac{1}{2} - \frac{\eta}{1.7}\right) + j(\eta + 1)\right]^2}\right]\right].$$

As before, $h_3(t)$ is obtained by resupplying the amplitude factors and the delay T. By way of examining the form of $h_3(t)$, (19) has been evaluated for $\eta = 0.1$ and 3.0. The resulting $h_3(t)$ is plotted in Fig. 8. Comparison with the previous response for $h_1(t)$ shows the results to be similar.

1.2.3 *Combined Frequency-Domain Responses*

The individual frequency-domain responses for $G(s)$ and $F_1(s)$ have been shown in Figs. 3 and 5 respectively. The combined response in the

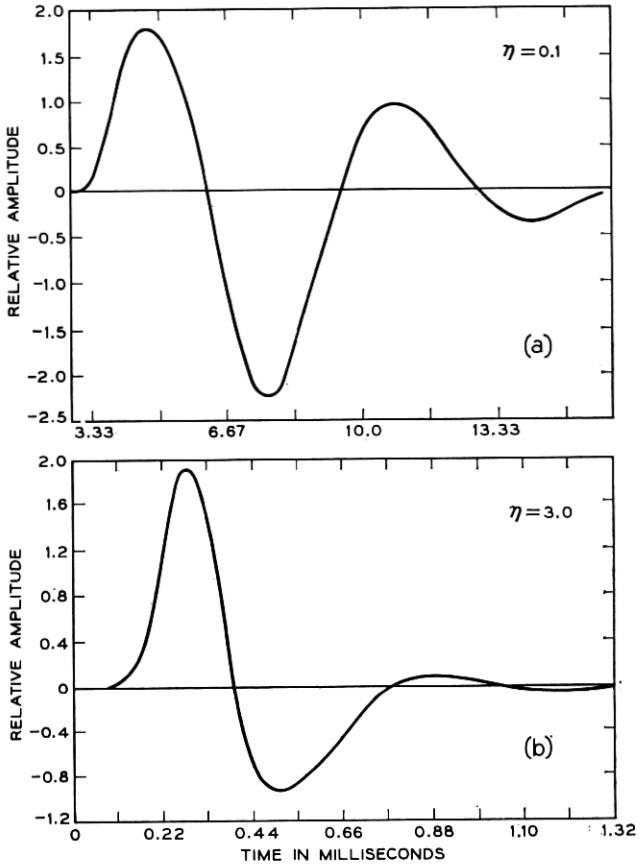

Fig. 8 — Impulse responses for apical and basal points computed from $[g(t) * f_3(t)]$.

frequency domain is simply the sum of individual amplitude (in db) and phase (in radians) responses. The combined amplitude and phase responses for the model $G(s)F_1(s)$ are shown in Figs. 9 and 10, respectively.

As already indicated by the impulse responses, one sees that the response of apical (low-frequency) points on the membrane is given essentially by $F(s)$, while for basal (high-frequency) points the response is considerably influenced by the middle-ear transmission $G(s)$. In particular, notice two things about the frequency response of the membrane model [i.e., $F(\omega)$]. One, the low-frequency skirt of the amplitude curve rises at about 6 db/octave. And two, the phase of the membrane

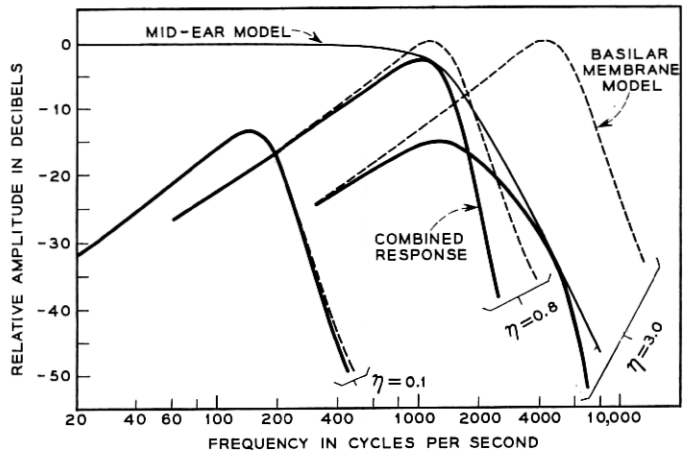

Fig. 9 — Frequency responses for the combined models $[G(\omega)F_1(\omega)]$.

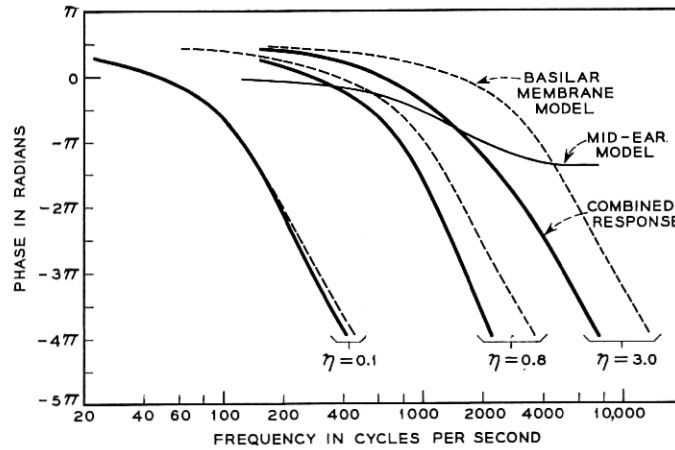

Fig. 10 — Phase responses for the combined models $[G(\omega)F_1(\omega)]$.

model [i.e., $F(\omega)$] approaches $+\ \pi/2$ radians at frequencies below the peak amplitude response. In other words, at frequencies appreciably less than its peak response frequency, the membrane function $F(\omega)$ behaves approximately as a differentiator.

Because the middle-ear transmission begins to diminish in amplitude at frequencies above about 1500 cps, the membrane displacement in the basal region is roughly the *time derivative of the stapes displacement.* The

waveform of the impulse response along the basal part of the membrane is therefore approximately constant in shape. Along the apical part, however, the impulse response oscillates more slowly as the apex is approached. This has already been illustrated in Fig. 7. [If the apical response is considered on a time scale normalized in terms of (βt), then the displacement waveform is constant in shape.] This relation can, and has been, supported by psychoacoustic measurements. These results will be discussed in the second part of the paper.

Notice one other thing from Fig. 9. Because the amplitude response of the middle ear declines appreciably at high frequencies, the amplitude response of a basal point is highly asymmetrical (for example, the combined response for $\eta = 3.0$.) The result is that a given basal point, while responding with greater amplitude than any other membrane point at its characteristic frequency (i.e., at β_l), responds with greatest amplitude (but not greater than some other point) at some lower frequency.

1.3 *Some Temporal and Spatial Relations For Membrane Displacement*

Certain results from physiological research[5] suggest that shear stresses along the basilar membrane may be as important in the mechanical-to-neural transduction as absolute displacements of the membrane. Accordingly, the spatial derivative of the displacement may be the mechanical factor of consequence. The computational tractability of the model permits a straightforward consideration of some temporal and spatial relations for the displacement.*

As a beginning, because they are easiest to talk about, consider only apical membrane points where the middle-ear transmission is essentially constant. In this case the displacement is nearly $f(t)$ [see (4) and (5)], and is only a function of t and the point parameter β. The variable β is a function of the distance along the membrane and can be so specified. (This functional relationship will be developed presently.) The impulse response is essentially a function of the product $\beta(t - T)$ and has a multiplicative factor involving β^{1+r} (i.e., $\beta^{1+r}g[\beta(t - T)]$). This fact points up a simple aspect of the dispersive nature of the basilar membrane.

If a disturbance is propagating in a nondispersive medium, the wave moves with a velocity which is the same for all frequency components, and the waveform is maintained undistorted. Let the wave for a one-dimensional situation be $p(t,x) = p(ct - x)$, where c is the velocity.

* See the further discussion of spatial derivatives (displacement gradients) in Section II.

Then,

$$\frac{\partial p}{\partial t} = c\,\frac{\partial p}{\partial(ct - x)}, \quad \text{and} \quad \frac{\partial p}{\partial x} = -\frac{\partial p}{\partial(ct - x)}, \tag{20}$$

and the time and space derivatives have the same waveform. The corresponding relations for the displacement responses of the membrane, however, must differ somewhat in time waveform. The model $f_3(t)$, Eq. (1.5), because of its simplicity, is particularly useful for illustrating this.

Again neglecting the amplitude constants which do not involve β or t, and which can be resupplied in the result, $f_3(t)$ reduces to:

$$f_3''(t) = \beta^{1+r} g[\beta(t - T)]; \quad t \geqq T$$

where

$$g[\beta(t - T)] = [\beta(t - T)]^2 e^{-\beta(t-T)/1.7} \sin \beta(t - T), \tag{21}$$

and

$$T = 3\pi/4\beta.$$

Then,

$$\frac{\partial f_3''}{\partial \beta} = \beta^r \left[\beta \frac{\partial g}{\partial \beta} + (1 + r)g\right]. \tag{22}$$

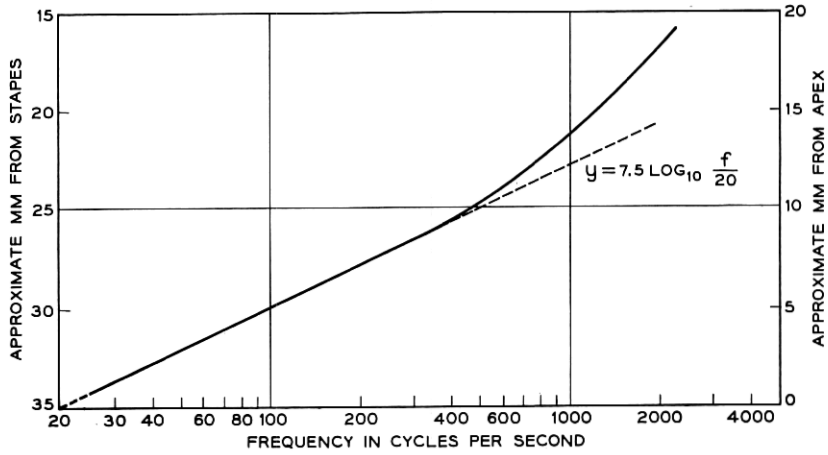

Fig. 11 — Location of peak displacement of basilar membrane as a function of frequency (after Bekesy[3]).

When the differentiation is carried out the result is:

$$\frac{\partial f_3''}{\partial \beta} = \beta^r[\beta(t - T)]^2 e^{-\beta(t-T)/1.7}\left[\beta t \cos \beta(t - T) + \left(1 + r + \frac{2\beta t}{\beta(t - T)} - \frac{\beta t}{1.7}\right)\sin \beta(t - T)\right]; \quad t \geqq T. \tag{23}$$

As indicated earlier, the functional relation between the frequency β and the distance along the membrane is needed to put (23) into the form of a space derivative. Bekesy[3] gives data on the place of maximal displacement along the membrane as a function of frequency. These data are replotted in Fig. 11. For purposes of the present discussion, the data for frequencies less than about 1000 cps are of main interest. If the basilar membrane is assumed to be 35 mm long, and if distance is now reckoned from the *apical* end, the low-frequency data are reasonably well approximated by:

$$x = 7.5 \log_{10} \frac{\beta}{40\pi}, \tag{24}$$

where x is the distance from the apex in mm. This line is drawn in Fig. 11.

It is now easy to compute

$$\frac{\partial f_3''}{\partial x} = \frac{\partial f_3''}{\partial \beta} \cdot \frac{\partial \beta}{\partial x},$$

where

$$\frac{\partial \beta}{\partial x} = \frac{\beta}{7.5 \log_{10} e}, \qquad = 0.31\beta. \tag{25}$$

Applying this result to (23) yields:

$$\frac{\partial f_3''}{\partial x} = 0.3\beta^{1+r}[\beta(t - T)]^2 e^{-\beta(t-T)/1.7}\left[\beta t \cos \beta(t - T) + \left(1 + r + \frac{2\beta t}{\beta(t - T)} - \frac{\beta t}{1.7}\right)\sin \beta(t - T)\right]. \tag{26}$$

Except for the constant amplitude factor, this is the spatial derivative of the impulse response for apical membrane points. It is plotted in Fig. 12. One notices its form is not radically different from the displacement.

The time derivative follows directly from (21), and is:

$$\frac{\partial f_3''}{\partial t} = \beta^{1+r} \frac{\partial g}{\partial t} = \beta^{2+r} \frac{\partial g}{\partial \beta t},$$

or,

$$\frac{\partial f_3}{\partial t} = \beta^{2+r}[\beta(t - T)]e^{-\beta(t-T)/1.7}\left[\beta(t - T)\cos\beta(t - T) + \left(2 - \frac{\beta(t - T)}{1.7}\right)\sin\beta(t - T)\right]; \quad t \geqq T. \tag{27}$$

This function, except for amplitude factor, is the time derivative of the apical impulse response. It is plotted in Fig. 13. One notices that for a given (apical) point on the membrane, the time derivative of displacement is not greatly different in form from the spatial derivative. As men-

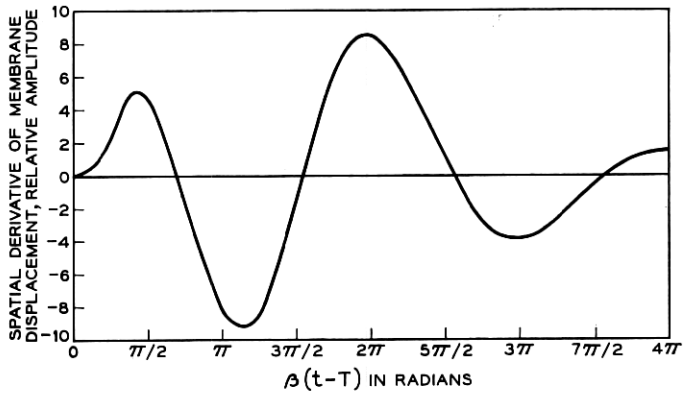

Fig. 12 — First spatial derivative of membrane displacement.

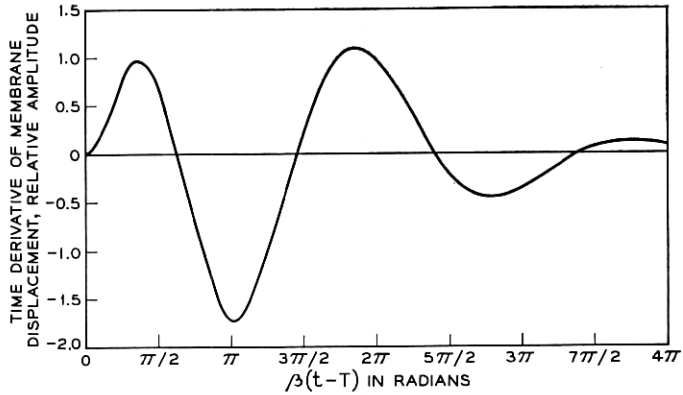

Fig. 13 — First time derivative of membrane displacement.

tioned earlier, the derivatives would have the same form if all frequency components propagated at the same velocity.

It also is of interest to consider the frequency-domain correlate of the spatial derivative. In this case it is equally easy to begin in a general way and not initially restrict the discussion to the apical region. For the model $H_1(s)$ the impulse response can be written in terms of its Laplace transform:

$$h(t) = \frac{1}{2\pi j} \int_{-j\omega}^{j\omega} H_1(s)\, e^{st}\, ds,$$

where

$$H_1(s) = G(s)F_1(s), \tag{28}$$

and where $G(s)$ and $F_1(s)$, the latter a function of the point parameter β, have been specified previously. The spatial derivative, in terms of the frequency parameter β, is therefore:

$$\frac{\partial h}{\partial \beta} = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} G(s)\, \frac{\partial F_1(s,\beta)}{\partial \beta}\, e^{st}\, ds. \tag{29}$$

The quantity of interest is $\partial F_1/\partial \beta$. From the previous discussion:

$$F_1(s,\beta) = c_1 \beta^{4+r} \left(\frac{2000\pi}{\beta + 2000\pi}\right)^r \left(\frac{s+\epsilon}{s+\gamma}\right) \left[\frac{1}{(s+\alpha)^2 + \beta^2}\right]^2 e^{-\frac{3\pi s}{4\beta}}. \tag{30}$$

Taking, as earlier indicated, $\beta = 2\alpha$, $\gamma = \beta$, and $\epsilon = 0$, and carrying through the differentiation gives:

$$\frac{\partial F_1}{\partial \beta} = F_1(s,\beta) \left[\frac{4+r}{\beta} - \frac{r}{(\beta + 2000\pi)} + \frac{3\pi s}{4\beta^2} - \frac{2(s+2.5\beta)}{(s+0.5\beta)^2 + \beta^2} - \frac{1}{s+\beta}\right]. \tag{31}$$

If we consider the result for real frequencies (i.e., $s = j\omega$) and normalize frequency by letting $\zeta = \omega/\beta$, then (31) becomes:

$$\frac{\partial F_1}{\partial \beta} = F_1(\zeta,\beta)\, \frac{1}{\beta} \left[\frac{[4\beta + 2000\pi(4+r)]}{(\beta + 2000\pi)} + j\zeta\, \frac{3\pi}{4} - \frac{2(2.5 + j\zeta)}{(1.25 - \zeta^2 + j\zeta)} - \frac{1}{1 + j\zeta}\right]. \tag{32}$$

This can be put in terms of the spatial derivative (at least for apical points) by applying (25). If this is done, then the spatial derivative be-

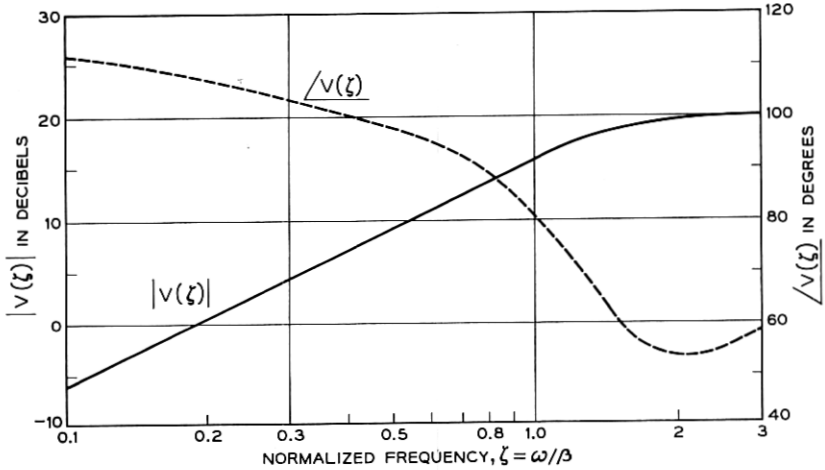

Fig. 14 — Frequency domain correlate, $V(\zeta)$, of the first spatial derivative.

comes:

$$\frac{\partial F_1}{\partial x} = (0.3)F_1(\zeta,\beta)\left[4 + \left(\frac{2000\pi r}{\beta + 2000\pi}\right) + j\zeta\frac{3\pi}{4} - \frac{2(2.5 + j\zeta)}{(1.25 - \zeta^2 + j\zeta)} - \left(\frac{1}{1 + j\zeta}\right)\right] \quad (33)$$

$$= 0.3F_1(\zeta,\beta)V(\zeta).$$

For points lying in the apical half of the membrane, therefore, (i.e., $\beta < 2000\pi$) the frequency-domain representation of the spatial derivative is simply $F_1(\zeta)$ (see Fig. 3) multiplied by the bracketed factor $V(\zeta)$ in (33). The phase and amplitude of this factor for $\beta \ll 2000\pi$ are plotted in Fig. 14.

One notices that for $\zeta < 1$ the bracketed term, to a crude approximation, is similar to a time differentiation. That is, the amplitude variation is +6 db/octave and the phase is $+\pi/2$. This indication is consonant with the previous time-domain results shown in Figs. 12 and 13.

1.4 *An Electrical Circuit for Simulating Basilar Membrane Displacement**

On the basis of the relations developed in the previous sections [Eqs. (6) and (30), for example], it is possible to construct electrical circuits

* The material in this section was presented orally before the Sixty-Second Meeting of the Acoustical Society of America, Cincinnati, Ohio, November, 1961. The abstract appears in J. Acoust. Soc. Am., **33,** 1961, p. 1670.

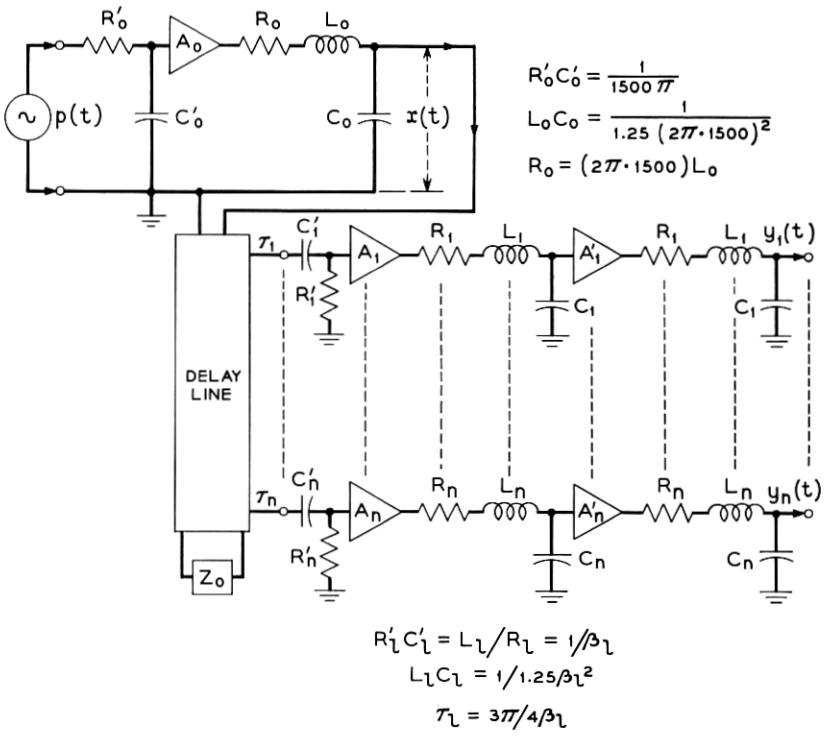

Fig. 15 — Electrical network representation of the model $[G(s)F_1(s)]$.

whose transmission properties are identical to those of the functions $G(s)$ and $F_1(s)$. This is most easily done by representing the critical frequencies in terms of simple cascaded resonant circuits. The additional phase delay can be supplied by means of an electrical delay line. A simulation of $G(s)$ as given in (6) and $F_1(s)$ as given in (30) (for $\epsilon = 0$) is shown in Fig. 15. The voltage at an individual output tap represents the membrane displacement at a specified distance from the stapes. The electrical voltages analogous to the sound pressure at the eardrum and to the stapes displacement are also indicated. The buffer amplifiers labelled A have fixed gains which take account of the proper multiplicative amplitude constants.

The circuit elements are selected according to the relations stated for $G(s)$ and $F_1(s)$. [See Eqs. (3) and (7).] For example, the procedure can be as follows. For the middle-ear simulation, choose a convenient R_0', say 10K. Then, because $b = 2a = 2\pi \cdot 1500$, and because $a = 1/R_0'C_0'$,

$$C_0' = 0.02\ \mu f.$$

Select a convenient L_0, say 2h. Then, because $1.25b^2 = 1/L_0C_0$ and $a = R_0/2L_0$,

$$C_0 = 0.005\ \mu f,$$

and

$$R_0 = 19\text{K}.$$

The components for the basilar membrane networks are chosen in the same way. In this case:

$$\beta_l = \frac{1}{R_l'C_l'}, \qquad 1.25\beta_l^2 = \frac{1}{L_lC_l},$$

and

$$\alpha_l = \frac{R_l}{2L_l}.$$

Consider, for example, the membrane point which responds maximally to 4500 cps (i.e., $\beta_l = 2\pi \cdot 4500$). For convenience take R_l' and L_l as 10K and 1h, respectively. Then:

$$C_l' = 0.0035\ \mu f$$

$$C_l = 0.001\ \mu f$$

$$R_l = 28K.$$

For each membrane point the relative gains of the amplifiers are set to satisfy the amplitude relations implied in Fig. 9. This takes account of the constant multiplying factors in the model specification.

Some representative impulse responses of the analog circuit of Fig. 15 are shown in Fig. 16(a). One notices the degradation in time resolution as the response is viewed at points more apicalward.

As indicated earlier, the spatial derivative may figure in the conversion of mechanical to neural activity. In previous psychoacoustic work[6] it was found useful to approximate the first spatial derivative by a finite difference. The present circuit can be used to provide such an approximation by taking the differences between the deflections of adjacent, uniformly spaced points. Fig. 16(b) shows first-difference approximations to the spatial derivative obtained from the analog circuit by taking:

$$\frac{\partial y}{\partial x} \cong \frac{y(t,x + \Delta x) - y(t,x)}{\Delta x},$$

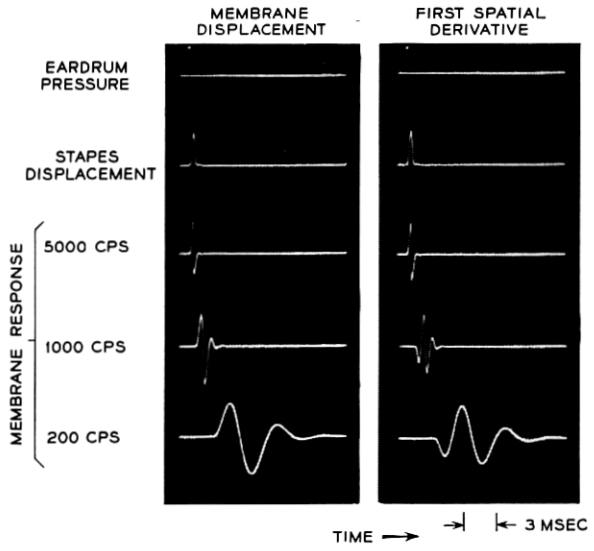

Fig. 16 — (a) Impulse responses measured on the network of Fig. 15; (b) first difference approximations to the spatial derivative measured from the network of Fig. 15.

with $\Delta x = 0.3$ mm. These responses can be compared (for apical points) with the calculated derivative in Fig. 12. Because of amplification, the polarity of the derivative traces in Fig. 16(b) is inverted from that shown in Fig. 12.

II. SOME RELATIONS BETWEEN SUBJECTIVE AND PHYSIOLOGICAL BEHAVIOR

The preceding discussion derived computationally tractable models for the operation of the middle ear and basilar membrane. Can these models be used to further our understanding of auditory subjective behavior? In particular can they help us to relate psychoacoustic phenomena to the physical operation of the peripheral ear?

The models describe only the mechanical functioning of the ear. Any comprehensive hypothesis about auditory perception must make provisions for the transduction of mechanical displacement into neural activity. The details of this process are not well understood and the assumptions that presently can be made must be of an approximate and simplified nature. Three such assumptions will be useful to us. Although

gross simplifications, they do not seem to violate known physiological facts.

The first assumption (actually a fact) is that sufficient local deformation of the basilar membrane elicits neural activity in the terminations of the auditory nerve at the organ of Corti. Such neural activity may be in the form of volleys triggered synchronously with the stimulus, or in the form of a signaling of place localization of displacement. Implicit in this is the notion that the displacement, or perhaps spatial derivatives of displacement,[5] must exceed a certain threshold before nerve firings take place.* The number of neurons activated depends upon amplitude of membrane displacement in a monotonic fashion. Psychological and physiological evidence suggests that the intensity of the neural activity is a power-law function of the mechanical displacement. A single neuron is presumably a binary (fired or unfired) device. It is refractory for a given period after firing; hence a limit exists upon the rate at which it can fire. Large populations of neurons, all of which are not refractory at the same time, can give rise to neural volleys at rates greater than the maximum rate of a single element.

Second, neural firings occur on only one "polarity" of the displacement, or of the spatial derivative.[7] In other words, some process like half-wave rectification operates on the displacement function, or on its spatial

* Earlier, in Section 1.3, it was suggested that the spatial derivative of displacement, as well as the displacement, may be important in the mechanical-to-neural conversion process. Further explication of this allusion and the present one is necessary.

Electrophysiological experiments on guinea pig [G. von Bekesy, J. Acoust. Soc. Am., **25** (1953) p. 786; H. Davis, Ann. Oto. Rhin. Laryn. **67** (1958) p. 789.] suggest that the outer and inner hair cells of the organ of Corti differ in their sensitivities to mechanical stimulation. The outer hair cells are sensitive to bending only in a direction transverse to the long dimension of the membrane. Further than this, only outward bending of the hairs (away from arch of Corti) produces an electrical potential in the scala media favorable for exciting the auditory nerve endings. This outward bending is produced on upward motions of the basilar membrane — that is, motions which drive it toward the tectorial membrane and produce a relative shear.

On the other hand, the inner hair cells, which reside between the arch of Corti and the axis of the cochlear spiral, are sensitive to bending in a direction parallel to the long dimension of the membrane. In this case only bending toward the apex of the cochlea produces a scala media potential favorable for stimulating the nerve. So far as a given point on the membrane is concerned, the inner hair cells are essentially sensitive to the longitudinal gradient of displacement — that is, to the spatial derivative in the long dimension. Furthermore, the inner cells fire only on that polarity of the gradient which corresponds to bending toward the apex. Threshold for firing of the inner cells apparently is about 20 db higher than that for the outer cells.

If this behavior is common to the human ear, displacement gradient, as well as displacement may be significant. As the results of Section 1.3 show, the displacement and the spatial derivative have gross features which are similar. For this reason the hypotheses and arguments to be put forward in this section generally can apply equally to displacement and gradient.

derivatives. Third, the membrane point displacing with the greatest amplitude originates the predominant neural activity. (More strictly, perhaps, this is the point experiencing the greatest transverse and longitudinal bending.) The latter may also operate to suppress or inhibit activity arising from neighboring points.

These assumptions, along with the results from the models, have in a number of instances been helpful in interpreting subjective auditory behavior. Without going into any case in great depth, let us consider several of these instances.

2.1 *Pitch Perception*

Pitch is that subjective attribute which admits of a rank ordering on a scale ranging from low to high. As such, it correlates strongly with objective measures of frequency. One important facet of auditory perception is the ability to assign pitch to sounds which exhibit time periodicity.

Consider first the pitch of pure (sinusoidal) tones. For such stimuli the basilar membrane displacement at any point is sinusoidal. The frequency responses given previously in Fig. 9 indicate the relative amplitudes of displacement versus frequency for different membrane points. At any given frequency, one point on the membrane responds with greater amplitude than all others. In accordance with the previous assumptions, the most numerous neural volleys are elicited at this maximum point. For frequencies sufficiently low (less than about 1000 cps) they are triggered once per cycle and at some fixed epoch on the displacement waveform. Subsequent processing by higher centers presumably appreciates the periodicity of the stimulus-locked volleys. For frequencies greater than about 1000 to 2000 cps, electro-physiological evidence suggests that synchrony of neural firings is not maintained.[9] Pitch is apparently perceived through a signaling of the place of greatest membrane displacement or displacement gradient. The poorer frequency resolution of points lying in the basal part of the basilar membrane probably also contributes to the psychoacoustic fact that pitch discrimination becomes less acute at higher frequencies.[10]

Suppose the periodic sound stimulus is not a simple sinusoidal tone but is more complex, say repeated sharp pulses. What pitch is heard? For purpose of illustration, imagine the stimulus to be alternately positive and negative periodic impulses. Such a pulse train has a spectrum which is odd-harmonic. Pulse rate and fundamental frequency are in the ratio of two-to-one. If the pulses occur slowly enough, the membrane

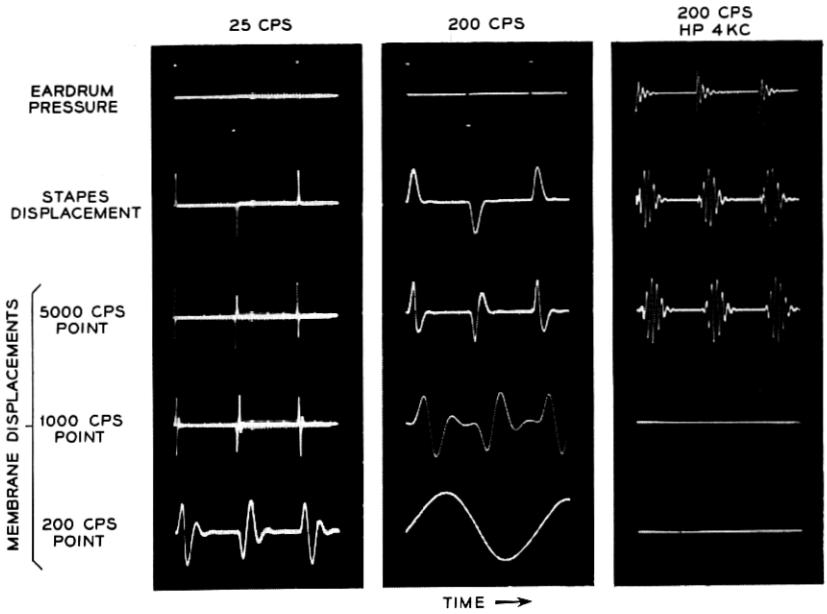

Fig. 17 — Displacement responses for alternate positive and negative pulses simulated by the network of Fig. 15.

displacements at all points along its length will resolve each pulse in time. That is, the membrane will have time to execute a complete, damped impulse response at all points for each pulse, positive or negative. If, however, the fundamental frequency of the train is sufficiently high, the fundamental component will be resolved (in frequency) at the most apically responding point. This situation is illustrated by the traces in the first and second columns of Fig. 17. These waveforms were measured on analog networks as illustrated in Fig. 15. The oscilloscope gain was adjusted for constant peak-to-peak amplitude to display the waveforms more effectively. The proper relative amplitudes are therefore not indicated in the traces.

For the low pulse-rate condition (25 cps fundamental) in the first column, one might imagine that neural firing synchronous with each pulse, regardless of polarity, would be triggered at all points along the membrane. The perceived pitch might be expected to be that of the pulse rate, and it is.[6] For such stimulation, the models indicate that the greatest membrane displacements occur near the middle portion, in the region maximally responsive to 1000 to 2000 cps.

In the second column, the fundamental frequency is 200 cps. This is

high enough for the apical end of the membrane to resolve the fundamental component. The displacement of the 200 cps point on the membrane is the fundamental sinusoid, while the more basal points continue to resolve each pulse in time. At the apical end, neural volleys might be expected to be triggered synchronously at the fundamental frequency, while toward the basal end the displacements favor firings at the pulse rate. For this condition, the apical fundamental-correlated displacements are generally of greater amplitude and subjectively more significant than the basal, pulse-rate displacements. The fundamental-rate volleys generally predominate, and a pitch is heard corresponding to 200 sec^{-1}.

If this same stimulus is high-pass filtered at a sufficiently high frequency, only the basal displacements remain effective in producing the pitch percept. If the present arguments continue to hold, this filtering should again give rise to a pulse-rate pitch because the time resolution in the basal end separates each pulse, whether positive or negative. Psychoacoustic measurements show this in fact to be the case.[11] Representative membrane displacements for this condition, as given by the models, are shown in the third column of Fig. 17.

A slightly more subtle effect is obtained if the high-pass filtering is made at a low harmonic number, for example, at the second harmonic so as to remove only the fundamental component. Under certain of these conditions the significant membrane displacement can be seen to exhibit displacements that favor fundamental-rate neural activity. The pitch percept would then be expected to be the fundamental, even though the fundamental is not present in the stimulus. Again psychoacoustic measurements give this result.[12] The effect is the so-called residue pitch.

Another of the many variations of pulse stimuli, but one which is diagnostically useful in exploring pitch perception, is the periodic, unipolar impulse train in which the equispaced pulses have amplitudes (areas) alternately a and b. Such a stimulus exhibits an infinite number of complex spectral zeros, the imaginary parts of which occur at every other spectral line. The spectral envelope is cycloidal and is described by:

$$K(s) = (a + be^{-sT/2}), \tag{34}$$

where T is the fundamental period. The spectral zeros lie at

$$s = -\frac{2}{T} \ln |a/b| \pm j\frac{2(2n+1)\pi}{T},$$

and the ratio of odd-line amplitude to even-line amplitude is:

$$R = \frac{|K(s)|_{s=j2\pi/T}}{|K(s)|_{s=j4\pi/T}} = \left[\frac{a-b}{a+b}\right]. \tag{35}$$

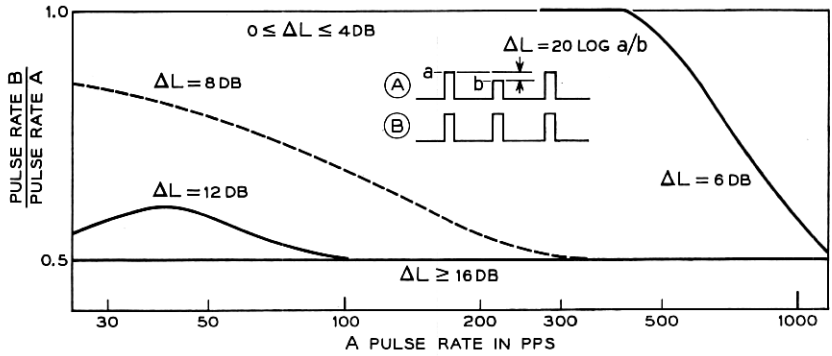

Fig. 18—Subjective pitch assigned to a periodic pulse train composed of sharp pulses with alternate amplitudes a and b. $\Delta L = 20 \log a/b$. The pitch equation is determined by matching the uniform pattern B to the test pattern A.

One psychophysical question that could be posed is "For a given sensation level, fundamental frequency and a/b ratio, what is the pitch?" When this question is answered by means of a pitch-matching experiment, the result for several values of the variables is shown in Fig. 18. These results are for a sensation level of approximately 45 db. When the a/b ratio is less than about 4 db, one never hears any pitch except the pulse rate. On the other hand, when the a/b ratio is greater than about 16 db, one never hears any pitch but one-half the pulse rate, i.e., the fundamental. Between these level differences, a transition from one pitch mode to the other takes place. The transition depends upon fundamental frequency as shown in the figure. As in the previous case, calculations and observations with the analog networks show the correlation between these modes and the displacement patterns of the basilar membrane. Unlike the situation depicted previously in Fig. 17(b), however, a pitch percept equivalent to half the pulse rate does not necessarily mean that the fundamental frequency is resolved by the membrane.

A somewhat different example of pitch stimulus is periodically interrupted random noise. Under certain conditions of interruption rate, duty factor and frequency content, chopped noise possesses a pitch.[13] It is relevant to consider how such a signal is represented in the mechanical displacements of the basilar membrane.

Because of the ear's frequency characteristics, a broad-spectrum noise would be expected to produce the greatest displacements somewhere near the middle of the membrane, around the 1000 to 2000 cps point. Let us look at these displacements for a flat-spectrum noise which is chopped with constant duty factor of 0.2. The response waveforms for

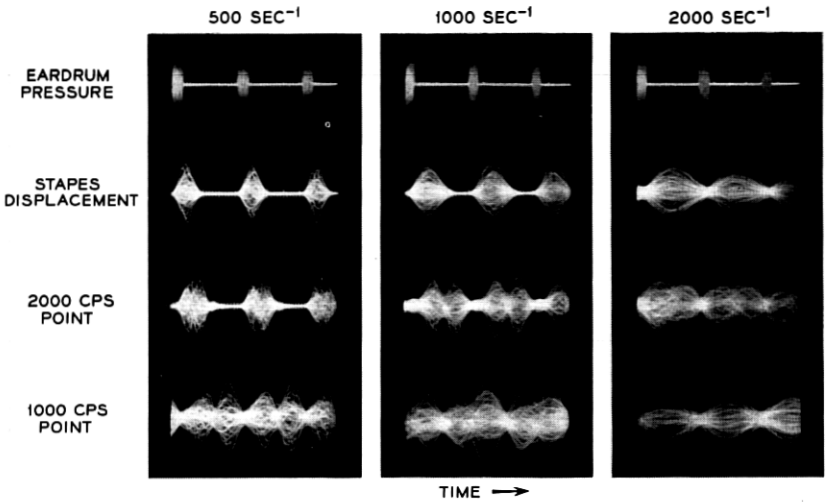

Fig. 19 — Displacement responses simulated by the network of Fig. 15 for periodically interrupted noise. Constant duty factor = 0.2.

several interruption rates are given in Fig. 19. For the slow rate, 500 sec^{-1}, the noise bursts are well resolved in time. Nerve volleys synchronous with the onset of the noise bursts might be expected for this stimulation. As the interruption rate is increased to upwards of 2000 sec^{-1}, however, neither the stapes nor membrane displacements resolve each burst separately in time. Stimulus-locked synchrony of the neural activity might be expected to be impaired or lost, even if the neural volleys could be elicited at this rate. Psychoacoustic observations bear this out. They show that it is difficult to ascribe a pitch to interrupted noise for rates greater than about 1000 sec^{-1} even under favorable conditions of low duty factor. It is not clear how much of this limit is determined by neural resolution, and how much by mechanical. Very likely both factors contribute to the resultant behavior.

2.2 *Binaural Lateralization*

Another aspect of perception is binaural lateralization. This is the subjective ability to locate a sound image at a particular point inside the head. The phenomenon is conventionally observed in earphone listening. If identical clicks (impulses of sound pressure) are produced simultaneously at the two ears, the average listener hears the sound image to be located in the center of his head. If one click is produced a littler earlier (or with slightly greater intensity) than the other, the sound image shifts

toward the earlier (or more intense) ear. This shift continues with increasing time difference until the image moves entirely to one side and eventually breaks apart. One then hears individual clicks located at the ears.

Naively we suppose the subjective position of the image to be determined by some sort of computation of coincidence between neural volleys. The volleys originate at the periphery and travel to higher centers via synaptic connections. The volley initiated earliest progresses to a point in the neural net where a coincidence occurs with the later volley to produce a subjective image appropriately off center. To the extent that intensity differences can shift the image position, intensity probably is coded, at least partially, in terms of the volley timing. As was the case in pitch perception, there are several areas in binaural phenomena where the ear models have been helpful in suggesting explanations of, and correlations with, subjective behavior. One of these areas is the effects of phase upon the binaural lateralization of clicks.

Suppose one produces impulses of pressure at the two ears, of equal intensity but of opposite polarity (i.e., one a rarefaction and the other a condensation). How would a listener adjust the times of occurrence of such pulses so that he hears the sound image exactly in the center of head? Let us consider what the displacement waveforms and the mechanical-to-neural conversion hypotheses would predict.

An impulse of pressure rarefaction draws the eardrum and stapes initially outward and causes the membrane displacement to be initially upward. A condensation pulse, on the other hand, causes an initially inward displacement of drum and stapes and consequently an initially downward movement of the membrane. At any given point on the membrane the waveforms of displacement produced by these two stimuli differ only in sign; that is, one is the negative of the other. Typical displacements of apical and basal points caused by rarefaction and condensation pulses are shown in Fig. 20. (These traces are essentially the impulse responses calculated previously in Fig. 7.)

The top diagram in Fig. 20 is illustrative of the displacement response of points lying in the apical (low-frequency) half of the membrane. The solid curve is the displacement for a rarefaction pulse, the dashed for a condensation. The abscissa at the top is in terms of the product βt, where β is the radian frequency of maximal response for the particular apical point. The lower abscissa on the top graph shows time scales appropriate to the specific points maximally responsive to 1200 and 600 cps, respectively. The lower graph shows the displacement appropriate to points lying in the basal (high-frequency) half of the membrane. As discussed

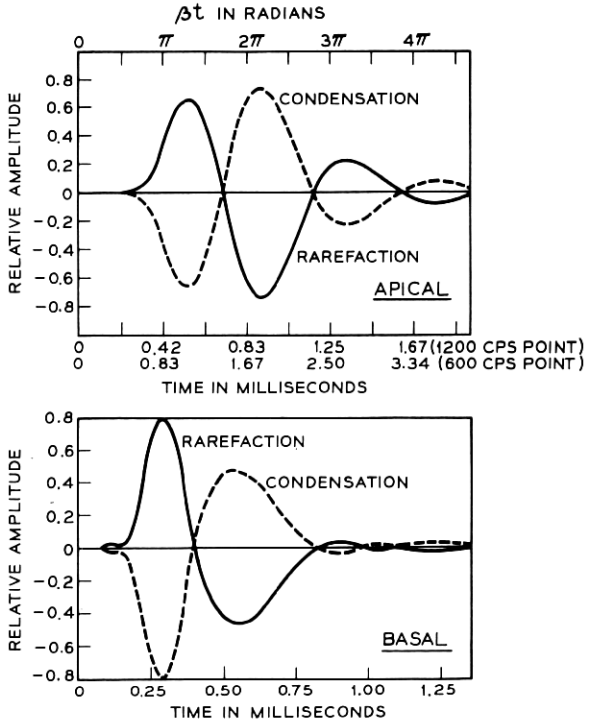

Fig. 20 — Apical and basal displacements of the basilar membrane for rarefaction and condensation pressure impulses at the eardrum. These responses from those computed in Fig. 7 for $[G(s)F_1(s)]$.

earlier, this waveform has essentially the same shape and time scale for all basal points.

Following our earlier assumptions, we suppose that neural firings (at least of the more sensitive outer hair cells) take place at some amplitude level on the upward deflections of the membrane. The curves suggest, therefore, that a time difference should exist between the firings for a rarefaction pulse and those for a simultaneous condensation pulse. The difference should be about one-half cycle on the displacement waveforms. The earlier results indicate that for broad-spectrum excitation the greatest deflections occur near the middle of the membrane, in the vicinity of the region maximally responsive to 1000 to 2000 cps. For such a place, the half-cycle intervals are of the order of 250 to 500 μsec. The time scale for the 1200 cps point in the top graph is indicative of this magnitude.

Assuming that simultaneity of neural firings at the two ears produces a central sound image, a rarefaction pulse and a condensation pulse should produce a centered image if the condensation is advanced in time to bring its positive displacement peak approximately into coincidence with that of the rarefaction. This means advancing the condensation, or letting it lead, by about 250 to 500 μsec. Furthermore, the periodic nature of the displacements suggests that multiple fusions of the sound image might occur by virtue of neural firings triggered at secondary positive excursions. These should occur for interaural times which are full-cycle increments of the principal fusion, including lead and lag shifts. A half-cycle lead of the condensation would represent a principal fusion; a half-cycle lag would be a secondary fusion.

If cophasic pulses are delivered to the two ears, that is, rarefaction-rarefaction or condensation-condensation, the same argument says that the principal fusion should obtain for zero interaural time difference, and secondary fusions for full-cycle shifts, either lead or lag.

The preceding remarks relate to broadband, unmasked pulses, where the neural response is likely to originate near the central portion of the basilar membrane. Suppose the dominant response were elicited from some other place on the membrane. The interaural time difference for lateralization ought to change in accordance with what the curves in Fig. 20 imply. Band-filtering of the pulse stimuli is an obvious means for confining membrane activity to specific regions. This has the disadvantage, however, that the stimulus signal is contaminated with the impulse response of the filter, so that it is inconvenient, if not difficult, to analyze the membrane displacement. The objective can be achieved more conveniently by selectively masking the membrane response with filtered random noise. The significant neural information can then be originated in a normally less responsive region by obscuring the maximally responding place with noise.

The top graph in Fig. 20 suggests that if the disparity between the interaural times for lateralizing cophasic and antiphasic pulses is to be increased, the significant response must originate from places more apicalward, that is, at points which respond maximally at lower frequencies. In such a case high-pass (HP) noise should be used to obscure activity in the basal part of the membrane. Low-pass (LP) noise, on the other hand, causes the coherent information to arise from the basal section. Here, because of the nature of the pulse response, the disparity between cophasic and antiphasic fusions is predicted to be roughly constant with place along the membrane, and should be of the order of 250 μsec. This, too, ought to be the minimum interaural disparity that can be produced for the antiphasic situation.

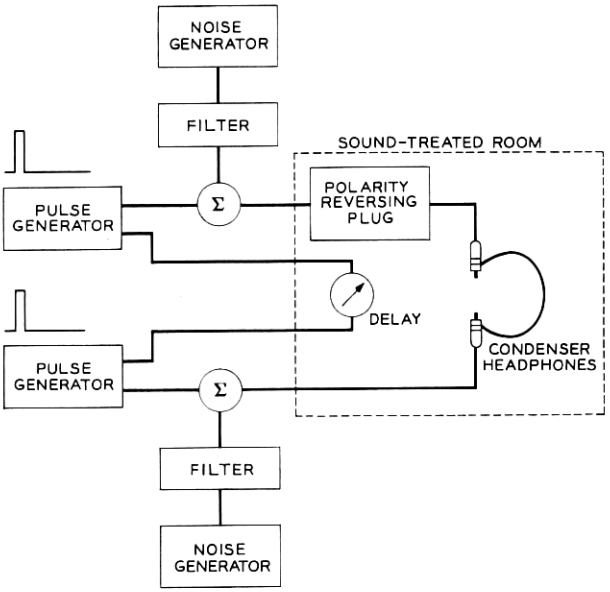

Fig. 21 — Equipment arrangement for binaural lateralization of cophasic and antiphasic clicks.

An experiment was performed to determine whether the predicted phenomena are in fact manifest.[14] The arrangement to measure the effects is shown in Fig. 21. Twin pulse generators produced identical 0.1-msec rectangular pulses in separate channels at a sensation level of 40 db. The repetition rate of the pulses was 10 sec^{-1}. Random noise from two uncorrelated generators was filtered by identical filters and added to the signal channels. This noise level completely masked the selected portions of the pulse spectra. HP and LP noise cutoff frequencies of 600, 1200, and 2400 cps were used in addition to no masking. Condenser microphones fitted with ear-insert plugs were used as earphones to provide good transduction of the pulse signals. The subject was provided a delay control which permitted continuous adjustment of the time of occurrence of one pulse relative to the other over the range ± 5 msec. A switch could reverse the polarity of the pulse delivered to one earphone.

The results of this experiment for three listeners are summarized in terms of median responses in Fig. 22. For HP masking, Fig. 22(a), the interaural time for the principal antiphasic lateralizations is seen to increase as the cutoff frequency of the HP noise is lowered. For these conditions the maximally responding unmasked place on the membrane

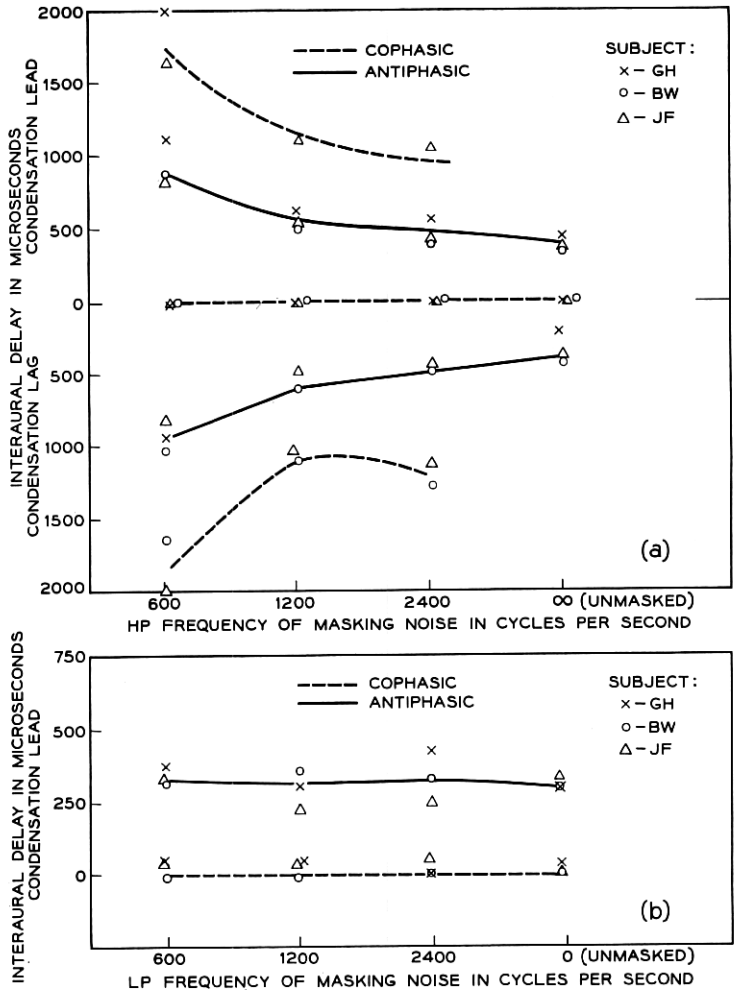

Fig. 22 — Effects of masking upon the lateralization of cophasic and antiphasic pulses.

should be that just below the cutoff frequency, f_c, of the filter. The interaural time for the antiphasic fusion ought then to be about $\pm 1/2f_c$. The data follow this value reasonably well.

The secondary antiphasic fusions (condensation lag) are roughly a reflection of the principal ones in the x-axis. The time separation between the principal and secondary points is approximately the predicted full-

cycle shift, or about $1/f_c$. The principal cophasic fusions fall along the axis for zero interaural times, and the secondary cophasic fusions (highest and lowest curves) fall at about the right value for the full-cycle shift.

The results for the LP masking, Fig. 22(b), indicate that the hypotheses about fusions of basal-end information are essentially sustained. The cophasic-antiphasic disparity is roughly constant at about 300 μsec. Secondary images, however, are not easily heard because the LP noise is a more potent out-of-band masker. Both masking conditions make it clear that the significant neural timing information can be made to originate from different points along the membrane. Further, the neural timing is intimately related to the individual mechanical excursions of the membrane at the significant point.

Some electrophysiological evidence also exists to support these psychoacoustic results and the assumptions made earlier. Peake[15] measured the latency of the gross neural component, N_1 , in cat's ear for stimulation by rarefaction and condensation pulses. For moderately high signal levels, the difference in latencies is found to be of the order of 200 to 300 μsec with condensation pulses giving the greater latencies. In addition, very recent data by Kiang[7] on the activity of single, peripheral nerve units suggest that the firings are synchronized with the individual unipolar displacements of the membrane, as conjectured here.

2.3 *Time-Intensity Trade*

In other binaural experimentation it has been observed that the position of a sound image can be maintained stationary by trading relative intensity against relative time of occurrence of pulses at the two ears. That is, the movement of the sound image towards the ear receiving a leading click can be offset by an increase in intensity of the pulse at the lagging ear. To a certain extent such a trading relation is implied in the calculated membrane responses and in the simple hypotheses about conversion of mechanical to neural activity. It is worth considering the degree to which the experimentally observed trade can be explained by the membrane relations.

Consider again binaural excitation of the ears by short, unipolar pulses. The earlier assumptions about neural firings on upward displacements of the membrane, in excess of a fixed threshold, imply a time-intensity trade. For ease in illustration, consider that the form of the impulse response of the membrane for middle to apical points is essentially specified by the model $f_3(t)$ given in (5). For simplicity this can be written

without constants as

$$f_3(\theta) = (\theta)^2 e^{-\theta/1.7} \sin \theta, \tag{36}$$

where $\theta = \beta t$. The waveform of this function has already been plotted in Fig. 4.

Imagine that the amplitude of a rarefaction stimulus has been set to threshold value so that neural firings are produced exactly at the first positive crest of the displacement. Now if the intensity of the stimulating pulse is steadily increased, the threshold level will be crossed at successively earlier times on the initial quarter cycle of the wave. In terms of the model, the advance in time of the threshold crossing is a simple function of the stimulus amplitude, and we can compare it with experimentally measured figures.

For reasons that will be obvious presently, we take

$$\ln f_3(\theta) = 2 \ln \theta - \frac{\theta}{1.7} + \ln \sin \theta. \tag{37}$$

Differentiating with respect to θ,

$$\frac{d[\ln f_3(\theta)]}{d\theta} = \frac{2}{\theta} - \frac{1}{1.7} + \cot \theta \tag{38}$$

and taking the partial with respect to time,

$$\frac{\partial[\ln f_3(\theta)]}{\partial t} = \beta \left(\frac{2}{\theta} - \frac{1}{1.7} + \cot \theta \right). \tag{39}$$

Equation (39) gives, in effect, the time-intensity trade for the wave in terms of nepers per second, as a function of the epoch θ at which threshold is crossed. In psychoacoustic tests the trade has customarily been specified in terms of msec/db—that is, the number of milliseconds by which the stimulus in one ear must be advanced to offset a relative intensity increase of one db in the other ear. Equation (39) can be put in terms of msec/db by taking its reciprocal, and multiplying by $10^{-3}/8.7$. One often sees this trade plotted as a function of intensity or sensation level of the stimulus. Let us arbitrarily take the positive maximum, $f_3(\theta_{+\max})$, as the threshold level of displacement. An increase in intensity of X db will then cause the threshold to be crossed at an epoch, $0 \leqq \theta \leqq \theta_{+\max}$ that satisfies: $-8.7 \ln \dfrac{f_3(\theta)}{f_3(\theta_{+\max})} = X$ db. We can therefore plot (39) [converted to msec/db] versus (37) [converted to db *re* $f_3(\theta_{+\max})$] for common values of the threshold crossing θ. This function, for three different apical points on the membrane, is shown in Fig. 23(a).

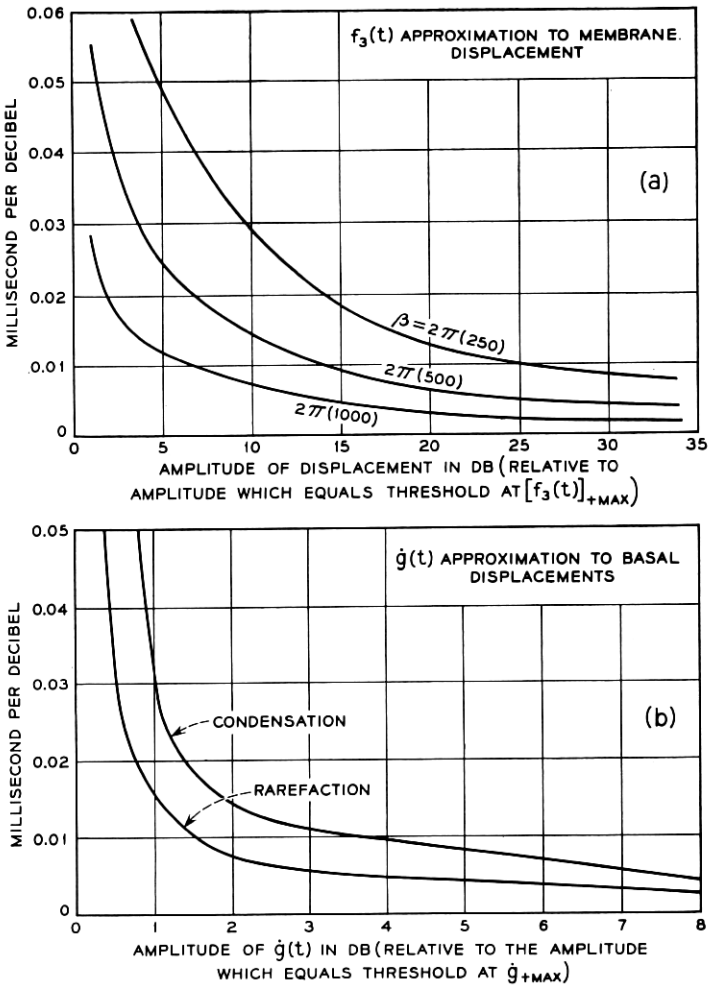

Fig. 23 — Time-intensity trade predicted for (a) apical points by membrane model $f_3(t)$; (*b*) for basal points by stapes derivative $\dot{g}(t)$.

The curves suggest that the trade of msec/db is greatest for low signal levels and diminishes for higher levels. It also indicates that the trade in msec/db is greater for points closer to the apex, that is, for lower β. Broadband pulse excitation of the model, as previously stated, produces greatest displacements near the middle of the membrane. The 1000-cps point is representative of this region. Low-level values of the trade for this point are on the order of 0.03 msec/db.

The earlier arguments also indicated that the impulse responses of basal points were similar to the time derivative of stapes displacement. A time-intensity trade computed from these responses ought to be suggestive of the minimum msec/db value that could be expected if the trade were based upon basal activity. We can use (11) and approximate the basal displacements by the stirrup derivative, $\dot{g}(t)$. Letting $(bt) = \Phi$,

$$\ln \dot{g}(\Phi) = -\frac{\Phi}{2} + \ln (2 \sin \Phi + \cos \Phi - 1). \tag{40}$$

And,

$$\frac{\partial t}{\partial \ln \dot{g}(\Phi)} = \frac{2(2 \sin \Phi + \cos \Phi - 1)}{b(\cos \Phi - 3 \sin \Phi + 1)}. \tag{41}$$

Again, expressing (40) in db relative to $\dot{g}(\Phi_{+\max})$ and (41) in msec/db, the two can be plotted for common values of Φ. When this is done the trading relation obtained is shown in Fig. 23(b). Because of the substantial asymmetry in $\dot{g}(\Phi)$, the function is computed for the initial quarter of the positive deflection and initial quarter of the negative deflection (that is, the initial positive deflection if the displacement phase were reversed). The former would be appropriate for rarefaction pulse excitation; the latter for condensation. These figures are, of course, susceptible of the uncertainty connected with the value b and the approximation of the basal displacements by $\dot{g}(\Phi)$. Nevertheless, the trading values thus obtained fall reasonably close to those for the

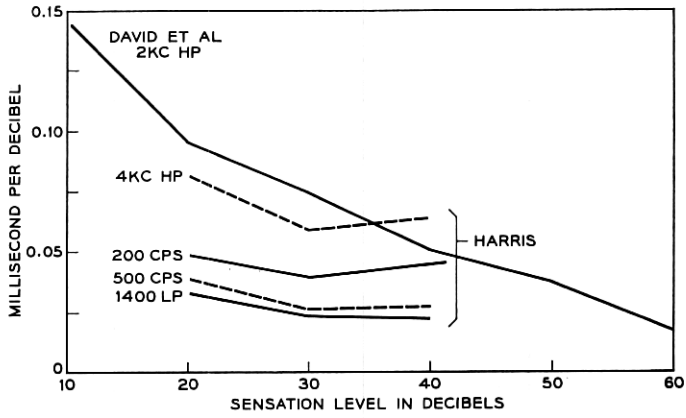

Fig. 24 — Measured time-intensity trade for several stimuli (after David et al[16] and Harris[17]).

1000-cps point shown in Fig. 23(a). To the extent that $\dot{g}(t)$ is a reasonable approximation of the basal displacements, the results indicate that the trade for condensation pulses should be slightly greater than those for rarefactions.

Some psychoacoustic data are available for comparison with these calculations. David, Guttman and van Bergeijk[16] used 2-kc high-pass clicks to measure the trading function and obtained a median result shown in Fig. 24. Harris[17] used both HP and LP filtered pulses and pure tones in a related investigation. Several of his results are also shown in Fig. 24. The subjective data for the HP pulses are clearly greater than the predictions from the model. The results for 1400 LP, however, are more nearly of the magnitude suggested by the computations. The previous computations also suggest that the msec/db trade should increase in magnitude as the significant neural information is elicited from more apical (low-frequency) points. Harris found, however, that for LP clicks with cutoffs between 200 to about 1000 cps, the trade was nearly constant at about 0.03 msec/db. One difficulty in comparing the computed and measured data is that we do not know how to equate values on the abscissae of Figs. 23 and 24. That is, we do not know what sensation level corresponds to the zero-db reference amplitude of the displacement wave. Only general directions and trends can therefore be legitimately compared.

Another difficulty in comparing the data is that the computed time-intensity trades assumed ideal impulse excitation of the ear. The experimental measurements, on the other hand, used pulses which were HP or LP filtered. The effect of the filter response upon that of the membrane is somewhat uncertain. The experimental determinations and the computations may not therefore be strictly comparable. To attempt to obviate this difficulty, we made some cursory measurements of the trade using the masking technique described earlier in Section 2.2. For a sensation level of 40 db, and with unmasked rarefaction clicks, one trained subject from the previous lateralization experiment made the $\Delta t - \Delta I$ swap plotted as the lower curve in Fig. 25. A binaural masking by 600-cps HP noise presumably constrains the coherent neural activity to come from a more apical point (somewhere near the 600-cps point). For such a masking the same subject made the trade indicated by the upper curve, giving values about twice as great as the unmasked trade. The slope of the unmasked function at the origin is approximately 0.03 msec/db. That for the 600 HP masking is about 0.05 msec/db.

Clearly these data for one subject are tentative and must be confirmed

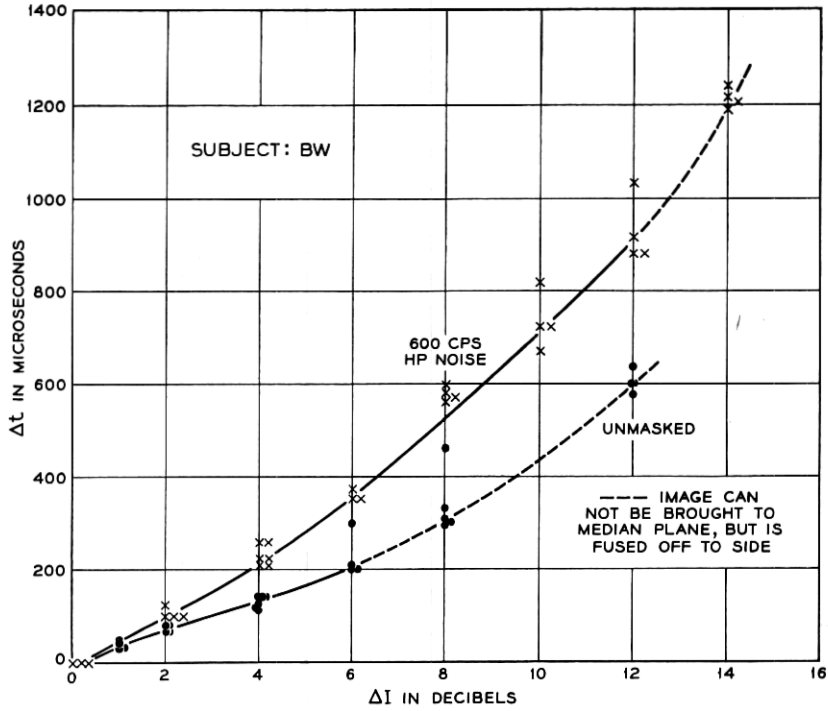

Fig. 25 — Effects of masking upon time-intensity trade for broadband, co-phasic pulses.

or refuted by additional experimentation. To the extent that they are correct, however, they support the general predictions of the model as to the frequency (or membrane place) dependence of the time-intensity trade. They do not agree well with absolute magnitudes of the computed trade, and there is still the question of how to equate abscissae. It is highly probable, too, that the time-intensity trade involves neural mechanisms not here included. Even so, the mechanical operations appear to go a long way in contributing to an explanation of the phenomenon. A time-intensity trade has also been observed at the neural level. In the cat's ear, Peake[15] finds that an intensity change of about 40 db in a stimulating pressure click causes a reduction in the latency of the N_1 neural component by roughly 0.6 msec. A simple division of these figures gives 0.015 msec/db for the trade. This figure falls within the range predicted by the model for human hearing.

As a final point in this theme, the same arguments can be made for

pure tones. For such stimuli the membrane displacements are also sinusoidal (at least over a large intensity range) and have the form

$$f_\beta(t) = K_\beta(\omega) \sin \omega t, \tag{42}$$

where $K_\beta(\omega)$ is an amplitude versus frequency factor appropriate to the membrane point maximally responsive to radian frequency β; $K_\beta(\omega)$ is largest, of course, for $\omega = \beta$. The previous argument gives

$$\ln f_\beta(t) = (\ln \sin \omega t + \ln K_\beta), \tag{43}$$

and,

$$\frac{d}{dt}(\ln f_\beta) = \omega(\cot \omega t). \tag{44}$$

A plot of this last result in terms of msec/db versus amplitude in db (relative to the amplitude for threshold crossing at $\omega t = \pi/2$) is shown in Fig. 26. In order of magnitude and frequency dependence, these values seem to compare reasonably well with results of Harris for 200- and 500-cps pure tones, previously shown in Fig. 24.

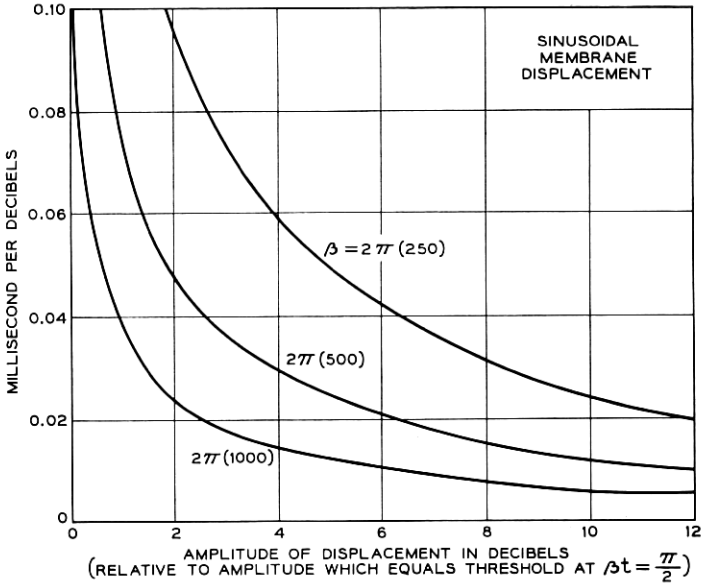

Fig. 26 — Time-intensity trade predicted for sine wave stimuli.

2.4 *Threshold Sensitivity*

The combined response curves in Fig. 9 indicate that the ear is more sensitive to certain frequencies than to others. This is well known to be subjectively true. To what extent, then, are the variations in the threshold of audibility accounted for by the mechanical sensitivity of the ear? We can use the model responses to examine the question.

The envelope of the peak responses in Fig. 9 can be compared with the subjectively determined minimum audible pressure for pure (sine) tones. Fig. 27 shows this comparison. The agreement is quite poor, although the gross trends are similar. The model responses here are on the basis of a 1500-cps critical frequency for the middle ear. The earlier discussion has pointed up the uncertainty of this value. The middle-ear critical frequency chosen to illustrate the computational technique was that derived from Zwislocki's data. The latter, in turn, were based upon one of Bekesy's investigations. In other investigations, Bekesy also found middle ear cutoffs higher than 1500 cps, so some uncertainty exists as to where this number should be fixed. Obviously the choice of this constant does not alter the computational method or analytical

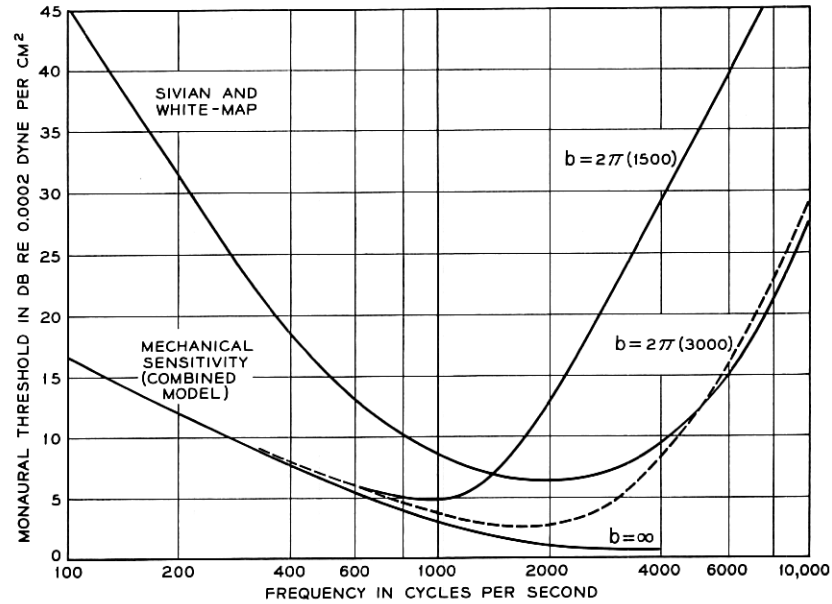

Fig. 27 — Relation of mechanical sensitivity of model to subjective monaural threshold for pure tones.

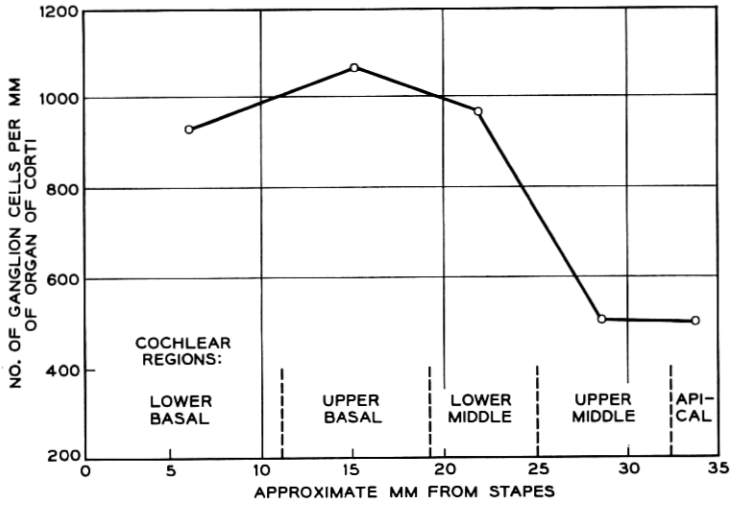

Fig. 28 — Average number of ganglion cells per mm length of organ of Corti (after Guild et al[18]).

technique. If we choose instead a critical frequency of 3000 cps for the middle ear, the fit to the threshold curve at high frequencies is more respectable. The match at low frequencies, however, is not improved, but we are less concerned about this for a different reason.

For the low frequencies, the disparity between mechanical and subjective sensitivity probably is a neural effect. According to our earlier assumptions, the number of neurons activated bears some monotonic relation to amplitude of membrane displacement. Perception of loudness is thought to involve possibly temporal and spatial integrations of neural activity. If a constant integrated activity were equivalent to constant loudness, the difference between mechanical and subjective sensitivities might be owing to a sparser neuron density in the apical (low-frequency) end of the cochlea. There is physiological evidence to this effect.

In histological studies Guild et al[18] counted the number of ganglion cells per mm length of the organ of Corti. Their results for normal ears are summarized in Fig. 28. These data show a slight decrease in the number of cells at the basal end and a substantial decrease in the density as the apex is approached. The innervation over the middle of the membrane is roughly constant.

One can pose the same questions about threshold sensitivity for short pulses or clicks of sound. For brief pulses of sufficiently low repetition

rate, the maximal displacements of the membrane, as stated before, are near the middle. According to the model, this continues to be the case for pulse rates well in excess of several hundred per second. The resonance properties of the membrane in this region are such as to resolve in time each individual exciting pulse. If, then, the predominant displacement takes place at one point for a large range of pulse rates, polarity patterns, and pulse durations, how might the subjective threshold vary and how might it be correlated with the membrane motion. One investigation of this question has led to a model for pulse threshold loudness.[19] These results can be partially summarized.

Thresholds of audibility for a variety of periodic pulse trains with various polarity patterns, pulse rates and durations are shown in Fig. 29. One notices that the thresholds are relatively independent of polarity pattern. For pulse rates less than 100 pps, the thresholds are relatively independent of rate, and dependent only upon pulse duration. Above 100 pps, the thresholds diminish with increasing pulse rate. Amplitude of membrane displacement would be expected to be a function of pulse duration and to produce a lower threshold for the longer pulses, which is the case. For rates greater than 100 sec^{-1}, however, some other non-mechanical effect apparently is of importance. The way in which audible pulse amplitude diminishes suggests a temporal integration with a time constant of the order of 10 msec.

Using the earlier assumptions about conversion of mechanical to

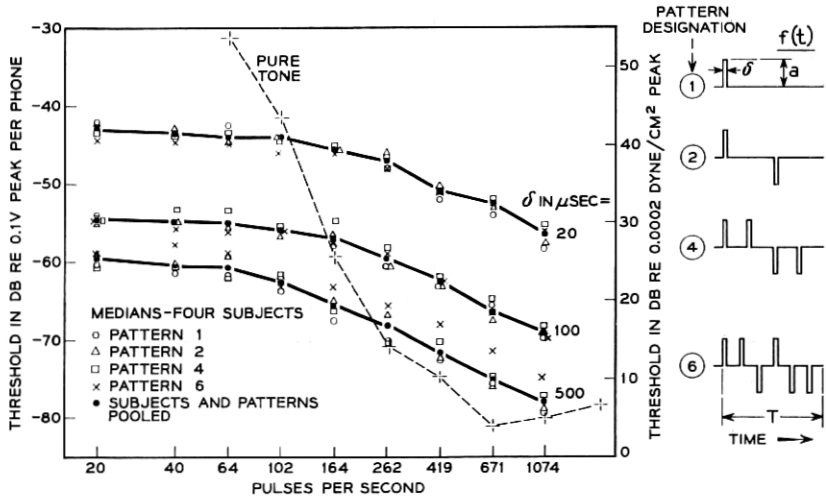

Fig. 29 — Thresholds of audibility for periodic pulses (after Flanagan[19]).

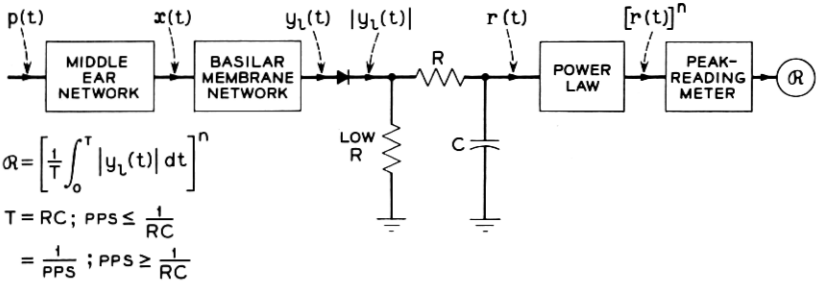

Fig. 30 — Model of the threshold for pulse data shown in Fig. 29.

neural activity, one might ask "What processing of the membrane displacement at the point of greatest amplitude would reflect the constant loudness percept at threshold?" A possible answer is suggested by the operations illustrated in Fig. 30.[19] The first two blocks represent middle-ear transmission [as specified in (6)] and basilar membrane displacement [vicinity of the 1000-cps point, as specified in (31)]. The diode represents the half-wave rectification associated with neural firings on unipolar motions of the membrane. The RC integrator has a 10-msec time constant, as suggested by the threshold data. The power-law element (exponent = 0.6) represents the power-law relation found in loudness estimation.* A meter indicates the peak value of the output of the power-law device. When all stimulus conditions represented by points on the threshold curves in Fig. 29 are applied to the circuit, the output meter reads the same value: that is, threshold.

One can also notice how this model might be expected to perform for sine wave inputs. Because the integration time is 10 msec, frequencies greater than about 100 cps produce meter readings proportional to the average value of the half-wave rectified sinusoid. In other words, the meter reading is proportional to the amplitude of the sine wave into the rectifier. Two alterations in the network circuitry are then necessary. First, the basilar membrane network appropriate to the point maximally responsive to the sine frequency must be used. This may be selected from an ensemble of networks. And second, to take account of the sparser apical innervation, the signal from the rectifier must be attenuated for the low-frequency networks in accordance with the difference between the mechanical and subjective sensitivity curves in Fig. 27. The power-law device still operates to simulate the appropriate growth of loudness with sound level.

* The power-law device is not necessary for threshold indications of "audible-inaudible." It is necessary, however, to represent the growth of loudness with sound level, and to provide indications of subjective loudness above threshold.

2.5 *Pure-Tone Masking*

Masking is defined as the increase in the threshold of audibility of one sound caused by the presence of another. The models with which we have been dealing describe the mechanical frequency sensitivity of the ear and hence ought also to imply something about the masking of one pure tone by another.

The significant neural information for a pure-tone stimulus is assumed to come from the membrane point which responds maximally (mechanically) to that frequency. The ability to detect activity correlated with such a tone ought likewise to be related to the amplitude of displacement produced at this same point by any interfering (masking) sound. In other words, the relative amplitudes of displacement caused at the point by the tone and masker might be expected to be related to the shift in threshold of the tone. From the model we can determine the relative amplitudes produced by tone and masker at the membrane point which responds maximally to the tone. For low frequencies, where middle-ear attenuation is not appreciable, this can be done simply from the membrane response curves such as shown in Fig. 3.

Let us take, for example, a masker of 400 cps (because there are corresponding subjective data for this condition). The relative levels of maskee and masker are shown by the dashed curve in Fig. 31. (These levels are read on the righthand ordinate.) Subjective threshold measurements for the same conditions produce the solid curve.[20] One sees

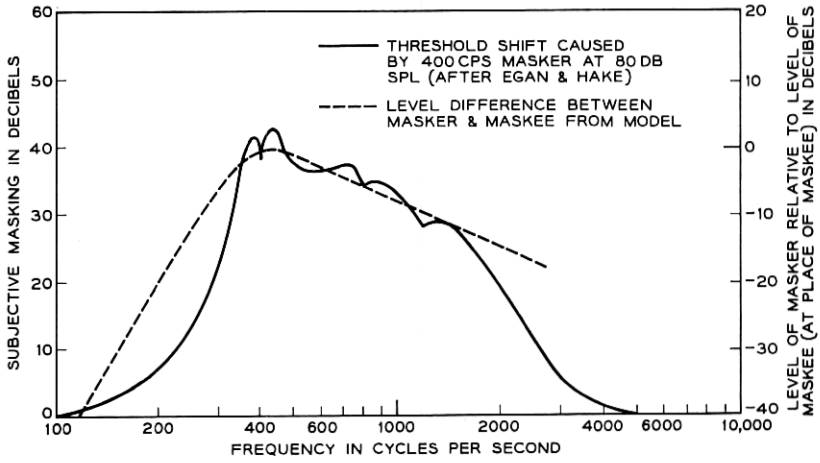

Fig. 31 — Masking of one tone by another (a) predicted by the model; and (b) measured by Egan and Hake.[20]

that the agreement is not particularly close, although the curves have similar gross shapes. The psychoacoustic measurement indicates less masking at frequencies removed from the masker than the mechanical response implies. This might suggest at least two possibilities: one, that the upper and lower skirts of the membrane resonances are a little steeper than we think; or two, that when the maskee has a relative level as much as about 10 db or so greater than the masker (at the maskee point) some neural inhibitory mechanism functions to suppress the masker even more.

One notices irregularities in the subjective masking curve at frequencies where the tone is an integral multiple of the masker. These are produced by beats and subjectively generated harmonics. One notices, too, that when tone and masker are the same frequency, the measurement is essentially a determination of the intensity limen. For example, the masking at 400 cps is 40 db, which means that a 400-cps tone must be raised 40 db above its unmasked threshold to be just audible in the presence of another 400-cps tone at a sound-pressure level of 80 db (*re* 0.0002 dyne/cm^2). The unmasked threshold (minimum audible pressure) for a 400-cps tone is approximately 10 db spl (see Fig. 27). The maskee is just detectable, therefore, when its level is about 50 db spl, or 30 db less than the masker. For an in-phase (or out of phase) condition, the maskee could maximally increase (decrease) the intensity of the masker by about 0.3 db. This is roughly the size of the intensity limen measured at this sensation level.[21]

When this same masking comparison is made for higher frequencies, the middle-ear transmission must be considered. If the middle-ear cutoff used in the model calculations is used, the agreement between subjective and mechanical results is poor at high frequencies. This again argues that the normal critical frequency for the middle ear is somewhat higher than that used to illustrate the model calculations.

The mechanical response also shows why a lower-frequency tone is a more effective masker than a higher-frequency tone. The reason is simply that the frequency response of a given point on the basilar membrane has a low-frequency skirt less steep than its high-frequency skirt. This same fact also suggests why low-frequency hearing is so difficult to impair by local injury or disease in the ear. The shallow low-frequency skirts of the response of all points along the membrane show that even basal points can respond appreciably to low-frequency stimuli. Even if the apical end of the basilar membrane were destroyed, basal locations could provide some low-frequency response.

Essentially in the same vein, these relations suggest why high-fre-

quency hearing is so susceptible of impairment. The high-frequency skirt of the frequency responses is quite sharp. Damage to any basal location leaves no other point capable of responding substantially to that frequency.

2.6 *Conclusion*

It seems clear that the extent to which subjective behavior can be correlated with, identified in, and predicted by the mechanical operation of the peripheral ear is rather substantial. The models developed here have been found to be useful computational tools in the analyses of a number of different psychoacoustic problems. They have, in fact, precipitated several experiments by predicting hearing phenomena which were later confirmed by the experiments. Further, electrophysiological data obtained recently link neural activity intimately with the individual mechanical excursions of the membrane. These findings also lend support to the simple assumptions about the conversion of mechanical to neural information.

The models do not, of course, account for higher-order neural functions and hence describe only a peripheral part of the hearing process. Even so, they seem in many cases to contribute substantially to physiological explanations of subjective behavior. As more knowledge is gained about mechano-neural conversion and about neural processing, analytical specification of the mechanical operation, such as developed here, may prove increasingly useful.

III. ACKNOWLEDGMENT

I wish to thank B. J. Watson of the Speech and Auditory Research Department for his capable technical assistance in the experimental phases of this work.

REFERENCES

1. Flanagan, J. L., Models for Approximating Basilar Membrane Displacement, B.S.T.J., **39,** 1960, pp. 1163–1191.
2. von Bekesy, G., Über die Resonanzkurve und die Abklingzeit der verscheidenen Stellen der Schneckentrennwand, Akust. Zeit., **8,** 1943, p. 66; J. Acoust. Soc. Am., **21,** 1949, p. 245.
3. von Bekesy, G., Über die Schwingungen der Schneckentrennwand beim Präparat und Ohrenmodell, Akust. Zeit., **7,** 1942, p. 173; J. Acoust. Soc. Am., **21,** 1949, p. 233.
4. Zwislocki, J., Some Impedance Measurements on Normal and Pathological Ears, J. Acoust. Soc. Am., **29,** 1957, p. 1312; Electrical Model of the Middle Ear, J. Acoust. Soc. Am., **31,** 1959, p. 841 (A); also personal communication.
5. Huggins, W. H., and Licklider, J. C. R., Place Mechanisms of Auditory Frequency Analysis, J. Acoust. Soc. Am., **23,** 1951, p. 290.

6. Flanagan, J. L., and Guttman, N., Pitch of Periodic Pulses, J. Acoust. Soc. Am., **32,** 1960, p. 1308.
7. Kiang, N., Watenabe, T., Thomas, E., and Clark, L., *Stimulus Coding at the Periphery of the Auditory System*, book in preparation.
8. Hughes, J. R., and Rosenblith, W. A., Electrophysiological Evidence for Auditory Sensitization, J. Acoust. Soc. Am., **29,** 1957, p. 275.
9. Galambos, R., Neural Mechanisms in Audition, Laryngoscope, **68,** 1958, p. 388.
10. Shower, E. G., and Biddulph, R., Differential Pitch Sensitivity of the Ear, J. Acoust. Soc. Am., **3,** 1931, p. 275.
11. Guttman, N., and Flanagan, J. L. Pitch of High-Pass Filtered Periodic Pulses, J. Acoust. Soc. Am., **33,** 1961, p. 839(A); see also last footnote in Ref. 1.
12. Flanagan, J. L., and Guttman, N., Pitch of Periodic Pulses without Fundamental Component, J. Acoust. Soc. Am., **32,** 1960, p. 1319.
13. Miller, G. A., and Taylor, W .G., The Perception of Repeated Bursts of Noise, J. Acoust. Soc. Am., **20,** 1948, p. 171; Harris, G. G., Pitch in Gated Random Noise, J. Acoust. Soc. Am., **32,** 1960, p. 1506(A).
14. Flanagan, J. L., David, E. E., Jr., and Watson, B. J., Binaural Lateralization of Cophasic and Antiphasic Clicks, J. Acoust. Soc. Am., **33,** 1961, p. 840(A); Effects of Masking Upon the Binaural Lateralization of Cophasic and Antiphasic Clicks, J. Acoust. Soc. Am., **33,** 1961, p. 1670. (A).
15. Peake, W. T., An Analytic Study of Electric Responses at the Periphery of the Auditory System, Technical Report 365, Research Laboratory of Electronics, MIT, 1960.
16. David, E. E., Jr., Guttman, N., and van Bergeijk, W. A., Binaural Interaction of High-Frequency Complex Stimuli, J. Acoust. Soc. Am., **31,** 1959, p. 774.
17. Harris, G. G., Binaural Interactions of Impulsive Stimuli and Pure Tones, J. Acoust. Soc. Am., **32,** 1960, p. 685.
18. Guild, S. R., Crowe, S. J., Bunch, C. C., and Polvogt, L. M., Correlations of Differences in the Density of Innervation of the Organ of Corti with Differences in the Acuity of Hearing, Acta Oto-Laryngologica, **15,** 1931, p. 269.
19. Flanagan, J. L., Audibility of Periodic Pulses and a Model for the Threshold, J. Acoust. Soc. Am., **33,** 1961, p. 1540.
20. Egan, J. P., and Hake, H. W., On the Masking Pattern of a Simple Auditory Stimulus, J. Acoust. Soc. Am., **22,** 1950, p. 622.
21. Riesz, R. R., Differential Intensity Sensitivity of the Ear for Pure Tones, Phys. Rev., **31,** 1928, p. 867.