

A THEORY OF HEARING

by

WILLIAM HERBERT HUGGINS

B.S., Oregon State College
(1941)

M.S., Oregon State College
(1942)

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June, 1953

Signature of Author _____

Department of Electrical Engineering, May 20, 1953

Certified by _____ Thesis Supervisor

Accepted by _____
Chairman, Departmental Committee on Graduate Students

by

WILLIAM HERBERT HUGGINS

Submitted to the Department of Electrical Engineering, M.I.T.
on 25 May 53 in partial fulfillment of the requirements
for the degree of Doctor of Science.

ABSTRACT

A model of the inner ear has been developed which combines into a simple, plausible mechanism the mechanical, dynamical, anatomical and neurological details of the cochlear structure. This mechanism adequately accounts for the results of a variety of physical and psychological experiments. It suggests a functional purpose for certain of the anatomical details of the cochlear partition and it unifies into a single consistent construct the experimental facts concerning: the discrepancy between the sharpness of the excitation pattern along the basilar membrane, as inferred from masking data, and the sharpness of the vibration pattern, as observed by Bekesy; the localization of, and relation between, the microphonic and action potentials; the increase of sensation threshold at low frequencies; the abrupt inhibition of second-order neurons in the cochlear nucleus of a cat as observed by Galambos and Davis; the extraordinary pitch discrimination of the ear; and the beats and other phase effects heard when listening to mistuned unison or consonant tones.

Voiced

The mechanism described here utilizes a phase principle which, from theoretical and experimental considerations, is shown, to be particularly well suited for the recovery of the characteristic frequencies of the damped oscillations which comprise many common sounds, such as those of speech. The use of a phase principle requires the measurement of the phase difference between two related waves. Since phase may be measured by detecting the instants at which the wave passes thru a value of zero, a phase principle permits "all-or-none" elements, such as the individual neurons of the nervous system, to utilize the neural phenomena of facilitation, inhibition and summation in the accurate measurement of phase differences over a wide range of stimulus intensity. Furthermore, the two related waves needed to utilize a phase principle may be generated by plausible functions ascribed to the inner and outer hair cells and to the tectorial and basilar membranes. The action of the tectorial membrane is hypothesized to be such that the excitation of the inner hair cells is directly proportional to the displacement of the basilar membrane whereas the outer-hair-cell excitation is proportional to the curvature of the curvature of this displacement. With the additional assumption that the outer-hair-cell neurons excite, and the inner-hair-cell neurons inhibit, certain of the second-order neurons, a model is obtained which accounts for the experimental facts summarized above.

Thesis Supervisor: J. C. R. Licklider
Title: Associate Professor of Economics
Massachusetts Institute of Technology

ACKNOWLEDGEMENTS

I am indebted to the Air Force Cambridge Research Center, for allowing me, as an Air Force employee, to work toward my Doctorate at the Massachusetts Institute of Technology.

As my thesis supervisor at M.I.T., Professor J. C. R. Licklider encouraged and stimulated the main lines of thought developed and, through many invaluable discussions, endeavored to orient me in the field of psycho-acoustics. Much of the work was done within the Research Laboratory of Electronics at M.I.T. and I wish also to acknowledge the continued support and encouragement given by Prof. J. B. Wiesner, Director of that Laboratory.

Finally, I wish to thank all my friends who have given encouragement and direct assistance throughout this work, including the preparation of this thesis. I am especially grateful to Dr. O. H. Straus whose penetrating suggestions and criticisms have enhanced the major ideas presented here.

Title Page	
Abstract	i
Acknowledgment	ii
Table of Illustrations	v
I. Statement of Problem	1
II. Background	4
III. The Representation of Meaningful Sounds	9
IV. The Analysis of Meaningful Sounds	15
V. The Phase-Principle and Complex Frequency Analysis	19
A. A Phase Filter for Obtaining Resonagrams	19
B. Random Noise Excitation	25
C. A simplified Form of Phase Filter	28
D. The Masking Effect of a Second Resonance or Noise.	31
E. Implications for Hearing	40
VI. A Phase-Mechanism for Cochlear Frequency Analysis.	41
A. Structural Properties of the Cochlea	41
B. The Mechanical Action of the Cochlea	51
(1) Sharpness of Response.	55
(2) Possibility of a Phase-sensitive Mechanism	55
(3) Correct Polarities of Stimulation	56
(4) Relation between Microphonic and Action Potentials	56
(5) The Increase of Sensation Threshold at Low Frequencies	57
VII. Discussion of Experimental Evidence	60
A. Cochlear Excitation Pattern Inferred from Masking Data	60
B. Localization of Source of Microphonics	60
C. Response of Second-Order Neurons	64
D. Frequency Discrimination	72

VIII.	Phase Effects in Monaural Perception	74
	A. Beats of Mistuned Unisons	75
	B. Beats of Mistuned Consonances	82
	C. Is There a Phase-Masking Effect? - A Decisive Experiment	87
	Appendix A. Analytic Representation of the Mechanical	91
	Response of the Cochlea	
	Appendix B. Solution of the Cochlear equation by Milne's	104
	Numerical Method	
	Appendix C. Summary of Equations for $W(x)$	109
	Appendix D. A Neural Sharpening Mechanism	113
	References	122

TABLE OF ILLUSTRATIONS

PAGE

- Fig. 1 Representation of the phase characteristic of a sample tuned filter upon the complex-frequency plane. 19
- Fig. 2 Phase difference $\theta_1 = \theta_A - \theta_B$ as a function of the frequency of an input wave having: zero decrement (solid line); decrement one-half that of the tuned circuits (broken line). 21
- Fig. 3 Sharpened response curve corresponding to $\left| \theta_1 + \theta_2 \right|^{-1/3} \left| \theta_1 - \theta_2 \right|$. 23
- Fig. 4 Comparison of resonagrams obtained with the filtering scheme of Fig. 3 with the ordinary sonagram obtained with the conventional wide-band filter of the sonagraph. 24
- Fig. 5 Generalized phase filter. 25
- Fig. 6 Response of the simplified phase filter as a function of the frequency of the source resonator when subjected to impulsive excitations. A source damping of zero and one-half that of the filter are considered. 29
- Fig. 7 Response of the simplified phase filter as a function of the frequency of the source resonator when subjected to impulsive excitation. A source damping of zero and of one-half that of the filter are considered. 30
- Fig. 8 Same circuit as in Fig. 7 except that source is now excited by white noise. 32
- Fig. 9 Response of simplified phase filter as a function of filter frequency when the input signal consists of two sinusoidal components differing in frequency by 5 units. The relative amplitude of one is taken as a parameter. 35

- Fig. 10 Comparison of masking thresholds, based on the curvature hypothesis, with the experimental data of Wegel and Lane for the sensation thresholds of a higher frequency tone in the presence of a 1200-cps tone at 80 db sensation level. The solid curve marked $\delta^2 r/\delta \omega^2 = 0$ is based on the average correlation, whereas the broken curve is for detection at the peak of the "short-time correlation". One frequency-unit is equal to a frequency difference of 64 cps. 38
- Fig. 11 The anatomy of the human ear. (Max Brodel). (Frontispiece from ref. W-1). 42
- Fig. 12 A midmodiolar section of the cochlea. (Max Brodel, Fig. 2 of W-1). 43
- Fig. 13 The cochlear duct, in cross section (Fig. 3 of W-1). 44
- Fig. 14 Phase and amplitude variation of the transverse pressure, displacement, and fourth place derivative of the displacement for a 1,000-cps tone. 46
- Fig. 15 Relative instantaneous displacement (solid line) and force (dashed line) waves along the tectorial membrane as functions of the distance from the stapes. The excitation is a 1,000 cps tone. 47
- Fig. 16 A. Cross-section of the organ of Corti. 48
 B. Side view showing how the tectorial membrane may act as a beam to exert upon the outer hair cells forces proportional to the fourth place-derivative of the displacement of the basilar membrane.
- Fig. 17 Schema for mechanical action of the cochlea. 52
- Fig. 18 Magnitude of fourth space derivative of the displacement compared to the stimulation pattern on the basilar membrane determined from data on masking by Fletcher. 61

- Fig. 19 A. Regions of response and inhibition along the organ of Corti for a 1,000-cps tone. 66
- B. Estimated effect of the change in response of a second-order to a fixed tone T_1 at -78 db intensity caused by variation of the frequency of a second tone at sound intensities of -58, -44, and -24 decibels.
- Fig. 20 Experimental data after Galambos and Davis for response of a single second-order neuron in the cochlear nucleus of a cat, at the same relative stimulation levels as represented by the theoretical curves of Fig. 19. 68
- Fig. 21 Effect of duration of tone on frequency discrimination. 73
- Fig. 22 Variation of the intensities and relative phase of the excitation of the inner and outer-hair-cell excitations at the point $x = 2.4$ cm, when the 1,000-cps stimulating tone is 100% amplitude modulated at 50 cps. 79
- Fig. 23 Same as Fig. 22 except that carrier has been shifted 90 degrees to give a stimulating tone that is frequency modulated. 80
- Fig. 24 Variation of the phase of firing of the 1,000-cps neuron as a function of the phase of a second harmonic at 2,000 cps. of relative amplitude A. The stimulation level is well above threshold so that $T = 0$. 86
- Fig. 25 Variation of critical phase angle of second harmonic for which no change in pitch should occur, as a function of stimulation level. 90

	PAGE
Fig. A-1 Comparison of $ Y(x) $ as given by Eq. (A-23) with the "exact" solution calculated by numerical methods described in Appendix B.	102
Fig. A-2 Phase data corresponding to the amplitude curves shown in Fig. A-1.	103
Fig. D-1 A. The path followed by a single nerve fiber innervating the outer hair cells. B. A simple model of a nerve fiber.	115
Fig. D-2 Distribution of voltage across end twigs of outer-hair-cell neurons which run basalward for a distance of l millimeters. The external potential distribution $ W(x) $ is shown for comparison.	120
Fig. D-3 Phase data associated with the amplitude curves of Fig. D-2.	121

A THEORY OF HEARING

I. STATEMENT OF PROBLEM

Although the phenomenon of hearing has attracted scientific attention for centuries, there still remains to be found an adequate explanation of how we are able to perceive even the simplest of sounds, such as the click and the steady sinusoidal tone, with the acuity that psychophysical data show possible. For example, binaural measurements using clicks reveal a temporal resolution between the two ears of as little as 12 microseconds (K-1), and measurements using steady tones show that changes in frequency as small as 1/2 percent can be discerned throughout much of the range of audible tones. Whereas the known facts concerning the nervous system do enable one to construct certain neural mechanisms that would permit the perception of the simultaneity of the stimuli, the best available facts concerning the broadly tuned resonance phenomena that have been observed in the inner ear by Bekesy and others have not appeared to be compatible with the acute perception to slight changes in pitch displayed by the human listener.

The major problem, then, is to find a plausible mechanism for auditory frequency analysis that will be consistent with experimental psychophysical data and will use the known facts concerning the anatomical, mechanical, and neurological properties of the ear. Since the temporal aspects of neural behavior appear to offer the most precision, it is desired that this mechanism should utilize the mechanical resonance

properties of the ear in such a way as to reduce the question of pitch resolution to one of temporal resolution for which the available neural phenomena of temporal inhibition, facilitation and summation may be utilized. If such a mechanism can be found, the process of frequency analysis will have been translated into the time domain and the temporal and spectral aspects of hearing will have been unified.

II. BACKGROUND

Several recent publications cover the historical development of various theories of hearing (W-1, S-1) as well as the current state of our knowledge concerning auditory phenomena and structural details of the ear. (L-1, S-2) It is desired to review here only enough of this history to show how certain ideas, such as Fourier's resolution of a signal into sinusoidal components, have resulted in a biased and unnatural interpretation of the essential phenomena involved.

The history of efforts to explain how the ear can detect the pitch of a sound begins in 1605 with the formulation of the first resonance theory of hearing by Caspar Bauhin, a Professor of Anatomy in the University of Basle, Switzerland. The resonance was thought by Bauhin to occur within the various cavities and labyrinths of the middle and inner ear, which at that time (and, indeed, for 150 years following) were all believed to be filled with air. By 1683, knowledge of the anatomy of the inner ear had advanced to the point where DuVerney was able to publish a treatise in which he gives a surprisingly good description of the cochlea with its two scalae, bony lamina, and basilar membrane. However, DuVerney reasoned that the bony lamina, to which he wrongly assigned the resonance property, would vibrate near the base of the

cochlea for low tones and near the apex for high tones. It was not until 1760 when Cotugno, adopting the view that it was the membranous rather than the bony lamina which resonated much as strings in a harp, concluded that the low tones should be localized near the apex of the cochlea. Thus, for almost two centuries there has existed a theory, which must now be accepted as a fact, that a resonance phenomenon occurs within the ear such that a type of frequency analysis is possible on the basis of the spatial distribution of the response within the cochlea.

By the middle of the 19th century, a number of ideas necessary for the further refinement of a theory of hearing had developed. With the perfection of the compound microscope around 1830, rapid progress was made in the study of the finer anatomy of the cochlea; particularly outstanding was the work of Corti. About the same time, Johannes Muller presented his doctrine of "specific nerve energies" -- that "sensation consists in the sensorium receiving through the medium of the nerves, and as a result of the action of an external cause, a knowledge of certain qualities or conditions, not of external bodies, but of the nerves of sense themselves." Contemporary to these ideas were the mathematical notions of Fourier, who deduced that a periodic wave could uniquely be resolved into sinusoidal harmonics, and of Ohm who asserted that the ear actually performs the type of analysis defined by Fourier's theorem. It was the genius characteristic of Helmholtz that in 1857 he was able to integrate these several different ideas into a resonance theory that has occupied a dominant place in the theory of hearing to the present day.

Helmholtz conceived of the cochlea as being functionally equivalent to a series of sharply tuned resonators with their frequencies of resonance varying progressively, that tuned to the highest frequency being

at the base and the lowest at the apex of the cochlea. This arrangement differed little from that offered by Cotugno 100 years earlier. What made Helmholtz' theory significant was that the anatomical evidence available at that time appeared to indicate that the cochlea did contain the discrete resonators required for his theory and that each could have associated with it a specific nerve fiber. The perception of sound could thus be reduced to the perception of the mechanical response of a specific set of resonators within the ear and, by Ohm's law and Muller's, this conception of the mechanism of hearing was, and still is, appealing because of its simplicity.

Despite its appealing simplicity, Helmholtz' notion of independent resonators and specific nerve fibers associated with each pitch seems to have been contradicted by nearly every new fact that has been learned about the dynamics of the cochlea. By the beginning of the twentieth century, the theory, which enjoyed almost universal acceptance 25 years earlier, had become subject to vigorous attack from all sides, and various other theories were, and have continued to be, proposed. (However, the mechanism to be described in this dissertation does account for a high degree of specificity within the cochlea itself so that, neurally at least, the situation could resemble that conceived by Helmholtz.)

The failure to find specific resonant elements within the cochlea having the high selectivity required by the Helmholtz theory has encouraged the notion that, perhaps, the analysis of frequency is not performed in the peripheral organ of the ear at all, but rather, that the auditory nerve transmits the details of the sound -- telephone fashion -- to the brain where the frequency analysis is completed. When Rutherford presented in 1886 what might be called the first of

these "telephone" theories, little was known about the limitations of the auditory nerve. In the light of present day knowledge of neurophysiology as well as Békésy's direct evidence that a degree of place-analysis does in fact occur within the cochlea, the "telephone" theory in its simple form must be discarded. Nevertheless, modern psychoacoustic data on binaural space localization of sound sources seem to indicate a telephone mechanism for some aspects of the sound and Licklider's recent "Duplex Theory of Pitch Perception" (L-2) suggests a neural mechanism capable of performing the type of analysis required in such a theory.

Now, Ohm's acoustic law, relating to the resolution of a complex wave into its sinusoidal frequency components, was deeply implanted in hearing theory by Helmholtz and has remained there ever since. In his recent book "Theory of Hearing", (W-1), Ernest Glen Wever gives an account of the many theories that followed Helmholtz. Most of these theories are the result of efforts to account for the analysis required by Ohm's law in terms of the known anatomy and physiology of the ear with the result that both theory and experiment since the time of Helmholtz have been predominantly directed toward the response of the ear to sinusoidal tones. And, since the time of Helmholtz, serious difficulties have arisen because the "resonant" properties of the ear -- whether this "resonant" property be ascribed to the ordinary interchange of kinetic and potential energies or to the selective attenuation or interference of traveling waves -- just do not seem to be capable of explaining adequately the observed frequency discrimination of the hearing mechanism.

For historical reasons, the phenomenon of "resonance" has come to be associated almost inseparably with sinusoidal excitation as have also

our notions relating to "frequency". The notion of "resonance" has, in recent years, been broadened to include other waveforms. In its most general form, "resonance" is fundamentally nothing other than the "principle of resemblances" employed by the ancient Greek philosophers to account for sensory perception of whom Wever (W-1, p. 4) has this to say:

"They began by accepting the naive (sic!) view that perception is a mirroring of the external world, yet they soon found it necessary to complicate this outlook in order to account for the obvious differences among the senses. They solved this problem of sense variety by reference to the magic formula that 'like is perceived by like'. This is a formula handed down from primitive thought, just as the mirroring principle was; and though it may have seemed a more sophisticated idea at the time, it still was an expression of the same elementary kind of thinking, in which all things are accounted for by direct analogies."

Now, resonance of a system may be said to occur when the exciting wave approximately matches the natural (impulse) response of the system, for then "like is perceived by like" and the forced response is enhanced. Indeed, this principle finds its expression in modern times as the matched-filter criterion of Van Vleck and Middleton.* A tuning fork when sharply

* Middleton has recently shown that under very general conditions the best possible detector of any signal imbedded in noise consists of a suitable linear, matched filter followed by an appropriate non-linear operator. If the signals to be detected consist of highly-damped oscillatory components, then the best linear filter, for the case of Brownian noise, must also possess high damping if optimum detection is to be achieved. (D-2)

struck produces as its characteristic transient a wave which, except for a slight damping, is sinusoidal -- and it is, therefore, resonant to a sine wave of this same frequency. An electrical capacitor when struck by a current pulse produces as its characteristic transient a constant voltage difference between its terminals -- and it is, therefore, resonant to a constant current impressed upon it.

Perhaps during the early development of man it was vital to his safety that he should possess acute perception, not of the steady tones so leisurely studied by his distant progeny, but rather of those sounds of short duration caused, possibly, by the breaking of a twig, the grunt of an animal, or a padded foot striking the ground. It would be to these sounds that Empedocles would have expected the ear to possess sharp "resonances". And I, for one, am inclined to agree with Empedocles. The theme developed throughout this dissertation is that sound should be analyzed into exponential rather than sinusoidal components, and that the structure of the ear is adapted to this kind of analysis. If in the process of describing a mechanism whereby the ear can perform a generalized frequency analysis of this type we can account as a byproduct for known facts about the response of the ear to sinusoidal stimulation, this accomplishment would appear to provide further justification of this underlying idea.

The difficulty encountered here is a most common one that generally arises because the elements into which a thing is decomposed for analytical reasons may not be "natural" ones nor reveal the essentially significant structure or pattern of the thing itself. Once this decomposition has been made, it is all too easy to concentrate upon these elements so exclusively that the really important attributes of the thing are

overlooked or lost. Certainly it would seem unreasonable to assert that, because the sound pressure acting upon the ear drum arises from the impact of individual molecules, we should make the study of the auditory effects of single molecular impacts the basis for our science of hearing. Such a study would exclude the most important aspect of a sound wave which has to do with the structural coherence, both in time and space, between the many molecules striking the ear drums. It is this coherent structure that is important -- not the effects of the individual molecules.

Yet, to decompose a given sound wave into an infinite spectrum of infinitesimal sinusoidal components, and then to give major attention to the auditory effects of each of these components with only slight attention to the structural relationships between them, is to make a very similar kind of error. Fortunately, the modern notion of a system function provides just the concise method needed to represent these structural aspects of a sound wave. (H-1) It reminds us that the sense of hearing must obviously have as its main purpose not the detection of the frequency components of the sound wave at the ear drum, but rather the sensing of the temporal and structural aspects of the agency that creates the sound, the sound itself being only a carrier and, in a sense, no more important to the phenomena of hearing than the existence of the individual gas molecules is to the important correlates of a sound wave.

If it is granted that the purpose of the sense of hearing is to recover the significant information from the sounds that we hear, it follows that the precise nature of this information must play a significant role in the development of any theory of hearing. Accordingly, we first ask ourselves, "What is the meaningful information content of

a typical sound wave and how can it be represented and measured?"

III. THE REPRESENTATION OF MEANINGFUL SOUNDS

Information theory has given us some useful criteria for evaluating the capacity of a channel for transmitting an information-bearing signal. This capacity is expressed as an average rate over a very long time interval T and the appropriate measure is, therefore, in terms of the frequency bandwidth B of the channel which, by definition of the Fourier spectrum, is also an average over all time. If we consider the set of all possible signals that could be transmitted over a channel having a frequency bandwidth B of, say, 5000 cps and a signal-to-noise ratio of 30 db, we find that the maximum rate at which information could be conveyed would be, from $I = 2B \log_2(1 + \frac{S}{N})$, about 50,000 bits per second. The total number of distinguishably different waves, each of 0.1 second duration, that would be transmitted would be about $10^{1,500}$. If in a hypothetical experiment a listener were presented with sounds corresponding to each of these waves it is obvious that he would be able to distinguish between only a tiny fraction of them. Just what this fraction would be, I do not know, but it is likely that the rate at which the information can be perceived may not be so very much in excess of the rate at which it can be generated -- in the order of 20 bits per second for man. It is apparent, then, that meaningful (i.e. distinguishable) sounds form a very small subset of the set of all possible sounds that may be created from sinusoidal components lying within a given frequency band. For this reason, we are led to suspect that the representation of meaningful sound can be accomplished in a less wasteful manner, perhaps by a representation that seeks to specify the intrinsic pattern

and structure of the sound wave rather than the 2BT sampling points dictated by the measuring instrument. (S-4).

We have already remarked that a sound wave should be regarded as a carrier that has encoded within it pertinent information about the agency which creates it. To ask about the significance of a sound wave is to ask about the significance of its source, for, at most, the significant information in a sound wave can be no more than the information required to describe its source. But the sources of many sounds are vibrating dynamic systems that obey a few well-defined physical laws. Thus, the nature of the physical world itself tends to limit the class of possible sounds. If the sources of sounds commonly occurring in nature may be shown to possess to within an acceptable approximation a simple, compact representation, there would be some reason to suspect that the function of the ear would be to analyze the sound so as to deliver up to the nervous system of the listener a set of neural messages corresponding to this compact representation of the source. Certainly, such an assumption is no less acceptable than the common notion that the ear delivers to the brain a frequency analysis of the sound.

Now, the class of sound sources in which we are interested are those which during the evolutionary process will have been pertinent in determining the survival of the species. Obviously, unnatural sounds, such as the informationless sinusoidal waves so frequently used in modern acoustical measurements, are of little concern here. Instead, the important sounds are those which were created by "something happening" to an object or structure common to the organism's environment. The information contained in the sound may be separated into two types: structural

information that specifies what the object is, and temporal information that specifies when and how violently "something" happened to the object.* These two kinds of information may be considered independently, provided the physical structure possesses the properties of a linear dynamical system for which superposition of various forcings produces a response that is the superposition of the individual responses. The assumption of linearity is carried throughout this paper. Although many sounds are created by systems that are not linear, nevertheless this approximation provides a useful conception.

Only part of the structural information characterizing the source can be implanted in the sound wave. The source may be described by many independent parameters, such as its physical dimensions, shape, and the mechanical properties of its materials. However, dimensional analysis requires that its dynamic vibrations, and hence the radiated sound wave, be determined by parameter groups each of which must have only the dimension of "time". The number of such groupings is generally fewer than the number of original parameters. For many of the parameters of like kind, e.g. the linear dimensions of different parts of the system, combine in such a way that a small change in any one may have but little effect upon the dominant vibration pattern of the source. Thus, the encoding process whereby the characteristics of the source are translated into sound is generally accompanied by a loss of information. The structural information derivable from the sound cannot exceed the information

* The adjective 'temporal' is here used to designate those aspects of the sound wave which are peculiar to definite instants in time as distinguished from the 'structural' aspects which are relatively invariant under translation in time.

content of this smaller set of derived parameters. However, the restriction that these derived parameters all have the dimension of "time" does not prevent them from conveying spatial information because the listener may learn from experience something of the velocity of sound in different media, and, with this information, the derived parameters do convey an implicit notion of the size and shape of the sound source.

To express these ideas more concisely, we resort to mathematical symbolism and denote by $h_{ij}(\tau)$ the value of the instantaneous sound pressure observed at position "i" τ seconds after an impulsive forcing of unit magnitude has been impressed upon the system at position "j". This function $h_{ij}(\tau)$ will be called the system function because it is completely determined by the parameters of the system. It represents all of the structural information that may be deduced from sound produced by excitation of the system. It is, however, dependent upon the particular positions that are used for exciting and observing the system. To obtain a set of parameters that are invariant under changes in the points of excitation and observation, we note that the system function can often be represented by expressions of the form

$$\begin{aligned}
 h_{ij}(\tau) &= \sum_k A_{ij}^{(k)} \exp(p_k \tau) && (\tau > 0) \\
 &= 0 && (\tau < 0)
 \end{aligned} \tag{1}$$

where the $p_k = \sigma_k + j\omega_k$ are the complex frequencies and the A_{ij} are the initial (complex) amplitudes of the exponential components of the system function. Whereas the set of amplitudes ($A^{(1)}, A^{(2)}, A^{(3)}, \dots$) will vary with changes in the points at which the system is driven and observed, the natural frequencies (p_1, p_2, p_3, \dots) are invariant under such changes

and therefore represent a most important property of the linear system. It is precisely these natural frequencies that, having the dimension of T^{-1} , epitomize the various grouping of the physical parameters specifying the sound source. By specification of the structural content of a sound wave, we shall mean, then, the specification of the p_k primarily, and the $A_{ij}^{(k)}$, secondarily.

This structural information provides a name, or signature, for the physical system responsible for the sound. To complete the specification of a meaningful sound, the excitation of the system must also be given. Most systems are passive and ordinarily at rest. They emit a sound only as a result of a forcing externally imposed upon them. The values of a sound wave $S_i(t)$ at "i" resulting from an arbitrary forcing wave $f_j(t)$ impressed upon the system at "j" may be represented in terms of the system function by

$$S_i(t) = \int_0^{\infty} f_j(t - \tau) h_{ij}(\tau) d\tau \quad (2)$$

provided the system is linear.

The relation, expressed by Eq. (2), governing the generation of a sound wave, suggests that the ear in analyzing the sound may first separate it into its structural and temporal components, (H-1). However, this relation alone will not permit unique factorization. The problem involved here is rather like the problem of finding the two factors, x and y, of some prescribed number, say 10; for every x there will be a $y = 10/x$ that satisfies the prescribed relation. Nevertheless, a unique solution is possible if we regard the problem as one of statistical inference in which likelihood values have been assigned a priority to the various permissible forcing and system functions. The problem is then

solved by determining that particular pair which have the greatest joint likelihood of occurrence.

Now, the essential characteristic of the forcing function is its unpredictability in time. Who is to say at what instant the maid will drop the dishes in the sink? Back of most acoustical events is a trigger that sets them off. These trigger actions are characterized by their time of occurrence, intensities, and nothing else. A forcing wave which activates a sound source should include in its representation a specification of the epochs and intensities of these trigger events. If these triggers act through an auxiliary physical system, the forcing function itself will have acquired structural content which, for our purpose, may be considered to arise from an intermediary system function $h_f(t)$ characterizing the forcing system. If we designate the epoch and magnitude of the n 'th trigger impulse by t_n and F_n , respectively, the forcing wave may be represented by an expression similar to Eq. (2):

$$f(t) = \int_0^t \left\{ \sum_n F_n \delta(t-t_n-\tau) \right\} h_f(\tau) d\tau \quad (3)$$

where $\delta(t)$ represents a unit impulse occurring at $t = 0$.

By substituting Eq. (3) into Eq. (2), it may be shown that the two system functions may be combined and the integration performed to obtain for the sound,

$$s(t) = \sum_n F_n h(t-t_n) \quad (4)$$

where $h(\tau)$ includes the structural information of the source and is given in terms of previously defined quantities by the convolution integral,

$$h(\tau) = \int_0^{\infty} h_f(\tau - \xi) h_{ij}(\xi) d\xi \quad (5)$$

Equation (4) is the principal result of this section and is the basis for further analysis of the sound. We next inquire how the values of the (F_n) and (t_n) parameters appearing in Eq. (4), and also the (A_k) and (p_k) parameters characterizing $h(t)$, may be determined by direct analysis of the sound wave.

IV. THE ANALYSIS OF MEANINGFUL SOUNDS

The temporal information represented by the set of t_n 's is perhaps the easiest to recover from the sound. To accomplish this, we first conclude that, by Eq. (4), the sound wave is merely the superposition of the sound waves that would have been created had each of the "trigger" actions occurred separately. Now, a most important property of an $h(\tau)$ represented by Eq. (1) is that it will be everywhere repeatedly time-differentiable except at $\tau = 0$. At that point an impulsive singularity will almost always develop after a few differentiations. The number of differentiations required will depend upon the $A_{ij}^{(k)}$ of the system function and, hence, will provide a measure of the position of the excitation with respect to the rest of the system including the listener. But, the significant fact is that repeated differentiation of $S(t)$ with respect to time will eventually lead to a series of impulses whose epochs are the t_n and whose magnitudes are proportional to the F_n . In this manner, the temporal information as represented by the (t_n) and (F_n) parameters may, in principle, be completely recovered from the sound wave.

The same representation for describing the impulsive excitations may also be used to describe random or white-noise excitations. This can be accomplished by letting the number of impulses per second, ν , become very large and the amplitudes of the individual impulses become very small in such a way that: (a) the interval between any consecutive epochs, t_k and t_{k+1} , is a random variable having a Poisson distribution with a mean value of $1/\nu$, and (b), the F_n all have the same magnitude but with random polarities that are equally likely to be either positive or negative. An argument similar to that just given for recovery of the impulse excitation from the sound wave may be given for the detection of random noise excitation. It is concluded that by forming successive time derivatives of the sound wave, the essential information may be recovered for either impulse or random noise excitation of the system.

There is little need to describe how the analysis outlined above can be accomplished with electronic circuits. However, it is pertinent to mention that the simple mechanism of the inner ear to be developed in this dissertation is capable of accounting for a fourth place-derivative along the basilar membrane. Assuming that the basal end of the cochlea serves as a delay line, this mechanism together with a neural gradient mechanism may conceivably provide neural stimulation proportional to the sixth time derivative. (H-2). This would be adequate to recover the excitation characteristics of most common sounds.

Is it possible that the structural information represented by the set (p_k) of natural frequencies may likewise be recovered from a segment of the sound wave? We have already stated that the system function $h(\gamma)$ possesses analytic properties of such a nature that exact specification of its values over any small interval of positive γ is sufficient to definite it for all values of γ . Since differentiation

of the system function with respect to time leaves these natural frequencies unchanged and simply changes the original set of amplitudes (A_k) into a new set ($p_k A_k$), exact specification of any one of the time derivatives of the system function over the corresponding interval should also be adequate to determine the (p_k). (This result may explain why time-differentiated speech is still highly intelligible.)

Consider that segment of sound wave beginning just after the impulsive excitation at t_g and ending just before t_{g+1} . By Eq. (4), this particular segment may be represented by

$$\begin{aligned} S(t) &= \sum_{n=1}^g F_n \sum_k A_k \exp p_k (t-t_n) \\ &= \sum_k A_k' \exp p_k (t-t_g) \end{aligned} \quad (6)$$

$$\text{where } A_k' = A_k \sum_{m=1}^g F_m \exp p_k (t_g-t_m) \quad (7a)$$

In other words, the sound wave occurring during the interval between any consecutive pair of impulsive excitations is composed of exponential components that have exactly the same complex frequencies as those of the system function. And, just as with the system function, detailed analysis of only this segment of a sound wave should permit determination of the (A_k') and the (p_k) that represent it. So, the answer is "Yes, the values of the natural frequencies are, in principle, determinable from a finite segment of the sound." (The presence of noise or uncertainty in the data of the sample will obviously limit the accuracy with which this determination can be made.)

Furthermore, the (A_k') amplitude coefficients obtained from this analytic sample of a sound wave will closely resemble the (A_k) para-

meters of the system function if the time interval is sufficiently great compared to the damping time of the exponential components. The explanation for this is apparent from consideration of Eq. (7a), which may be summed for the special case of repetitive impulsive excitation occurring at regular intervals, Δ ,

$$A_k' = \frac{F_m}{1 - \exp(p_k \Delta)} \cdot A_k \quad (7b)$$

If the damping, $-\zeta_k$, of the k'th natural mode is so large that $\exp(p_k \Delta) \ll 1$ it follows that A_k' is nearly equal to $F_m A_k$.

Perhaps the most remarkable thing about these exponential components is their indestructibility when subjected to linear operations. For instance, if $h(t)$ is passed through a linear filter, whose own system function is $h_f(t) = \sum_n B_n \exp(p_n t)$, the output of this filter will be

$$\begin{aligned} h_o(t) &= \int_0^t h(t-\tau) h_f(\tau) d\tau \\ &= \sum_k \sum_n \frac{A_k B_n}{p_k - p_n} \left[\exp(p_k t) - \exp(p_n t) \right] \end{aligned} \quad (8)$$

and we see that the output consists of a set of these same natural frequencies (p_k) plus a new set (p_n) associated with the filter. The amplitude A_k of the k'th component will be changed by the filter from A_k to $A_k \sum_n B_n / (p_k - p_n)$. By making all of the filter (p_n) about equal to a particular p_k of the input signal, the amplitude of that component may be greatly enhanced over those of all other components appearing in the output of the filter. This initial enhancement will continue to be effective, however, only if the damping of each of the (p_n) natural

frequencies of the filter is in excess of the particular p_k chosen, for then these other components will die away more rapidly leaving after a short time only the desired component in the output of the filter. Thus, an ordinary linear filter, each of whose natural frequencies has a larger decrement than the natural frequency to be measured, will provide a means for selecting any one of the (p_k) frequency components provided the others are sufficiently separated in frequency. Assuming that this preliminary filtering has been accomplished, we next seek a filtering principle that will enable us to measure directly the complex frequency $\sigma_k + j\omega_k$ of this single exponential component.

V. THE PHASE-PRINCIPLE AND COMPLEX-FREQUENCY ANALYSIS

Any device for measuring the complex frequency of a damped wave should obviously yield the correct frequency value when applied to a sine wave since the latter is but a special case of the former. Now, the common feature of both sine waves and damped waves is their phase. Accordingly, a device that utilizes the phase of a signal, as determined by its zero crossings, should function just as well for damped waves as for sine waves. It remains to show how the phase characteristics of simple tuned filters may be employed to achieve the desired filtering action.

A. A Phase Filter for Obtaining Resonagrams

Consider a simple tuned filter represented in Fig. 1 by its natural

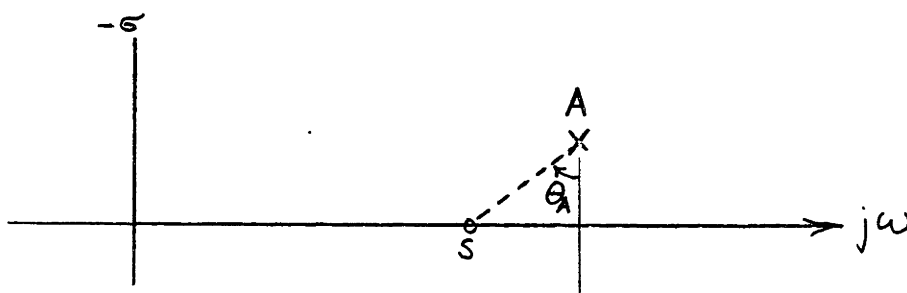


Fig. 1. Representation of the phase characteristic of a simple tuned filter upon the complex-frequency plane.

frequency A . If a sinusoidal input signal of frequency S is suddenly applied, the response of the filter will consist of a component of frequency S plus a component of frequency A . After a time determined by the excess decrement of A over that of S , the output will consist for practical purposes only of a wave having a frequency S . It is well known that, within the accuracy of the narrow-band approximation, the phase angle between the phases of the input and output waves of the filter is simply θ_A , the angle between the σ -axis and the line drawn from the excitation frequency to the natural frequency of the tuned filter. This angle, illustrated in Fig. 1, will continue to be a correct representation of the input-output phase relation even when the input is a complex-frequency wave which is growing or decaying. This simple geometrical construction enables one to visualize easily the phase angle θ_A associated with an excitation wave of any complex frequency.

It is a fact of particular significance that increasing the damping of the input wave has exactly the same effect upon the phase characteristic as decreasing the damping of the filter. That is, the change in phase with frequency will be more abrupt in the region of resonance if the input wave is moderately damped than if it is undamped or growing in amplitude. This suggests that a phase principle may prove particularly effective for analyzing sound into damped-wave components.

Next, consider a signal fed into two simple filters A and B , each having the same bandwidth but A tuned slightly higher in frequency than B . Each of these circuits will have a phase characteristic similar to that just described. Since the same signal is fed into both, the phase angle between the two output waves will be $\theta_1 = \theta_A - \theta_B$ and will vary with the frequency of the input signal as shown in Fig. 2. The variation of θ_1 with frequency resembles rather closely the typical resonance curve of a double-tuned

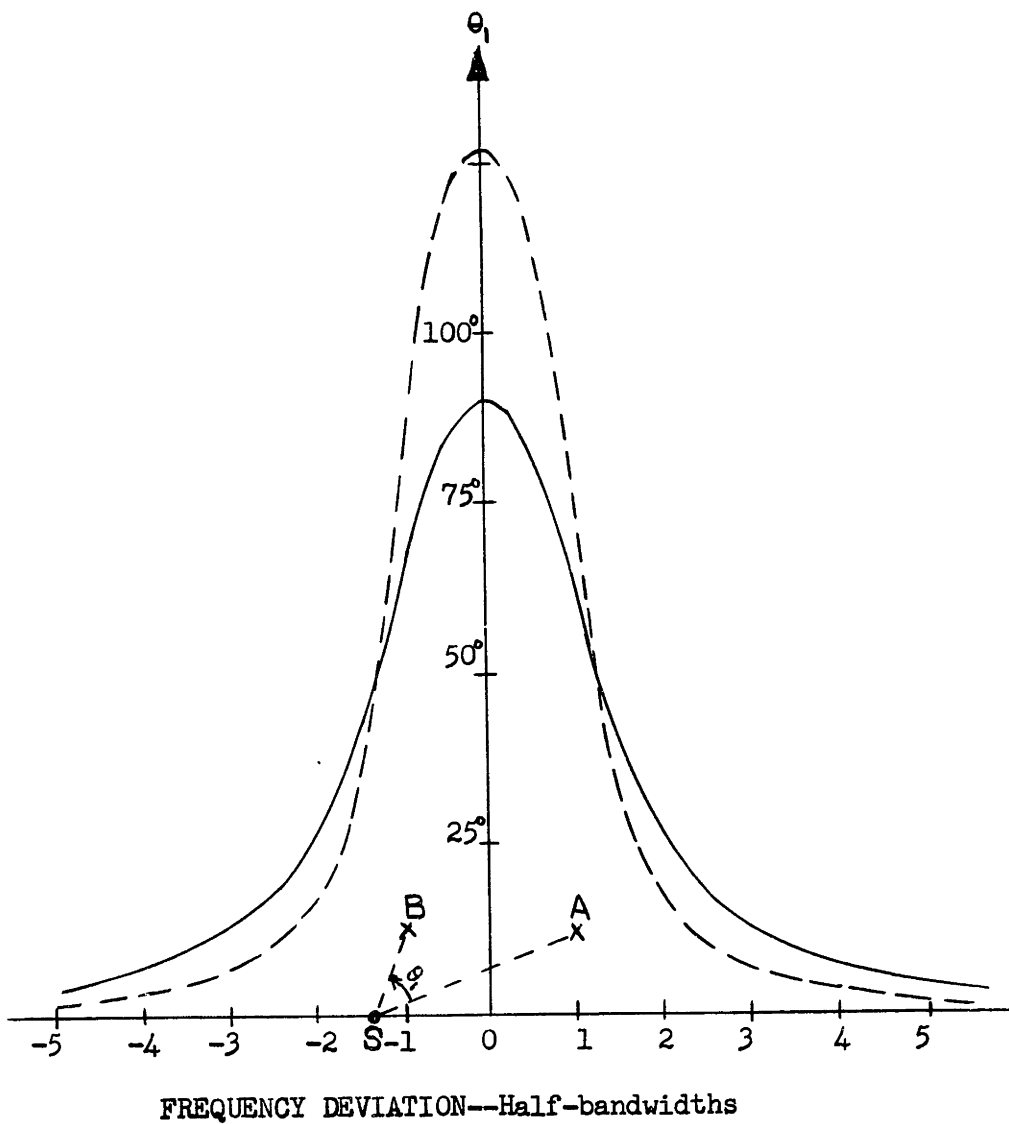


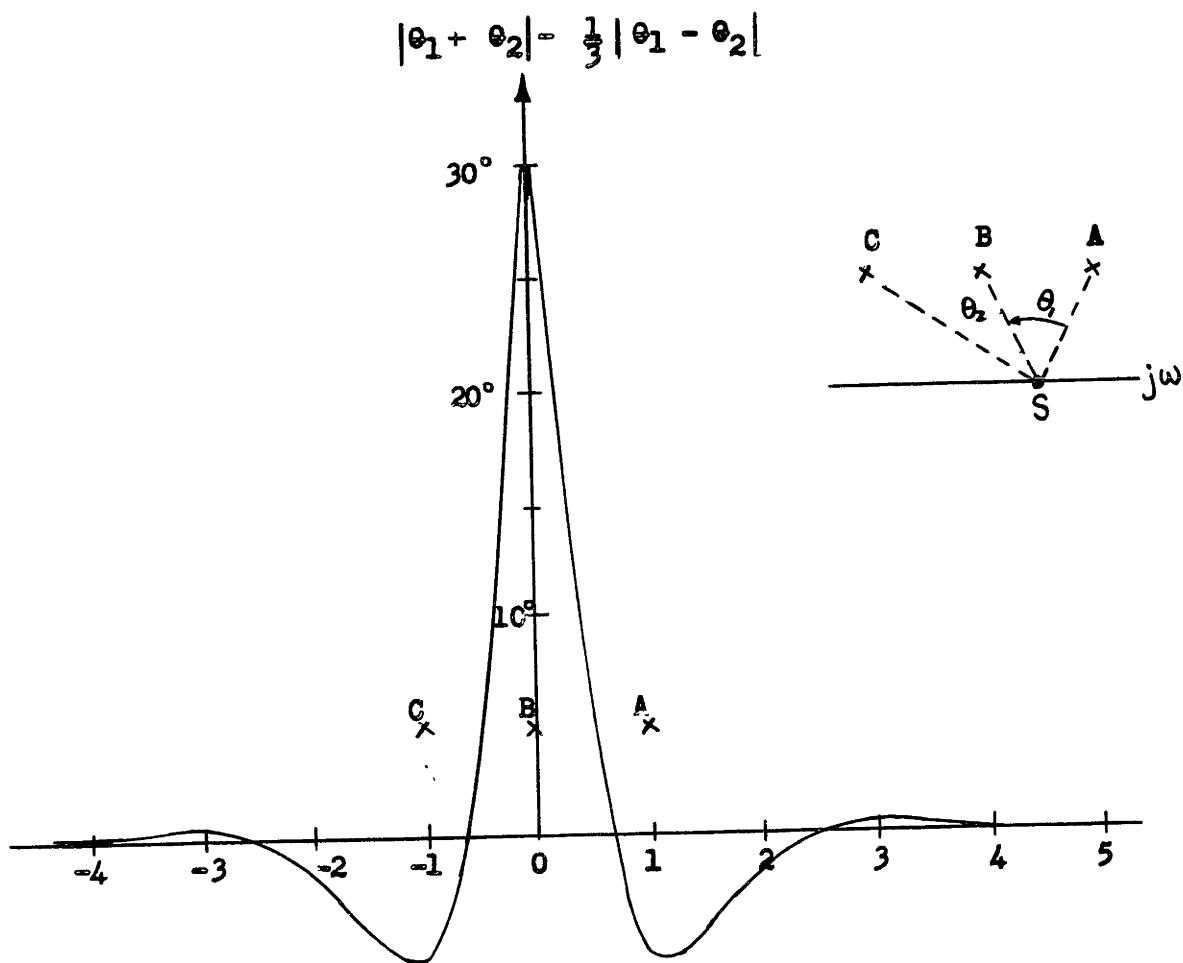
Fig. 2. Phase difference $\theta_1 = \theta_A - \theta_B$ as a function of the frequency of an input wave having zero decrement (solid line); decrement one-half that of the tuned circuits (broken line).

circuit except that θ_1 is independent of the amplitude of the forcing signal:

Figure 2 also shows the variation of θ_1 as a function of the frequency of the damped input wave having a decrement one-half that of the filters. As expected, the selectivity curve for these damped waves is indeed sharper than for sine waves. But even more significant, the maximum value of θ_1 may now be calibrated directly in terms of the decrement of the input wave. We have thus found a principle for measuring the natural frequency $\sigma + j\omega$ of an input wave: the frequency at which the maximum θ_1 occurs gives ω and the value of this maximum gives σ .

The sharpness of the selectivity characteristic may be improved still further by the addition of a third tuned filter. Let the natural frequencies of these three circuits be arranged as indicated in Fig. 3. The quantity $|\theta_1 + \theta_2|$ will resemble the resonance curves shown in Fig. 2. The quantity $|\theta_1 - \theta_2|$ will have the shape of a rectified discriminator curve and, by subtracting some fraction of it from $|\theta_1 + \theta_2|$, a much sharper selectivity characteristic may be obtained as illustrated in Fig. 3.

A filter utilizing the phase principle illustrated in Fig. 3 has been built and used in place of the ordinary filter in the sound spectrograph which normally operates on an amplitude principle and is, therefore, incapable of separating the structural information from the temporal. This filter was found to perform as expected. The output from each of the three simple-tuned circuits was converted to a square wave by repeatedly amplifying and limiting the wave. These square waves were then applied to coincidence circuits in which voltages proportional to $(\theta_1 + \pi/2)$ and $(\theta_2 + \pi/2)$ were developed. (The $\pi/2$ phase shift was added to the middle filter to make the quiescent phase more nearly equal to the average phase for random input noise.) These two voltages were filtered and combined in the desired



FREQUENCY DEVIATION—Half-Bandwidths

Fig. 3. Sharpened response curve corresponding to

$$|\theta_1 + \theta_2| = \frac{1}{3} |\theta_1 - \theta_2|.$$

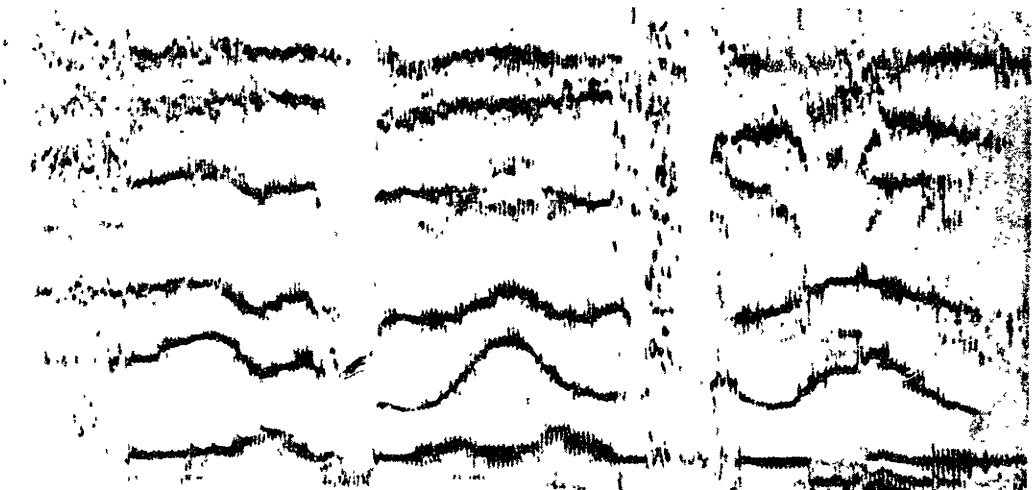
Sonagram



Resonagram



Resonagram



S a y I B o y O u t G o N e w

Figure 4. Comparison of a resonagram obtained with the filtering scheme of Figure 3 with the ordinary sonagram obtained with the wide-band filter of the sonagraph.

manner to control the marking amplifier of the sonograph. Figure 4 compares the sonagram and resonagram representations of the same speech sample, "Say I Boy Out Go New," for two different amounts of sharpening. It is apparent that the natural frequencies are revealed far more vividly in the resonagrams than in the regular sonagram. On the other hand, the resonagrams provide very little information about the intensity or excitation of the source—which is what was intended.

B. Random Noise Excitation

In the preceding discussion, the single complex-frequency wave applied to the input of the filter may be considered to have been produced by the impulsive excitation of a resonator whose natural frequency is $\sigma_s + j\omega_s$. We wish next to inquire what the output of the phase filter would be if the source resonator were excited by random white noise excitation instead of by discrete impulses. It is also desirable to introduce at this time a generalized representation for the three system functions* as shown in Fig. 5. In practice, the filter would be made to scan in frequency by a frequency converter, also shown in Fig. 5. The phase relation would be measured simply

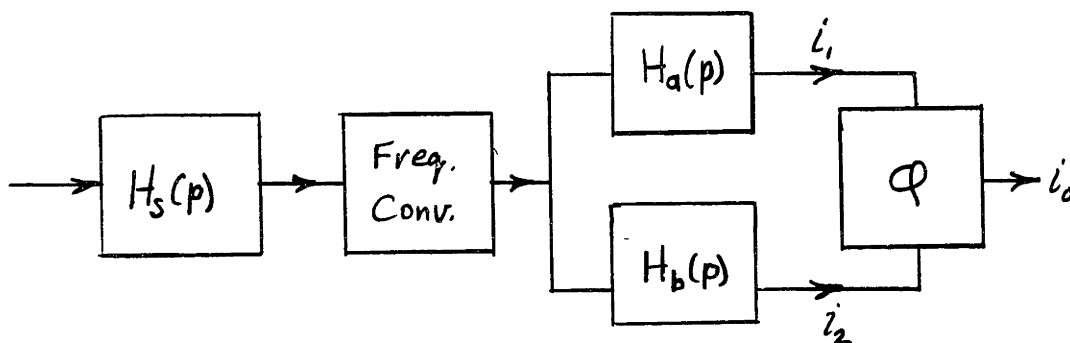


Fig. 5. Generalized Phase Filter.

*From laplace transform theory, if $h(t) = \sum_k A_k \exp(p_k t)$ then $H(p) = \sum_k A_k / (p - p_k)$ so either expression specifies the system transformation.

at some higher frequency by a coincidence circuit " ϕ " in which the output current is bivalued:

$$i_0 = 1 \text{ if both } i_1 \text{ and } i_2 \text{ are positive;}$$

$$i_0 = 0 \text{ if either } i_1 \text{ or } i_2 \text{ is negative.}$$

It follows that the average value of i_0 will be a measure of the average phase difference between i_1 and i_2 .

Now, it is well known that when the excitation of the source is a random gaussian noise, the output currents i_1 and i_2 will likewise be random and characterized by gaussian statistics, (R-2, M-2). In particular, the joint probability that both currents will lie within the intervals $(I_1, I_1 + dI_1)$ and $(I_2, I_2 + dI_2)$ is $W(I_1, I_2)dI_1dI_2$, and the average value of i_0 is simply the expectation, that both i_1 and i_2 will simultaneously be positive. For a bivariate gaussian distribution, this may be shown to be

$$\langle i_0 \rangle_{AV} = \int_0^{\infty} \int_0^{\infty} W(I_1, I_2) dI_1 dI_2$$

$$= 0.25 - (1/2\pi) \sin^{-1} (-r) \quad (9)$$

in which r is the correlation between the output currents and is given by the real part of the complex correlation. $\langle i_1 i_2 \rangle_{AV} / (\langle i_1 i_1^* \rangle_{AV} \langle i_2 i_2^* \rangle_{AV})^{1/2}$ where the asterisk indicates that the complex conjugate of the quantity is to be taken. Incidentally, addition of a fixed phase shift to one of the filters adds the same phase shift to the complex correlation coefficient.

Thus, the correlation of a sine wave with a cosine wave of the same frequency has a complex value of "j"; its real part is zero, but if a 90-degree phase shift were added to one of the channels, the correlation could be made unity.

To evaluate the correlation obtained for various source system functions, we utilize the fact that the second-order moments characterizing the bi-variate distribution for white-noise excitation are proportional to the moments calculated from the response to a single unit impulse, for which case,

$$\langle i_1 i_2^* \rangle_{Av} = \int_0^{\infty} i_1 i_2^* dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} H_1(j\omega) H_2^*(j\omega) d\omega \quad (10)$$

where $H_1(p)$ and $H_2(p)$ are the laplace transforms of the $i_1(t)$ and $i_2(t)$ resulting from this impulsive excitation. The integral on the right may usually be evaluated by contour integration around the left half of the p-plane so that

$$\langle i_1 i_2^* \rangle_{Av} = \sum \text{Res. } H_1(p) H_2^*(-p) \quad (11)$$

where only the residues at the poles associated with $H_1(p)$ are to be taken.

To illustrate with the simple case where the system functions are each characterized by a simple resonance $H_x(p) = 1/(p - p_x)$, we have $H_1 = H_s \cdot H_a$ and $H_2 = H_s \cdot H_b$. Evaluating Eq. 11, we find

$$\langle i_1 i_2^* \rangle_{Av} = \frac{1}{p_b^* - p_s^*} \left[\frac{1}{(p_a + p_b^*) \cdot (p_s + p_b^*)} - \frac{1}{(p_a + p_s^*) \cdot (p_s + p_s^*)} \right],$$

$$\langle i_1 i_2^* \rangle_{Av} = \frac{1}{p_a - p_s} \left[\frac{1}{(p_a + p_s^*) \cdot (p_a + p_a^*)} - \frac{1}{(p_a^* + p_s) \cdot (p_s + p_s^*)} \right], \quad (12)$$

with a similar expression for $\langle i_2 i_2^* \rangle_{Av}$ obtained by interchange of subscripts. If all frequencies are normalized so that the half-bandwidth (i.e. the decrement) of each filter is unity, and the usual bandpass to low-pass transformation is invoked so that the center of the band corresponds to $p = 0$, the various natural frequencies become: $p_a = (-1 - j)$, $p_b = (-1 + j)$, $p_s = (\sigma + j\omega)$, and we may calculate from the above equations the variation of $\langle i_0 \rangle_{Av}$ as either σ or ω is varied. Since the resonance curves ob-

tained for variation of ω are symmetrical, only the right half is shown in Fig. 6. The curve for $\epsilon = 0$ is identical to that for sinusoidal excitation shown in Fig. 2. As the damping of the source resonator is increased, the maximum value of $\langle \dot{i}_o \rangle_{Av}$ decreases rather than increases as in Fig. 2. However, the value of this maximum still depends upon the damping of the source resonator. The left half of Fig. 6 shows this maximum value, and also the minimum value obtained with large detuning, as a function of ϵ . Thus, the phase filter provides a direct measurement of the bandwidth of the source resonator when excited by either impulses or white noise--only the calibration is different in the two cases.

C. A Simplified Form of Phase Filter

Since completing the experimental work on the resonograph, it has become evident that a phase filter of simpler form might offer superior performance. Although this simplified form has not been tested experimentally, it appears to be sufficiently straightforward that, in view of the successful performance of the more complicated form, it deserves description here.

In this simplified form, the frequency-sensitive phase shifter is produced by a single all-pass section having a system function $H_b(p) = (1 - p)/(1 + p)$. For $p = 0$, $H_b(p) = 1$ and there is no phase shift. But for any wave for which $|p| \gg 1$, $H_b(p) \simeq -1$, and the all-pass section, in effect, simply reverses the polarity of the wave. Thus, discrete impulses or white noise applied to this phase filter will produce a vanishingly small $\langle \dot{i}_o \rangle_{Av}$ since i_1 and i_2 will nearly always be of opposite polarity. On the other hand, when the input signal has acquired structural content approximating that of the all-pass section, these phase relations are modified and a non-zero value for $\langle \dot{i}_o \rangle_{Av}$ is obtained.

Figure 7 shows the selectivity characteristics of this filter for a source resonator of zero damping, and also for damping equal to one half that

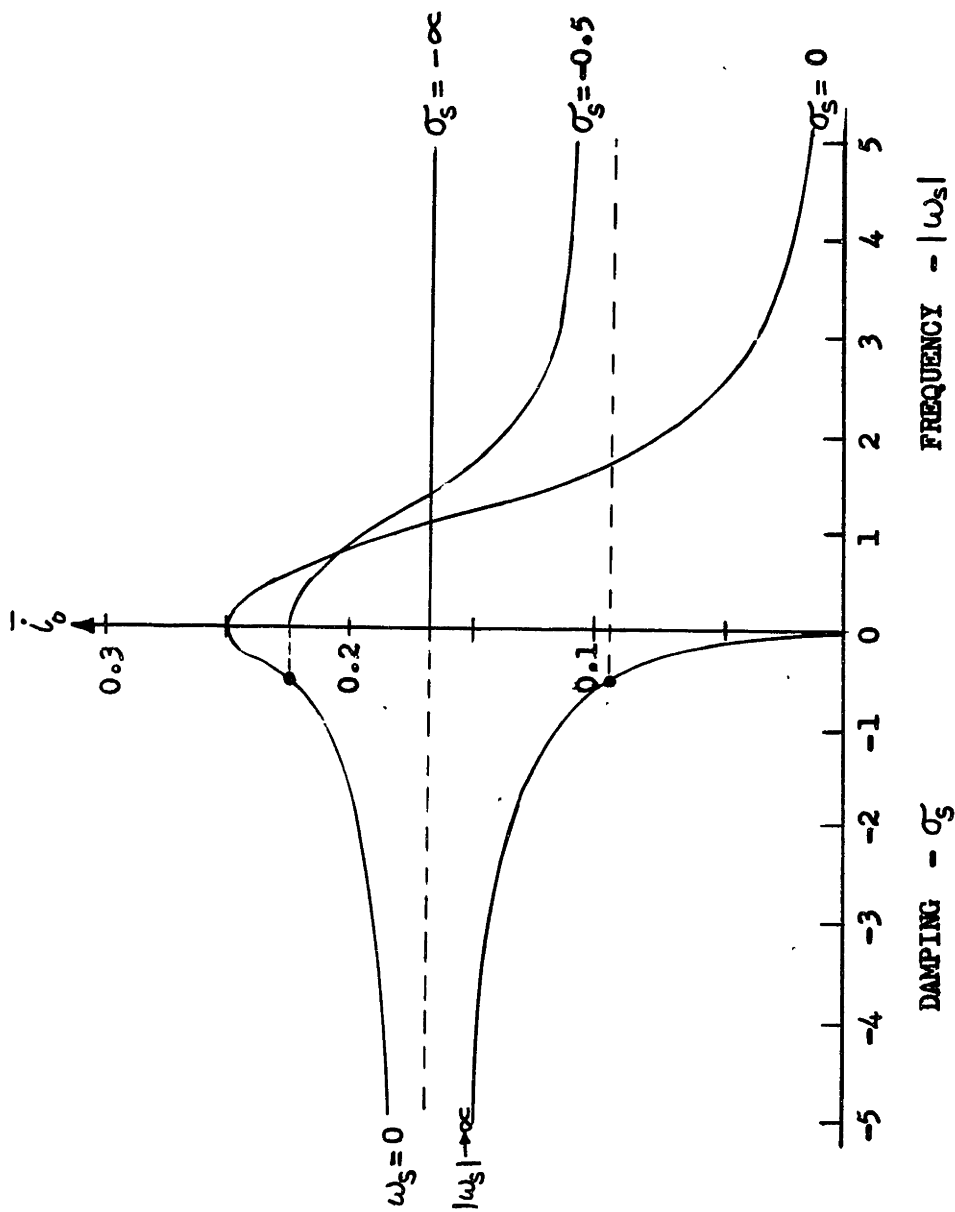


Fig. 6. Response of the filter of Fig. 3 as a function of the tuning and the damping of the source resonator when excited by white noise.

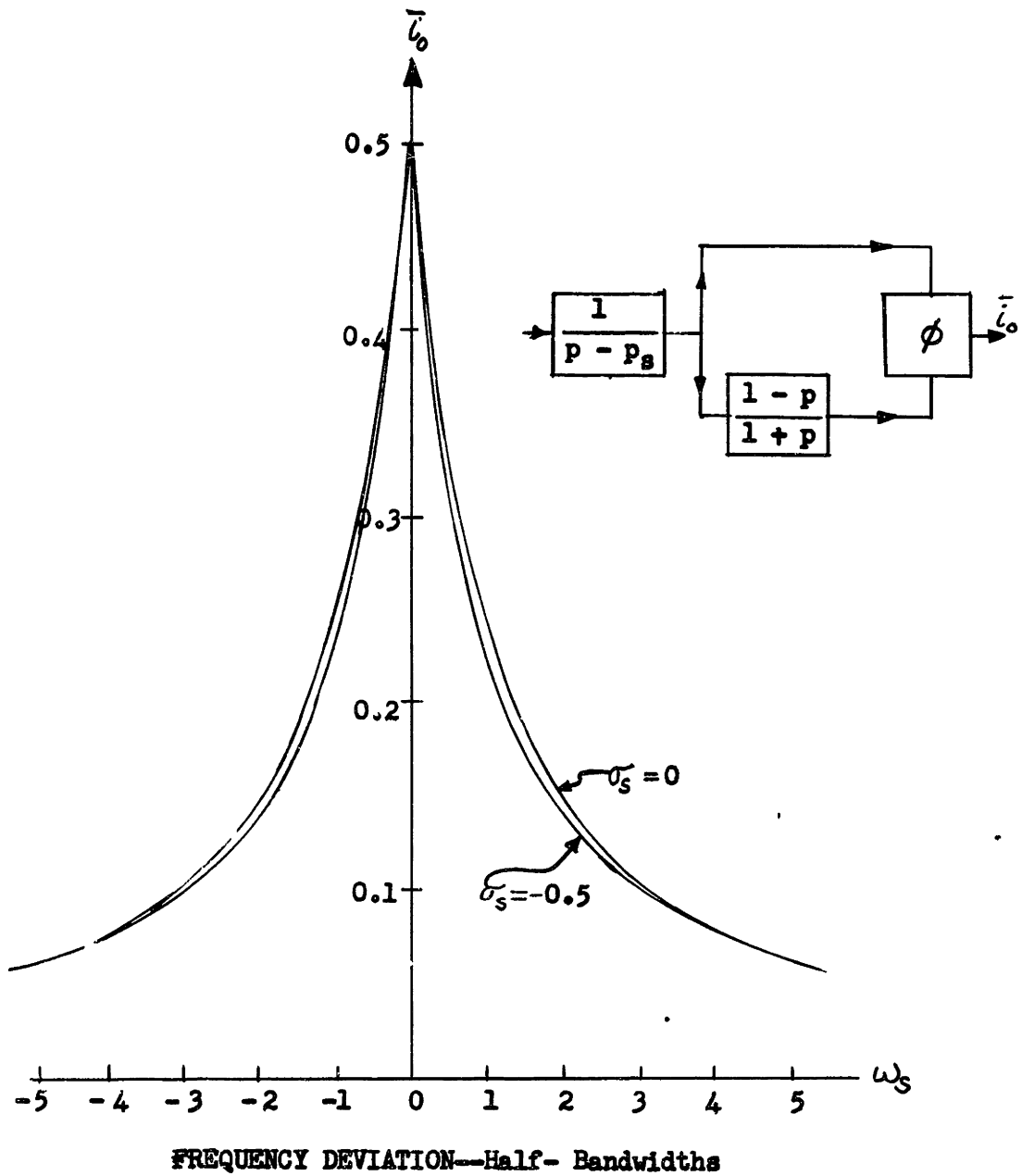


Fig. 7. Response of the simplified phase filter as a function of the frequency of the source resonator when subjected to impulsive excitation. A source damping of zero and of one-half that of the filter are considered.

of the filter. These curves correspond to the curves shown in Fig. 2. for impulsive excitation of the source resonator. To measure successfully the phase resulting from impulsive excitation, it would be necessary to build into the phase meter a sampling mechanism so that the phase angle is measured only during a brief interval of time whose beginning is delayed sufficiently after the epoch of the impulse to permit the filter transients to die away, and whose termination occurs before the damped signal component has vanished into noise. The sampling mechanism could be triggered by the energy spike and appears to be practical of realization. As indicated in Fig. 7, the simplified phase-filter gives directly a sharply peaked response, the maximum value of which is independent of the damping of the input resonator--a feature that may be useful when driving the marking amplifier of a sonagraph.

The output of the simplified phase-filter, when the source resonator is excited by white noise, may be calculated by the equations of the previous section. Figure 8 shows the selectivity curves for a source damping of zero, one-half, and equal to the filter damping. The maximum value of $\langle i_o \rangle_{Av}$ is related to the decrement of the source resonator by

$$\bar{i}_{o \max} = 0.5 - \frac{1}{2\pi} \cos^{-1} \left[\frac{1 + \sigma_s}{1 - \sigma_s} \right] \quad (13)$$

and this form of phase filter should thus provide a more accurate method of measuring the source bandwidth because $\bar{i}_{o \max}$ varies over its full range of 0.5 to 0 as the source damping is increased from 0 to ∞ .

D. The Masking Effect of A Second Resonance or Noise.

Thus far, the source has been considered to have a simple resonance. In this section, we investigate the effect of a second source resonance occurring simultaneously at a nearby frequency. Since there is no amplitude selectivity in the simplified form of phase filter, it will be used to

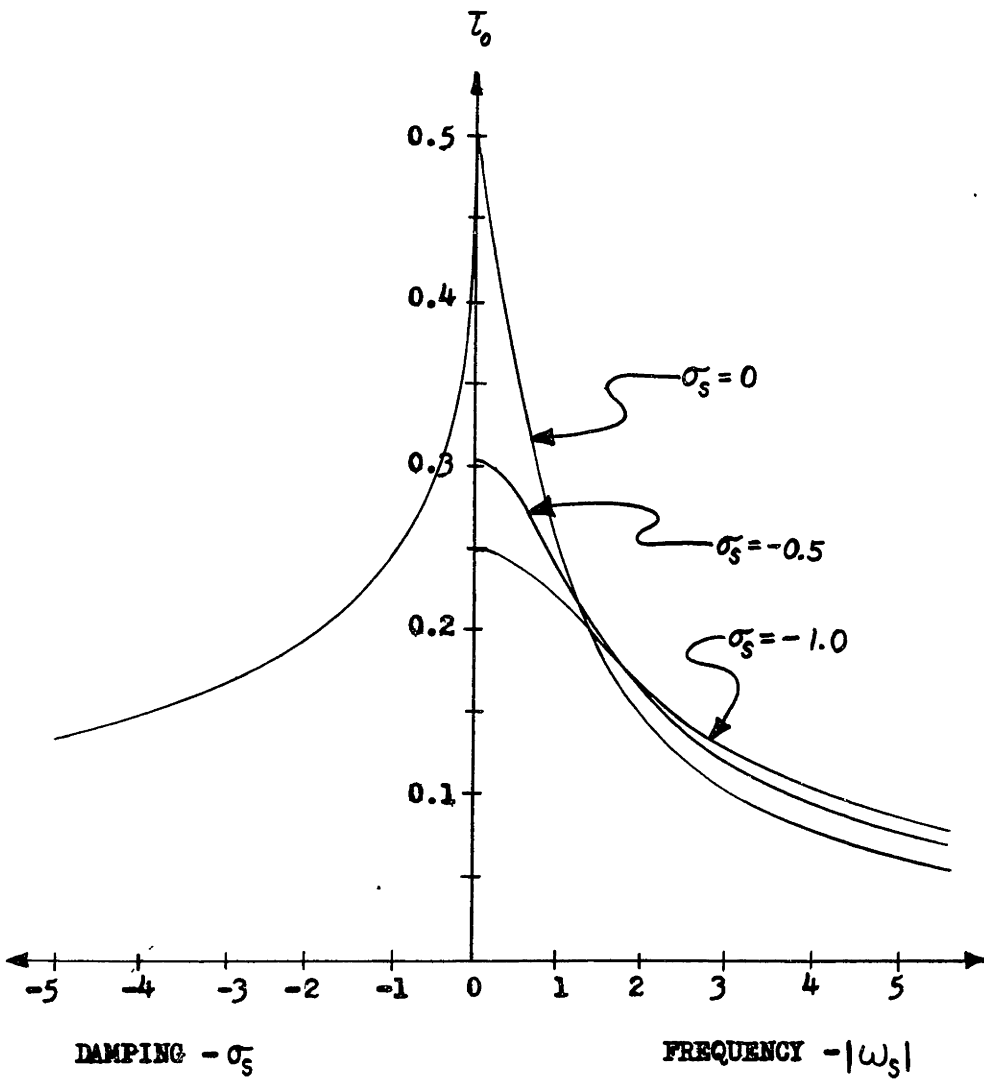


Fig. 8. Same circuit as in Fig. 7 except noise source is excited by white noise.

demonstrate these effects in their most serious form.

We assume that the source resonator is characterized by a principal resonance at a complex frequency p_s and a secondary resonance having a relative amplitude A and a complex frequency p_4 . That is, $h_s(t) = \exp(p_s t) + A \exp(p_4 t)$, and from the equations of the previous section, one finds that

$$\begin{aligned} \langle i_1 i_1^* \rangle_{Av} &= \langle i_2 i_2^* \rangle \\ &= -\frac{1}{2\sigma_s} - \frac{|A|^2}{2\sigma_4} - 2 \operatorname{Re} \left(\frac{A}{p_4 + p_s^*} \right), \end{aligned} \quad (14)$$

and

$$\begin{aligned} \langle i_1 i_2^* \rangle_{Av} &= - \left[\frac{1}{2\sigma_s} + \frac{A^*}{p_4^* + p_s} \right] \left[\frac{1 + p_s}{1 - p_s} \right] + \\ &\quad - \left[\frac{|A|^2}{2\sigma_4} + \frac{A}{p_4 + p_s} \right] \left[\frac{1 + p_4}{1 - p_4} \right]. \end{aligned} \quad (15)$$

To obtain the response of the filter to two sine waves, one of unit amplitude and frequency $j\omega_s$, and the other of amplitude A and frequency $j\omega_4$, we may set $\sigma_s = \sigma_4 = \sigma$ and let $\sigma \rightarrow 0$. This gives for the correlation coefficient,

$$r = \frac{1}{1 + |A|^2} \left[\frac{1 - \omega_s^2}{1 + \omega_s^2} + |A|^2 \frac{1 - \omega_4^2}{1 + \omega_4^2} \right]. \quad (16)$$

Equation (16) is easily generalized to include any number of incoherent sinusoidal components--the resulting correlation is simply the weighted average of the correlations of the individual components, each being weighted in proportion to the square of the amplitude of that component. From this fact, it may be shown that the correlation resulting from a band of white noise, Δ units wide and centered at w , applied directly to the phase filter will be

$$r_n = (2/\Delta) \left[\tan^{-1}(w + \Delta/2) - \tan^{-1}(w - \Delta/2) \right] - 1 \quad (17)$$

Thus, the masking effect of this band of noise, having a power N , upon a single sinusoidal signal of unit power and frequency ω_s will be represented by the resultant correlation

$$r = \frac{1}{1+N} \left[\frac{1-\omega_s^2}{1+\omega_s^2} + N r_n \right], \quad (18)$$

Figure 9 shows the variation of the output of the phase filter as it is scanned in frequency for the particular case where the two undamped resonances of the source differ in frequency by 5 units. The relative amplitude of the second resonance is taken as a parameter. These curves clearly show the masking effect of a strong signal upon a weaker one, as well as the reduction and broadening of the response to a strong signal due to the phase modulation produced by a weaker one. It should be noted that the shape of the response curves is determined by the ratios of the signal characteristics in that the masking effect is proportional to the difference of intensities in decibels of the two signals—a fact bearing upon the application of the phase-principle in auditory theory.

Examination of Fig. 9 shows that the best indicator of resonance is the marked downward curvature of the response curve at resonance. As Licklider and I have already suggested, formation of the 2nd frequency derivative of the response curve will considerably sharpen its selectivity in frequency (H-2). This operation may be instrumented by employing three of the simplified phase filters, the first tuned slightly higher and the second tuned slightly lower in frequency than the third, and then combining their output \bar{i}_s , weighted by -1, -1, and 2 respectively. This composite filter demonstrates how it may be possible by the use of phase alone to achieve resolution of frequency with a structure that does not possess any of the amplitude selectivity properties of the usual filter.

According to this curvature hypothesis, the presence of a weak signal will be noted only if it strong enough to produce a downward curvature

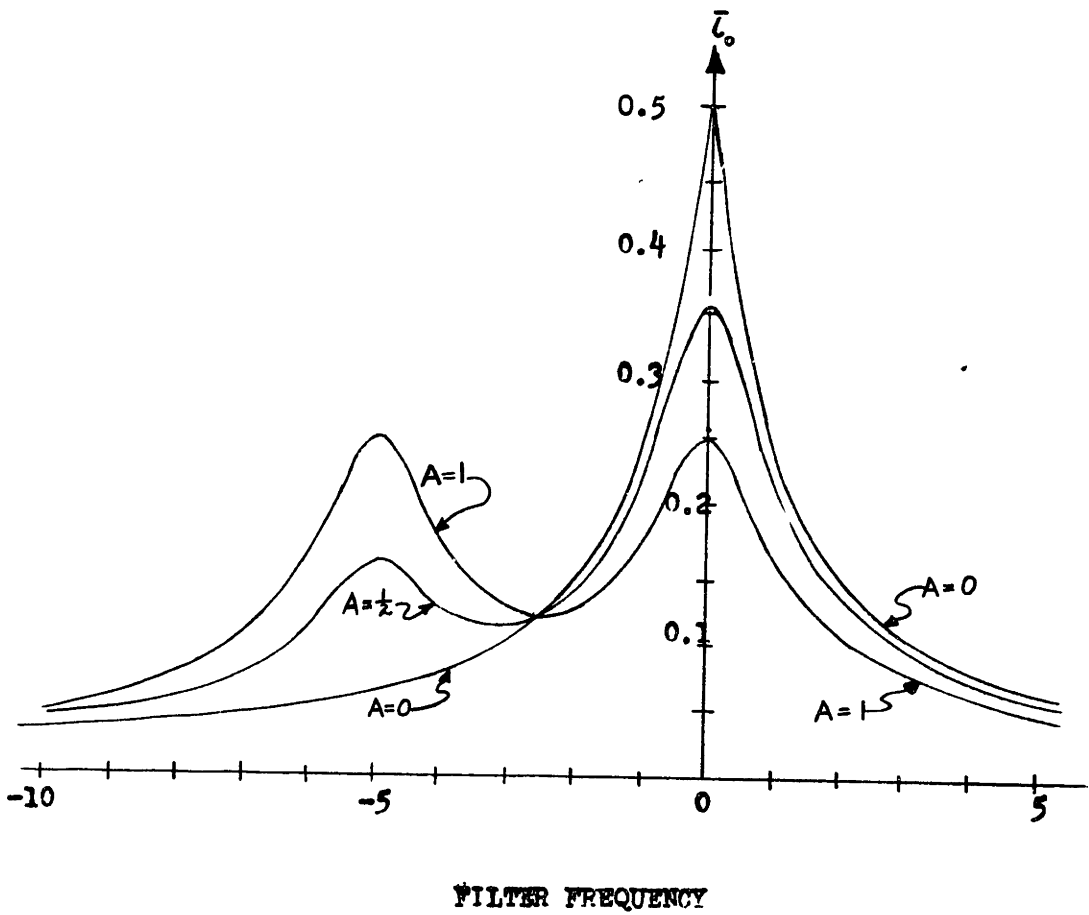


Fig. 9. Response of simplified phase filter as a function of filter frequency when the input signal consists of two sinusoidal components differing in frequency by 5 units. The relative amplitude of one is taken as a parameter.

component which is in excess of the upward curvature of the skirt of the strong signal. Since \bar{l}_0 and the correlation r are monotonically related by Eq. (9), we may think of the correlation r as being a measure of the response of the filter, and curves of r versus frequency similar to those of Fig. 9 may be plotted. As the second signal becomes weaker, the curvature at the frequency of the weak signal will eventually become zero. If this is assumed to be the masking threshold of the weak signal, we may study the variation of this threshold as a function of the frequency separation between the two tones as follows:

The correlation for a single frequency input at $\omega = 0$ is related to the frequency ω of the filter by Eq. 10 with $A = 0$, or

$$r = \frac{1 - \omega^2}{1 + \omega^2} \quad (19)$$

If the correlation is plotted as a function of the filter tuning ω , the curvature of this curve will be

$$\frac{d^2 r}{d\omega^2} = -4 \frac{1 - 3\omega^2}{(1 + \omega^2)^3} \quad (20)$$

The maximum curvature, achieved at the frequency of the signal, is simply 4.

Consider now the effect of adding a second weak signal of relative amplitude A at a frequency differing from the first by δ . By Eq. (16), the resultant correlation is given by the weighted average of the individual correlations. Similarly, the resultant curvature may be found by forming the weighted average of the individual curvatures. Accordingly, at the frequency of the weak signal, the resultant curvature in the correlation spectrum will be

$$\frac{d^2 r}{d\omega^2} = \frac{-1}{1 + A^2} \left[4A^2 + 4 \frac{1 - 3\delta^2}{(1 + \delta^2)^3} \right] \quad (21)$$

Masking threshold occurs, according to our hypothesis, when this curvature vanishes, or when

$$A = \sqrt{\frac{3\delta^2 - 1}{(1 + \delta^2)^3}} \quad (22)$$

In order for the curvature to vanish, the weak signal must be sufficiently separated in frequency to lie on the upward curvature parts of the strong signal "skirts". Thus, to be distinguished as a separate signal $|\delta|$ must exceed $1/\sqrt{3} = 0.576$.

The fact that there is a region, extending plus and minus 0.576 units to either side of the frequency of the strong signal within which a second signal cannot be discerned as a separate region of downward curvature, suggests that this region may correspond to the critical bandwidth observed by Fletcher in his study of the masking effects by noise in the ear. From p. 1009 of ref. S-2, this critical bandwidth is seen to be 73 cps at 1200 cps. Thus, the half-bandwidth of the equivalent analyzing filter would be $0.576(BW/2) = 73 \text{ cps}/2$ or $BW/2 = 64 \text{ cps}$. Using this frequency calibration, we may compare in Fig. 10 the masking threshold, as calculated from Eq. (22), with the experimental data taken by Wegel and Lane (S-2, p. 1006) for the masking effect of a 1200 cps tone at 80 db sensation level upon a weaker second tone of somewhat higher frequency. Two observations are pertinent:

a) The frequency calibration obtained from the critical-bandwidth datum appears to give agreement as to frequency separation at which maximum masking effect occurs and also the slope of the threshold curve at wider frequency intervals.

b) The threshold for the ear appears to be nearly a constant 10 decibels less than that based on the correlation model.

It is possible to explain the difference in masking threshold predicted by this simple model and that actually observed in the ear by noting that the correlation value r is the average over all time. Actually, when the

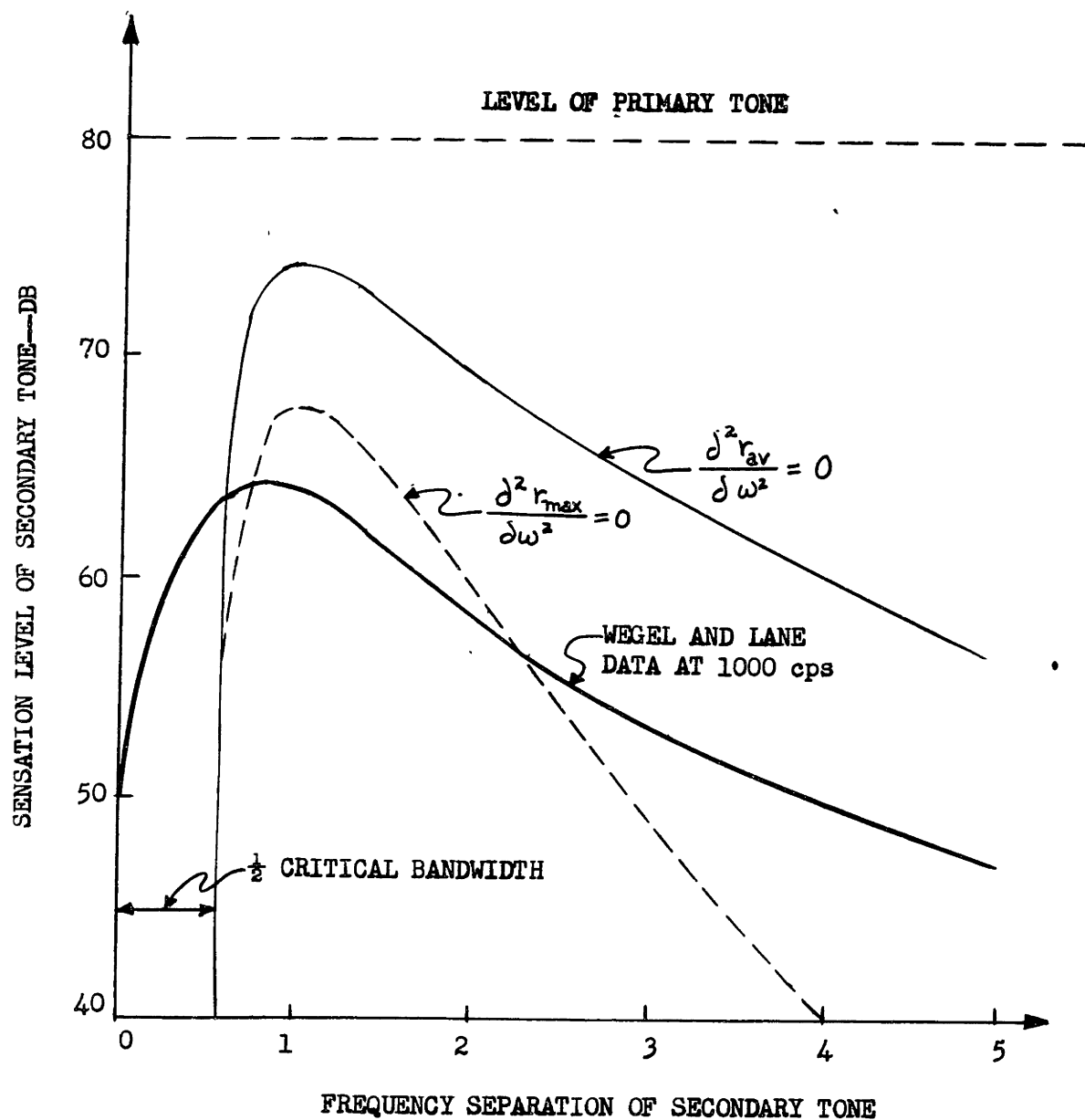


Fig. 10. Comparison of masking thresholds based on the curvature hypothesis with the experimental data of Wegel and Lane for the sensation thresholds of a higher frequency tone in the presence of a 1,200 cps tone at an 80 decibel sensation level. The solid curve marked $\frac{\partial^2 r}{\partial \omega^2} = 0$ is based on the average correlation, whereas the broken curve is for detection at the peak of the "short-time correlation." One frequency unit is equal to 64 cps.

two tones are close together, the apparent correlation observed during any small time interval will at times be greater and other times smaller than this average value by an amount that fluctuates in time with a frequency equal to the difference between the frequencies of the components. If we then ask "what is the maximum curvature obtained at the peak of one of these beats?", we discover by evaluation of $i_1 i_2^*$ that the following term must be added within the brackets of Eq. (21):

$$\frac{-4A}{(1+\delta^2)^3} \left[(2 + 3\delta^4 + \delta^6) \cos \delta t + \delta (1 - \delta^2) \sin \delta t \right] \quad (21a)$$

If one picks that instant in the beat cycle for which (21a) has the greatest negative value, one finds that the lowest possible masking threshold varies as shown by the broken line in Fig. 10. This curve may be expected to have some validity provided the beats are sufficiently slow that the short-time correlation is well defined. However, as the frequency separation of the second tone is increased to about 2 units (i.e. to about 130 cps), each beat period will include only about 10 cycles of the basic tones, of which number only about one-quarter, or roughly 2 cycles, will provide high correlation. For this beat frequency and higher, it is apparent that the notion of "peak correlation" becomes increasingly meaningless and the average correlation becomes the only measure having significance. For a frequency of $\delta = 0.576$, the average curvature vanishes for $A = 0$ because at that point the correlation curve for the masking signal alone has no curvature. For still smaller frequency separations the average correlation for the masking signal alone already has a strong downward curvature (since this region is near the peak of the response curve) and the addition of a second signal can only increase this downward curvature. However, if one takes into account the temporal fluctuation represented by Eq. 21a, it may be shown that during certain parts of the beat (i.e. for some values of δt in Eq. 21a), the curvature will pass through zero and reverse its sign.

However the interpretation of the correlation model becomes rather dubious here; a more satisfactory explanation for beats of mistuned unisons is given in Section VIII of this thesis.

E. Implications for Hearing.

It is believed that the phase principle has important implications in auditory theory: It ascribes functions to certain of the structural details of the ear and offers explanations for certain rather puzzling facts. If the ear did use a phase principle, the high damping of the cochlear partition would become a virtue and a necessity rather than a conundrum.

Furthermore, a mechanism utilizing a phase principle would permit effective use of the available neural elements and their well-known mechanisms of synaptic facilitation and inhibition. The measurement of phase may be accomplished by measuring the time intervals between related zero crossing of two waves. Since each time interval is independent of the amplitudes of the wave, its accurate measurement is possible over wide dynamic ranges of signal intensity provided only that the zero crossing can be detected accurately. Regardless of the type of analysis occurring in the ear, the results must ultimately be delivered to the central nervous system in the form of "all or none" impulses. Such neural messages may convey temporal date, such as the instants of zero crossings, with good accuracy; they are poorly suited for conveying amplitudes and intensities. The classical place theory of maximal stimulation has been difficult to reconcile with the wide dynamic range of the ear for, at the point of maximal stimulation, the vibration amplitude would not be changing with place. But it is precisely in this region of resonance (where the round top of the amplitude envelope makes exact determination of the position of the maximum so difficult) that the phase of the vibration is generally changing most rapidly. Thus, a phase principle could account for the ability of the ear

to discriminate slight changes in frequency consistently over a wide range of stimulus intensities.

Finally, in any mechanism utilizing a phase principle the masking effect of one tone on another is a function of the ratio of their amplitudes. This logarithmic-like property is precisely that needed to account for the masking effects produced in the ear by two tones of nearly the same frequency. Furthermore, as will be shown subsequently, when the tones are harmonically related the phase of the zero crossings is modulated in such a manner as to account for the observed beats and other effects such as the dependence of the quality of the sound upon the phase of the harmonic. (In this connection, our theory predicts that this phase modification should produce an apparent change in the pitch of the fundamental as the relative phase of the second harmonic is varied).

In the light of these considerations, we next inquire whether the known structure of the cochlea does exhibit the elements necessary to utilize a phase principle.

VI. A PHASE-MECHANISM FOR COCHLEAR FREQUENCY ANALYSIS

Only a brief and, consequently, incomplete summary of the mechanical and structural properties of the inner ear will be given here. For additional details the reader may consult references W-1 and S-2.

A. Structural Properties of the Cochlea

The inner ear or cochlea, in which the first analysis of sound occurs, consists of a spiral-shaped fluid-filled tube into one end of which the sound vibrations are transmitted from the ear drum via the mechanical linkages of the middle ear, parts of which may be discerned in Fig. 11. This bone-imbedded tube is separated along most of its length into two

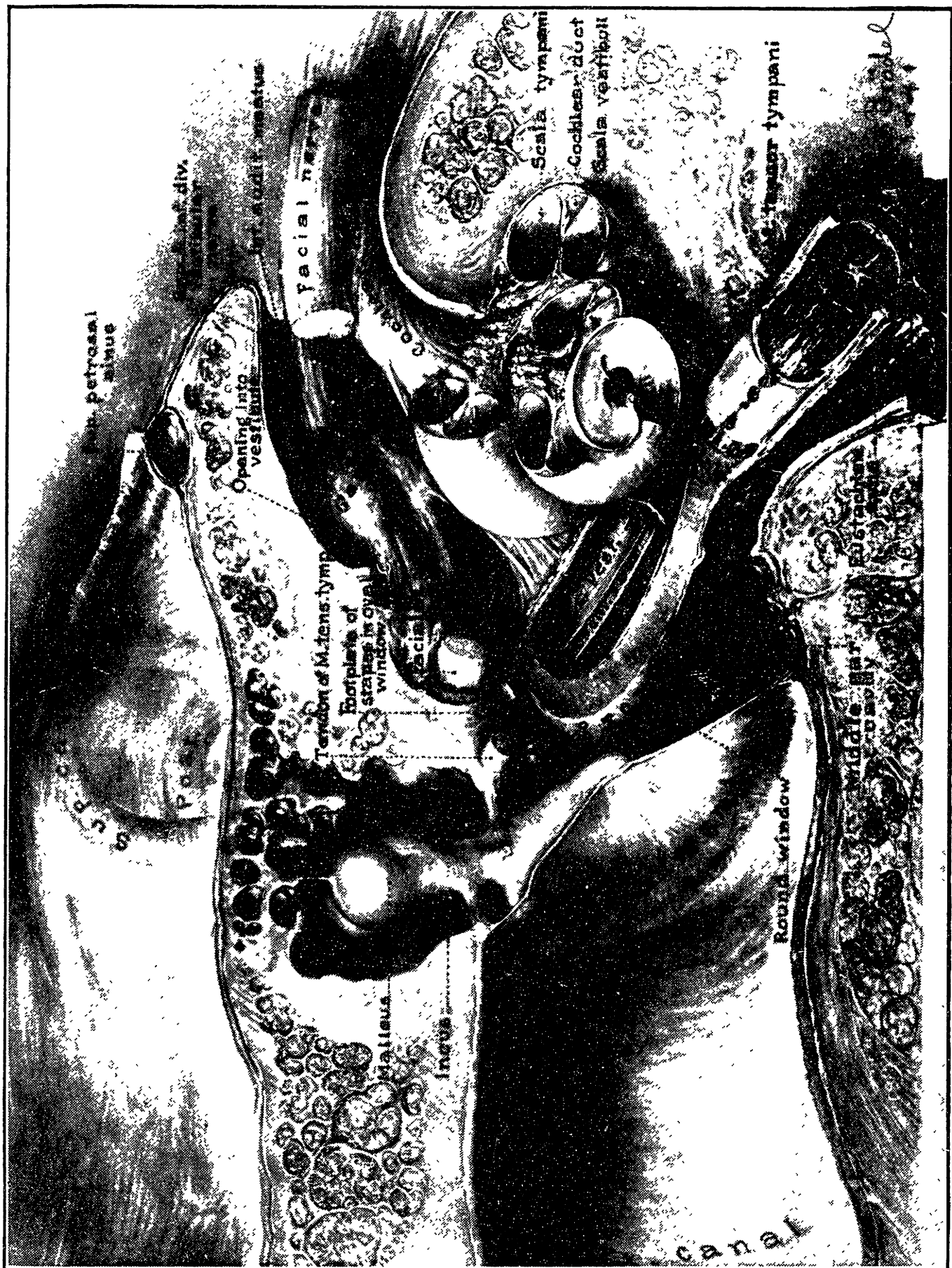


Figure 11. The anatomy of the human ear. (a frontispiece drawing by Max Brödel from reference W-1.)

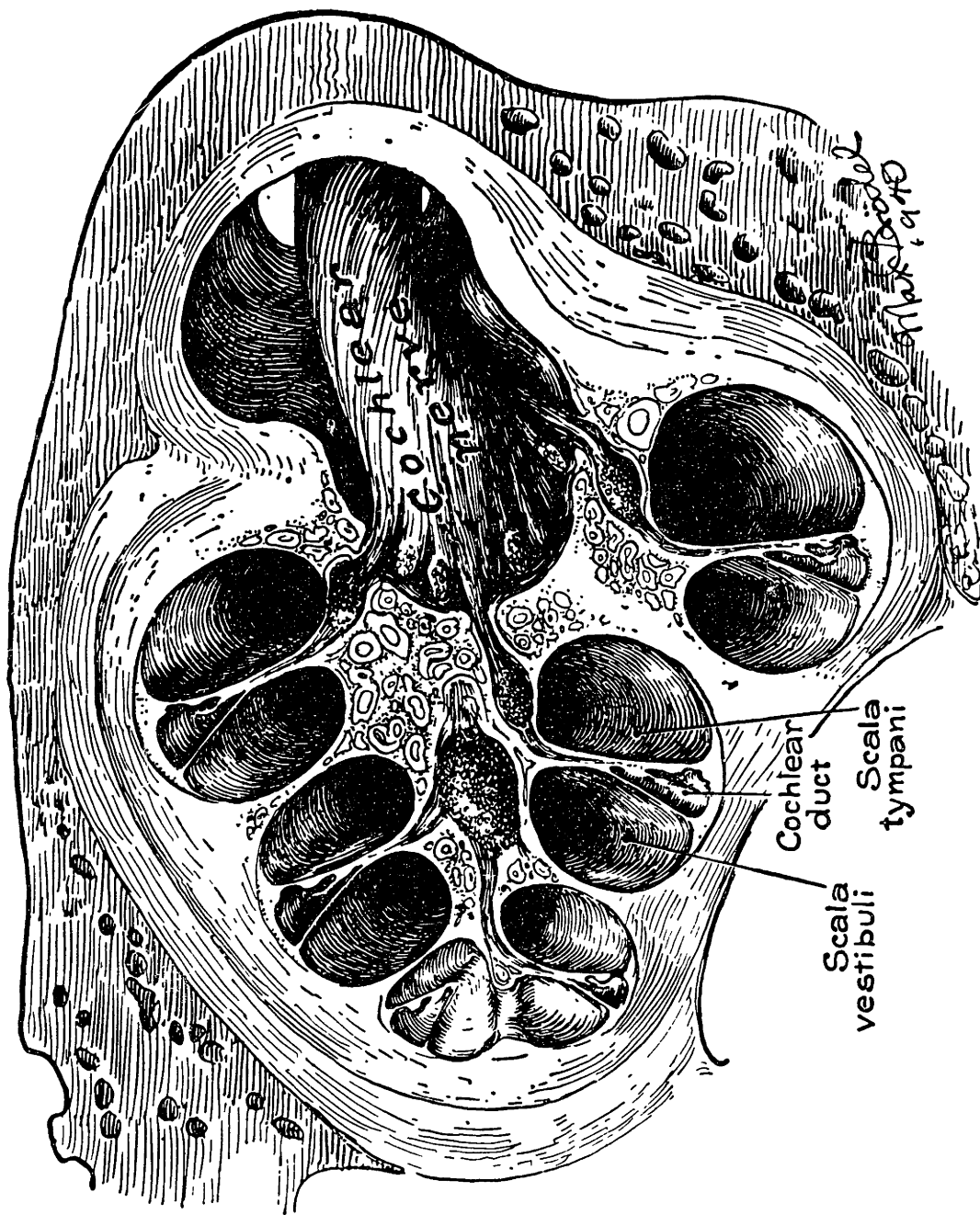


Figure 12. A midmodiolar section of the cochlea. (Drawing by Max Brödel, from the 1940 Year Book of the Eye, Ear, Nose and Throat.)

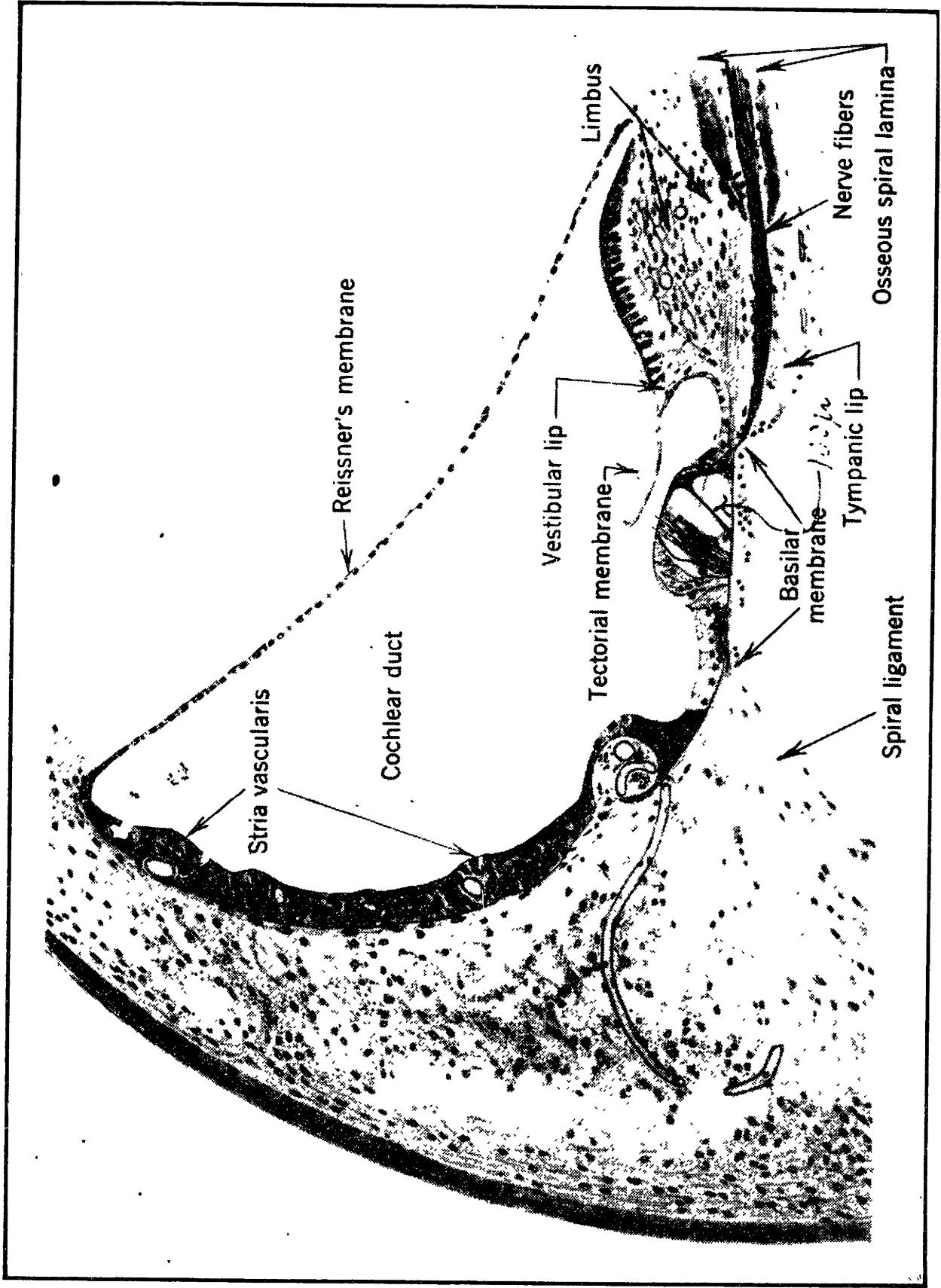


Figure 13. The cochlear duct, in cross section. For general orientation, see Figure 12. (From Held, in *Bethe's Handbuch der normalen und pathologischen Physiologie*, Vol. 11, 1926, J. Springer, Berlin.)

parts by a partition which is actually a fluid-filled cut bounded on one side by Reissner's membrane and on the other side by the basilar membrane as shown in Figs. 12 and 13.

The elasticity and width of the basilar membrane increase markedly as one progresses along the tube. As a consequence, different portions of the membrane exhibit a hydrodynamical resonance effect at different frequencies, the resonant frequency starting at around 15,000 cps at the basal or input and decreasing in a roughly exponential manner with distance to about 500 cps at the apical or far end. The resonance effect along the basilar membrane is not sharply localized and the problem of reconciling this fact with the acute perception of slight changes in pitch displayed by the human listener is one of the principle tasks of this dissertation.

The dynamics of the vibration of this cochlear partition may be visualized by considering its transverse displacement at different points along its entire 35-mm length when excited by a sinusoidal tone of 1000 cps. If the distance from the input end is represented by x , the sinusoidal displacement of the basilar membrane at the point x may be represented by the complex quantity $Y(x)$, the absolute magnitude $|Y(x)|$ being the amplitude of the vibration at that point and the argument of $Y(x)$ being the phase of the vibration relative to some reference phase. An approximate analytical solution for this vibration was made several years ago (H-3), and the mathematical details are reported here in the appendices. The principle results of these computations are summarized in Figs. 14 and 15.

Figure 14 shows the variation of the amplitude and phase of this vibration along the entire length of the membrane when it is excited by a 1000-cps tone. The resonance condition for this frequency is most marked in the region of $x \approx 20$ mm. The traveling-wave character of these vibrations is

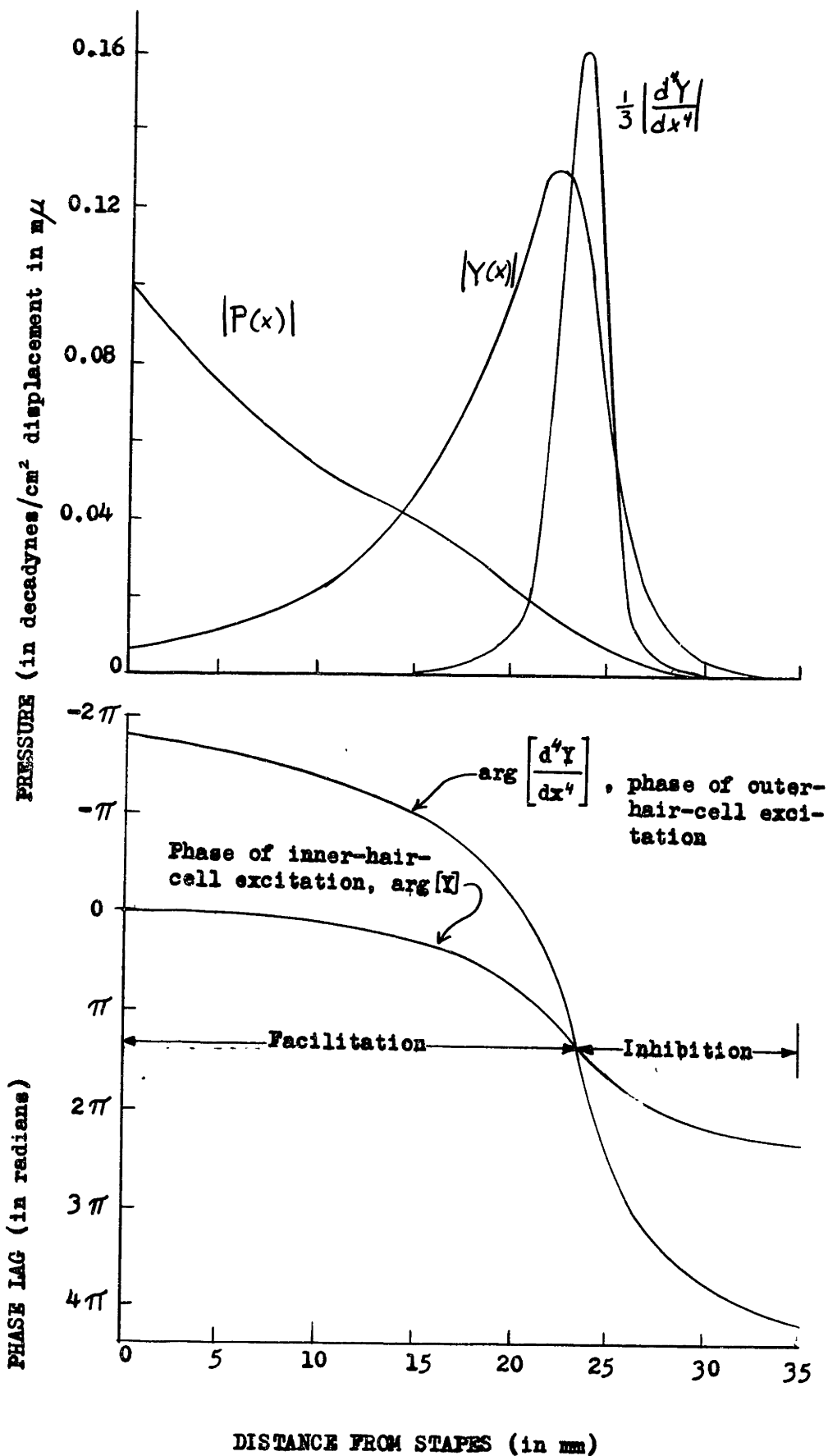


Fig. 14. Phase and amplitude variation of the transverse pressure, displacement, and fourth space derivative of the displacement for a 1000-cps tone.

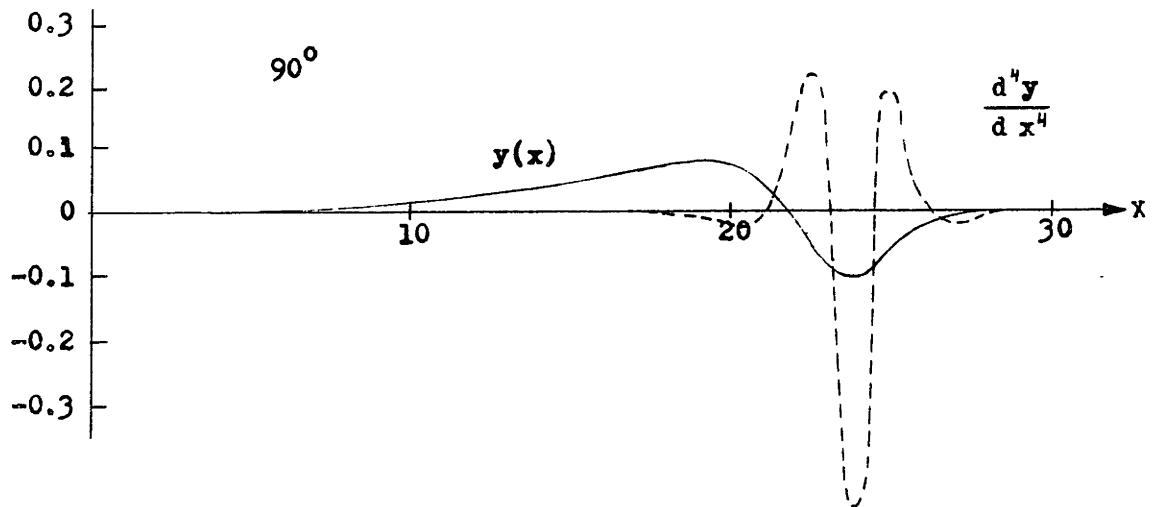
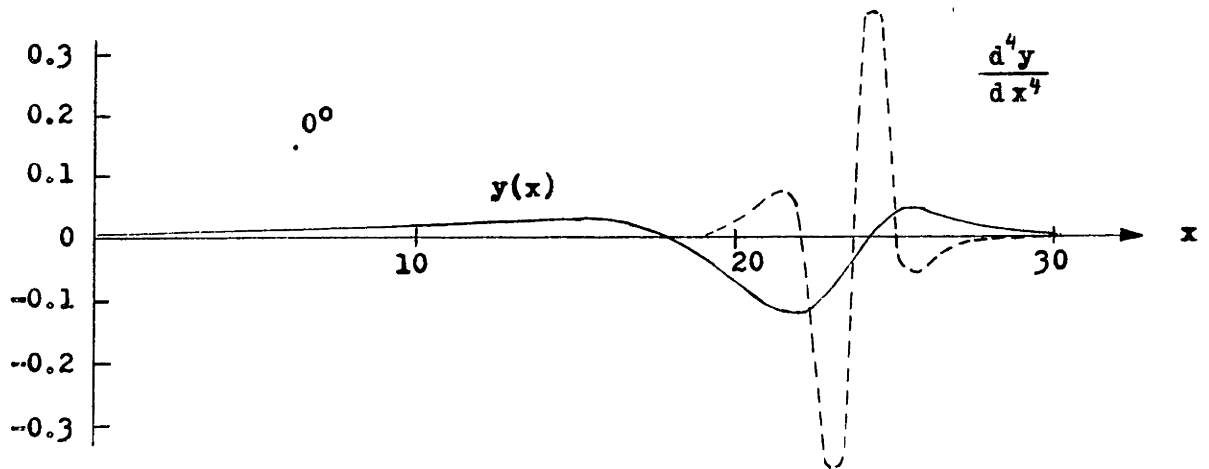


Fig. 15. Relative instantaneous displacement (solid line) and force (dashed line) waves along the tectorial membrane as functions of the distance from the stapes. The excitation is a 1000-cps tone.

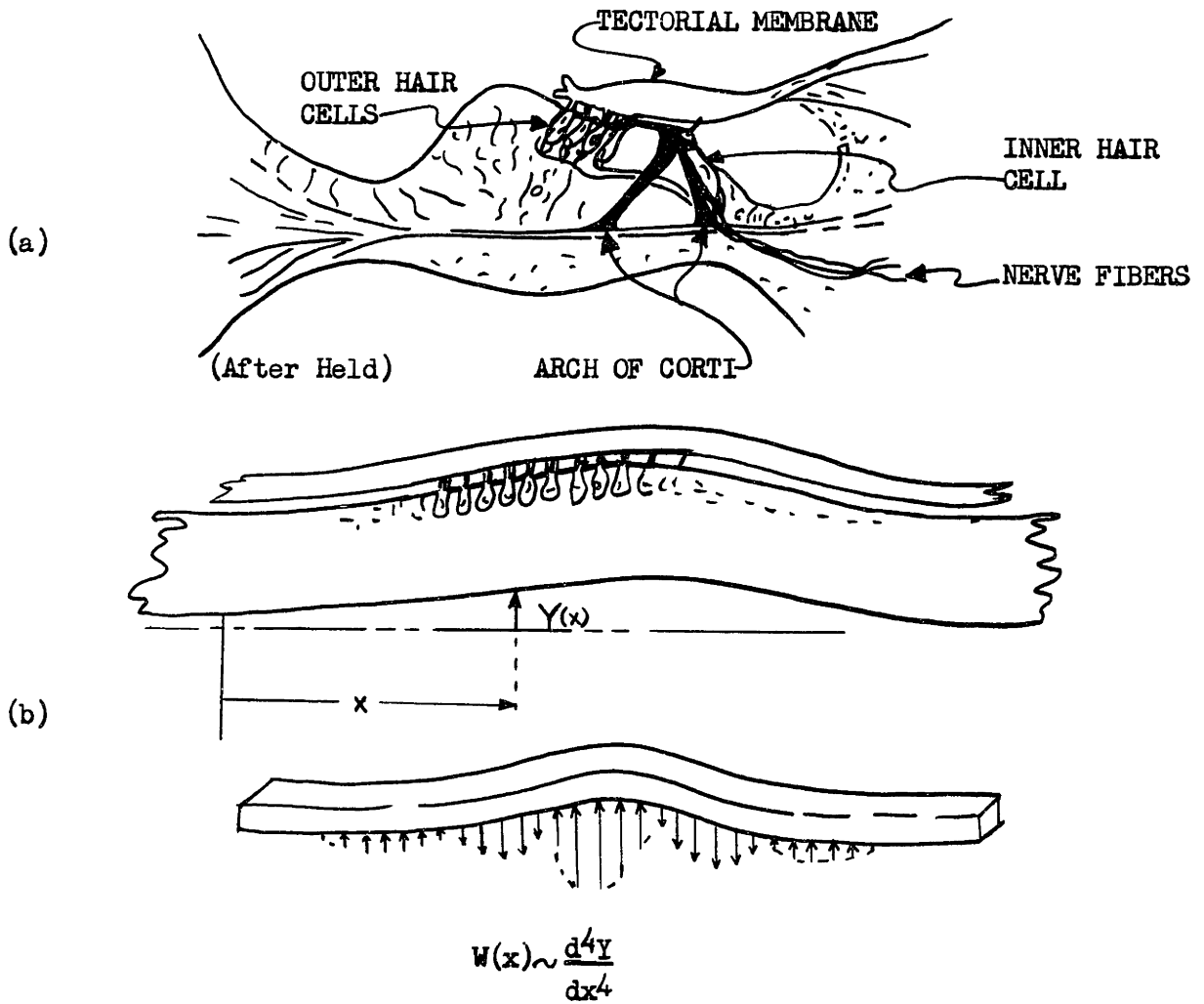


Fig. 16(A) Cross-section of the organ of Corti. (B) Side view showing how the tectorial membrane acts as a beam to exert upon the outer hair cells forces proportional to the fourth place-derivative of the displacement of the basilar membrane.

revealed by replotting in Fig. 15 the data of Fig. 14 for two instants of time separated by one-quarter of a vibration period. The solid curve represents the instantaneous displacement of the cochlear partition. It shows that the wave travels at first with a very high velocity until it reaches the region of resonance where it moves more slowly and is rapidly dissipated.

In Fig. 16(a) is sketched a cross-section of the cochlear partition which shows some of the finer anatomical detail needed for our theory. The basilar membrane is given additional mass by some rather large fatty cells, fastened to its top. Near the axial edge of this membrane, there is also a truss-like arrangement of rods called the "arch of Corti". Arrayed along the innermost edge of this arch is a single row of hair cells which, in some way provide the means for converting the mechanical stimulation of their cilia into excitation of the nerve fibers terminating around their base. Several other closely packed rows of hair cells are also found on the outward side of the arch of Corti. It is believed that the cilia of these hair cells extend upward and are imbedded in another membrane that also runs along most of the length of the cochlear duct. This is the tectorial membrane and it plays a most important role in our theory.

The tectorial membrane has been reported by Bekesy (S-2, p. 1095) to exhibit different elasticities in different directions. "It seems to exhibit stiffness in the longitudinal direction as noted by the fact that upon tearing it away from its attachment a considerable segment comes off, whereas the other membranes may be removed in quite small sections..... This membrane is noteworthy for the ease with which it may be moved in a direction perpendicular to the basilar membrane. Together with its supporting structure it represents essentially a flat, delicate, very thin-walled tube, filled with liquid, which can be rotated about the edge that is attached to the bony shelf of the cochlear partition."

Now, H. de Vries has studied the microphonic voltages arising from stimulating of the lateral line organs of fishes (V-1). These organs also contain hair cells and de Vries' conclusions concerning the excitation of these hair cells should also be applicable to those of the cochlea to which they are ontogenetically related. de Vries has suggested that these cells resemble a nerve fiber in that they build up an internal negative charge through migration of sodium ions into the cell. The cell surface around the hairs is believed to be particularly sensitive to mechanical disturbance which allows the cell to discharge its ions with the results that "the voltages are related to the tension in the hairs and that this tension is accompanied by a negative voltage at the upper side of the organ." Using this model, de Vries was able to account for the double-frequency microphonic observed when the line organ was stimulated by a sinusoidal water current. He also remarks that "when the tectorial membrane moves relative to the organ of Corti, the only possibility is a gliding motion (because of the incompressibility of the endolymph). When the cochlear partition bends upward (i.e. toward Reissner's membrane), the tectorial membrane will glide over the organ of Corti to the left. This means that the outer cells will be stretched. Since they are already in an inclined (inward) position they will be stretched only once every period."

De Vries continues, "Undoubtedly there is some meaning in the fact that the inner hair cells are inclined in the opposite direction. They will be compressed when the outer hairs are stretched and vice versa. Up to now no special attention has been paid to this, and it would take us too far to discuss here all of the possible implications of this situation."

These remarks of de Vries are very suggestive. I had the privilege of discussing some of these questions with him while visiting Holland in October, 1952. Based on his valid and helpful criticism of the beam hypo-

thesis, which Licklider and I had previously published (H-2), I wish to present now a description of a plausible mechanical action within the cochlea that appears to agree with the peculiar structure of the organ of Corti and to account for the necessity of having two kinds of hair cells located in the particular positions that they occupy.

B. The Mechanical Action of the Cochlea

We consider first the mechanical action resulting from uniform displacements of the basilar membrane such as might occur for very low-frequency sounds. The mechanical elements have been schematized as shown in Fig. 17 (a). The organ of Corti, with its inner and outer hair cells, is assumed to pivot about the point "A" near the place where the basilar membrane fastens onto the bony shell. The cilia of the inner and outer hair cells are inclined in opposite directions and enter the tectorial membrane at an acute angle so that any sliding displacement of the tectorial membrane will immediately be effective in stretching their cilia. If the cilia entered at right angles, small sliding displacements would produce negligible stretching and mostly bending which would account, as de Vries has shown, for a double-frequency microphonic.

The inner edge of the tectorial membrane is assumed to be anchored on the top of the spiral limbus which provides an elastic support that will permit the tectorial membrane to move in a radial direction.

The reason de Vries states that the tectorial membrane can only move in a sliding, tangential manner with respect to the surface of the organ of Corti is that the thin layer of gelatinous fluid trapped in between is incompressible and sufficiently viscous at audio frequencies to fix quite rigidly the spacing between the two surfaces. But it should also be noted that the gelatinous mass trapped in the internal spiral sulcus is likewise

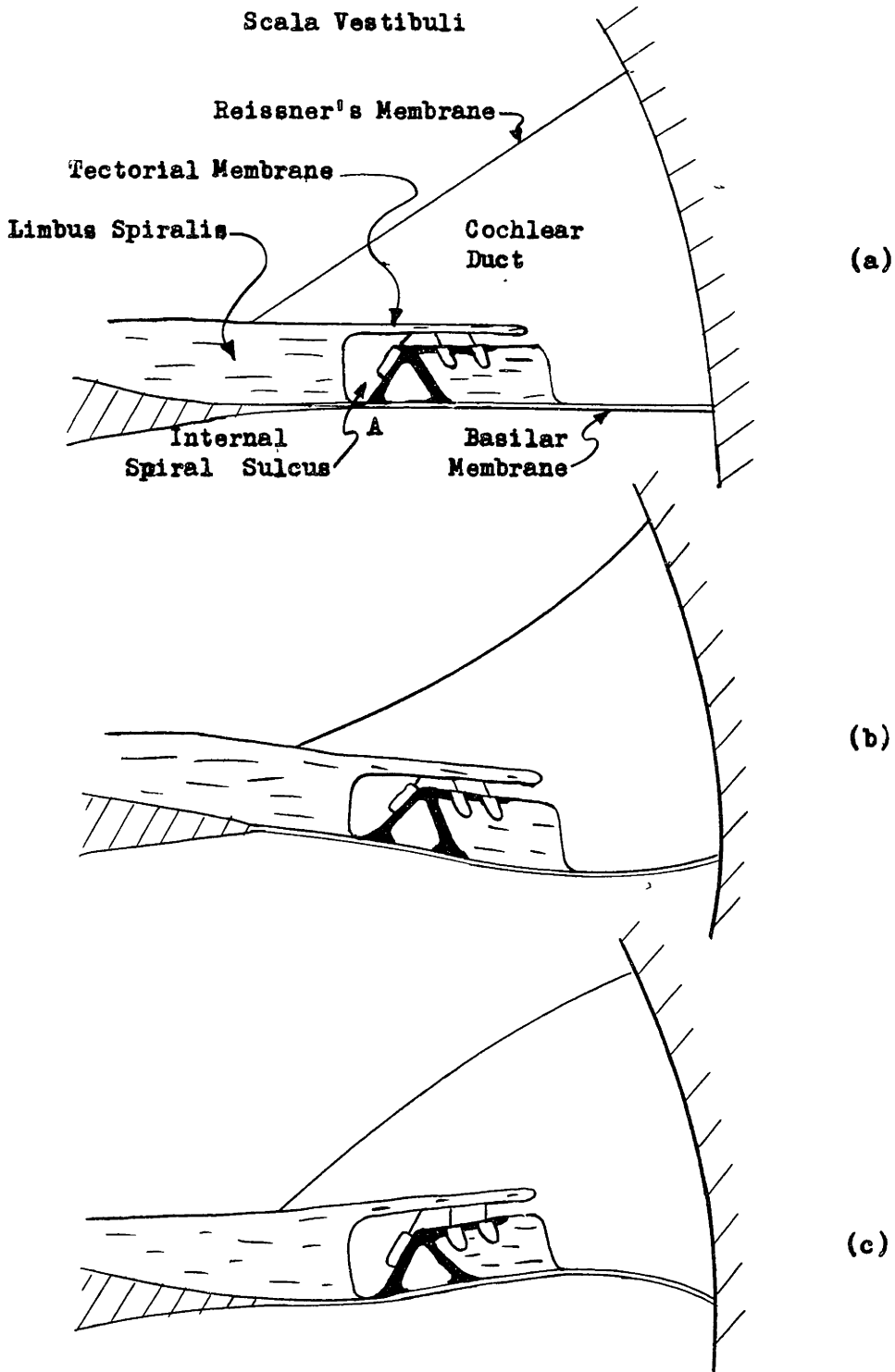


Fig. 17. Schema for the mechanical action of the cochlea.

incompressible. As a result, displacements of the organ of Corti that would otherwise tend to change cross-sectional area of internal spiral sulcus, will cause corresponding distortions of the limbus and bordering portions of the tectorial membrane so as to keep the volume constant. Thus, a downward displacement of the basilar membrane, as in Fig. 16(b), will cause the inner hair cell to move outward and create a drop in pressure within the fluid of the sulcus at this point. This pressure differential causes the limbus and inner edge of the tectorial membrane also to move outward and downward. Similarly, an upward displacement of the basilar membrane, with its accompanying inward displacement of the inner hair cell, pushes the limbus and inner edge of the tectorial membrane inward and upward, as indicated in Fig. 16(c).

The mechanical action just described differs markedly in its excitation of the hair cells described by de Vries. In the case of the downward displacement, illustrated in Fig. 16(b), the outward forces exerted upon the limbus by the fluids of the sulcus, and perhaps, even by the deformation of Reissner's membrane, cause a radially outward motion of the tectorial membrane. If this outward motion is nearly equal to the outward movement of outer hair cells, there will be little excitation of these cells. This is assumed to be the case. Similarly, it is assumed that an upward displacement will cause a radially inward movement of the tectorial membrane that is just about equal to the inward motion of the outer hair cells.

Thus, for low-frequency stimulation, where essentially the whole length of the basilar membrane moves in phase, there will be only slight excitation of the outer hair cells.

Where, then, do the microphonics observed for low-frequency excitation come from? If not from the outer hair cells, they must be caused by excitation of the inner hair cells. Now, we have just described how an upward movement of the basilar membrane could lead to an upward movement

of the inner edge of the tectorial membrane, the outer edge of which is rigidly fastened via the entrapped fluid surrounding the cilia of the outer hair cells to the organ of Corti. From Fig. 16(c), it is apparent that the tectorial membrane can act much like a lever to stretch the cilia of the inner hair cell. Similarly, a downward movement tends to compress the cilia of the inner hair cell. Thus, uniform displacements of the basilar membrane can produce a directly proportional tension upon the cilia of the inner hair cell. Positive tension, corresponding to outward movement of the stapes, should produce, according to de Vries' model of the hair cell, a negative potential on the vestibular side of the cochlear duct. This agrees with the facts.

We next consider the function of the outer hair cells. The mechanical action that is visualized here produces little force upon the cilia of these cells at low frequencies where the entire membrane is displaced in phase. However, as the driving frequency is increased, the time required for the displacement wave to propagate down the cochlea becomes large in comparison to the period of the vibration until, as shown in Figs. 14 and 15 for a driving frequency of 1000 cps, nearly two undulations may exist along the basilar membrane at any instant. It is apparent that the upward transverse displacement of the basilar membrane produces, through rotation of the organ of Corti about the pivot point "A", shown in Fig. 16(a), a proportional inward displacement (to the left) of the outer hair cells.

When the displacement was everywhere uniform, the tectorial membrane was also displaced inward. However, for higher frequency oscillations, such as shown in Fig. 15, one section of the tectorial membrane is required to move radially inward at the same time that an adjoining section is moving outward. This causes the tectorial membrane to bend in the plane for which it has the greatest stiffness, in a manner rather similar to the normal deformation of an

"I" beam. The spiral limbus also appears to provide further stiffening, for Bekesy has reported (S-2, p. 1097) that "the cochlear partition possesses near its bony edge considerable rigidity in the longitudinal direction, (When under pressure of a probe) a considerable section of this zone, up to 1/8 of a turn, is displaced simultaneously."

It is hypothesized, therefore, that the cilia of the outer hair cells exert upon the tectorial membrane forces that tend to make it follow at each point the radial displacement of the outer hair cells. If the force acting in a little region Δx at the point x is represented by $W(x) \cdot \Delta x$, we have from the well-known relation for the bending of a beam that

$$W(x) \sim d^4 Y(x) / dx^4 \quad (22)$$

Thus, our mechanism indicates that the outer hair cells will be excited in proportion to the curvature of the curvature of the displacement.

We have thus described a relatively simple mechanism whereby mechanical waves traveling down the cochlear partition can create two entirely different waves of neural excitation--the first wave produced by forces, acting on the inner hair cells, which are proportional to $Y(x)$, and the second wave produced by forces, acting on the outer hair cells, which are proportional to $d^4 Y / dx^4$. This simple mechanism offers several interesting features.

(1) Sharpness of response--It is noted from Figs. 15 and 16 that the region of intense excitation of the outer hair cells is far more localized than has been previously believed possible in view of the broad nature of the displacement.

(2) Possibility of a phase-sensitive mechanism--The existence of two distinct waves of neural stimulation now permits us to utilize a phase principle for sharpening the frequency analysis. As indicated in Fig. 14,

there is only one point along the entire membrane where these two waves will be exactly in phase. And this point is operationally definable, using accepted neural mechanisms, for damped waves, as well as for growing waves, provided only that the damping of the forcing waves is somewhat less than that of the cochlear partition. Furthermore, to the extent that the excitation is well above threshold, this point of phase coincidence is essentially independent of the amplitude of the excitation, and accurate analysis of damped waves is possible.

(3) Correct polarities of stimulation--Examination of Fig. (16) indicates that the tension on the inner hair cells is proportional to the displacement $y(x)$ in the upward direction. Assuming that an inward motion relative to the organ of Corti corresponds to a positive deflection of the tectorial membrane regarded as a thin beam, we see that this positive deflection is proportional to $-Y(x)$, and that the tension (i.e. $-W(x)$), which must be applied along it to produce this deformation, will be proportional to $-d^4(-Y)/dx^4$ or, simply, d^4Y/dx^4 . Thus the polarities of the excitation waves represented in Figs. 14 and 15 are correct.

(4) Relation between microphonic and action potentials--A slight movement of the tectorial membrane relative to motion of the organ of Corti at low frequencies may actually be beneficial. For small, uniform, upward deflection of the basilar membrane, this will cause the cilia of the outer hair cells to be bent toward the vertical direction and no excitation will occur. For downward deflections, a small microphonic voltage may be produced which would tend to neutralize the positive microphonic potential created by the inner hair cells. This voltage could account for the observed distortion concerning which Wever says (W-1, p. 330), "For low (frequency) tones, the responses show appreciable distortion even at moderate levels

of stimulation--levels of the order of the human threshold." Wever continues to point out that "complications in the action of the low-tones appear also in the neural responses. When the imposed stimuli are strong these responses may contain more than one volley of impulses per cycle. In the simplest case there are two discharges in a cycle, one larger than the other, one near a peak and the other near a trough." Wever assigns the double firings to the generation of a second-harmonic distortion product, but as these occur at low amplitudes, the alternate excitation of the outer and inner hair cells by the action just described appears to be a more probable explanation.

At higher frequencies, this effect could be beneficial and possibly result in a sharpening of the frequency selectivity of the ear. It will be noted from Fig. 14 that at a point about $1\frac{3}{4}$ millimeters either side of the resonant (or in-phase) point, the phase of $Y(x)$ and of $W(x)$ differ by 180 degrees. Thus, the additional force due to $Y(x)$ will cause the skirts of the outer-hair cell selectivity curve to fall off more rapidly than that indicated by $W(x)$ alone. This effect is mentioned here only as being one possibility; whether or not such an effect is present is immaterial to the theory.

(5) The increase of sensation threshold at low frequencies--The place of maximum displacement indicated in Fig. 15 is located about 2.6 millimeters closer to the stapes than the place of maximum stimulation of the outer hair cells. But, as Békésy has pointed out (B-1, Fig. 13), the localization of "hearing" as determined from pathological observations is also displaced apicalward--about 3.5 millimeters from the point of maximum vibration. This fact, together with the greater selectivity exhibited by the frequency response curves of the outer hair cells, leads us to

assign to the outer hair cells the basic neural excitatory function. On the other hand, the inner hair cells shall be assumed to create neural waves that have primarily an inhibitory function within the cochlear nucleus.

According to this hypothesis, as the frequency becomes very small, the inhibitory neural waves generated by the inner hair cells dominate the increasingly weaker neural waves created by the outer hair cells. This would account in part for the fact that the sensation threshold of the ear seems to vary as the cube of the frequency at very low frequencies despite the fact that 60-cps cochlear microphonics and action potentials may be measured at intensities which are 30 decibels below the threshold of hearing (e.g. see W-1, Figs. 102 and 103). At these low frequencies, there would be only weak excitation of the outer hair cells and, hence, little neural response at the higher nerve centers because of the inhibiting effect of the strong excitation of the inner hair cells.

Fletcher, in a recent paper (F-1), attributes the drop in response with decreasing frequency to the imperfect transformer action of the middle ear. Included in his analysis are the simple assumptions that the neural excitation at a given point will be proportional to the energy of the vibration (i.e. to the velocity squared) and also to the frequency (below 300 cps). These two together account for a variation proportional to the cube of the frequency. Fletcher, however, extrapolates Békésy's experimental data into the frequency range below 100 cps. According to his Fig. 1, the ratio of the pressure difference, P_O , across the basilar membrane near the stapes, to the external sound pressure, P_{ED} , acting on the eardrum continues to fall off indefinitely with decreasing frequency, for both an infinite inner-ear impedance and for the "actual inner-ear impedance". This obviously cannot be so when the ear impedance is infinite for, as the frequency is indefinitely decreased, the ratio of these two pressures would not become

zero but would approach a constant value. The correct curve for infinite impedance at the stapes should resemble that of a low-pass amplifier having simple shunt high-frequency peaking rather than as drawn by Fletcher.

For the actual ear impedance, Fletcher assumes that P_o/P_{ED} continues to fall off at a rate of 30 db per decade decrease in frequency. Here, the indefinite decrease in P_o with decrease in frequency is reasonable if one assumes that the *perilymph* is a viscous fluid that can flow through the helicotrema. However, if this is so the curve should fall off only at the rate of 20 db per decade. In fact, careful examination of Békésy's data (Fig. 8 of ref. B-2) suggests that at the lowest frequency point (120 cps) the slope of the curve is indeed beginning to decrease.

Thus, although Fletcher assumes a transmission through the middle ear which falls off too rapidly with decreasing frequency, he still must make the additional assumptions of the excitation being proportional to the velocity squared and to the frequency. The reduction in the excitation of the outer hair cells at low frequencies indicated by our theory may provide an explanation for Fletcher's additional assumptions.

The fact that at very low frequencies the threshold for the appearance of cochlear microphonics and action potentials seems to fall off less rapidly than does the auditory threshold suggests that an experimental comparison of these various thresholds should settle the question as to whether this is caused by poor transmission through the middle ear as proposed by Fletcher, or whether the sound is transmitted to the cochlea where it stimulates the inner hair cells but, then, is ineffective in stimulating the outer hair cells, as suggested by our theory.

However, much more conclusive evidence that the inner and outer hair cells provide inhibitory and excitatory neural stimuli, respectively, is

found in the important experiments of Galambos and Davis. These experimental results will now be examined in some detail, together with other experimental data in support of the mechanism described here.

VII. DISCUSSION OF EXPERIMENTAL EVIDENCE

Fortunately, there exists a large body of experimental data concerning the phenomena of hearing against which we may test the mechanism postulated in the previous section.

A. Cochlear Excitation Pattern Inferred from Masking Data.

The first important fact is that the excitation pattern of the outer hair cells (i.e., $W(x)$) agrees quite closely with estimates of this excitation as inferred from psychophysical data on the masking of one tone by another. Figure 18 shows the excitation function $W(x)$ of Fig. 14 replotted on a logarithmic scale and compared with the stimulation pattern inferred by Fletcher (traced from S-1, Fig. 93, 55-db curve).

B. Localization of Source of Microphonics.

An even more important property of this mechanism is that the phase of the excitation of the outer hair cells changes much more rapidly with respect to place along the basilar membrane than does the phase of the excitation of the inner hair cells.

Because of this very rapid variation of phase with respect to position, as illustrated in Figs. 14 and 15, the potential fields established by the excitation of the outer hair cells must be extremely localized. At any instant, a strong positive excitation of the outer hair cells at one place will be accompanied by a strong negative excitation at adjoining places separated by only a millimeter or two. Hence, the cochlear microphonic as observed from more remote points must arise substantially from the inner hair cells. Experimental data have recently been published that seem to

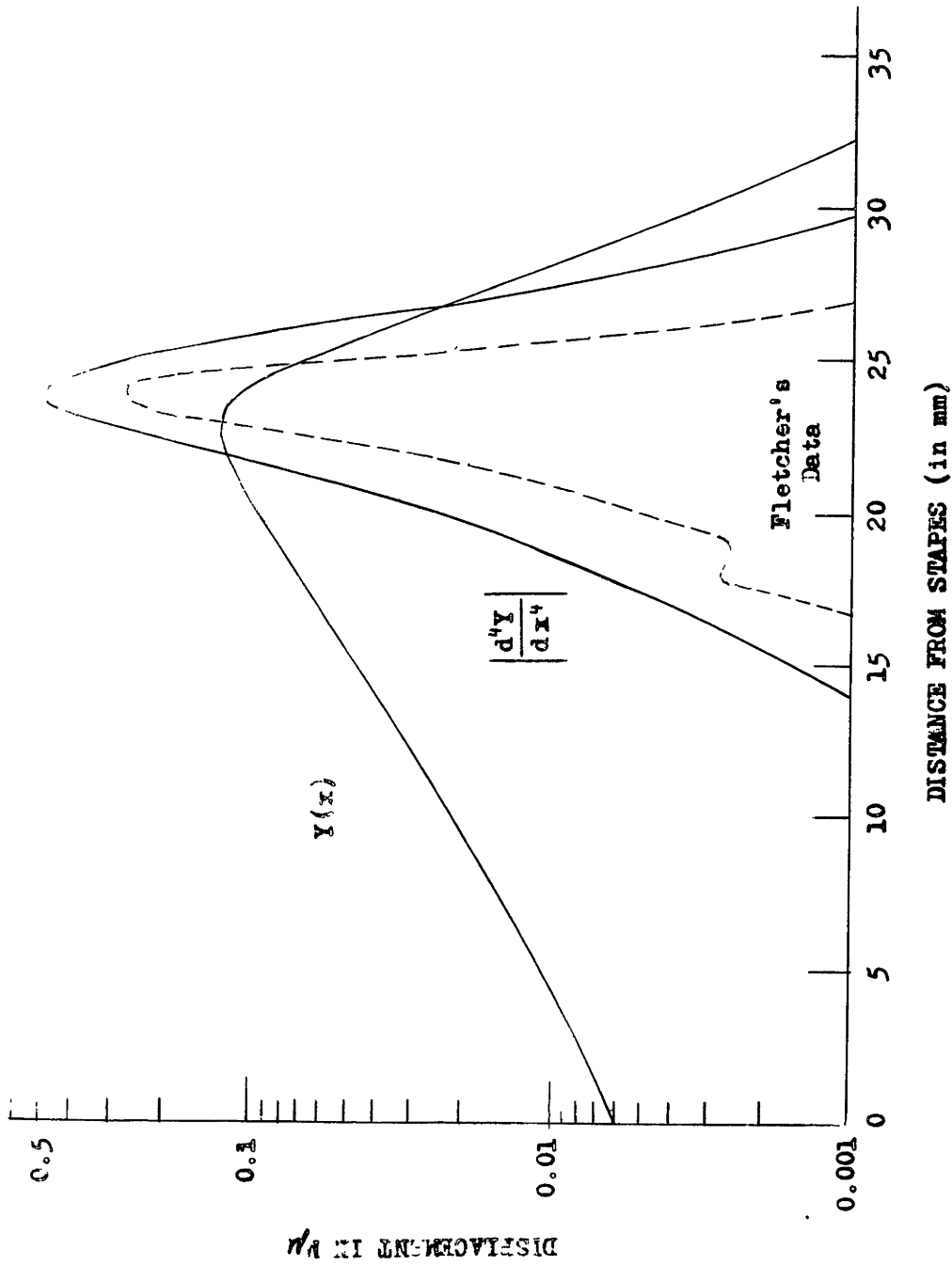


Fig. 18. Magnitude of the calculated force wave d^4y/dx^4 along the tectorial membrane as compared with the stimulation pattern on the basilar membrane inferred from data on masking by H. Fletcher. (Both are for a 1000-cps tone.)

confirm this theoretical expectation.

Tasaki, Davis, and Legoux have studied the cochlear microphonics in the guinea pig with a pair of differential electrodes located at each of several different positions on opposite sides of the cochlear duct (T-1). For a given location along the basilar membrane, the curve showing the variation of microphonic voltage with frequency showed an abrupt change in slope at a frequency approximately equal to the resonant frequency usually associated with that location. The phase of the microphonic at this critical frequency was consistently observed to be about 2π radians. But from Fig. 14, it is apparent that the phase of the inner-hair-cell excitation is about 2π radians at the point where the displacement amplitude is beginning to fall off abruptly. Thus, these experimental data are in agreement with our model.

Tasaki et al. also observed that a marked, but localized, change in the microphonics was produced by pressing a thread of hair against the basilar membrane or stria vascularis. Such an irregularity introduced along the cochlear portion will obviously cause a discontinuity in the traveling wave and consequently create a force density $W(x)$ which need not integrate out to zero, as in the case of the more usual force density function. Békésy also observed that when a tiny vibrating electrode was brought in contact with the outer lip of the tectorial membrane, a very large microphonic voltage was obtained (B-3). He further showed that when the electrode was moved to the opposite side of the cochlea, so that the radial motion created by the vibration would be reversed in phase, the phase of the microphonic also reversed. If this microphonic were due to excitation of the inner hair cells through ordinary transverse displacement, the phase would not have been reversed. The concentrated force exerted upon the tectoria by the vibrating electrode adds an impulsive (i.e., delta function) discontinuity to $W(x)$; its integral

over the place of resonance need no longer vanish--and a cochlear microphonic arising from these external hair cells may now be observed at a remote point.

If one assigns to the outer hair cells a quadratic (i.e., double-frequency) component of excitation, it would follow that an appreciable d-c component would appear in the microphonic because the rectified contributions of each of the hair cells would all have the same polarities and would not cancel at a distant point, as in the case of the fundamental component in the microphonic under these circumstances. However, the rapid phase change along the membrane would tend to suppress this second-harmonic component even more effectively than the fundamental. Thus, non-linearity in the outer-hair-cell excitation characteristic will tend to produce a large d-c potential in the microphonic but very little else. This may be the source of the "summation" potentials described by Davis, Fernandez and McAuliffe (D-1).

Is there any direct experimental evidence of the very rapid phase change along the outer hair cell. In the discussion of a so-called "ghost microphonic" (on page 505 of Ref. T-1, Tasaki remarks "The fact that we could (by the differential method) record cochlear microphonics having a phase difference of 90° at two points in the third turn separated by about one millimeter is strong support for our argument."

This is about the value that would be expected from our theory if the electrodes had been placed in a position where they could pick up these outer-hair-cell microphonics. Our notion that the inner and outer hair cells may respond to different aspects of the wave motion in the ear has not been widely publicized, so it is likely that no effort was made to distinguish between the two in most experiments. We can say, then, that phase changes of the order of the predicted by our theory have been observed experimentally (T-1).

G. Response of Second-order Neurons.

Because of the difference between the phase characteristics of the inner- and outer-hair cell excitations, there is only one point along the basilar membrane where the excitations of a given frequency are in phase. This important feature of our theory suggest three assumptions concerning the neural phenomena associated with the ear:

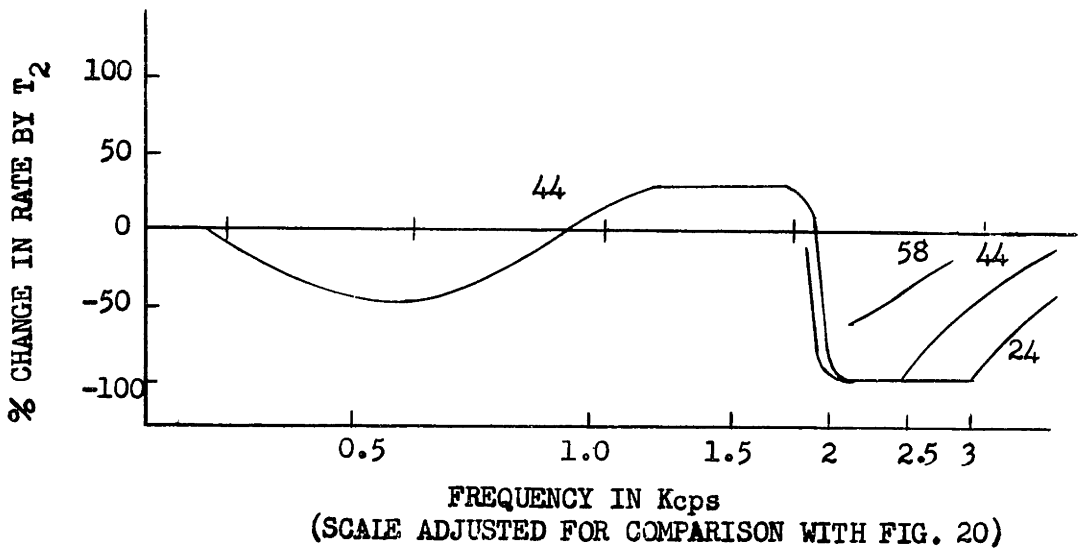
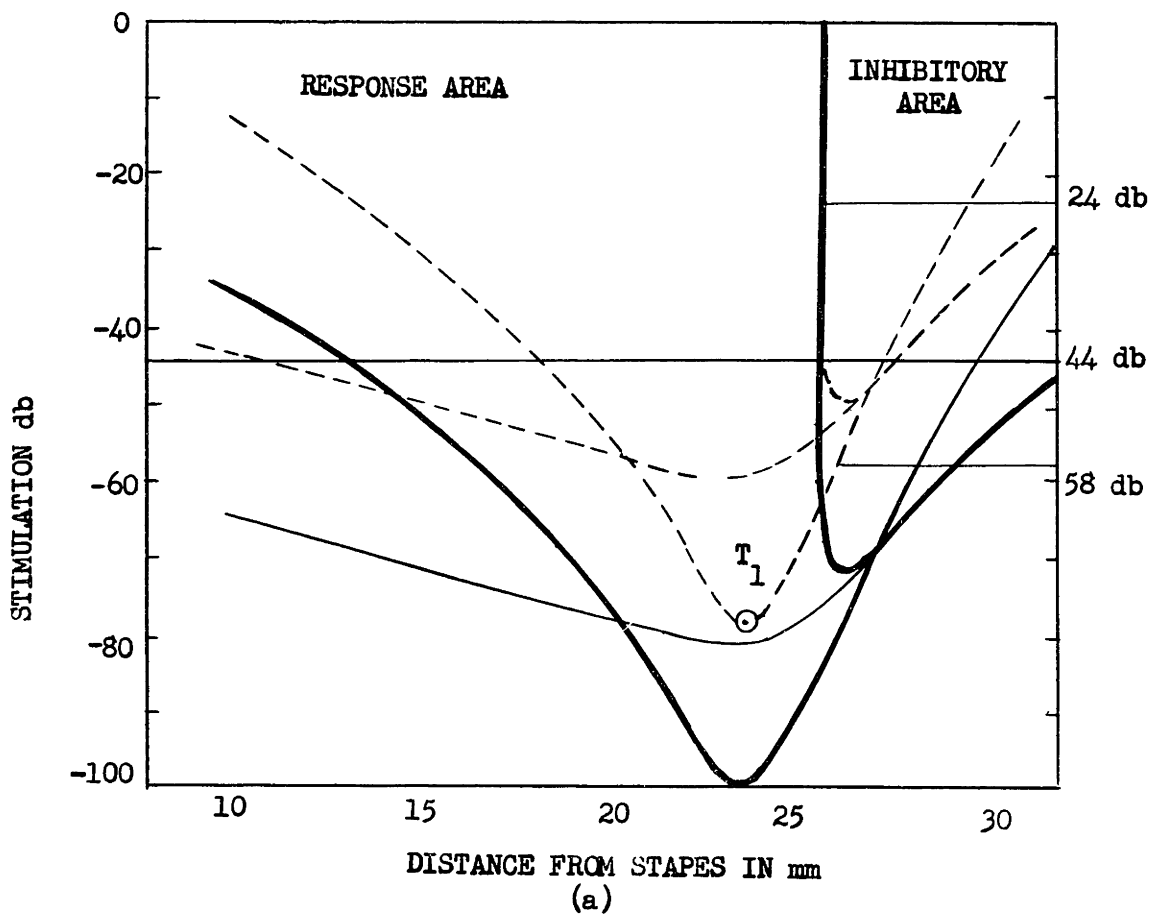
- a) The firing of the nerve fibers arborizing on the inner hair cells temporarily inhibits the response of second-order neurons to certain of the first-order neurons that terminate on the outer hair cells. Thus, for periodic excitation, the second-order neural response corresponding to the region of the basilar membrane extending from the point of resonance (and phase coincidence) toward the apex will be inhibited as shown in Fig. 14. Response will be obtained from the basal portion because here the excitatory outer cells will fire slightly earlier than the inhibitory inner cells.
- b) The stimulation threshold for firing of the inner-hair-cell nerve fibers is 20 db higher than that for the outer-hair-cell fibers.
- c) The nerve fibers arborizing on a given inner hair cell act as inhibitors for those nerve fibers terminating on the group of outer hair cells that are 1.5 mm basalward. This assumption is consistent with the observation that the fibers innervating the outer hair cells issue from the habenula perforata, pass the inner hair cells and hence, after reaching the outer-hair-cell region, run basalward for a considerable distance before arborizing around adjacent hair cells.

The data of Galambos and Davis provide nearly ideal evidence in support of our theory. Their observations were made with microelectrodes inserted into

the eighth nerve of anesthetized cats. The data were originally interpreted as action potentials of the axons of the first-order neurons running from the cochlea to the medulla (G-1). Later, however, evidence was found that the unitary action potentials recorded were those of second-order neurons separated by at least one synapse from the nerve cells which are excited by the inner and outer hair cells (G-2). Hence, the data permit direct testing of assumption (a) stated above.

To compare the theory with the experimental results of Galambos and Davis, the data of Fig. 18 have been replotted in Fig. 19 to show the acoustic stimulus required to stimulate the nerve fibers terminating on either the inner or outer hair cells. The threshold for stimulation of the outer-hair-cell neurons is assumed to be -100 db which, by assumption (b) above, fixes the minimum threshold for the inner hair cells at a value 20 db higher. By assumption (c) above, the entire inner-hair-cell curve is displaced 1.5 mm to the right to bring the corresponding inner- and outer-hair-cell data into line.

Whether or not the inhibiting nerve fiber associated with the inner hair cell will fire before or after the corresponding outer-hair-cell fiber depends not only upon the phase angle between the two excitations but also upon the amplitudes of the stimulation at the two hair cells relative to their respective thresholds. For stimulation by very loud tones, inhibition will occur when the phase of the inner-hair-cell excitation leads that of the outer by even a very small angle and the simple situation illustrated in Fig. 14 applies. But for weak tones, approaching the threshold of the inner-hair-cell fibers, the phase angle of the inner-hair cell excitation may lead the outer by angles up to 90° without inhibiting the response, for the more sensitive outer-hair-cell reaches threshold first, despite its lag in phase.



(b)

Fig. 19. A. Regions of response and inhibition along the organ of Corti for a 1,000-cps tone. B. Estimated effect of the change in response of a second-order neuron to a fixed tone T_1 of -78 db intensity caused by variation of the frequency of a second tone at sound intensity levels of -58, -44, and -24 decibels.

This boundary point, at which simultaneous firing of inner and outer nerve fibers occurs, is easily determined as a function of stimulus intensity using simple geometry and the amplitude and phase data of Fig. 14. The theoretical results indicate "response" and "inhibitory" areas, bounded by the heavy lines in Fig. 19(a). Since 'place' along the basilar membrane is roughly interchangeable with 'frequency,' these theoretical results may be compared with the experimental data obtained by Galambos and Davis, reproduced in Fig. 20. The agreement is satisfactory. The theory even offers an explanation for the cusp-like response area at the bottom end of the inhibition area. This must have appeared to Galambos and Davis as an unaccounted-for irregularity in their experimental data since they extrapolated the inhibitory region downward using the broken lines shown in Fig. 20(a). Also shown in Fig. 19(b) are theoretical estimates, based on the calculated sensitivity curves given immediately above, of the effect of a second tone in inhibiting the activity of a single second-order nerve fiber already subjected to the steady excitation of a 1000-cps tone, 22 db above threshold (indicated by the point T_1 in Fig. 19(a)). When the ear is driven simultaneously by a second tone T_2 of different frequency and amplitude, the average phase of the resultant excitation at any point along the basilar membrane will be controlled by that frequency component which has the greater amplitude. The smaller component produces a phase modulation having a frequency equal to the difference between the frequencies of the two components and an amplitude given by the size of the smaller component relative to the larger. In other words, our mechanism exhibits a "capture effect" just as in the interference between two f-m radio transmissions.

As the frequency of T_2 is varied, the effect will be similar to a displacement along the basilar membrane. Examination of Fig. 19(a) shows that

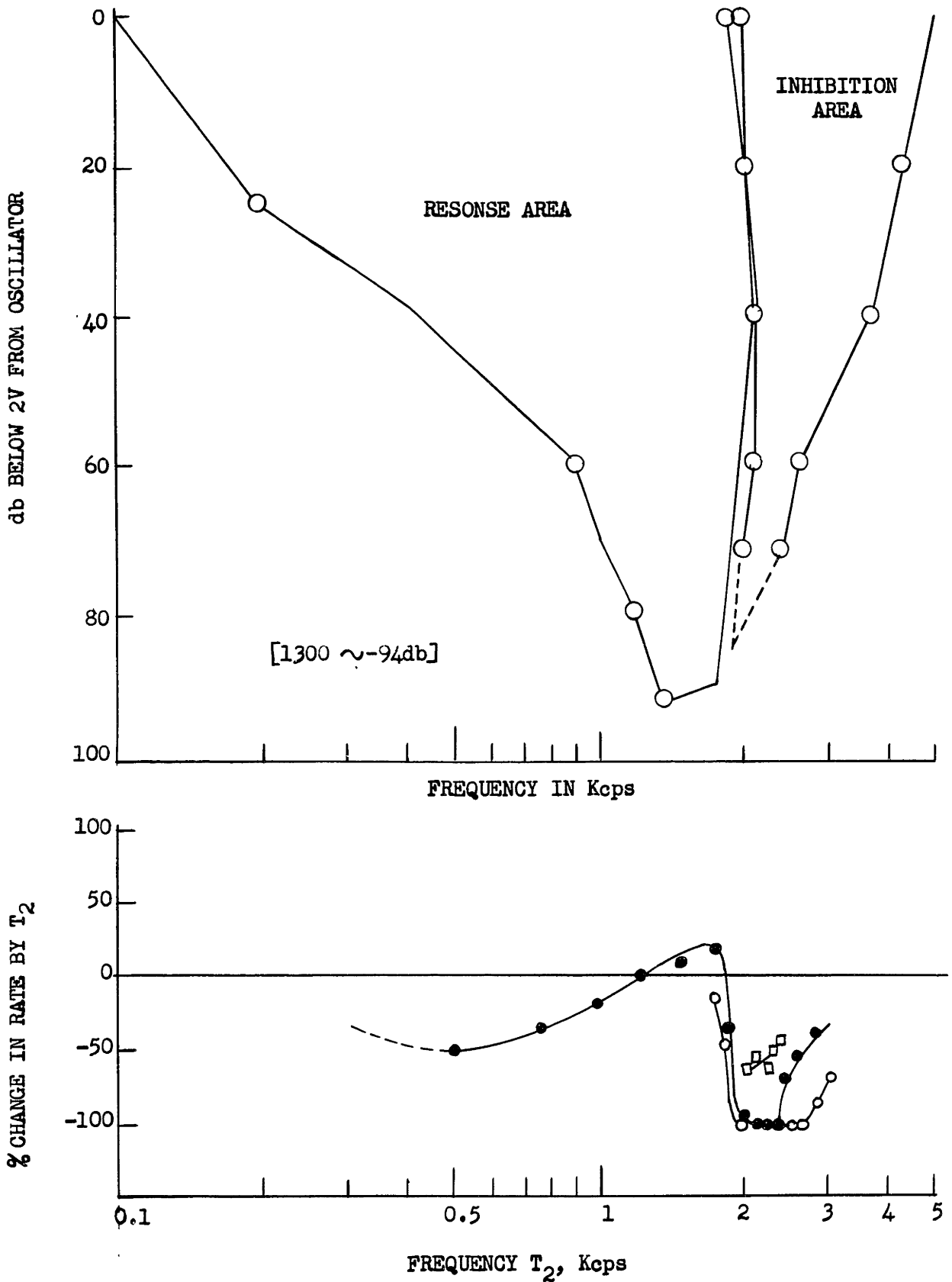


Fig. 20. Experimental data after Galambos and Davis for the response of a single second-order neuron in the cochlear nucleus of a cat under the same relative stimulation conditions as represented by the theoretical curves of Fig. 19.

even with the strength of T_2 held constant, changes in its frequency can cause widely different effects. For a T_2 at 44-db intensity and frequencies corresponding to the region between 26 and 27 mm, the excitation caused by T_2 will have greater amplitude than that caused by T_1 at both the inner and outer hair cells.

Consequently, the response of the second-order neuron will be inhibited because T_2 lies in the inhibitory area. At other frequencies, corresponding to a T_2 in the region of 16 mm, the inner-hair-cell excitation will be controlled by T_2 whereas the outer-hair-cell excitation is controlled by T_1 . Since the firing of the inhibitory nerve fibers is then no longer synchronized with the out-hair-cell excitation, it is reasonable to expect partial, but not complete, inhibition. The theoretical estimates shown in Fig. 19(b) were determined for the same relative stimulus levels as the experimental data, obtained by Galambos and Davis and reproduced in Fig. 20(b). (The threshold of the particular nerve fiber studied by Galambos and Davis is 6 db higher than that assumed in the theoretical curves, and all intensity levels have, therefore, to be modified by that amount.)

It is apparent that the theoretical model accounts not only for the abrupt transition between the response and inhibitory areas but also it provides a reasonable explanation for the observed variation of the response of this second-order neuron under double-tone stimulation in which the frequency and intensity of the second tone are varied. The author knows of no other theory that will account for these experimental results.

Professor E. G. Wever has pointed out, in personal communication, that other data of Galambos show different phenomena than that represented by Fig. 20. For instance, in some neurons investigated, the inhibitory area was found at a lower frequency. Sometimes, two inhibitory areas on either

side of the critical frequency would be observed. These data may be accounted for with our mechanism if it is assumed that the cilia on certain of the outer hair cells are inclined in the outward direction, thus shifting the phase of outer neural excitation by 180° . Reference to Fig. 14 will show that, with the phase of $W(x)$ modified by $\pm 180^\circ$ (or $\pm \pi$ radians), there will be produced a narrow inhibitory region at a lower frequency (i.e., closer to the stapes) and that the original inhibiting region will have been shifted to a still higher frequency (i.e., closer to the helicotrema) where it may escape observation. On the other hand, there is no logical reason why the rôles of some of the inner and outer fibers should not occasionally be reversed, thus reversing the areas of facilitation and inhibition.

Galambos and Davis also reported that, as the intensity of the pure tone stimuli was increased from its sub-threshold value, at first the spontaneous activity present in some neurons (even in the absence of a stimulus) is sharply decreased and then, as the tone continues to grow in intensity, the neural activity rapidly increases. This may be accounted for by the assumption that at large amplitudes, some of the outer hair cells will begin to be excited with a double-frequency component. This produces, in effect, another excitatory impulse which, because it is shifted in phase by 180° , may succeed in stimulating the second-order neuron whereas the normal impulse was inhibited.

These simple possibilities, all of which are inherent in the mechanism thus far described, may be combined in various ways to account in qualitative way for much of the experimental data on the neural response of the cochlea to steady tones.

Since our mechanism depends on accurate phase relations between the neural excitations, the rather large jitter that seems to be present in the

time within the microphonic cycle at which action potentials occur could be a most serious limitation. However, it is believed that this jitter is not nearly so large as it may at first appear. From Figs. 14 and 15, it is observed at 1000 cps (where action potentials appear throughout most of the cycle), that over the stimulated regions of both the inner and outer hair cells, there are phase variations considerably in excess of 180° .

Thus, at any instant there will likely be some hair cell, the phase of whose excitation is suitable for firing its nerve. Each of these neural responses shows up as an individual response. The cochlear microphonic, on the other hand, consists of contributions having different phases that arise from the different portions of the cochlear partition. But since each of the contributions is of the same frequency and roughly sinusoidal in wave shape, they all combine into a single microphonic which does not reveal its components as in the case of the action potentials.

Furthermore, it should also be noted that the microphonic potential will be generated mostly by the inner hair cells. Little contribution can be made by the microphonic potentials created by the outer hair cells for the simple reason that $\int_0^{3.5} \frac{d^4 y}{d x^4} dx = \frac{d^3 y}{d x^3} \Big|_0^{3.5} = 0$, since the third derivative essentially vanishes at both limits of the integration. Expressed in another way, the electric field set up by the force wave must be highly localized because a positive force acting upon the tectorial membrane must necessarily be almost exactly balanced at every instant by nearby negative forces. Consequently, one may expect high voltage gradients to be established in the region of resonance where maximum forces of opposite polarities may occur simultaneously within a distance 1.5 mm. The fact that the nerve fibers innervating the outer hair cells run basalward for at least this distance suggests that the high local potential gradients may be utilized to stimulate these nerves. In appendix IV it is shown that a

very simple model indicated that these fibers may be stimulated by the curvature of the potential field in which they are immersed--thus giving, in conjunction with the beam hypothesis, a neural excitation that varies as the sixth place derivative of the instantaneous displacement. This would provide a frequency selectivity somewhat sharper than that indicated in Fig. 14. However, even without this additional sharpening action, our mechanism easily accounts for the ability of the ear to discriminate slight changes in the pitch of a sound, as we shall now show.

D. Frequency Discrimination

From Fig. 14(b), it is apparent that the phase difference between the two waves of neural excitation is changing most rapidly with respect to place in the vicinity of resonance, specifically at the rate of 110° per millimeter. Relating this change in place to an equivalent change in frequency, one finds that the phase difference between the excitation of the inner and outer hair cells at a given point near the region of resonance will change by about 10° for each percent change in frequency of a steady, 1000-cps tone.

Now, it seems reasonable to assume that each cycle of the wave will provide the nervous system with a single estimate of the in-phase point or, what is equivalent, the frequency. There will be irregularities in the neural response which will limit the accuracy with which this in-phase point can be determined, but it seems plausible that available neural elements could determine this point of phase coincidence to within 30° (or $1/12$ of a cycle). Each estimate thus provided would measure the frequency to an accuracy of about 3 percent.

But it is also reasonable to believe that the nervous system is capable of computing the average value of a large number of such samplings extending over a time interval as great as 0.2 seconds. For a 1000-cps tone, this would correspond to 200 estimates of the frequency. The accuracy of the mean value

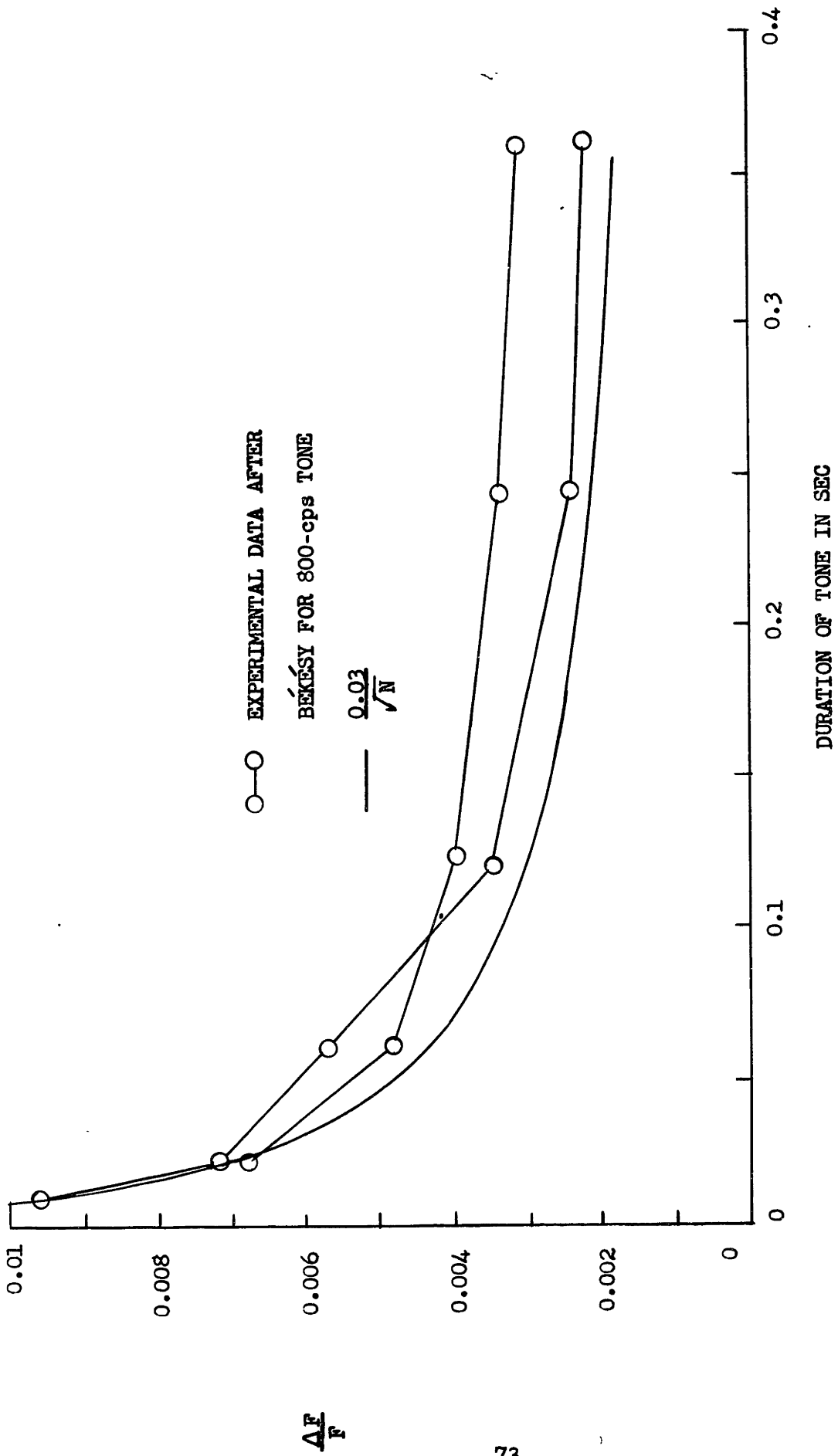


Fig. 21. Effect of duration of tone on frequency discrimination.

of these observations would be in the order of 3 percent $/\sqrt{N}$, where N is the number of measurements, or about 0.002. If a short tone burst were presented to the ear, the number of samples would be reduced and the ability of the ear to detect slight changes in pitch would also decrease. Békésy has studied experimentally the smallest frequency change ΔF that may be detected as a function of the duration of a 800 cps tone. His data, taken from Fig. 38 of Ref. S-1, for two different subjects, are compared in Fig. 21 with values calculated from the equation $0.03/\sqrt{N}$ derived from our theory. Here, also, the theory seems to be in substantial agreement with experimental fact. Although the accuracy with which phase could be measured was assumed rather arbitrarily, threshold considerations suggest that the error could not be much larger than 30 degrees. However, even with this large value, the proposed mechanism has no difficulty in accounting for the observed data.

VIII. PHASE EFFECTS IN MONAURAL PERCEPTION

In this section we show how our cochlear mechanism can account for the phase effects and beats that occur when two or more tones of different frequency are presented simultaneously to the ear.

Despite Ohm's acoustic law, which states that the ear analyzes a complex sound into simple tones independent of each other, it was apparent to Helmholtz, and should now be to everyone, that when the frequency components lie so close together that the periodicity of the beats occurring between various pairs may be resolved in time by the ear, then modification of the phase of these components will certainly change the temporal pattern of the sound as heard by the ear. This could not be made clearer than by the simple observation that the Fourier amplitude spectrum of, say, a 1-second sample of speech, is unchanged if the signal is reversed in time--only the phase spectrum is changed in sign. Yet, subjectively,

the one-second speech sample would sound quite differently if reproduced backwards in time. The frequency components would be spaced by only one cycle per second and the change in the one-second temporal pattern of the sound would certainly be resolved by the ear. On the other hand, if the structural content of the sound is characterized by widely separated natural frequencies, the beats would be indiscernible to the ear. Furthermore, if the temporal content is periodic, as for example in a steady-voiced vowel sound, reversal in time would leave the temporal content unchanged. The auditory perception of such a sound would not be altered if time were reversed.

A. Beats of Mistuned Unisons.

Mathes and Miller studied the perception of small groups of tones, equally spaced in frequency, whose amplitudes and phases could be systematically adjusted ($M-1$). Some of their most interesting studies were made with a sinusoidally modulated tone in which either amplitude or frequency modulation could be selected by simply shifting the phase of the carrier relative to the upper and lower side bands. The amplitude of the carrier could be varied independently to control the depth of the modulation. The following effects were observed:

- a. For a 100-cps tone, amplitude-modulated by a 50-cps signal, flutter at the modulation frequency first appears at a low value of the modulation index M . As M is increased, the quality of the tone becomes increasingly raucous "until it is judged to have a very rough quality for values of M between 0.85 and 1.5. The degree of roughness falls off quite rapidly for smaller values of M , is a maximum for a value between 1.1 and 1.2, and diminishes for large values of M to that associated with the beating of two equal pure tones at a spacing of twice the modulation frequency".

b. By shifting the phase of the carrier 90 degrees, an approximate frequency modulation is produced. The AM and FM types of waves produce two quite different sequences of aural sensation as the modulation frequency was increased from a very low value, for which "the actual fluctuation in the frequency as a function of time for the FM wave and the similar swelling and diminishing in the loudness in the AM case are readily followed. By the time the signal (modulating frequency) has reached 7 to 10 cycles, the FM sensation is that of a warble while the 100-percent AM case has an interrupted character tending toward raucousness.

"It is in the—region from 25 to 75 cycles, which provides the greatest sense of roughness in the 100 percent AM case, that the most striking difference is noted. In this region the switch from the AM to the FM case is marked by the disappearance of the roughness, which has a characteristic beat at the signal frequency rate, to what sounds like a combination of pure tones accompanied by an apparent pitch (i.e. modulation rate) sensation of twice the signal frequency. The latter effect correlates with the rate at which amplitude maxima appear in the (envelope) of the FM case. In these tests the carrier need not be harmonically related to the (modulation) signal."

These experimental observations of Mathes and Miller also provide corroboration of our theory. Using the data of Fig. 14, we may calculate the stimulation of the organ of Corti for both the AM and FM modulated waveforms that were employed in the experiment and show that the theoretical excitation patterns correlate with the subjective sensations.

The analytic representations for the pressure wave having a carrier frequency, ω_c , modulated at a frequency, Δ , are, for the AM and FM

cases respectively,

$$p(t) = e^{j\omega_0 t} \left[1 + \frac{M}{2} (e^{j\Delta t} + e^{-j\Delta t}) \right] \quad (23a)$$

and

$$p(t) = e^{j\omega_0 t} \left[j + \frac{M}{2} (e^{j\Delta t} + e^{-j\Delta t}) \right] \quad (23b)$$

Now, for the assumed tuning of our model of the cochlea (see appendix A), a change, Δx cm., in place along the basilar membrane is nearly equivalent to a change in the frequency by the factor $-1.15 \Delta x$. Thus the data of Fig. 14 may also be regarded as the variation with frequency of the intensity and relative phases of the excitation of the inner and outer hair cells. The stimulus is taken to be a 1000-cps tone which is modulated at 50 cps. Assuming the 1000-cps carrier to be at $x = 24$ mm, the upper side band at 1,050 cps would correspond to $x = 24.43$ mm and the lower side band to $x = 23.57$ mm. By interpolation from the data of Fig. 14, one obtains the values tabulated in Table 1.

Frequency	x (mm)	Y(x)	W(x)
	23.	0.123 $\angle -220^\circ$	0.351 $\angle -166^\circ$
950	23.57	0.115 $\angle -224^\circ$	0.412 $\angle -237^\circ$
1000	24.00	0.107 $\angle -267^\circ$	0.465 $\angle -294^\circ$
1050	24.43	0.089 $\angle -281^\circ$	0.375 $\angle -352^\circ$
	25.0	0.067 $\angle -304^\circ$	0.246 $\angle -430^\circ$

Table 1. Relative stimulation of inner and outer hair cells at $x = 2.4$ cm caused by a sinusoidal pressure wave at the stapes of unit amplitude for frequencies 950, 1000 and 1050 cycles per second.

The amplitude and phase factors of Table 1 may now be applied to the frequency components of Eqs. (23) to obtain mathematical representations for the stimulation at the inner and outer hair cells at the point $x = 2.4$.

Letting $\omega_1 t = 0$, we have, for $M = 1.00$,

$$\begin{aligned}
 y(\theta) &= 0.107 e^{j(\theta - 263^\circ)} + 0.057 e^{j(0.95\theta - 244^\circ)} + \\
 &\quad + 0.049 e^{j(1.05\theta - 281^\circ)}, \\
 &= 0.107 e^{j(\theta - 263^\circ)} \left[1 + 0.989 \cos(0.05\theta - 18^\circ) - \right. \\
 &\quad \left. - j 0.77 \sin(0.05\theta - 18^\circ) \right], \quad (25d)
 \end{aligned}$$

$$\begin{aligned}
 w(\theta) &= 0.465 e^{j(\theta - 294^\circ)} + 0.206 e^{j(0.95\theta - 237^\circ)} + \\
 &\quad + 0.187 e^{j(1.05\theta - 352^\circ)}, \\
 &= 0.465 e^{j(\theta - 294^\circ)} \left[1 + 0.842 \cos(0.05\theta - 58^\circ) - \right. \\
 &\quad \left. - j 0.042 \sin(0.05\theta - 58^\circ) \right], \quad (25b)
 \end{aligned}$$

To study the effect of various amounts of modulation we need only to divide the "1" in the brackets by M . Also, to consider the FM case, simply replace the "1" with a "j". Thus, the remaining terms in the brackets may be evaluated once and for all, thereby simplifying the computations.

The bracketed terms are of principle interest since they represent the temporal effect of the modulation upon the relative intensity and the phase of the excitations at the inner and outer hair cells at the point $x = 2.4$ cm. These terms have been evaluated and are plotted in Fig. 22 for the AM case, and in Fig. 23 for the FM case.

In both instances, the envelope of the excitation temporal pattern at the outer hair cells lags to the envelope of the pattern at the inner hair cells. This is because of the sharper selectivity and more rapid change in phase with frequency of the former. Since this excess phase change is, for the data of Table 1, about 7.8 degrees per percent frequency change, the equivalent envelope delay (i.e. $\frac{d\phi}{d\omega}$) is about 2.2 milliseconds. For a 50-cycle modulation rate, this corresponds to the 40-degree shift that is observed in Figs. 22 and 23. As a result for the AM case, at the instant that the inhibitory stimulation at the inner cells has all

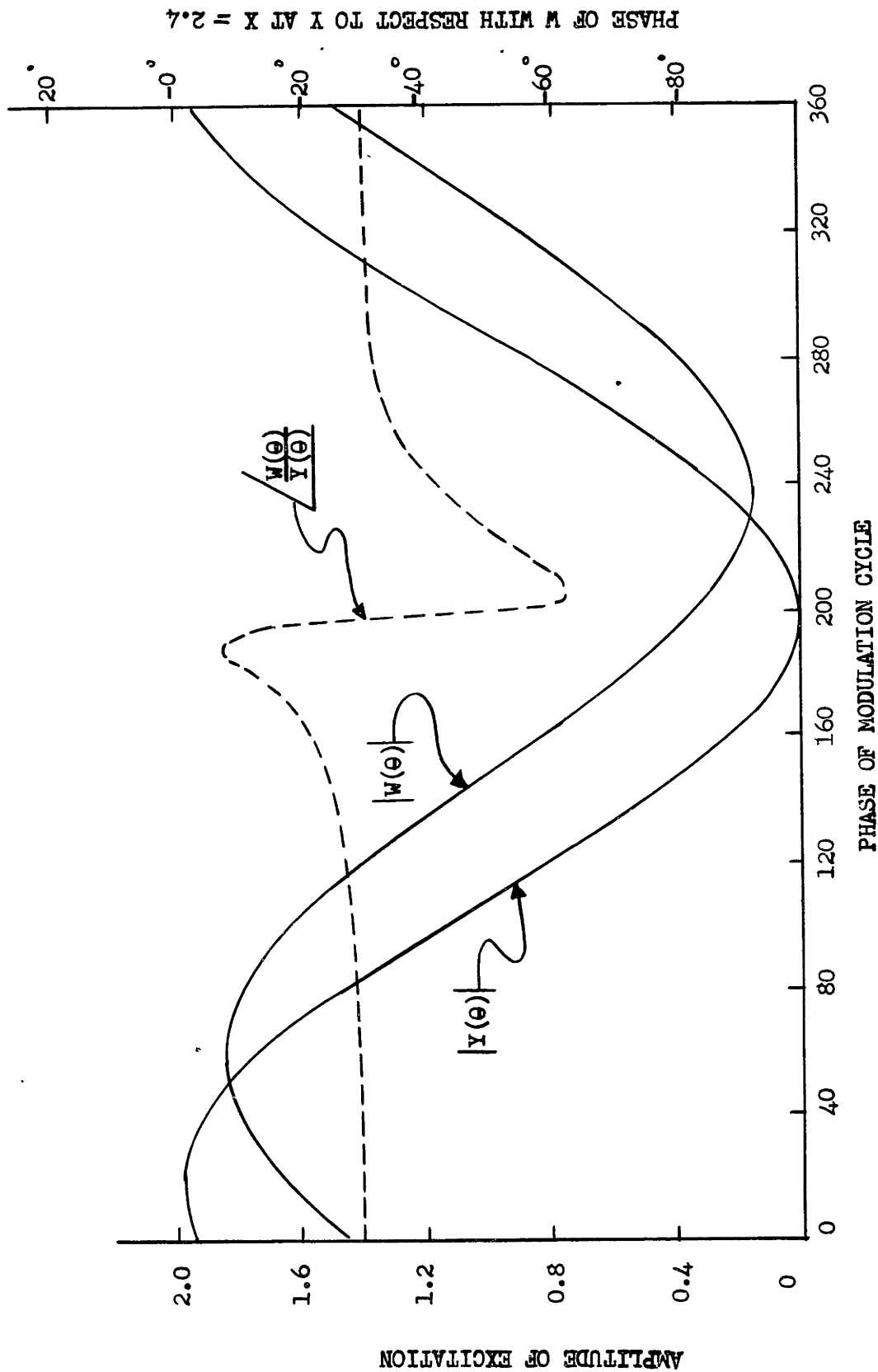


Fig. 22. Variation of the intensities and relative phase of the inner and outer hair cell excitations at the point $x = 2.4$ cm, when the 1,000-cps stimulating tone is 100% amplitude modulated at 50 cycles per second.

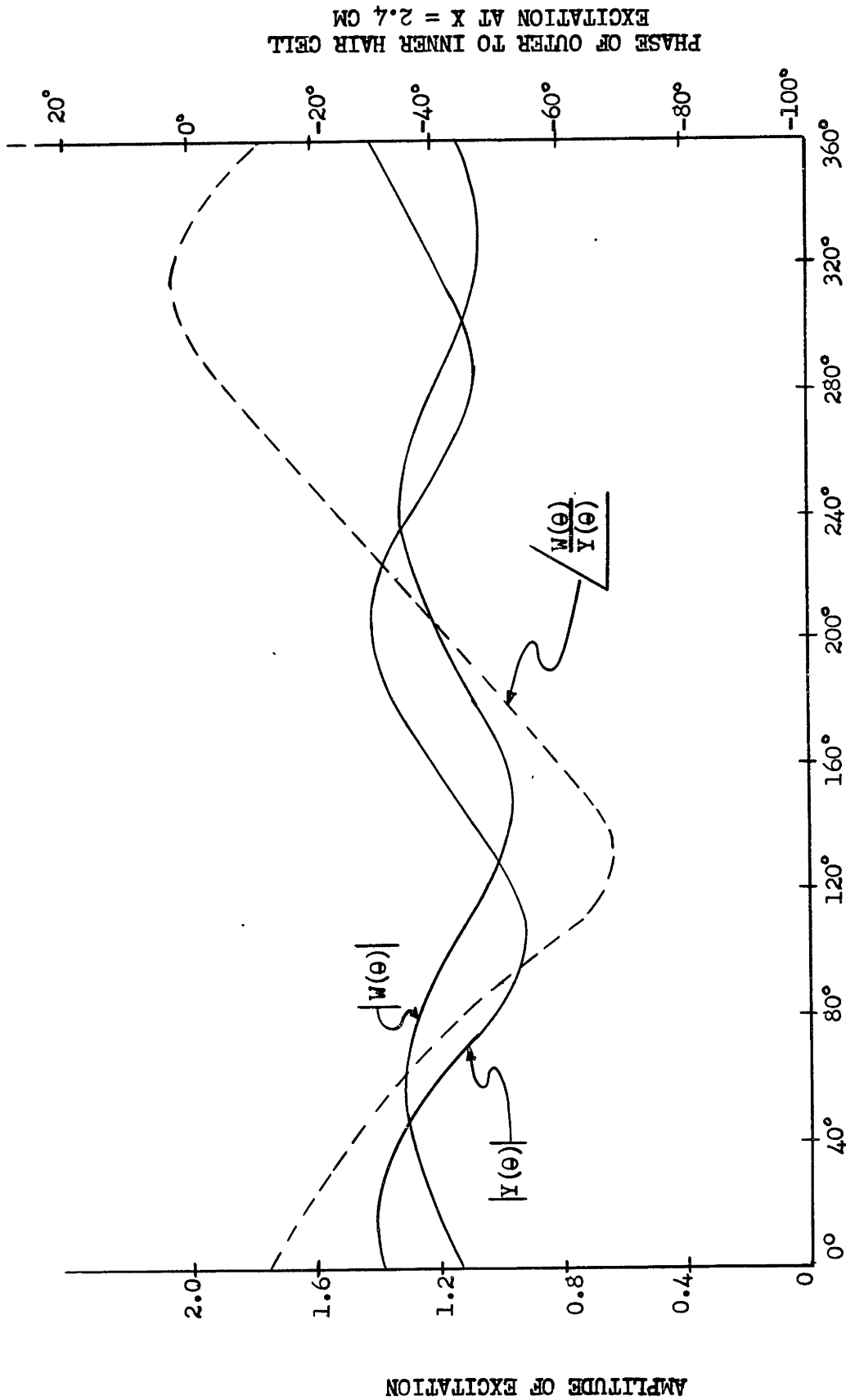


Fig. 23. Same as Fig. 22 except that carrier has been shifted 90 degrees in phase to give a stimulating tone that is frequency modulated.

but ceased, there remains a considerable stimulation of the outer excitatory cells. Thus, during the trough of the modulation, groups of second-order neurons that had previously been inhibited because the inner fibers were firing first, can now fire for a short fraction of the modulation period. The fact that the phase of the outer excitation, relative to that at the inner, changes abruptly during this time, permits firing of still other second-order neurons which during most of the cycle are inhibited.

If the modulation were increased to $M = 1/0.842 = 1.185$, then the excitation at the outer hair cells would be completely modulated and would reach zero for a short instant at the 240-degree point of the modulation cycle. This would, in effect, allow the outer neurons a brief resting period so that when the envelope again builds up to threshold, large numbers could fire in synchrony. In addition, because the excitation at the inner cells is now overmodulated, the phase of the inner excitation is captured by the side bands with the result that the phase curve is no longer bounded but is joined to the next branch, 360 degrees away, and continues to migrate through ever increasing phase angles in an irregular but periodic manner. This essentially means that bands of second-order excitation step abruptly along the mapping of the place domain in the higher neural centers, resting for a greater part of the time in the region associated with the pitch of the carrier. This effect, I suggest, accounts for the sensation at maximum raucousness encountered at $M = 1.2$.

When M is increased still further, the excitation wave of the outer hair cells also becomes overmodulated--its phase now migrates at the same rate as that of the inner hair cells and we return to a situation where the relative phase between the two appears more like that of

Fig. 22. Thus, the roughness decreases for M over 1.2, as observed by Mathes and Miller.

For the FM case, the phase difference between the inner and outer cells oscillates between 0 and -70 degrees in a fairly regular way, as shown in Fig. 23. The place of the in-phase point along the organ of Corti likewise oscillates between $x = 24$ mm and $x = 24 - \frac{70^\circ}{90^\circ/\text{mm}} = 23.22$ mm, since it is seen from Table 1 that the relative phase is changed by about 90 degrees per millimeter change in place. This should cause a peak-to-peak fluctuation in perceived frequency of (1.15) (2.4 = 2.322) or about 9 percent, a value easily discernible to the ear. The fluctuations of amplitude of the FM wave occur at twice the frequency of the modulation, so it is not surprising that the listener perceives a doubling of modulation periodicity.

B. Beats of Mistuned Consonances

The beating effects observed in the stimuli of the Mathes and Miller experiment could be interpreted as the classical beats of imperfect unisons. Another kind of beat arises when the frequency components depart slightly from some interval relationship, as for example the octave relationship. The classical resonance theory of hearing has had a hard time explaining the beats of these mistuned consonances and has had to assume the existence of overtones in the primary tones themselves or the generation of such overtones through distortion within the ear. However, as Wever points out (W-1, p. 377), "It is easy to dispute this explanation of the beats of mistuned consonances. As Koenig stoutly maintained in opposition to Helmholtz, and as modern experiments prove beyond question, these beats arise in the use of tones in which physical overtones are absent or negligible. They arise also when the level of stimulation is so low that no additional tones are generated within the ear."

Accordingly, we next examine our theory to see what explanation it offers regarding these beats of mistuned consonants. Without incurring loss of generality, we shall investigate the simplest consonance relationship, the octave, and shall show that, indeed, such beats are produced without the need for mechanical non-linearities. Furthermore, the theory indicates a new effect that has never been reported—the apparent pitch of the fundamental should depend upon the relative phase of the 2'nd harmonic. Since the theory gives quantitative relations between this change in pitch and the intensities and relative phase of the two tones, it suggests a psychophysical experiment that should be decisive.

To show how beats may arise, we first consider a 1000-cps tone plus and a second-harmonic tone of 2000 cps which has a definite phase relation with reference to the fundamental. To establish the vibration pattern of the organ of Corti for this sound, we first estimate the behavior of the cochlea at 2000 cps by making the reasonable approximation that the "1000-cps" cells at $x = 2.4$ cm will have the same amplitude and phase relations for a 2000-cps tone as the "500-cps" cells have for a 1000-cps tone. From Equation A-7 of Appendix A, the "500-cps" cells are found to be those at $x = 3.00$ cm. From the data of Fig. 14, (with the level of $W(x)$ adjusted so that $|W(2.4)|$ is 20 db greater than $|Y(2.4)|$ to account for the difference in sensitivity between the inner and outer hair cells) we find that at $x = 3.00$ cm., the ratio of inner to outer excitations is $2.3 / 286^\circ$. Assuming, then, that the excitation of the inner-hair cells is composed of a fundamental component of unit amplitude and a second harmonic of relative amplitude A and relative phase ϕ , we may write for the stimulation of the inner cells at the "equal phase" point

$$y(\theta) = \sin \theta + A \sin(2\theta + \phi) \quad (26)$$

where $\theta = 2\pi ft$ as before. The stimulation of the outer cells at this same place will be given by

$$w(\theta) = 10 \sin \theta + 0.435 A \sin(2\theta + \alpha - 286^\circ) \quad (27)$$

The neurons associated with either the inner or outer hair cells will fire when the stimulation first reaches the threshold value. This threshold value, measured relative to the amplitude of the fundamental component at the inner hair cell will be represented by T . Thus, increasing the stimulation level is equivalent to decreasing T .

According to these assumptions, the inner neuron will fire at the phase, θ_i , which satisfies the equation

$$y(\theta) = \sin \theta_i + A \sin(2\theta_i + \alpha) = T. \quad (28)$$

(Only those phase angles, θ_i , at which the stimulus is increasing through the value T have meaning--the other solution θ_i , for which the stimulus is dying away should be discarded). Similarly, the outer neuron will fire at the phase angle θ_o which satisfies the equation

$$w(\theta) = 10 \sin \theta_o + 0.435 A \sin(2\theta_o + \alpha - 286^\circ) = T. \quad (29)$$

It is apparent from Eq. (28) that the presence of a second harmonic will change the phase angle θ_i at which the inner neuron fires. The amount of change will be a function of the relative amplitude A of the second harmonic, its relative phase α , and the threshold T . To see just how big a change in phase of θ_i is produced by adding a second harmonic of amplitude A , we expand Eq. (28) in a power series in A about the value $\theta_i = \sin^{-1} T$ for $A = 0$. This gives for the leading terms:

$$\theta_i \approx \sin^{-1} T - A \left[2T \cos \alpha + \frac{1-2T^2}{\sqrt{1-T^2}} \sin \alpha \right]. \quad (30)$$

The corresponding expression for θ_o is,

$$\theta_o \approx \sin^{-1}\left(\frac{T}{10}\right) - 0.0435 A \left[0.2T \cos(\alpha - 286^\circ) + \frac{1-0.02T^2}{\sqrt{1-0.01T^2}} \sin(\alpha - 286^\circ) \right],$$

$$\theta_0 \approx \sin^{-1}\left(\frac{I}{I_0}\right) = 0.04A \left[0.2T \cos(\alpha - 286^\circ) + \sin(\alpha - 286^\circ) \right]. \quad (31)$$

Because of the greater frequency selectivity of the outer hair cells, the amplitude of the second harmonic is attenuated to a very small fraction, i.e. 4 percent. Therefore, for practical purposes, the outer-hair-cell excitation at $x = 2.4$ will be unaffected by the addition of a second harmonic.

However, since the phase of the inner-hair-cell excitation is greatly modified, the point where inner and outer neurons fire simultaneously will shift to a different place along the basilar membrane and this should result in a change in pitch at the fundamental as perceived by the listener.

Figure 24 shows the variation of θ_i as a function of the phase α of a second harmonic having an amplitude of 1, $\frac{1}{2}$, or $\frac{1}{4}$. The subjective pitch change that should result from this shift in phase may be estimated from the theoretical figure that the change in phase with frequency is about 0.78 degrees per cycle in this region. From Eq. (30) for $T = 0$, we then find that the perceived change in frequency of the 1000-cps tone should be roughly

$$\Delta f \approx 74 A \sin \alpha, \quad (\text{cps}) \quad (32)$$

for small A .

As A becomes larger and approaches 1, the second harmonic begins to control the firing of the inner nerve fibers. For instance, Fig. 24 shows that with $A = 1$ (and greater), the inner neuron will fire twice during each 1000-cps oscillation. Although not illustrated here, one may show that when A lies between 1 and $\frac{1}{2}$ and/or the stimulation level is near threshold, the double-frequency firing will occur if the phase angle of the second harmonic lies within certain ranges. Outside of these ranges, at angles differing by 180 degrees, only one firing will occur per period. It is suggested that this effect may account for the statement in Ref. S-1, pg 203 that "a given

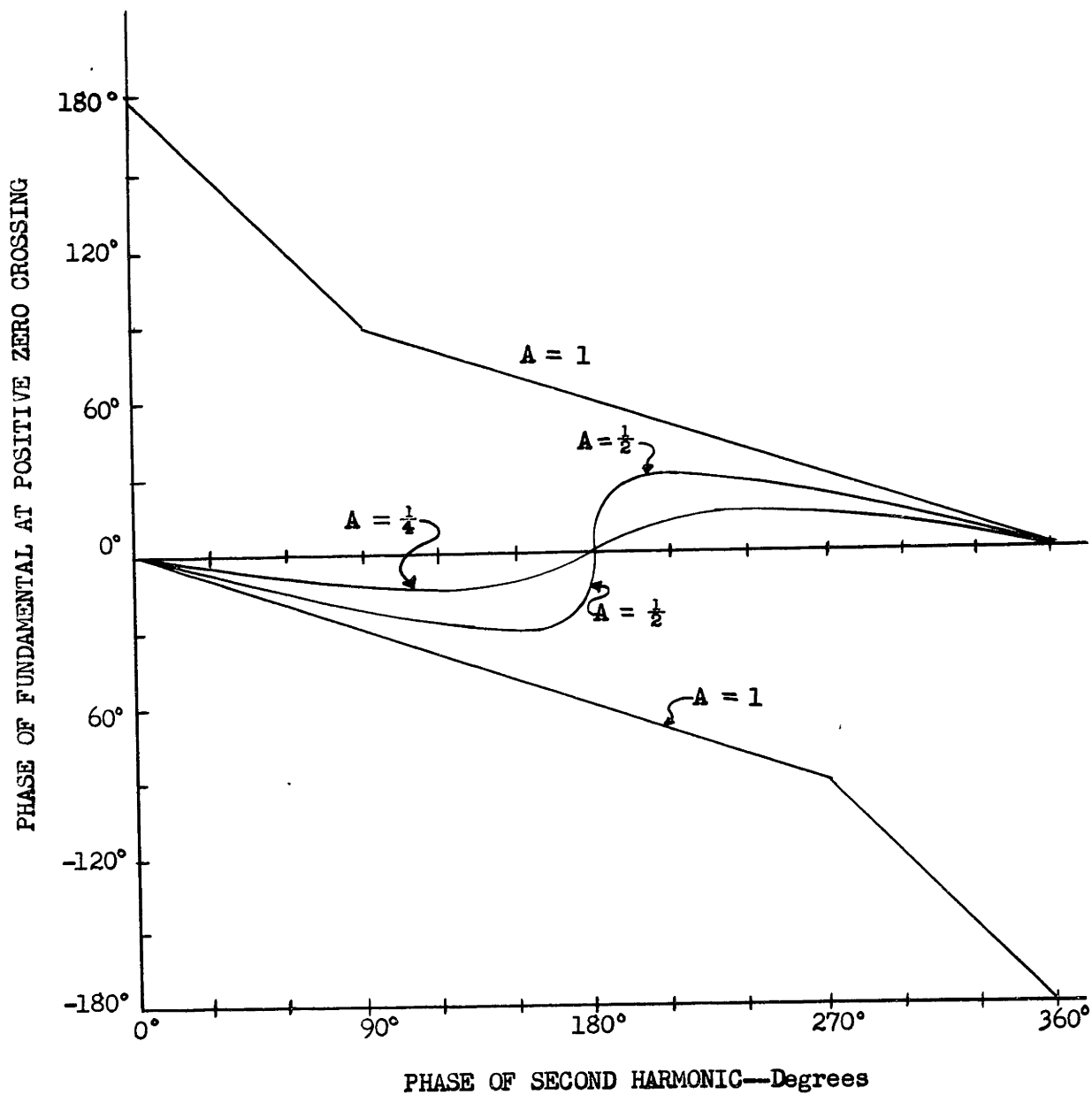


Fig. 24. Variation of phase of firing of the 1,000-cps inner-hair-cell neuron as a function of the phase of a second harmonic at 2,000-cps of relative amplitude A . (Curves are for the large signal case well above threshold, i.e. $T = 0$.)

tone, plus another tone of exactly twice the frequency, may sound either louder or less loud than the fundamental alone. The phase (of the harmonic) yielding maximal loudness differs from that giving minimal loudness by 180° ." And on pg 205, "Not only does a change of phase alter the loudness of a harmonic, and of the total experience, but it produces noticeable differences in quality, provided the fundamental is a low-tone of 100 cycles. The phase-relation giving minimal loudness is characterized by smoothness, whereas the opposite phase, which leads to maximal loudness, carries with it a rough or dissonant element." The similarity of the neural excitation pattern obtained here to that obtained with the type of stimuli used by Mathes and Miller, who also observed the sensation of roughness, provides further indication that abrupt and discontinuous movements of the "in-phase" point along the organ of Corti may be the neuro-physiological correlate of the sensation of tonal roughness.

When A is unity or larger, the second harmonic captures control of the phase of the inner-hair-cell excitation at $x = 2.4$ cm. The outer-hair-cell excitation at this point is, of course, still controlled by the fundamental tone. Consequently, if the frequency of the second-harmonic is detuned slightly, its phase, ϕ , will change slowly but continuously with time. As indicated in Fig. 24 for $A=1$, this will cause bands of "in-phase" regions to move continuously along the organ of Corti; toward the stapes when the second harmonic is tuned high and away from the stapes when it is tuned low. The reversal in the direction of movement caused by mistuning on the other side of consonance suggests that the beats may have a different quality for the two cases.

C. Is There A Phase-Masking Effect? -- A Decisive Experiment.

A more interesting possibility is that this sweeping motion of the in-phase points caused by a large, slightly detuned, second-harmonic tone

may be used to prevent the ear from using a phase principle in perceiving the pitch of the 1000-cps tone. If the second harmonic is detuned by about 10 cps, the sharp localization of pitch, such as that illustrated in Fig. 21 should no longer be possible because the "in-phase" points would continuously move across the resonance region and the sharp demarcation assumed in the mechanism of Fig. 21 could not be obtained.

In other words, our theory suggests that a kind of phase-masking may occur. The usual masking effect, where one tone cannot be heard in the presence of a stronger second tone, is probably associated with the amplitude pattern of the stimulation along the outer hair cells (See Fig. 18). However, it is apparent that a second tone widely separated in frequency can disturb the phase pattern of the inner nerve excitation long before it is sufficiently intense to produce any significant change in the outer-hair-cell excitation pattern.

This phase masking would influence only those aspects of hearing which are dependent on the phase mechanism. Thus, we may predict that as the amplitude of the mistuned octave tone is increased, the accuracy with which the ear can determine the pitch of the 1000-cps tone will change from its small value of about 4 cps for $A \ll 1$, to about 60 cps, which is the critical bandwidth of the ear as determined by its amplitude selectivity, for $A > 1$.

One possible way in which these effects could be tested experimentally would be to present the fundamental and its harmonic to one ear, and ask the subject to tune an oscillator supplying the tone to his other ear so as to match the fundamental frequency. With this set up, one could investigate both the apparent changes of pitch as a function of the phase angle of the harmonic as well as the predicted widening of the just-noticeable-pitch differences caused by the "phase-masking" effect described above.

In order to obtain a value of $A = 1$ at the inner hair cells, the sound pressure wave at the eardrum must have a second-harmonic component which is 2.6 times as intense as the fundamental component. This is necessary because the point being considered ($x = 2.4$ cm) is resonant at 1000 cps, instead of 2000 cps. From Fig. 14, comparison of $Y(x)$ for $x = 24$ and $x = 30$ mm. indicates that the response will be attenuated by a factor of 0.38 below its peak value.

Finally, if the phase of the second harmonic is such that it has a value of zero at the instant that the fundamental component is just reaching the threshold value, then the phase of firing will be independent of the amplitude of the second harmonic. This occurs when the bracketed expression in Eq. (30) vanishes, or when

$$\phi = -\tan^{-1} \left[\frac{2T\sqrt{1-T^2}}{1-2T^2} \right] = -2 \sin^{-1} T \quad (33)$$

Using this relationship, we may design a test that will avoid diplacusis errors. The subject would first match the fundamental tones in his two ears. The second-harmonic tone would then be introduced into one ear and its phase would be adjusted by the subject until the pitch of the fundamental tones in the two ears is again matched. The relationship between this critical phase angle, given by Eq. (33), and the intensity of the stimulation is plotted in Fig. 25.

It is hoped that these experiments may be conducted in the near future. The outcome will be of considerable importance because our theory clearly indicates that such changes in pitch of the fundamental should occur—yet, the literature on beats nowhere indicates that such pitch changes have been observed at moderate stimulation levels. At very high intensities, wavers of both the fundamental and beating harmonic are observed, but this may be attributed to mechanical non-linear effects within the ear.

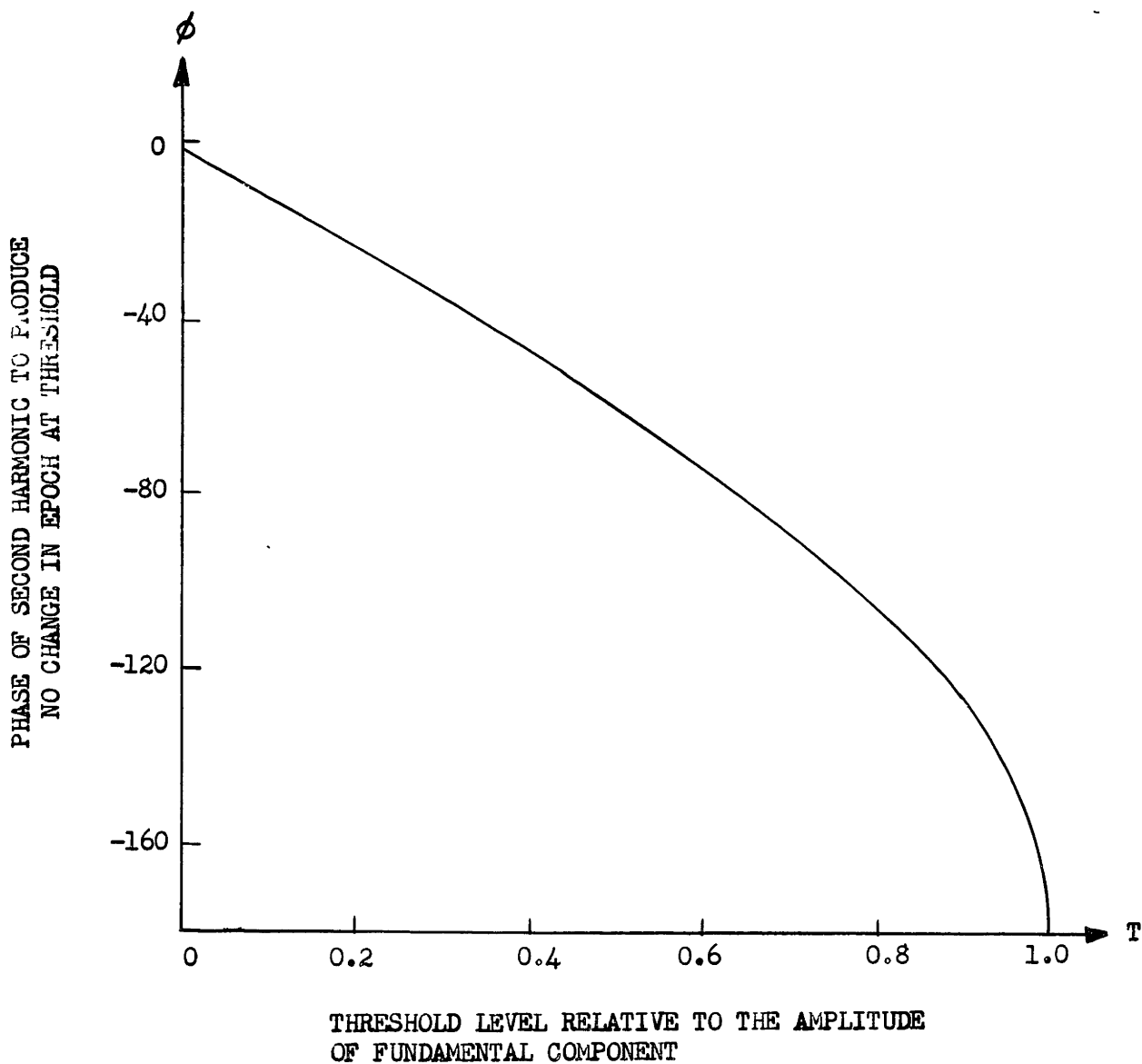


Fig. 25. Variation of critical phase angle of second harmonic for which no change in pitch should occur, as a function of the stimulation level relative to the threshold of the inner-hair-cell neuron.

APPENDIX A

Analysis of the Mechanical Response of the Cochlea

A number of mathematical models of the cochlea have been described in the literature, the most realistic of which allow for the motion of the fluid within the scalae and may be classified together, therefore, as hydrodynamical models. The solution of the equations representing these models is in general very difficult and exact analytic expressions can be given only when rather extreme simplifications are made as, for example, J. Zwislocki's neglect of the mass (Z-1). O. F. Ranke has insisted for years that the simple one-dimensional model of the cochlea isn't adequate and that the motion of the fluid in both transverse and longitudinal directions must be considered. He claims to have obtained solutions for the motion of the cochlear portion when subjected to sinusoidal sounds but remarks that "This work leads to very unpleasant equations (which) are very difficult to handle.", (R-1).

Now, it is not justified to argue here the relative merits of these various models, nor is it reasonable, in view of the rather broad scope of this dissertation, to attempt to develop here still another mathematical model of the hydrodynamics at the cochlea: this problem alone has been and will likely continue to be the subject of other Doctoral dissertations. Nevertheless, since the present thesis is concerned with phenomena subsequent to and influenced by the mechanical displacement at the cochlear portion, some single specification of its vibration is necessary if only for the purpose of achieving consistency between the various derived results. We seek therefore a mathematical expression which, primarily, is in reasonable agreement with the vibration patterns of the cochlear partition as actually

observed by G. von Békésy and which, only secondarily need be derivable from the more sophisticated hydrodynamical considerations and differential equations. In short, we would be satisfied with an empirical representation for Békésy's data, but lacking even that, it is better to deduce the form of such a representation by obtaining an approximate solution of one of the more precise mathematical models.

In 1950, Peterson and Bogert formulated a hydrodynamical model of the cochlea and solved by numerical methods the associated equations for the special case where dissipation is ignored (P-2). The effect of damping was later investigated experimentally by measurements made on a 175-section electrical-network analogue of the cochlea (B-4). Recently, Fletcher solved numerically essentially the same fundamental differential equation arising in the Peterson-Bogert study but with dissipation included (F-2). The displacement amplitudes and phases so obtained appear to be in good agreement with the experimental results of Békésy. We are encouraged, therefore, to seek an approximate analytic solution of these equations which will enable us to derive such things as $d^4y(x)/dx^4$ by differentiating the analytical expression for $y(x)$. Such a solution was found in August 1950 and the numerical values shown in Figs. 14 and 15 were calculated at that time. It is based on the simple transmission-line model of the basilar membrane proposed by Wegel and Lane in 1924 which leads to an equation that differs by a negligible amount from the differential equation given by Peterson and Bogert for the transverse pressure mode. In this model, the series impedance $Z(x)$ of the transmission line at a distance x from the stapes represents the inertial effects of the current $I(x)$ flowing inward through the scala vestibula and outward through the scala tympani;

in electrical terms $Z(x)$ is an inductance. The shunt admittance $M(x)$ represents the compliance of the cochlear partition. In electrical terms, it consists of capacitance (representing the elastic effect of the basilar membrane) in series with an inductance (representing the mass of organ of Corti and fluid loading) and a resistance representing the viscous damping. The pressure differential $P(x)$ across the cochlear partition corresponds to voltage across the transmission line.

We start with the same basic data for the physical dimensions and mechanical properties of the cochlea as used by Peterson and Bogert, but we shall approximate certain of these data by functional forms which lead to a simpler form of differential equation that may then be solved approximately by the methods applicable to transmission lines having continuously varying parameters (S-3). The accuracy of this approximate solution is improved by including the dissipation since the "reflections", caused by the rapid variation of $M(x)$ and $Z(x)$ along the line, are damped out. Thus, the more realistic model which includes damping happily offers improved accuracy in the approximation.

We describe only inertial effects to the longitudinal flow of the fluid in the two scalae, leaving all viscous dissipation to be accounted for in the motion of the cochlear partition. $Z(x)$ is thus analogous to an inductance whose value is equal to the momentum associated with a unit current through the cross section $S_0(x)$ of the scala vestibuli. Because there is an equal and opposite flow in the two scalae at every point, the effective density is twice the density ρ_0 of the perilymph and

$$Z(x) = \left(\frac{2 \rho_0}{S_0} \right) \cdot P \quad (A-1)$$

where p is the time-derivative operator.

Petersen and Bogert have already examined the available experimental data on the variation of the cross section $S_0(x)$ with x , as well as the variation of the width $b(x)$ and the elasticity $k(x)$ of the basilar membrane. They concluded that the ratio of width to cross section may be approximated by

$$\frac{b \rho_0}{S_0} \approx 0.661 e^{0.548x} \quad (\text{A-2})$$

where x is in centimeters throughout.

Instead of using their linear expression for $b(x)$, we shall use an exponential approximation

$$b(x) \approx 0.019 e^{0.3x} \quad (\text{A-3})$$

Taking the density ρ_0 of the perilymph as 1g/cm^3 , we obtain finally

$$Z(x) = 69.6 e^{0.248x} p \quad (\text{A-4})$$

as the series impedance of the transmission line.

To evaluate the shunt admittance $M(x)$ of the line, we start with the empirical expression for the volume elasticity given by Petersen and Bogert.

$$k(x) \approx 1.72 \times 10^9 e^{-2x} (\text{dyne/cm}^2) \quad (\text{A-5})$$

Thus, the 'capacitance' $C(x)$ of a strip dx long and b wide is

$$C(x) dx = \frac{b(x) dx}{k(x)} \quad (\text{cm}^3/\text{unit pressure})$$

$$\text{or} \quad C(x) = \frac{b(x)}{1.72 \times 10^9} e^{2x} \quad (\text{A-6})$$

We further assume that the mass of the cochlear partition varies with in such a manner that the resonant frequency is also an exponential

function of x with a value of 5kcps at $x = 1$ cm. , and 500 cps at $x = 3$ cm.

That is,

$$f \cong e^{1.15(2.4-x)} \quad (\text{kcps}) \quad (\text{A-7})$$

$$\text{or } w(x) = 2\pi f = 10^5 e^{-1.15x} \quad (\text{radians/sec}) \quad (\text{A-8})$$

The "inductance" of the cochlear partition is, therefore, by Eqs. (6) and (3).

$$L(x) = \frac{1}{w^2(x)C(x)} \cong 9 \quad (\text{dyne-sec}^2/\text{cm}^4) \quad (\text{A-9})$$

The damping $R(x)$ is assumed to give everywhere a logarithmic decrement of about 0.5π so that the shunt admittance may be represented by

$$M(x) = \frac{1}{9} \frac{P}{p^2 + 0.5pw + w^2} \quad (\text{A-10})$$

where, of course, w is a function of x given by Eq. (8).

The transmission line equations for the cochlea are

$$\begin{aligned} \frac{dP}{dx} &= -Z \cdot I \\ \frac{dI}{dx} &= -M \cdot P \end{aligned} \quad (\text{A-11})$$

By differentiating the first equation and using the second to eliminate I , one obtains a second-order linear differential equation for the pressure

$$\frac{d^2P}{dx^2} - \frac{d \ln Z}{dx} \cdot \frac{dP}{dx} - \Sigma M P = 0 \quad (\text{A-12})$$

It should be noted in passing that this equation may also be written as

$$Z \left(\frac{P^1}{Z} \right)^1 - \Sigma M \cdot P = 0 \quad (\text{A-12a})$$

which, considering that $Z \sim 1/S_0$, is found to be of form identical to Petersen and Bogert's Eq. (16) for the transverse pressure mode:

$$\frac{1}{S_0} (S_0 F_-)' + \frac{\omega^2}{c^2} \left(1 + \frac{2b\rho_0 c^2}{S_0 Z} \right) P_- = 0 .$$

The two expressions are identical except for the "1" appearing in the second parenthesis which accounts for the compressibility of the fluid itself and, according to Fig. 3 from the Petersen Bogert paper, is at most but a few percent of the value, $\frac{2b\rho_0 c^2}{S_0}$. Hence its neglect in setting up the M of Eq. (A-12a) will yield negligible error, and we may consider our Eq. (A-12) to be equivalent for practical purposes to the Petersen-Bogert equation.

Our procedure will be to find from Eq. (A-12) an approximate solution for the pressure wave in response to a 1 dyne/cm^2 , 1000-cps pressure differential at the stapes. Then, in any infinitesimal length, dx , the shunt current will be $M \cdot P \, dx$; the volume displacement will be the time integral of this current; and the displacement $Y(x)$ of the partition will be this volume displacement divided by the area $b(x) \, dx$ of the partition-segment. Thus, the displacement of the membrane will ultimately be found from

$$Y(x) = \frac{1}{pb(x)} M(x) \cdot P(x) \quad (\text{A-13})$$

Now, by Eqs. (9) and (10), Eq. (12) may be written as

$$\frac{d^2 P}{dx^2} - 0.248 \frac{dP}{dx} - EM P = 0 \quad (\text{A-12b})$$

We define $\gamma = \sqrt{EM}$ and assume a solution of the form

$$P = P_0(x) e^{-\int_0^x \gamma dx} \quad (\text{A-12c})$$

in hope that by substituting Eq. (A-12c) into (A-12b) we shall obtain a new differential equation in $P_0(x)$ which will have a simpler solution. This substitution leads to the following equation in $P_0(x)$

$$\frac{d^2 P_0}{dx^2} - (2\gamma + 0.248) \frac{dP_0}{dx} + (0.248\gamma - \frac{d\gamma}{dx}) P_0 = 0 \quad (A-14)$$

There is some justification for assuming that P_0 will change so slowly with x that the second derivative term may be neglected. Our procedure will be to make this usual approximation and obtain an analytic expression for P_0 . However, we shall later justify this approximation by comparing the solution thus obtained with an 'exact' solution obtained from Eq. (A-12b) by numerical integration.

By ignoring the 2-nd derivative, we may write Eq. (A-14) as

$$\frac{\frac{dP_0}{dx}}{P_0} \approx \frac{0.248\gamma - \frac{d\gamma}{dx}}{2\gamma + 0.248} \quad (A-14a)$$

Now, over the range of x where $\gamma \gg 0.124$, Eq. (A-14a) may further be approximated by

$$\left(\frac{\frac{dP_0}{dx}}{P_0} \right) \approx 0.124 - \frac{1}{2} \left(\frac{\frac{d\gamma}{dx}}{\gamma} \right) \quad (A-14b)$$

which has as a solution

$$P_0 \approx C_0 e^{0.124x} \gamma^{-\frac{1}{2}} \quad (A-14c)$$

with C_0 an arbitrary constant to be determined.*

The expression (A-12c) for the pressure wave may now be written

$$P(x) = C_0 \gamma^{-\frac{1}{2}} e^{0.124x} - \int_0^x \gamma dx \quad (A-15)$$

and it remains to evaluate the integral $\int_0^x \gamma dx$

Since $\gamma = \sqrt{2M}$, it follows from Eqs. (A-4) and (A-10) that

$$\gamma = 2.79 e^{0.124x} \cdot \frac{p}{\sqrt{W}} = 9.63 \frac{p}{w^{0.108} \sqrt{W}} \quad (A-16a)$$

where

$$W = p^2 + 0.5pw + w^2 \quad (A-16b)$$

$$w = 10^5 e^{-1.15x} \quad (A-16c)$$

Replacing x by w in the integral, one finds

$$\int_0^x \gamma dx = 8.36 p \int_{w(x)}^{w(0)} \frac{dw}{w^{1.108} \sqrt{W}} \quad (A-17)$$

Integrating by parts,

$$\int_{w(x)}^{w(0)} [w^{-0.108}] \left[\frac{dw}{w \sqrt{W}} \right] = \left[w^{-0.108} \int \frac{dw}{w \sqrt{W}} + 0.108 \int w^{-1.108} \cdot \left[\frac{dw}{\sqrt{W}} \right] dw \right]_{w(x)}^{w(0)} \quad (A-18)$$

*Evaluation of $\gamma(x)$ as a fn of x shows that γ is 0.176 at the stapes but that by $x = 1.0$ cm, γ has increased to nearly 1 and reaches a magnitude of about 5. in the region of resonance so that the error of neglecting the constant term in the denominator of (A-14c) is significant only in the region near the stapes -- a region which is relatively unimportant in our study.

Over the range of x extending from the stapes to the point of resonance, w is large, the value of the first integral will be small, and the value of the second part of the bracket will be much smaller because of the extra factor $0.108/w$. Thus, only near the helicotrema ($x \rightarrow 3.5$ cm) will the contribution of the second part have appreciable value, and since we are interested primarily in the region near resonance (i.e. $x = 2.4$ cm), we shall neglect this second part. The first integral is easily found to be (Pierce 182).

$$\frac{dw}{w\sqrt{W}} = -\frac{1}{p} \log \left[\frac{\sqrt{W} + p}{w} + 0.25 \right] \quad (\text{A-19})$$

so that finally,

$$\int_0^x \gamma dx = 8.36 [Q(x) - Q(0)] \quad (\text{A-20})$$

$$\text{where } Q(x) = w^{-0.108} \log \left[\frac{\sqrt{W} + p}{w} + 0.25 \right] \quad (\text{A-21})$$

By absorbing the $Q(0)$ in the constant multiplier C_0 of Eq. (A-15), the approximate expression for the pressure at x becomes

$$P(x) \approx C_0 \gamma^{-\frac{1}{2}} e^{0.124x} - 8.36 Q(x) \quad (\text{A-22})$$

and the displacement $Y(x)$, by Eq. (A-13), becomes

$$Y(x) \approx \frac{5.85 e^{-0.3x}}{p^2 + 0.5pw + w^2} \cdot P(x) \quad (\text{A-23})$$

Eqs. (A-16), (A-21), (A-22) and (A-23) thus give an analytic expression for the magnitude and phase of the displacement of the basilar membrane at a point x which, because of the approximations involved in the derivations, must be remote from both stapes and helicotrema. Rather than attempt to analyze mathematically the errors involved in these approximations, we shall next carry through, by numerical

integration, an accurate solution of the pressure differential equation (A-12b). By comparing this accurate numerical solution with the results calculated from Eq. (A-22) we will be able to verify the range of x over which the approximations are valid and, furthermore, we shall be able to determine the constant multiplier C_0 appearing in Eq. (A-22).

Because Eq. (A-12b) yields a solution which oscillates, a numerical integration requires that much smaller increments be used than would be necessary were the solution non-oscillatory. This suggests that a change in variable should first be made such that the new variable is not oscillatory and changes more slowly than the $P(x)$ variable. A new variable (Γ) defined by

$$\Gamma(x) = -\log_e P(x) \quad (A-24)$$

has the desired properties. Its real part represents the amplitude of the oscillation expressed in nepers below some reference amplitude and its imaginary part is the phase of the oscillation in radians. It is apparent that $\Gamma(x)$ changes far more smoothly than $P(x)$ and, hence, a numerical solution for $\Gamma(x)$ will be easier to obtain.

Observing that

$$\begin{aligned} P &= e^{-\Gamma} \\ P' &= -\Gamma' P \\ P'' &= [-\Gamma'' + (\Gamma')^2] \cdot P \end{aligned} \quad (A-25)$$

we easily find that Eq. (a-12b) becomes

$$\Gamma'' - (\Gamma' + 0.248)\Gamma' + \Sigma M = 0 \quad (A-26)$$

where
$$ZM = 93 \frac{w^{-0.216} p^2}{p^2 + 0.5 pw + w^2} \tag{A-27}$$

Equation (A-26) is now a non-linear differential equation in T^1 , the integration of which may be easily carried out by numerical methods for any given frequency and boundary condition at the helicotrema. The significant details of this numerical solution for a 1000-cps. driving frequency are outlined in Appendix (B).

Figure A-1 compares the approximate solution for $|Y(x)|$ given by Eq. (A-23) with the "exact" solution obtained by numerical methods and Table II of Appendix A, for a frequency of 1000 cps.

Figure A-2 compares the phase lags of the two solutions. Both solutions are quite similar, particularly in the important region of resonance around $x = 24$ mm. where the amplitudes differ only by a nearly constant factor of about 2. As was remarked in the derivation of the approximate expression, errors were to be expected near the stapes and near the helicotrema. These are, in fact, the regions where appreciable errors do occur, as the figures show. However, the error is quite small in the important region of resonance, and we are therefore justified in using our approximate as an analytic expression for the vibration in this region.

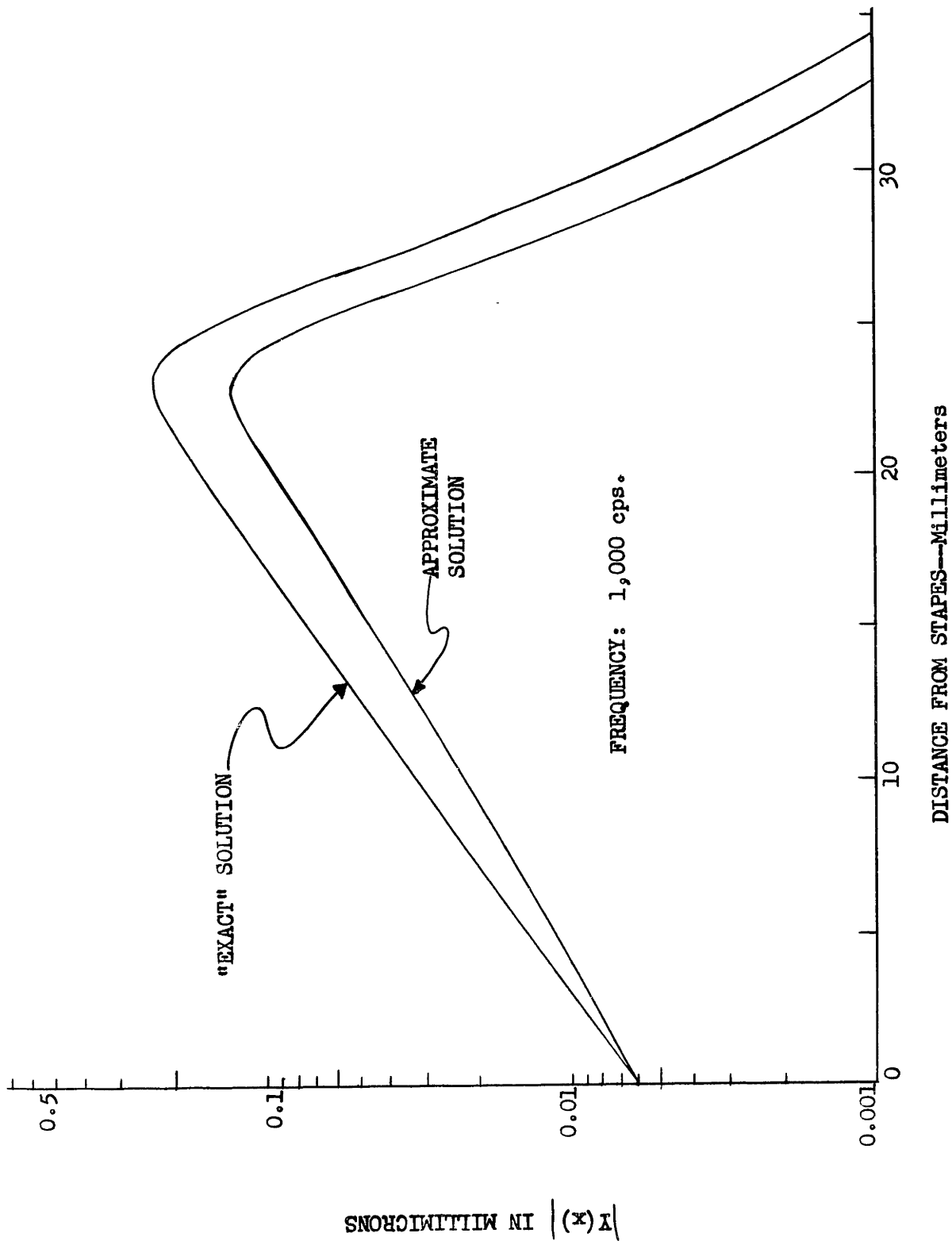


Fig. A-1. Comparison of $|Y(x)|$ as given by Eq. (A-23) with the "exact" solution calculated by numerical methods described in Appendix B.

DISTANCE FROM STAPES---Millimeters

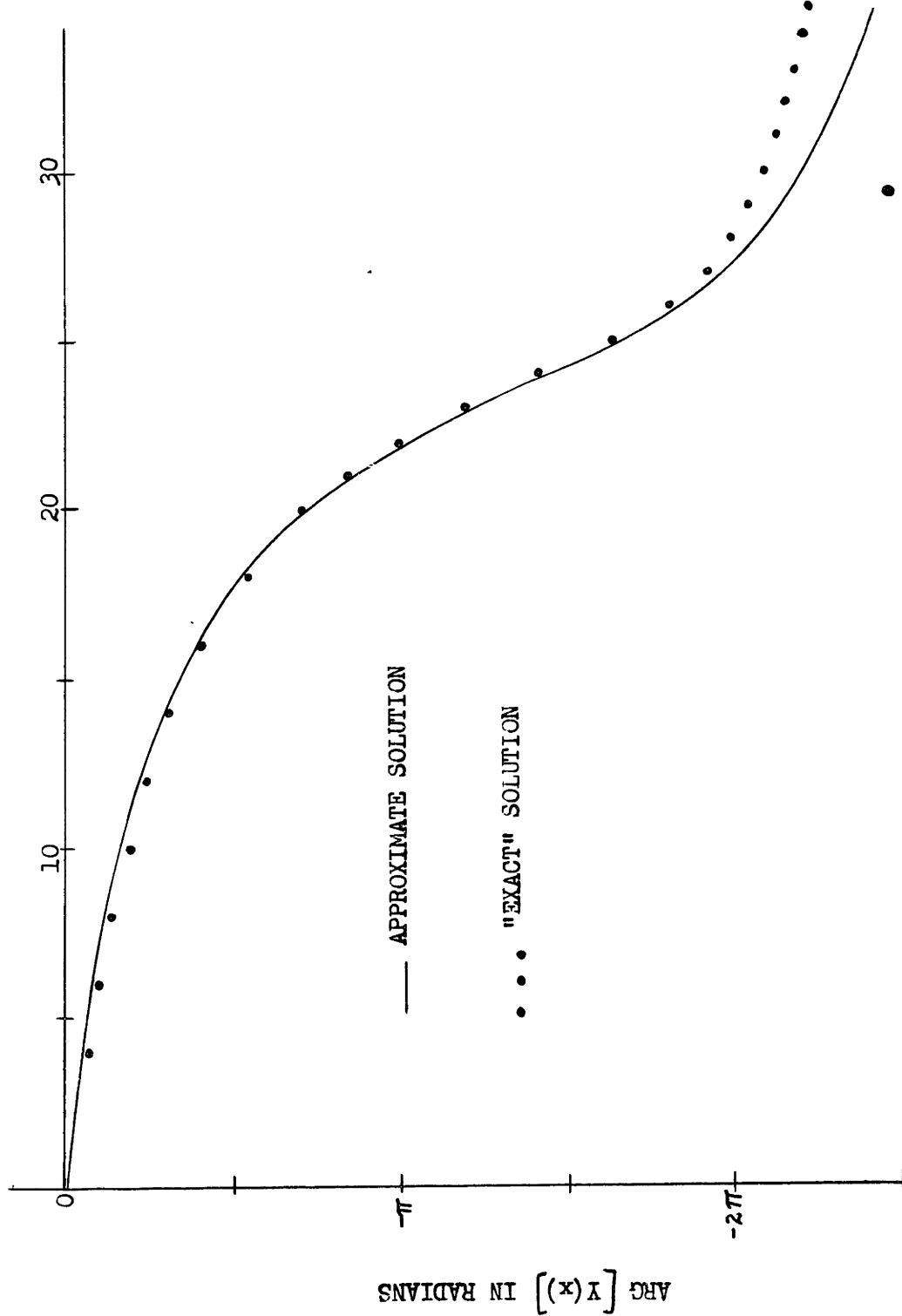


Fig. A-2. Phase data corresponding to the amplitude curves shown in Fig. A-1.

APPENDIX B

Solution of Equation (A-26) by Milne's Numerical Method

Equation (A-26) is a first-order, second-degree non-linear differential equation in $\gamma = \Gamma'$;

$$\gamma' - (\gamma + 0.248)\gamma + \epsilon M = 0. \quad (\text{A-26})$$

(Note that this γ represents a different quantity than the γ of Appendix A).

Provided the boundary value of γ at the helicotrema is specified, this equation may be integrated numerically in a step-by-step process starting at the helicotrema and working back towards the stapes.

The particular numerical method due to W. E. Milne (M-3) which is used here gives values of γ at discrete values of $x_n = nh$ which are separated by a small interval h . This is a step-by-step process using the formula

$$\gamma_{n+1} = \gamma_{n-3} + \frac{4h}{3} (2\gamma'_n - \gamma'_{n-1} + 2\gamma'_{n-2}) + \frac{28}{90} h^5 \gamma^{(5)} \quad (\text{B-1})$$

as a predictor, and Simpson's rule

$$\gamma_{n+1} = \gamma_{n-1} + \frac{h}{3} (\gamma'_{n+1} + 4\gamma'_n + \gamma'_{n-1}) - \frac{1}{90} h^5 \gamma^{(5)} \quad (\text{B-2})$$

as a corrector. The procedure is as follows: assuming that values

γ_{n-2} , γ_{n-1} , γ_n are known, the differential equation then gives values of the first derivative at these points: viz

$$\gamma' = (\gamma + 0.248)\gamma - \epsilon M \quad (\text{A-26})$$

From Eq. (B-1), the next value, γ_{n+1} , is predicted and the corresponding γ'_{n+1} computed from the preceding equation. Simpson's rule (B-2)

is then used to recompute the value of γ_{n+1} which should check closely with the predicted value provided the interval h is sufficiently small. This process is then repeated to determine γ_{n+2} , etc. The interval h is in practice so small that the 5th-derivative term $y^{(5)}$ appearing in (B-1) and (B-2) may be ignored.

In carrying through this computation, a rather interesting error of alternating sign arose which was unstable and because of its rapid growth threatened to destroy the solution. These alternations resulted in the values of γ_n for, say, n even being too large and for n odd being too small.

The cause of this difficulty is found in the predictor and corrector formulas which, it may be noted, use only the values of γ_{n-4} and γ_{n-2} in computing γ_n . Thus, the intermediate values γ_{n-1} and γ_{n-3} do not contribute in the same manner and the series of γ_n tends to disassociate into two interleaving series which are tied together only through the first derivatives γ'_n . Since the error tends to alternate, it is apparent that some sort of smoothing operation would reduce this error and damp out the instability.

A smoothing operator which is satisfactory for this purpose has been described by W. E. Milne and V. Rojansky. Assuming that the true value of the function is regular so that the fourth differences for equal arguments are zero; the sum of the squares of the errors may be reduced by $\frac{1}{2}$ by the following formula

$$\bar{\gamma}_{n-1} = \frac{1}{12} \left[-\bar{\gamma}_{n-3} + 4\bar{\gamma}_{n-2} + 6\gamma_{n-1} + 4\gamma_n - \gamma_{n+1} \right] \quad (B-3)$$

In operation, after γ_{n+1} has been predicted by Eq. (B-1), the smoothed value of $\bar{\gamma}_{n-1}$ is computed from Eq. (B-3), and it is this smoothed value

of $\bar{\gamma}_{n-1}$ which is used in Simpson's Rule (B-2) for correcting γ_{n+1} . Also, the values γ_{n-2} and γ_{n-3} might as well be the smoothed values determined on two previous steps. With this modification, the numerical process is perfectly stable and the error remained small throughout the range at integration.

The next point to be considered is the determination of the boundary condition at the helicotrema which we shall assume acts as an "inductive" orifice of 1 mm. length having the same inductance as the basilar membrane at that point but with an infinite elastance. Thus, the terminating admittance M_h representing the helicotrema is, from Eq. (A-10)

$$M_h = \frac{1}{9} \frac{p}{p^2 + 0.5wp} \quad (0.1) \quad (B-4)$$

For example, assuming a frequency of 1000 cps. or $p = j6,240$ and that the helicotrema corresponds to $x = 3.5$ cm., or $w(3.5) = 1,780$ from Eq. (A-8), we obtain

$$M_h = \frac{0.1}{9} \frac{1}{890 + j6,240} \quad \text{mhos}$$

as the terminating admittance. We must now translate this termination into the value of γ at $x = 3.5$ to establish the proper boundary value for our differential equation.

We note that the current at the helicotrema will be

$$I = M_h \cdot P \quad (B-5)$$

Now, from Eq. (A-11)

$$\frac{dP}{dx} = -g \cdot I = -g \cdot M_h \cdot P$$

But from Eq. (A-25), $\frac{dP}{dx} = -\gamma \cdot P$ so that

$$\gamma = \Gamma' = Z \cdot M_h \quad (\text{B-6})$$

This equation may be used to establish the value of γ at the helicotrema.

Setting $x = 3.5$ cm. and $f = 1000$ cps., we find

$$\gamma(3.5) = 1.81 / \underline{8.1^\circ} \quad \text{nepers/cm.}$$

In order to start the numerical method of integrating the differential equation, the first four consecutive values of γ must be determined. The method described by W. E. Milne (ref.M-3) was used here to find values of γ , γ' at $x = 3.55, 3.50, 3.45,$ and 3.40 , after which the step-by-step process was applied with $h = 0.05$ cm up to $x = 2.65$. In the region at resonance from $x = 2.65$ to $x = 2.15$ a smaller interval $h = -0.025$ cm. was used, and from $x = 1.40$ to $x = 0$, $h = 0.1$ cm. was sufficiently small. Integrating these values of $\gamma = \Gamma'$ by Simpson's rule gave the values for Γ indicated in Table B-1.

$x(\text{mm})$	$ \Gamma (\text{nepers})$	$\angle \Gamma(\text{radians})$	$x(\text{mm})$	$ \Gamma (\text{nepers})$	$\angle \Gamma(\text{radians})$
0	0.000	0	23	1.618	2.614
2	0.060	0.058	24	1.956	2.922
4	0.124	0.128	25	2.368	3.180
6	0.196	0.212	26	2.814	3.376
8	0.272	0.314	27	3.264	3.516
10	0.356	0.438	28	3.708	3.626
12	0.448	0.594	29	4.146	3.710
14	0.554	0.786	30	4.576	3.780
16	0.676	1.032	31	5.002	3.838
18	0.830	1.348	32	5.416	3.892
20	1.038	1.762	33	5.810	3.938
21	1.182	2.016	34	6.154	3.982
22	1.368	2.302	35	6.404	4.016

Table B-1. Attenuation and Phase lqg of Pressure wave at 1,000 cps.

To find the displacement $Y(x)$ at the point x , Eq. (A-13) may be evaluated to obtain

$$Y(x) = \frac{1}{pb(x)} \cdot M(x) \cdot P(x) \quad (B-13)$$

$$= - \frac{1.4818 \times 10^{-7}}{1 + \left(\frac{w}{p}\right)^2 + 0.5 \frac{w}{p}} e^{-0.3x - \Gamma} \quad (B-13a)$$

Here, also, it is convenient to express $Y(x)$ in logarithmic units;

$$\log Y(x) = \log \left\{ - \frac{1.482 \times 10^{-7}}{1 + \left(\frac{w}{p}\right)^2 + 0.5 \frac{w}{p}} \right\} - 0.3x - \Gamma \quad (B-13b)$$

in which the real part represents the amplitude of the oscillation in nepers below 1 cm. and the imaginary part is the phase of the displacement relative to the pressure at the stapes. Evaluating this expression at 1000 cps. and using values for x and Γ tabulated in Table B-1 gives values tabulated in Table B-2.

x	$\ln Y(x) $	$ Y(x) $	$\angle Y(x)$	x	$\ln Y(x) $	$ Y(x) $	$\angle Y(x)$
0	-5.148	0.0058	0.01 π	23	-1.453	0.234	1.19 π
2	-4.803	0.0082	0.03 π	24	-1.598	0.202	1.42 π
4	-4.458	0.0116	0.06 π	25	-2.010	0.134	1.64 π
6	-4.130	0.0161	0.09 π	26	-2.580	0.076	1.81 π
8	-3.794	0.0225	0.13 π	27	-3.180	0.0416	1.92 π
10	-3.464	0.0314	0.17 π	28	-3.762	0.0232	1.99 π
12	-3.135	0.0436	0.23 π	29	-4.316	0.0133	2.05 π
14	-2.809	0.0603	0.30 π	30	-4.843	0.0079	2.10 π
16	-2.476	0.0841	0.40 π	31	-5.351	0.0047	2.13 π
18	-2.140	0.1178	0.54 π	32	-5.835	0.0029	2.16 π
20	-1.800	0.165	0.71 π	33	-6.291	0.0018	2.19 π
21	-1.640	0.194	0.84 π	34	-6.689	0.0012	2.21 π
22	-1.506	0.222	0.99 π	35	-6.989	0.0009	2.24 π

Table B-2. Amplitude and phase log of displacement wave at 1,000 cps. Displacement is expressed in millimicrons (i.e. 10^{-9} meters).

APPENDIX C

Summary of Equations for $W(x) = \frac{d^4Y}{dx^4}$

In Appendix A, an approximate analytic expression for the displacement $Y(x)$ was obtained. Here, we wish to summarize the steps involved in differentiating $Y(x)$ four times with respect to x , the end result of which was an approximate analytic representation for the outer-hair-cell-force density $W(x)$.

Reference is made to Eq. (A-15) through (A-23) of the first appendix for definitions of symbols. The various derivatives of $Y(x)$ may be expressed as the product of two factors; one being simply $Y(x)$ itself, and the other indicating the sharpening action that results from the differentiation process. This factor, for the first derivative, will be represented by ξ , and it is defined as the first derivative of $\log Y$ with respect to x . Thus, since

$$\frac{d}{dx}(\log Y) = \xi = \frac{\frac{dY}{dx}}{Y}$$

we have,

$$\frac{dY}{dx} = \xi Y \tag{C-1}$$

By repeatedly differentiating this equation, one obtains the following expressions for the higher derivatives

$$\frac{d^2Y}{dx^2} = \left[\xi^2 + \frac{d\xi}{dx} \right] \cdot Y \tag{C-2}$$

$$\frac{d^3Y}{dx^3} = \left[\xi^3 + 3\xi \frac{d\xi}{dx} + \frac{d^2\xi}{dx^2} \right] \cdot Y \tag{C-3}$$

$$\frac{d^4Y}{dx^4} = \left[\xi^4 + 6\xi^2 \frac{d\xi}{dx} + 4\xi \frac{d^2\xi}{dx^2} + 3\left(\frac{d\xi}{dx}\right)^2 + \frac{d^3\xi}{dx^3} \right] \cdot Y \tag{C-4}$$

Equation (C-4) may be used to obtain an expression for $W(x)$ provided we can express ξ and its derivatives in terms of the frequency (p) and place (x) variables. To do this we first normalize certain of the variables by introducing the quantities z and D defined by

$$z = p/w, \quad D = W/w^2 \quad (C-5)$$

so that equation (A-16a) becomes

$$\gamma = 2.79 e^{0.124x} \frac{z}{\sqrt{D}} \quad (C-6)$$

and ξ may be closely represented by

$$= -0.18 + 1.38 \left(\frac{2+0.5z}{D} \right) - \gamma \quad (C-7)$$

Next,

$$\frac{d\xi}{dx} = A - \frac{d\gamma}{dx} \quad (C-8)$$

$$\text{where } A = -0.795 \frac{z}{D^2} [1 + z^2 + 8z] \quad (C-8a)$$

$$\frac{d\gamma}{dx} = \delta \cdot \gamma \quad (C-8b)$$

$$\delta = 0.124 + 0.575 \left(\frac{2+0.5z}{D} \right) \quad (C-8c)$$

Then,

$$\frac{d^2\xi}{dx^2} = B - C \cdot A - \frac{d^2\gamma}{dx^2} \quad (C-9)$$

$$\text{where } B = \frac{-0.914z(1+3z^2+16z)}{D^2} \quad (C-9a)$$

$$C = \frac{2.30z(2z+0.5)}{D} \quad (C-9b)$$

$$\frac{d^2\gamma}{dx^2} = \left[\frac{d\delta}{dx} + \delta^2 \right] \gamma \quad (C-9c)$$

$$\frac{d\delta}{dx} = \frac{0.661Z}{D} \left[0.5 - \frac{2+0.5Z}{D} \right] \quad (C-9d)$$

Finally,

$$\begin{aligned} \frac{d^3\xi}{dx^3} = & - \frac{1.05 Z(1+9Z^2+32Z)}{D^2} - C B + \\ & - 2.65 Z \left[\frac{4Z+0.5}{D} - \frac{Z(2Z+0.5)^2}{D^2} \right] \cdot A - \\ & - \left[\frac{d^2\xi}{dx^2} + \frac{d^2\gamma}{dx^2} \right] C - \frac{d^3\gamma}{dx^3} \quad (C-10) \end{aligned}$$

where

$$\frac{d^3\gamma}{dx^3} = \left(\frac{d^2\delta}{dx^2} + 3\delta \frac{d\delta}{dx} + \delta^3 \right) \gamma \quad (C-10a)$$

$$\frac{d^2\delta}{dx^2} = - \frac{2.30Z}{D} \left(\frac{d\delta}{dx} \right) + \frac{0.38Z}{D^2} (-3+3Z^2-Z). \quad (C-10b)$$

When these values are substituted into Eq. (C-4) numerical values for $W(x)$ are obtained. The results of the computations for $Y(x)$, $\frac{d^2Y}{dx^2}$ and $\frac{d^4Y}{dx^4}$ are given in Table C-1 for the case of a 1,000-cps. tone.

x	Y(x)		$\frac{d^2Y}{dx^2}$		$\frac{d^4Y}{dx^4}$	
	Amplitude	Phase	Amplitude	Phase	Amplitude	Phase
0	0.0058	-0.01 π	0.000392	-0.07 π	0.000029	-0.20 π
5	0.0111	-0.06 π	0.000806	-0.18 π	0.000077	-0.39 π
10	0.0217	-0.153 π	0.00183	-0.368 π	0.000254	-0.65 π
15	0.0462	-0.335 π	0.00576	-0.680 π	0.00147	-1.03 π
18	0.0740	-0.515 π	0.0161	-0.975 π	0.00777	-1.38 π
20	0.0985	-0.720 π	0.0334	-1.279 π	0.0272	-1.75 π
21	0.113	-0.858 π	0.0521	-1.488 π	0.0511	-2.06 π
22	0.128	-1.023 π	0.0885	-1.779 π	0.1610	-2.42 π
23	0.127	-1.227 π	0.128	-2.127 π	0.350	-2.94 π
24	0.107	-1.462 π	0.137	-2.598 π	0.470	-3.63 π
25	0.0673	-1.69 π	0.0782	-3.06 π	0.230	-4.38 π
26	0.0374	-1.86 π	0.0328	-3.31 π	0.0685	-4.90 π
27	0.0200	-1.98 π	0.0127	-3.64 π	0.0178	-5.26 π
28	0.0112	-2.08 π	0.00613	-3.78 π	0.0065	-5.46 π
30	0.0035	-2.22 π	0.00121	-4.04 π	0.00068	-5.82 π
32	0.0012	-2.32 π	0.000336	-4.19 π	0.000131	-6.02 π
35	0.0002	-2.44 π	0.000049	-4.35 π	0.000013	-6.23 π

Table C-1. Calculated values of the displacement and its second and fourth place derivatives. (Y is in millimicrons and x is in millimeters).

APPENDIX 7

A Neural Sharpening Mechanism

In the text, a mechanism for exciting the outer hair cells has been described which creates microphonic potentials that change in phase by roughly 100 degrees per millimeter increase in distance from the stapes. At the same instant that certain of the hair cells in the region of resonance are carrying a maximum positive potential at one place, the potential may have comparable negative values at adjoining places spaced only 1.5 millimeters on either side.

Consider, then, a nerve fiber that runs through these positive and negative regions. The simplest model of a nerve fiber, consists of a conducting core covered by an insulating sheath across which a d-c potential is developed. In the passive condition, the interior of the fiber takes on a negative potential of the order of tens of millivolts with respect to the outer medium. When the nerve fiber is stimulated at some point, by an externally applied cathodic current, a complex action sets in which, apparently, is little understood. But in effect, the membrane at that point of stimulation appears to break down electrically and positive current flows from the external surrounding region into the core of the fiber. This causes a decrease in the external potential of the adjoining regions of the nerve fiber. But the reaction of the membrane appears to be such that a decrease (i.e. negative change) in potential across the membrane increases its excitability and will lead to further breakdown. Thus, from the point of initial stimulation, two electrochemical waves propagate in both directions along the fiber.

By observing the potential at some point along the fiber with a sensitive amplifier and oscilloscope, one can establish the time pattern of these waves. It is found that they travel with a velocity which in

meters per second is roughly six times the diameter of the fiber in microns. Conduction velocities lying between 10 and 100 meters per second are easily obtained in mammalian nerve fibers.

Now, the velocities with which a wave is propagated along the cochlear portion lie in this same range. For example, Békésy (S-2, pg 1101, Fig. 30) gives data showing that when the stapes is suddenly displaced, a bulge travels along the partition toward the helicotrema, reaching the point $x = 2$ cm. in about 10^{-4} seconds and the helicotrema in about $2 \cdot 10^{-2}$ seconds. These correspond to average velocities in the order of 200 and 2 meters per second.

If it is assumed that the induced negative potential is a causative agent in the propagation of an action potential along a nerve fiber, then it would appear that an external source of negative potential waves that travel with the proper velocity along the fiber would tend to establish across the nerve membrane something of the same potential conditions that occur under actual propagation of a nerve spike. This type of external field could conceivably be a very effective stimulus for the nerve fiber, as had already been suggested by Ranke. (R-1).

In the absence of specific facts as to how these potential fields do, in fact, stimulate the nerve fiber, we may take as a working hypothesis that the effective excitation of any point is determined by the voltage across the insulating sheath. It is possible to estimate what this voltage will be for a nerve fiber running along the outer hair cells which are creating a local potential wave $W(x)$ previously defined.

As illustrated in Fig. D-1, the non-conducting fiber is assumed to consist of a core conductor; having resistance R per unit length and covered by an insulating dielectric sheath of capacitance C per unit

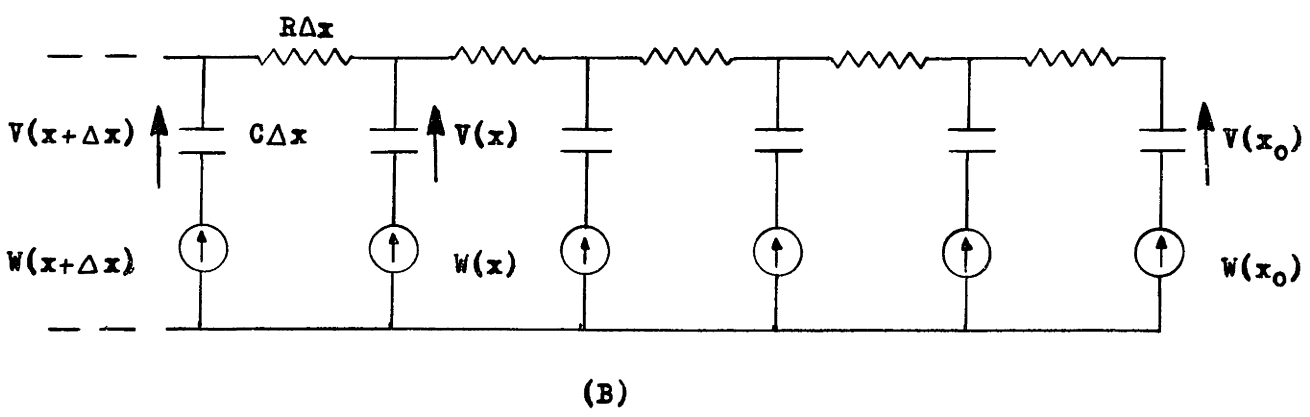
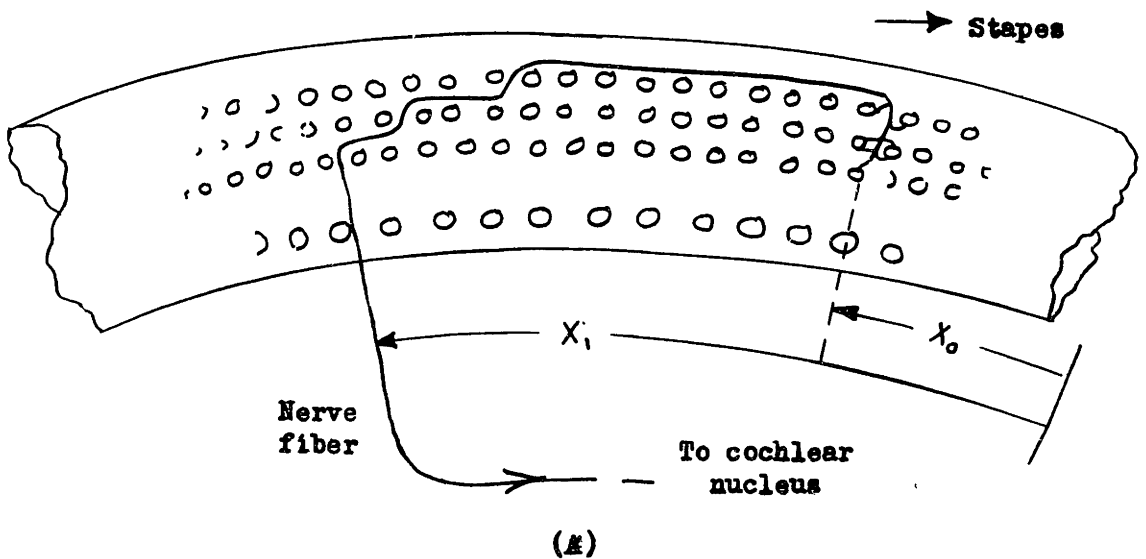


Fig. D-1. (A). The path followed by a single nerve fiber innervating the outer hair cells. (B) A simple model of a nerve fiber.

length. The potential $w(x)$ is impressed upon the outside of the nerve sheath and, because of the conduction of the core, a voltage $v(x)$ is developed across the sheath at the point x . By writing the difference equations for the voltage wave and letting $\Delta x \rightarrow 0$, it is a simple matter to show that the voltage across the sheath is related to the external potential by the partial linear differential equation.

$$-\frac{\partial^2 w}{\partial x^2} = \frac{\partial^2 v}{\partial x^2} = RC \frac{\partial v}{\partial t} \quad (D-1)$$

If we assume sinusoidal time variation, represented by $e^{j\omega t}$, the above equation becomes

$$-\frac{d^2 w(x)}{dx^2} = \frac{d^2 v(x)}{dx^2} - j\omega RC v(x) \quad (D-2)$$

and gives us a relation between the force density $w(x)$ and the sheath voltage distribution $v(x)$ for which we now solve D-2 explicitly.

If we assume that $-\frac{d^2 w(x)}{dx^2}$ vanishes at all x except at the point x_0 where it is equal to a unit impulse $\mathcal{U}(x - x_0)$, we may find the Green's function (i.e. space "system function" of the nerve) $n(x)$ from which we may by superposition, through the use of the convolution integral, obtain the resultant $v(x)$ for any given $-\frac{d^2 w(x)}{dx^2}$. This space weighting

function must satisfy (D-2), with $-\frac{d^2 w}{dx^2} = 0$, for all points except

at $x = x_0$ where the right hand side must reduce to $\mathcal{U}(x - x_0)$. That is, at $x_0 = x_0$, there will be a cusp in the distribution of $\frac{d^2 v}{dx^2}$. Solving

the homogeneous equations and invoking the physical requirement that $v(x)$ must die away as one recedes from x_0 , one obtains

$$n(x) = \frac{-1}{2\lambda(1+j)} e^{-\lambda(1+j)(x-x_0)}, \quad (x > x_0)$$

and

$$n(x) = \frac{-1}{2\lambda(1+j)} e^{\lambda(1+j)(x-x_0)}, \quad (x < x_0). \quad (D-3)$$

where $\lambda = \sqrt{\frac{\omega RC}{2}}$ represents both the attenuation in nepers and phase in radians per unit change in x .

Now, by superposition,

$$V(x) = \int_{-\infty}^{\infty} \frac{\partial^2 W}{\partial \xi^2} n(x-\xi) d\xi \quad (D-4)$$

where we assume, for simplicity, that the fiber is of infinite length so that the integrand vanishes at both limits. (The modification for finite-length fibers will be discussed later). Equation D-4 may then be integrated by parts, to obtain

$$V(x) = \frac{dW}{dx_0} \cdot n(x-\xi) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{dW}{d\xi} \cdot \frac{dn}{d\xi} \cdot d\xi \quad (D-5)$$

The first term vanishes. The second term may be split into two integrals,

$$V(x) = - \int_{-\infty}^x \frac{dW}{d\xi} \frac{dn}{d\xi} \cdot d\xi + \\ - \int_x^{\infty} \frac{dW}{d\xi} \frac{dn}{d\xi} \cdot d\xi, \quad (D-6)$$

where

$$\frac{dn}{d\xi} = \frac{1}{2} e^{-\lambda(1+j)(\xi-x)}, \quad (\xi > x_0)$$

$$= -\frac{1}{2} e^{\lambda(1+j)(\xi-x)}, \quad (\xi < x_0).$$

Integrating Eq. (D-6) again by parts, one finds that

$$V(x) = -W \frac{dn}{d\xi} \Big|_{-\infty}^x - W \frac{dn}{d\xi} \Big|_x^{\infty} + \int_{-\infty}^{\infty} W \frac{d^2n}{d\xi^2} d\xi \quad (D-7)$$

The first two terms in Eq. (D-7) are each equal to $W(x)/2$ which may finally be written as

$$V(x) = W(x) - \frac{\lambda(1+j)}{2} \int_{-\infty}^{\infty} W(\xi) e^{-\lambda(1+j)|x-\xi|} d\xi \quad (D-8)$$

The voltage across the sheath at the point x is thus seen to be the difference between the external potential $W(x)$ and the internal potential, which consists of contributions arising from all parts of the surface potential, each little part $W(\xi)d\xi$ being weighted by the factor

$$\frac{\lambda(1+j)}{2} e^{-\lambda(1+j)d} \quad \text{where } d \text{ is the distance from that part to}$$

the point in question. With this simple interpretation, we may now write down the expression for the voltage at the point x when the fiber terminates in an open circuit at the point x_0 , as shown in Fig. D-1. The open circuit creates a reflected voltage wave which at the point x_0 is just equal in value to the incident wave. Thus, in addition to the direct induced potential indicated by Eq. (D-8), there will be another component due to this reflected wave. The total path over which the reflected wave travels^{is} from ξ to x_0 and back to x or a total distance at $d = 2x_0 - x - \xi$. Thus, the voltage at the point x will be

$$V(x) = W(x) - \frac{\lambda(1+j)}{2} \int_{x_1}^{x_0} W(\xi) \left[e^{-\lambda(1+j)|x-\xi|} + e^{-\lambda(1+j)(2x_0-x-\xi)} \right] d\xi \quad (D-9)$$

where we have utilized the fact indicated in Fig. D-1 that $W(\xi)$ vanishes outside of the range $x_1 < \xi < x_0$.

At the terminal buds of the nerve fiber, $x = x_0$, and the voltage across the sheath is

$$V(x_0) = W(x_0) - \lambda(1+j) \int_{x_1}^{x_0} W(\xi) e^{-\lambda(1+j)(x_0-\xi)} d\xi \quad (D-10)$$

The integration required by Eq. (D-10) was performed numerically using Simson's rule and the data for $W(x)$ from Table 1 of Appendix C to obtain the voltage across the end twigs for nerve fibers that run basalward along the outer hair cells for distances $x_1 - x_0$ of 0.8 and 1.6 millimeters. The results are shown in Fig. D-2, which compares the amplitudes $|V(x)|$ with $|W(x)|$ and in Fig. D-3 which compares the phases of $V(x)$ and $W(x)$.

It is noted that basalward from the region of resonance, the phase continues to change more rapidly than does the phase of $W(x)$, and apicalward, the response falls off more rapidly. This greater selectivity in place is consistent with the fact that, by Eq. (D-2) when $\omega RC \gg \frac{d^2V}{dx^2} / V$, the voltage $V(x)$ across the sheath is essentially equal to the curvature $\frac{d^2W}{dx^2}$ of the external potential field. Hence, this effect may produce the equivalent, in conjunction with the beam hypothesis, of a neural excitation proportional to the sixth-place derivative of the displacement.

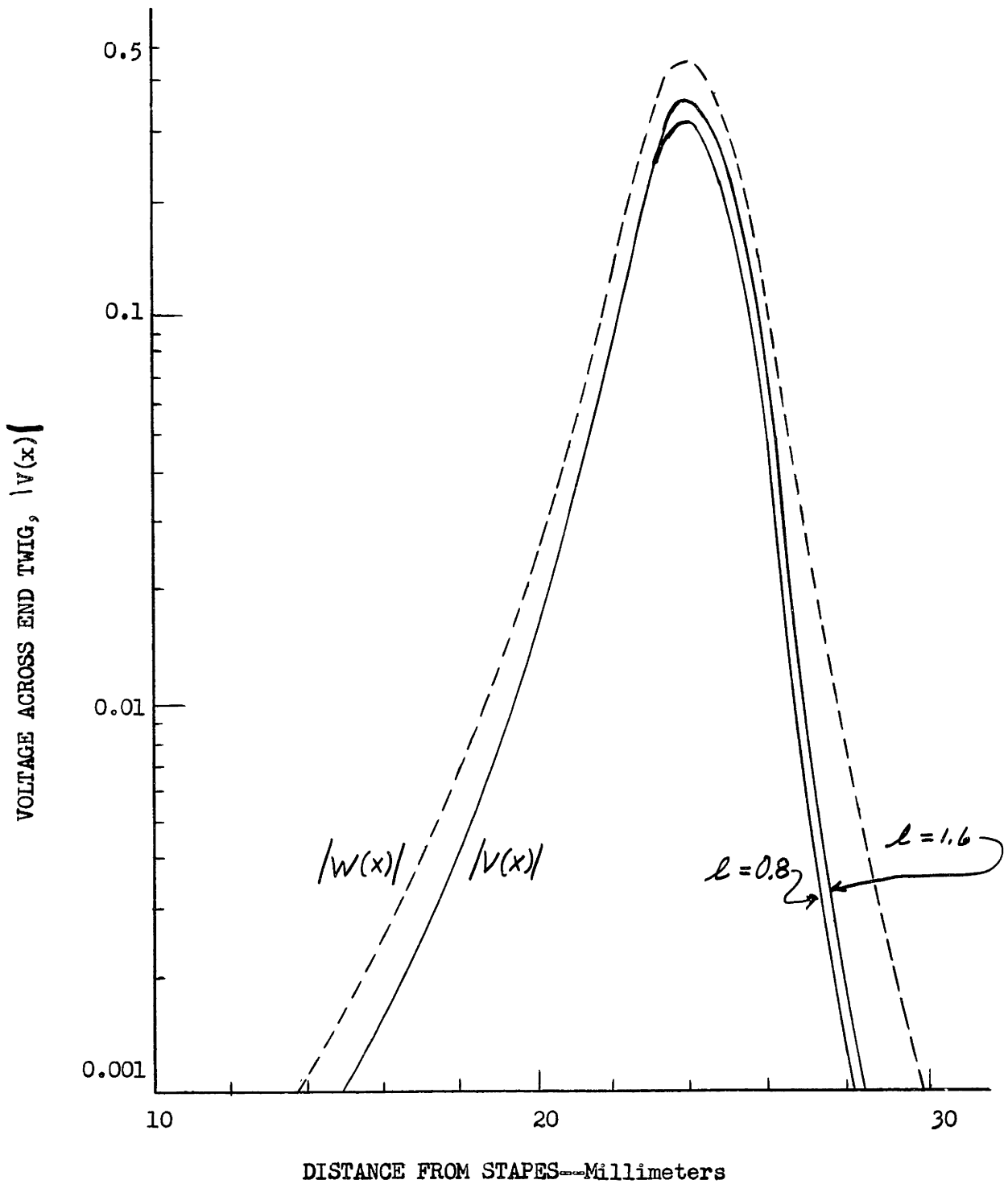


Fig. D-2. Distribution of voltage amplitude $|V(x)|$ across the end twigs of the outer-hair-cell neurons which run basalward for a distance of l millimeters. The external potential distribution $|W(x)|$ is shown for comparison. 1,000-cps. tone.

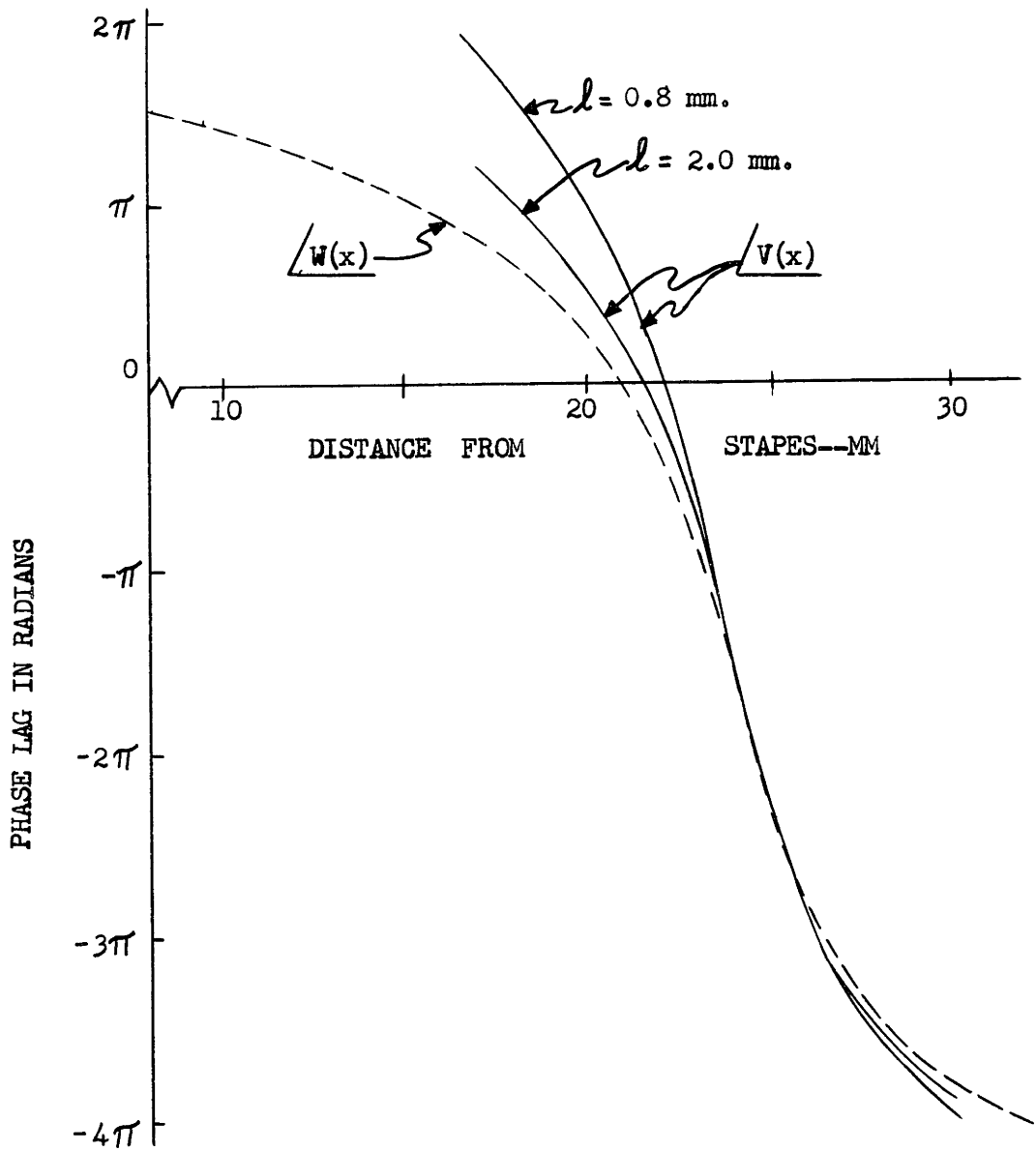


Fig. D-3. Phase data associated with the amplitude curves of Fig. D-2.

REFERENCES

- B-1 Békésy, G. V. "The vibration of the cochlear partition in partition in anatomical preparations and in models of the inner ear", J. Acoust. Soc. Am. 21, 233-245, (1949).
- B-2 Békésy, G. V. "On the resonance curve and the decay period at various points on the cochlear partition" J. Acoust. Soc. Am. 21, 245-254 (1949).
- B-3 Békésy, G. V. "Gross localization of the place of origin of the cochlear microphonics," J. Acoust. Soc. Am. 24, 399-409 (1952).
- B-4 Bogert, B. P. "Determination of the effects of dissipation in the cochlear partition by means of a network representing the basilar membrane," J. Acoust. Soc. Am. 23, 151-154 (1951).
- D-1 Davis, Fernandez, and McAuliffe, Proc. Nat. Acad. Sci. 36, 580 (1950)
- D-2 Dwork, B. M. "Detection of a pulse superimposed on fluctuation noise," Proc. IRE, 38, 771-774 (1950).
- F-1 Fletcher, H. "The dynamics of the middle ear and its relation to the acuity of hearing," J. Acoust. Soc. Am., 24, 129-131.
- G-1 Galambos, R. and Davis, H. "Inhibition of activity in simple auditory nerve fibers by acoustic stimulation" J. Neurophysiol., 7, 287-303, (1944).
- G-2 Galambos, R. and Davis, H. "Action potentials from single auditory nerve fibers?" Science 108 513 (1948).
- H-1 Huggins, W. H. "System function analysis of speech sounds" J. Acoust. Soc. Am. 22, 765-767 (1950)
- H-2 Huggins, W. H. and Licklider, J. C. R. "Place mechanisms of auditory frequency analysis" J. Acoust. Soc. Am. 23, 290-299 (1951)

- H-3 Huggins, W. H. "Theory of cochlear frequency discrimination"
Quarterly Progress Report for October 15, 1950, Research
Laboratory of Electronics, Mass. Inst. of Technology, pp 54-59.
- K-1 Klumpp, R. G. "Discriminability of Interaural time difference",
paper presented at the forty-fifth meeting of the Acoust. Soc. of
America, May 7, 8 and 9, 1953, Philadelphia, Penn.
- L-1 Licklider, J. C. R. "Annual Review of Psychology" vol 4, Annual
Reviews, Stanford, California, (1953).
- L-2 Licklider, J. C. R. "A duplex theory of pitch perception"
Experientia, 7, 128-134 (1951).
- M-1 Mathes, R. C. and Miller, R. L. "Phase effects in monaural
perception" J. Acoust. Soc. Am. 19, 780-797, (1947).
- M-2 Middleton, D., J. Appl. Phys. 22, 1143 (1951).
- M-3 Milne, W. E. "Note on the numerical integration of differential
equations" Amer. Math. Monthly, 48, 52-53 (1941).
- P-1 Potter, R. K. et al. "Technical aspects of visible speech,"
J. Acoust. Soc. Am. 17, 1-39, (1946).
- P-2 Peterson, L. C. and Boert, B. P. "A dynamical theory of the
cochlea," J. Acoust. Soc. Am. 22, 369-381 (1950).
- R-1 Ranke, O. F. "Theory of operation of the cochlea: A contribution
to the hydrodynamics of the cochlea" J. Acoust. Soc. Am. 22, 772-777
(1950).
- V-1 de Vries, H. L., Jlelof, R., Spoor, A. and Kuiper, J. W. "The
microphonic activity of the lateral line" J. Physiol. 116, 137-157
(1952). Also, "Brownian motion and the transmission of energy in
the cochlea" J. Acoust. Soc. Am., 24 527-533 (1952).

- W-1 Weaver, W. G. "Theory of Hearing", John Wiley and Sons, New York,
N.Y. (1949)
- Z-1 Zwislocki, J. "Theory of the acoustical action of the cochlea",
J. Acoust. Soc. Am. 22, 778-784 (1950)

BIOGRAPHICAL NOTE

W. H. Huggins was born 11 January 1919 at Rupert, Idaho, where he lived until his family moved to Eugene, Oregon, in 1930. He attended Oregon State College from 1937 to 1942, obtaining a B.S. degree in Communications Engineering in June 1941 and a M.S. degree in Power Engineering one year later. His thesis dealt with the problem of precipitation static and was entitled, "The Radio Influence Characteristics of High-Voltage Electrical Discharges."

After teaching for a year in the Army Specialized Training Program at Oregon State College, Mr. Huggins joined, in April of 1944, the wartime staff of the Radio Research Laboratory at Harvard University where he was engaged in the development of wide-tuning-range microwave oscillators.

Since 1946, he has been employed by the Air Force Cambridge Research Center, working first on receiver circuits and later on electronics plans. He is currently organizing a program directed toward the application of computer techniques to Air Force problems involved in the handling of information.

He is a member of Sigma Xi, RESA, Tau Beta Pi, Eta Kappa Nu, Pi Mu Epsilon, A.A.A.S., and is the recipient of the Browder J. Thompson Memorial prize awarded by the Institute of Radio Engineers in 1947.