Cochlear Modeling

Jont B. Allen

INTRODUCTION

A PROBLEM of fundamental interest to the speech processing community has been the subject of human communications via the transmission of acoustic information. This includes the production of the sound at the glottis, the formation of the speech code via the manipulation of the vocal tract, the subsequent propagation of the sound in the environment, digital speech coding and bandwidth reduction, algorithms to remove noise and reverberation at a receiving microphone, and speech recognition and understanding via extensive and sophisticated signal processing means. With only a few exceptions, most aspects of this process have been explored and exploited, time and time again. The successes of the FFT in speech processing, and the broad applicability of linear prediction to speech modeling, speech coding, and speech recognition, are abundant in IEEE Acoustic, Speech, and Signal Processing publications.

A common thread in this body of work are the principles, to the extent that we understand them, of the speech production model [13]. By this I mean the physical principles of the human production of the speech signal, namely transmission line theory using an acoustical transmission line. The success of linear prediction theory, as is well known, is largely due to the natural fit of the all pole model to the speech signal during voicing, which in turn is a result of the fact that a transmission line, driven at its end, has no zeros in its transfer function.

Models of the vocal tract provide us with the concept of a formant frequency, which is a natural resonance (eigenfrequency) of the vocal tract. The collusion of these various formants, as a function of time, is known to code much of the "information" that we perceive as the voiced speech signal.

What in fact actually happens to "information" as it is transferred from one person to another, in the form of an acoustic signal? Of course, there can be no exact answer to this question. However, based on an analysis of the physical representations of the signals at various points in the transmission path, it is possible to describe this process in some detail.

Through analysis of speech signals by signal processing methods, one may explain many aspects of the speech generation process. In an analogous manner, one may follow the speech signal through the auditory system as it is processed by the cochlea. In this article we shall do this to help explain how the information is carried and transformed on its way to the brain of the listener. We shall also attempt to describe the physical processes that take place along this locus. In doing this, hopefully the reader will gain an improved appreciation for the monaural perceptual aspects of human speech communication, to the extent that we presently understand them.

The paper is organized as follows. First we give a brief description of the function of the external ear and a description of degradations due to room reverberation. Next we present a simple model of the middle ear (see Fig. 1) with electrical analogues to the mechanical system. Using the model we discuss the input impedance of the cochlea, the loading of the middle ear by the cochlea, and the transfer function of the middle ear.

We then discuss the cochlea, or inner ear, the organ that converts acoustical signals into neural signals. The analysis of the cochlea is sub-divided into macromechanics, micromechanics, and transduction. Macromechanics is a well understood problem. Micromechanics, on the other hand, presents many unanswered questions. Thus this area is treated in greater detail and several recent models are described. Because of the uncertainties associated with cochlear micromechanics, we then discuss experiments on evoked echos and spontaneous cochlear emissions. These data exemplify the complexity of the problems that we face in hearing science today.

Next a model of transduction is described which matches many aspects of experimental neural data. Transduction is important because the cochlear detectors are nonlinear and strongly modify the properties of the signal.

Following the model description, we show the cochlear transformation on various signals, such as a tone burst in noise, and speech signals, as they pass through the complete model. We then describe masking-like phenomena which result from the nonlinearities of the hair cell.

THE EXTERNAL EAR

The effects of the external ear have been studied by Shaw and his colleagues [42], aided in part by measurements on physical models of artificial pinnas. It appears that we are able to use spectral cues introduced by the pinnas (e.g., zeros in the pinna transfer function), as shown in Fig. 2, to help us determine the elevation of the source of the incoming sound. Lateralization perception is believed to be derived from amplitude and delay differences between the ears plus cues obtained from head motions.

Animal studies of the spatial acuity of many species have been quite revealing, along with other physical data that have been collected over the years. Particularly interesting species are the bat, the barn owl, and the echo-locating

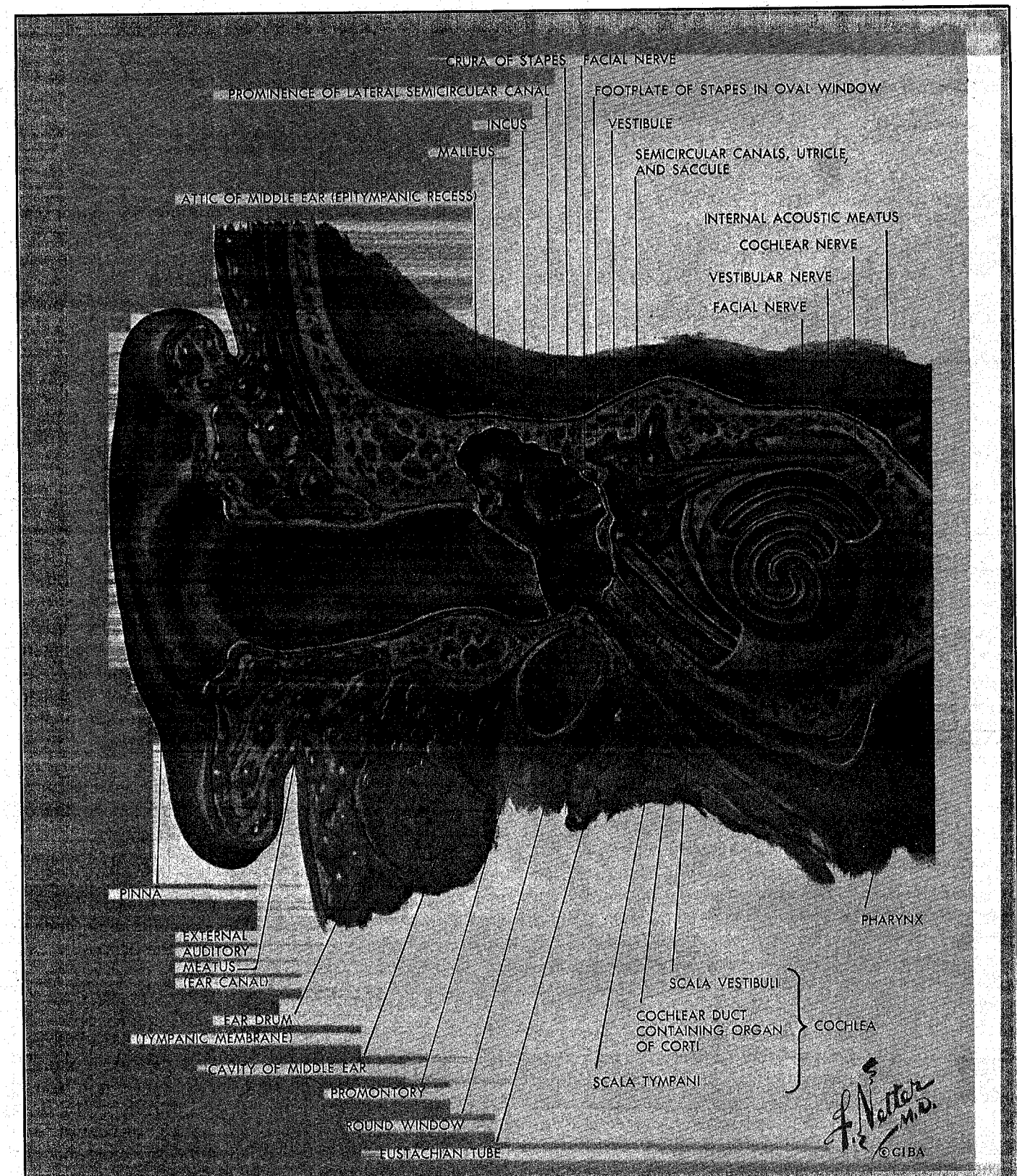

Figure 1. In this detailed drawing of the human cochlea, most of the major components are identifiable. Sound is directed by the pinna into the ear canal. As it passes down the ear canal, it becomes a plane wave due to the small diameter of the ear canal. Most of the energy delivered to the ear drum is absorbed, at least over the major portion of the audible spectrum. The sound is transmitted to the cochlea (inner ear) via the ossicles, which are the malleus, the incus and the stapes. The motion of the stapes displaces the fluid in the "upper" chamber of the cochlea. An equal amount of fluid is displaced at the round window since the net volume of fluid within the cochlea must remain constant. (©Copyrighted 1970 CIBA pharmaceuticals, CIBA GEIGY Corporation).

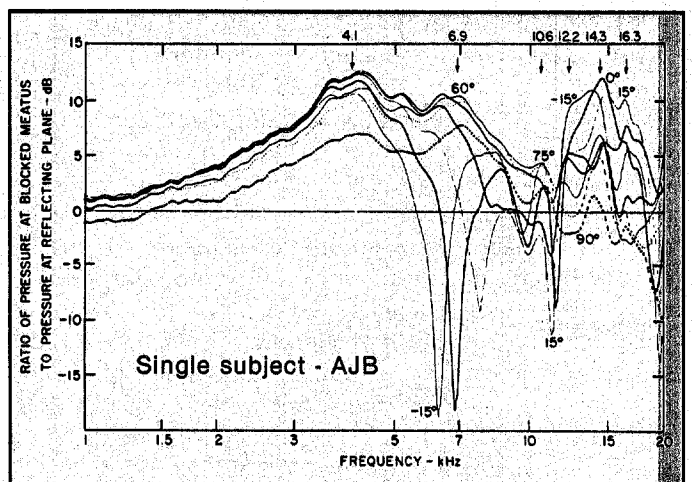

Figure 2. Above 6 kHz the sound is diffracted by the pinna in a way that depends on the incident sound angle in the median plane (the vertical plane that symmetrically bisects the head). The strong spectral zeros, such as the one at 7 kHz, are believed to provide helpful cues about the perception of direction in the cone of confusion, which is defined as all points having identical pathlength difference to the two ears. Head movements may also provide important information. Timing and amplitude differences account for localization in the horizontal plane. (Reprinted from Shaw, 1980).

porpoise. Several comprehensive overview articles have appeared in the last few years, and the interested reader is encouraged to pursue this topic in these sources [26,34,7].

It is useful to consider the effects of room reverberation on an incoming signal in order to fully appreciate the complexity of this problem. As a visual analogy, consider yourself in a room made up of highly reflective mirrors. In such a room, because of the reflective walls, it would be difficult to distinguish a real object from its reflections. The acoustic case is similar; however there are important differences. Because of the slow speed and large wavelength of sound, interaural time and amplitude difference information is obviously useful in the case of laterally displaced sources. However, as a simple analysis would show, these alone will not uniquely resolve an object from its reflections in three dimensions. The addition of the directionally dependent spectral zero cue created by the pinnas is believed to provide the resolution of this dilemma.

In Fig. 3a we show the impulse response of a typical office-sized room and in Fig. 3b, its frequency response. Since a room is a linear system, the impulse response characterizes the effect of the room on the speech signal. In Fig. 4a we see the waveform of "clean" speech and in 4b that of reverberant speech. Note how the room reverberation modifies the classical properties of the speech waveform. Yet when we hear such reverberant speech and when we communicate over the "room" channel, we are almost completely unaware of the presence of the reverberation! How can the information be coded in such a way that it is robust to the degrading effects of the room reverberation?

If digitally coded information were sent over this room channel, the information transmission rate would be quite small because of the reverberation. This important question of robustness appears to be a result of both the speech code and the (binaural) hearing system. One objective of hearing theory is to explain this amazing robustness.

In summary, the diversity provided by the two ear signals allows a certain amount of signal processing capability in the higher auditory centers, while the spatial frequency selectivity of the pinnas provide acoustic cues which are used by the binaural central nervous system (CNS) to give a spatial sense. The exact nature of this CNS processing is unknown, but is dramatically displayed in the barn owl, as described by Knudsen [26]. The physical acoustic transformations of the pinnas are better understood since they are accessible to precise measurement.

THE MIDDLE EAR

It is widely accepted that the middle ear acts as an acoustic transformer to match the airborne-sound impedance to the fluid-borne-sound impedance of the cochlea. Not as well known is that most of this transformer action is due to the ratio of the areas of the active parts of the ear drum and the foot plate of the stapes, rather than by a lever action (as in a classical lever). Once the sound wave is diffracted by the pinna cavities, it travels down the ear canal on its way to the ear drum. For frequencies below 27 kHz it is accepted that the sound becomes a plane wave in the human ear canal because the higher order modes are cut off by the small diameter of the ear canal [31]. The accepted average diameter of the human ear canal is .75 cm, while the accepted canal length is 2.25 cm.

When the sound field is a plane wave at the ear drum, it may be described (or modeled) by the use of an impedance. Because of its fundamental importance, many workers have made measurements of the acoustic impedance at the ear drum to characterize the acoustic signal properties there. At this point in time there is reasonable, but not perfect, agreement on the impedance properties at the drum below 5 kHz [19,42,43]. Some of the questions surrounding these impedance measurements are certainly due to the lack of a universally accepted set of procedures for making such measurements. A second problem is that the measurements are made under different conditions in different laboratories. Our present understanding is best summarized by the use of the model of the middle ear system [29] defined in Fig. 5.

I have simplified the middle ear model as shown here in order to reduce it to its basics. In Fig. 5 the mass of the middle ear bones is represented by the inductor M_s. This should also include the mass of other structures as well, such as the mass of the ear drum. The compliance, or spring, is shown as a capacitor C_{AL} in the electrical equivalent circuit, and mechanical friction is represented by a resistor. P_s is the pressure input at the ear drum, P_v is the cochlear pressure, and P_{RW} is the pressure across the

round window membrane. The major source of compliance is presently believed to be due to the tissue holding the stapes foot plate in the oval window. This conclusion is based on experiments on animal middle ears. Clearly, other components in the system might also contribute to the overall stiffness. However the exact source of the compliance is not important for this discussion. The impedance elements of the middle ear appear in series with the impedance of the inner ear, since the velocity of all the elements is assumed to be the same. In the analog used here, velocity plays the role of current and pressure the role of voltage in an acoustical circuit. Figures 6a and 6b compare the experimentally measured input impedance of a cat eardrum as measured at the end of a 0.5 cm length of ear canal with the model impedance under the same conditions. The impedance has been normalized by the characteristic impedance of an average ear canal.

Over the frequency region from 500 Hz to nearly 10 kHz the real part of the impedance of the inner ear dominates the input impedance. This was shown by Lynch et al. [29] by removing the cochlea (by draining the fluid out of the cochlea) from the acoustical circuit and remeasuring the mechanical impedance of the stapes. One may conclude that since the real part of the impedance of the cochlea dominates the eardrum input impedance over this frequency range, most of the incident sound energy will be delivered to the cochlea.

The transfer function between the displacement of the stapes and the ear canal pressure is shown in Fig. 7. These experimental results are consistent with an analysis of the middle ear model of Fig. 5. From measurements of the stapes displacement in the cat, 120 dB SPL (dB re 20μ Pa, a very loud sound corresponding approximately to 120 dB re our threshold of hearing) at 1 kHz, corresponds to an

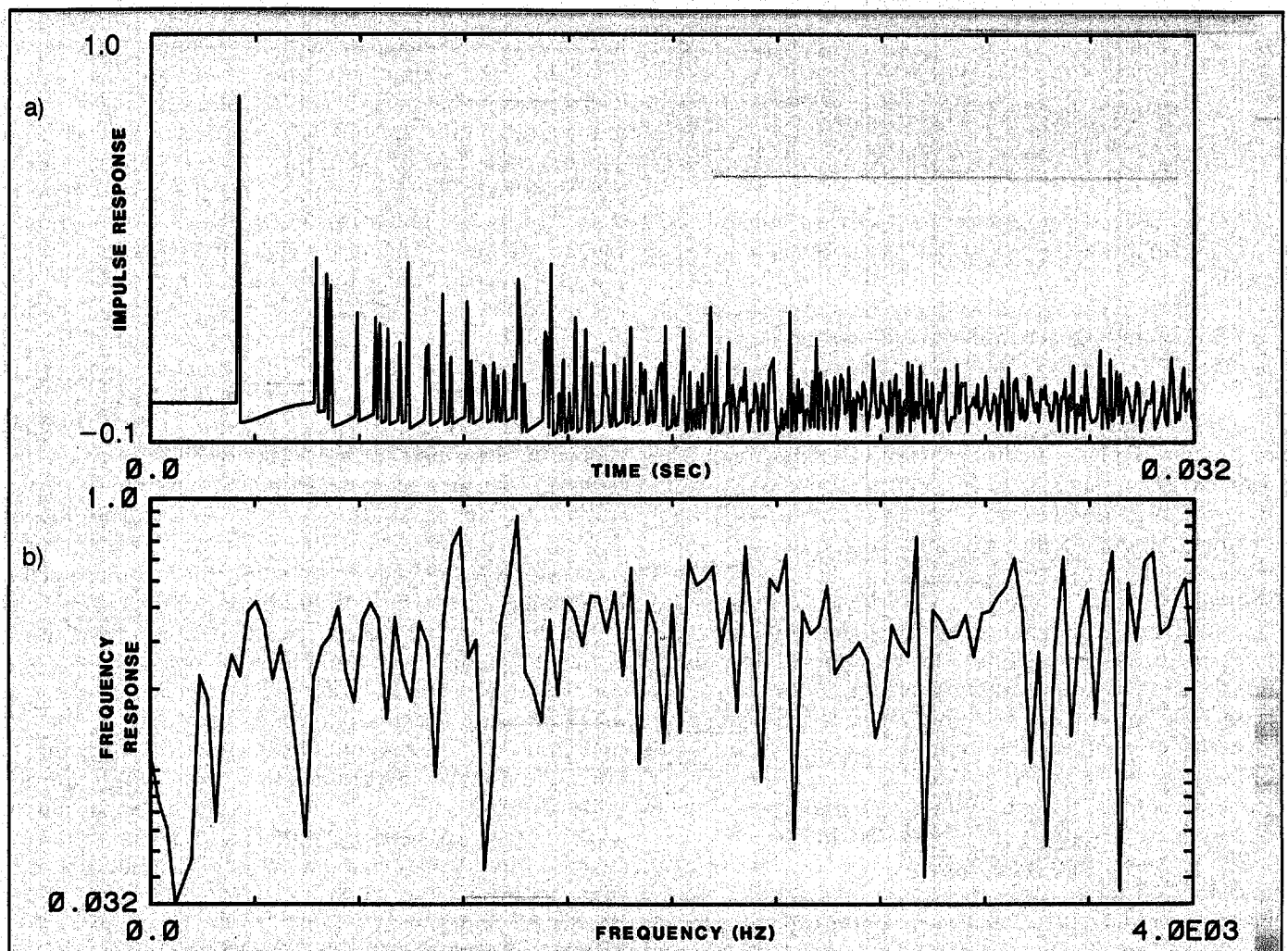

Figure 3. In part (a) of this figure we see a synthetic impulse response for a small room having slightly absorbing walls. The response has been truncated in the plot to 32 msec. The tail of the room impulse response, which is not shown here, is responsible for the rattling sound, or echoes, which are heard in a hard walled room. In the lower panel we see the frequency response of the model room as measured by Fourier transforming the impulse response. Note the deep nulls in the frequency response and the random nature of the transfer function due to the complexity of the wall reflections.

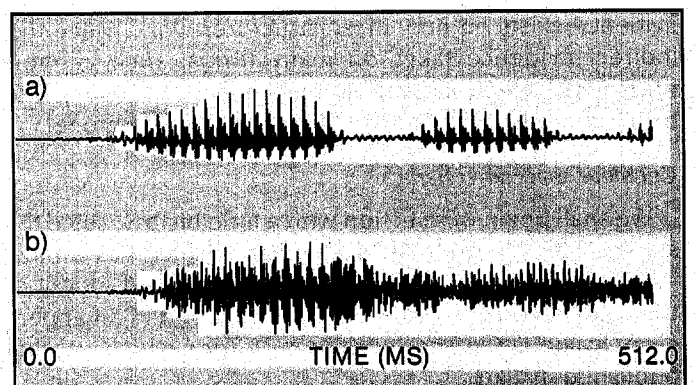

Figure 4. Clean speech, as seen in (a), when filtered by the room's impulse response, leads to a cluttered speech signal which looks like speech with additive noise. This reverberated speech (b) sounds much like that recorded by a microphone several feet from the speaker in a hard-walled office. When our two ears are presented with separately reverberated speech signals, we are almost unaware of this degradation. In fact we become aware of a pleasant spacious "presences" under this condition.

RMS displacement of the stapes of about 0.7 microns [16]. By way of comparison, the free field RMS displacement of air molecules under the same conditions is about 10 microns.

By placing an electrode on the round window, with the ground electrode near the cochlea, it is possible to use the animal's ear as a very sensitive low noise microphone. The signal to noise ratio of this "animal microphone" in fact is almost as good as the best commercially available microphone. Also the high frequency response is quite impressive, with a flat response to about 15 or 20 kHz. Below about 1 kHz, the low frequency response of the cochlear potential drops off at about 9 to 12 db/octave. Because of its properties, this cochlear potential has been called the cochlear microphonic, or CM. It is frequently used in hearing research to characterize the properties of the middle ear, and as an experimental control on the animal. The human CM can be measured with an electrode through the ear drum, however the procedure is not used due to the existence of less invasive procedures which give similar information. In Fig. 8 we have superimposed a measured cochlear microphonic response on the model response computed with the aid of the middle ear model [5].

It is interesting to note that direct measurements of the pressure in the ear canal by a probe microphone show a standing wave pattern due to reflections from the ear drum. These standing waves result from the increasing mismatch of the drum at higher frequencies due to the increasing inertial impedance at higher frequencies [43]. The smooth frequency response of the cochlear microphonic as a function of frequency above 1 kHz is a clear indication that the standing waves in the cat ear canal are not present in the eardrum or stapes displacement.

THE INNER EAR

Of clear significance in the auditory perception process are the signal processing operations performed by the cochlea. The cochlea has many outputs, with 30,000 neurons encoding 1,500 to 2,500 cochlear inner hair cell signals. Each neuron encodes a narrow band hair cell signal having a few hundred Hz of bandwidth, using a point process code, with the time between pulses coding the information being signaled into the neural network.

I shall describe this process by two separate means. First I shall give examples of the signal representation at various points in the system. Second I shall refer to models of the auditory system. These models are our most succinct means of conveying the results of years of difficult experimental work on cochlear function. An alternative way of describing our knowledge of the cochlea (which I shall not use) would be to describe the hundreds of experimental results that have been collected over the years. This body of experimental knowledge may be very efficiently represented (to the extent that it is understood) in the form of models. When the experimental results are at variance with the model, the model is not a useful description, and the more complex description using the experimental data base is necessary.

The inner ear may be naturally divided into several sub components anatomically. From Figs. 9 and 10, three major

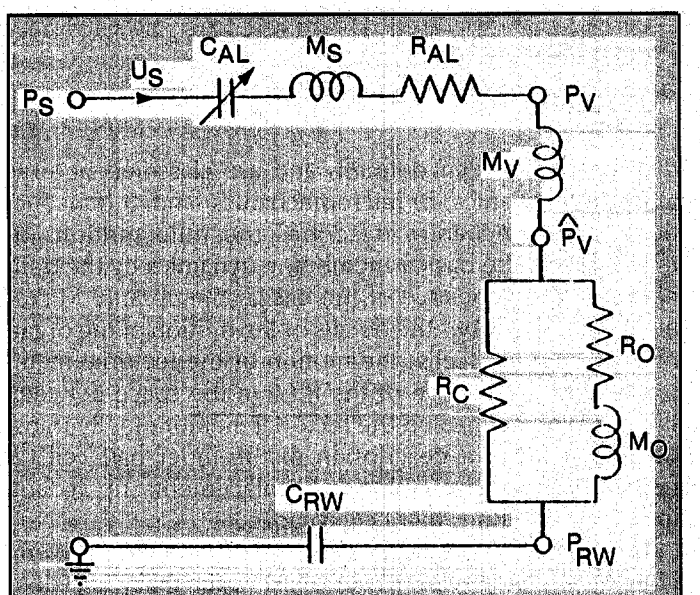

Figure 5. This model was first proposed to characterize the stapes, annular ligament, and cochlea. In fact the same model may be applied to the more complex system which includes the entire middle ear and the eardrum. Much more detailed models have been proposed; however this one describes essential features of the middle ear system. The annular ligament is represented by the non-linear capacitor (spring) C_{AL}, while the mass of the ossicles are represented by the inductor M_s. M_v is the mass of fluid behind the stapes, while the elements between nodes p_v and p_{RW} represent the input impedance of the cochlea. C_{RW} represents the stiffness of the round window. (Reprinted from Lynch et. al., 1982).

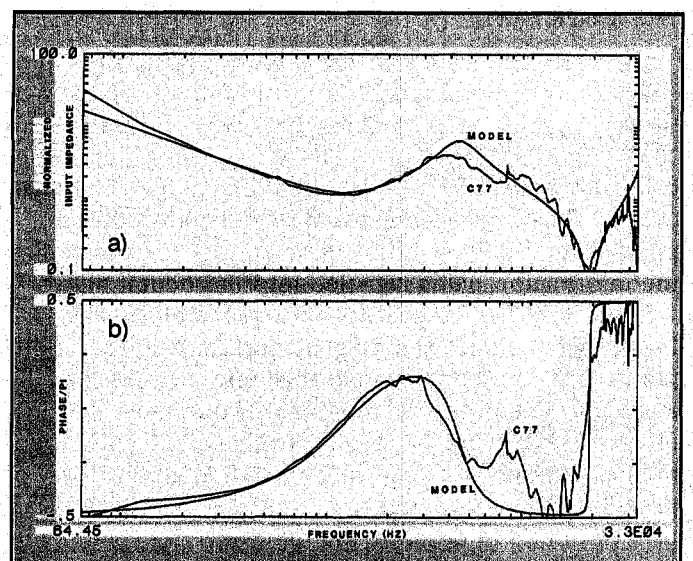

Figure 6. A comparison is made here between an experimentally measured cat ear drum impedance and the model of Fig. 5. The model response has been transformed by placing it at the end of a 0.5 cm tube, which has the effect of introducing the spectral zero seen at 20 kHz. While the magnitude response agrees quite well, the phase does not agree over the range from 4 kHz to 20 kHz. The phase of the experimental data is closer to zero degrees, indicating the lossey nature of the cochlear input impedance. The model, on the other hand, is reactive over this frequency interval. The experimental impedance has been normalized by the characteristic impedance of the sound delivery system, which is matched to an average cat ear canal.

divisions have been defined, and are classified here as a) macromechanics, b) micromechanics, and c) transduction. Macromechanics describes the fluid motions of the scala and the basilar membrane dynamics by the definition of an impedance of the basilar membrane. Micromechanics describes the details of the motion of the organ of Corti, the hair cells, the motion of the tectorial membrane, and the motion of the fluid in the space between the reticular lamina and the tectorial membrane. By transduction, I mean a description of the inner hair cell response to basilar membrane displacements up to and including the inner hair cell synapse.

The history of cochlear modeling is interesting. Several good books and review papers exist which make excellent supplemental reading [8, 30, 35, 39, 46, 51]. However, there is a great deal of diverse opinion about certain critical points. This diversity of opinion is, in part, due to uncertainty about key issues. For example, it seems to be very difficult to experimentally observe the motion of the basilar membrane in a functionally undamaged cochlea. Questions regarding the relative motion of the tectorial membrane are largely a matter of conjecture. Such questions are therefore best presently investigated by theoretical means. As a result, a variety of opinions exist as to the detailed function of the various structures.

On the other hand, firm and widely accepted indirect evidence exists on how these structures work. Since this indirect evidence takes on many forms, such as morphological, electrochemical, mechanical, acoustic, biophysical, etc., these data are best related via a model.

Cochlear macromechanics

We shall begin this section with a little history, since the study of the fluid mechanics of the cochlea has a most interesting history. The first widely recognized model of the cochlea, due to Helmholtz, is described in an appendix of his book *On The Sensations Of Tone* which was first published in 1862.

Helmholtz likened the cochlea to a bank of high Q resonators which are tuned to different frequencies, much like a piano. In actual fact, the model he proposed was not very satisfying since it left out many important features, such as the effects of the fluid which couple the mechanical resonators together.

It was not until the experimental observations of G. von Bekesy in 1928 on human cadaver cochleas that the nature of the basilar membrane traveling wave behavior was unveiled. Typical fluid motions in the cochlea are shown with arrows in Figs. 9 and 10. What von Bekesy found was that the cochlea was like a linear dispersive transmission line that disperses the different frequency components of the input signal out along the basilar membrane, thereby isolating those various components at different places

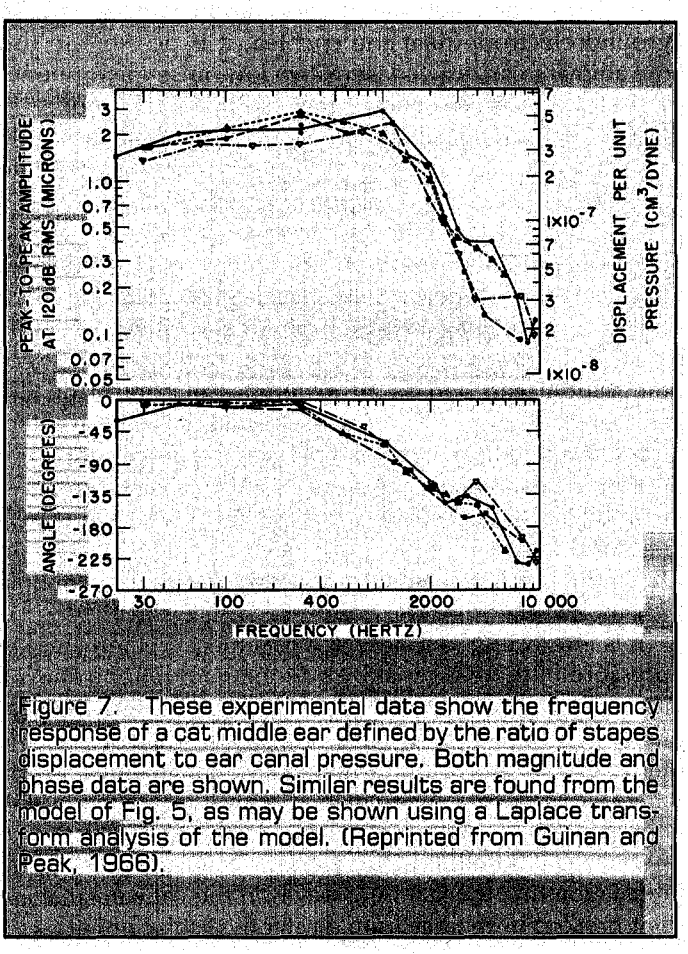

Figure 7. These experimental data show the frequency response of a cat middle ear defined by the ratio of stapes displacement to ear canal pressure. Both magnitude and phase data are shown. Similar results are found from the model of Fig. 5, as may be shown using a Laplace transform analysis of the model. (Reprinted from Guinan and Peak, 1966).

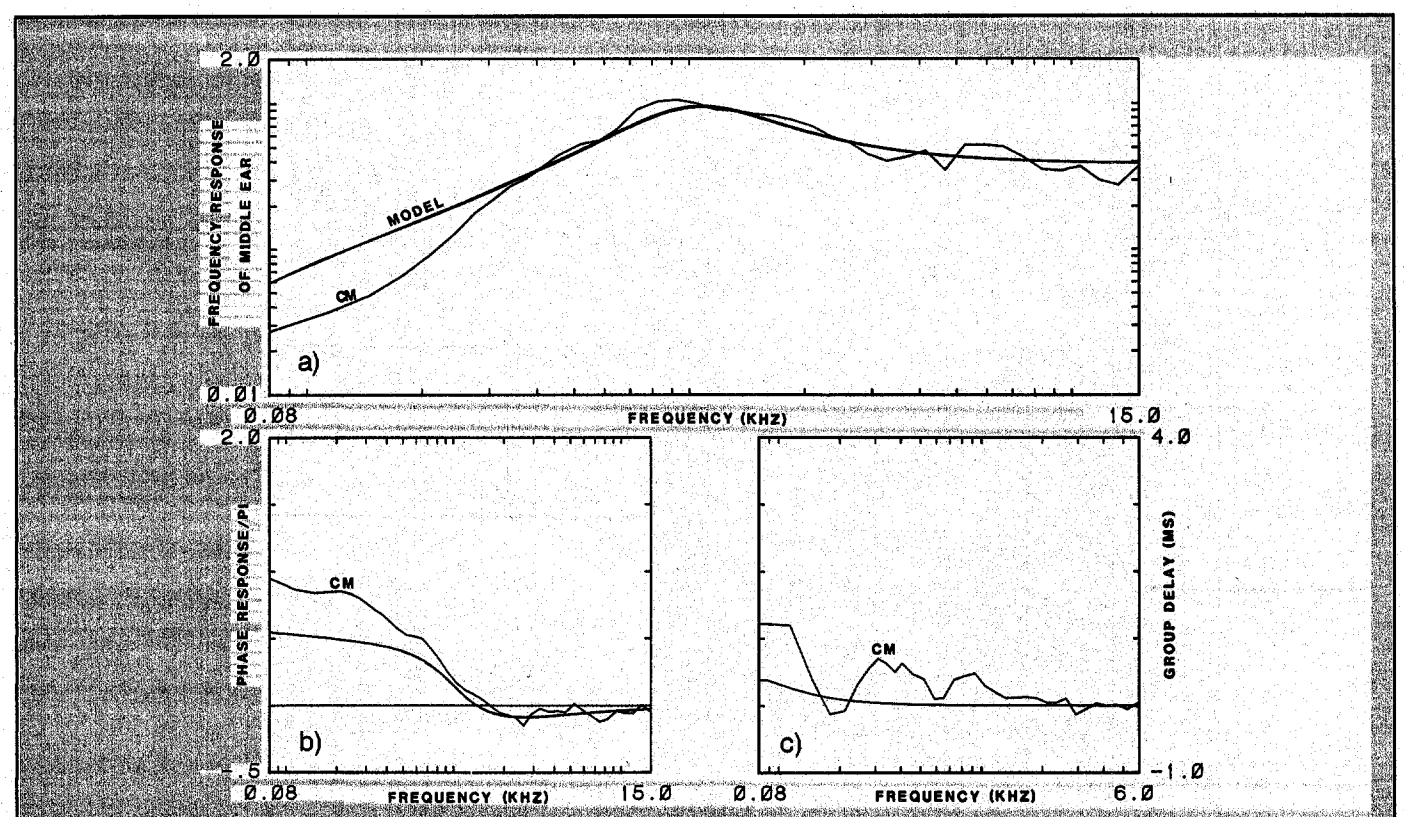

Figure 8. The next readily observable signal after the stapes displacement is the cochlear microphonic. It is a widely held belief that the cochlear microphonic (CM) is proportional to the displacement of the basilar membrane. Since the basilar membrane is stiffness-dominated at the stapes, the pressure across the basilar membrane and its displacement, and thus the CM, are proportional. Using the middle ear model, and the above assumptions, one may model the CM. At low levels, the CM is linearly related to the pressure input at the ear drum. However at larger levels the CM becomes nonlinear. These nonlinear effects have not yet been successfully modeled. (Reprinted from Allen, 1983b).

along the basilar membrane. He described this dispersive wave as a "traveling wave" on the basilar membrane. He observed this wave under stroboscopic light in a dead human cochlea, at sound levels well above our pain threshold, namely between 120 and 140 dB SPL and even above. Sound levels of this magnitude were required to obtain displacement levels that were observable. These experiments were so unusual, difficult, and important, that von Bekesy won the Nobel prize in 1961 for his experimental observations.

Over the years these experiments have been greatly improved, but von Bekesy's fundamental observation of the traveling wave still stands. His original measurements however are not characteristic of the responses seen in more recent experiments in several ways.

Today we find that the traveling wave has a very sharp cutoff frequency slope, much sharper than found by von Bekesy. In fact, according to modern measurements, the response on the basilar membrane to a pure tone can change by about five orders of magnitude per millimeter of distance along the basilar membrane. To describe this response it is helpful to call upon one of the early models of macromechanics, the transmission line model, first proposed by J. J. Zwislocki in 1948. This model is also frequently called the one-dimensional model, for reasons that will become clear later.

The transmission line model of the cochlea

The first model to capture the essential character of the traveling wave on the basilar membrane was the transmission line model of J. J. Zwislocki. This model is shown in Fig. 11. The stapes input is at the left, with the input current corresponding to the stapes velocity. Different points along the basilar membrane are represented as cascaded sections of the transmission line. The series inductors represent fluid inertia along the length of the cochlea, while the elements to ground represent the mechanical or acoustical impedance of an element of the basilar membrane. The inductor to ground represents the mass per unit length of the basilar membrane, while the capacitor represents the compliance (stiffness) of the basilar membrane. The compliance is assumed to vary in a systematic manner along the length of the cochlea, with a stiffness that decreases exponentially along the length of the cochlea. Thus each piece of basilar membrane is tuned to a different frequency, since the stiffness of the spring

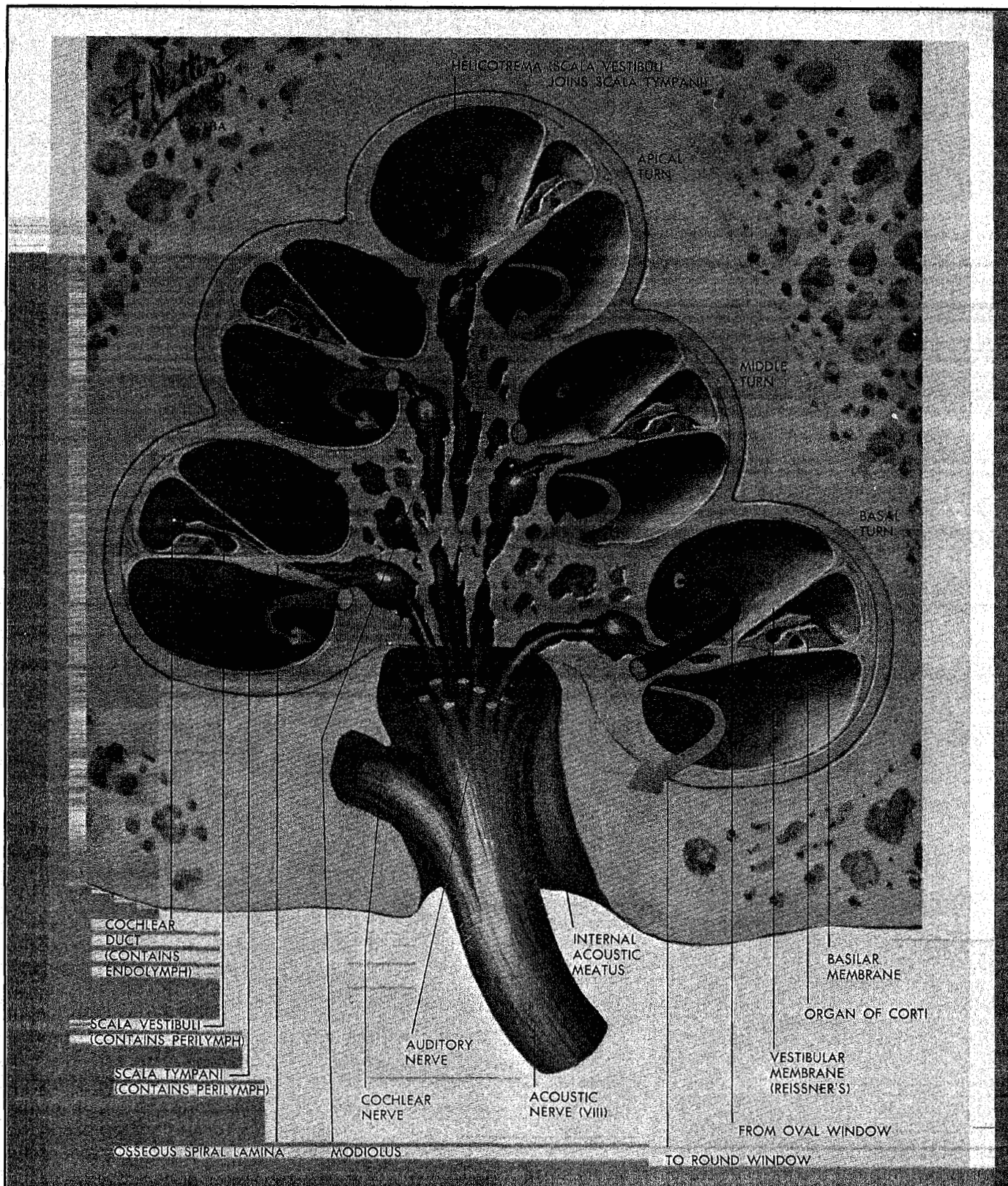

Figure 9. As we move into the inner ear we see the various fluid filled chambers. The arrows show the direction of fluid motion with the stapes pushed into the cochlea. The cochlear nerve forms the central core of the cochlea extending into the VIII nerve, which also comprises the facial and vestibular nerve bundles. The membrane labeled "Vestibular membrane (Reissner's)" is believed to be an electrical barrier which serves no direct mechanical function. The cochlear duct, defined as the space between Reissner's membrane and the basilar membrane, is at an 80 mV potential. This potential is important in the transduction process, as we shall describe later. (© Copyrighted 1970 CIBA pharmaceuticals, CIBA GEIGY Corporation).

changes with position (for convenience, we assume here that the mass of the basilar membrane remains constant along its length, which, roughly speaking, seems to be the case).

To understand the inner workings of this circuit, let's assume that we excite the line at the stapes with a constant current source of frequency f. Because of conservation of charge, the total current through the basilar membrane must equal the current at the stapes. The physical law being modeled in Fig. 11 of course is not conservation of charge, since the cochlea is not an electrical circuit, but conservation of fluid mass, namely the conservation of the fluid volume within the cochlea.

When the stapes is displaced, thereby producing a fluid volume change in the upper chamber, the integrated volume displacement of the basilar membrane (and helicotrema, Fig. 9) must take up an identical volume displacement. Simultaneously the round window membrane, connected to the scala tympani, must bulge out by an equal amount. In practice the motion of the basilar membrane is quite complicated, as we shall see. However the total volume displacement of the basilar membrane, at any instant of time, must be equal to the volume displacement of the stapes, or of the round window membrane.

Consider next where the current will flow, or where it can flow. At one point along the length of the cochlea the impedance is small, namely that point where the basilar membrane compliance reactance cancels its mass reactance, and the basilar membrane appears to have a "hole" in it. Thus the fluid current tends to go through the point of least resistance, or resonant point. To the left of the hole, the basilar membrane is very stiff (large capacitive reactance), and to the right of the hole, the impedance is a large mass reactance (inductive). In fact, beyond the hole the impedance is irrelevant since little current will flow past the shorted point.

Of course this picture is dependent on frequency, since the location of the "hole" is frequency-dependent. If we were to put a pulse of current (in the time domain) in at the stapes, the highest frequencies would be shorted out near the stapes, while the lower frequencies forming the pulse would propagate down the line. As the pulse travels down the line, the higher frequencies are progressively removed from the original pulse, until almost nothing is left when the pulse reaches the right end of the model (the helicotrema end, or apex of the cochlea).

From the above description it is easy to understand why the various frequency components of the signal are splayed out on the basilar membrane.

Let's next try a different mental experiment with this model. Suppose that the input at the stapes were a slowly swept tone. What would the response at a fixed point on the basilar membrane look like? In Fig. 12b we show the frequency response magnitude of this transfer function, namely the ratio of the response of the model basilar membrane to a constant pressure input at the stapes. The response is a bandpass response, with a shallow low frequency slope, and a very sharp high frequency slope. The impulse response of this transfer function is shown in the upper panel of the figure, Fig. 12a, and the group delay is in 12c.

Inadequacies of the one dimensional model

The transmission line model was a most important development since it was in agreement with the experimental evidence of the day, and it was based on a simple set of physical principles, namely conservation of fluid mass and a variable resonant basilar membrane. In fact, this model was the theory of choice until improved experimental observations were available in the late sixties and early seventies.

In 1976 George Zweig and his colleagues pointed out that the transmission line theory could be accurately integrated by the use of the method in physics called the "WKB" method [48]. This provided an approximate solution to the transmission line model, and, as may be shown by numerical methods, the error of the approximation is not significant for parameter values such as those typically chosen for cochlear models. As further model results became available, it eventually became clear that the one-dimensional theory was not totally satisfactory since that theory did not agree with more detailed and complete descriptions of the cochlear geometry. It is now possible to compute the response of a two-dimensional [1], and recently even the response of a three-dimensional geometry [9]. As the complexity of the models approached the physical geometry, the solutions tended to display steeper high frequency slopes and increased frequency selectivity.

A great deal of neural data is available which defines quite precisely the input-output properties of the cochlea. However, since the signals undergo significant transformations between the basilar membrane and the neural measurement point, one cannot directly compare neural response curves with the basilar membrane model, at least not without careful consideration. We ultimately seek a model which describes the response of the 1,500 hair cell outputs and 30,000 neural channels. One such model description of these transformations will be discussed in the section on cochlear transduction.

What is important here is that the frequency response, as computed by the mechanical transmission line model of the basilar membrane motion, is quite different from the response as estimated from the nerve fiber measurements. The difference can be on the order of 20 to 40 dB, and appears to be even greater under some conditions. Thus when the two-dimensional models showed sharpened responses, relative to the transmission line model, the hope was that these more detailed models would converge to the response measured in the nerve fiber. This convergence has not yet occurred.

A second area where the existing one-dimensional theory seemed inadequate followed from the nonlinear phenomena which have been experimentally observed. Since the transmission line theory is a linear theory, many researchers have studied ways of realistically making the

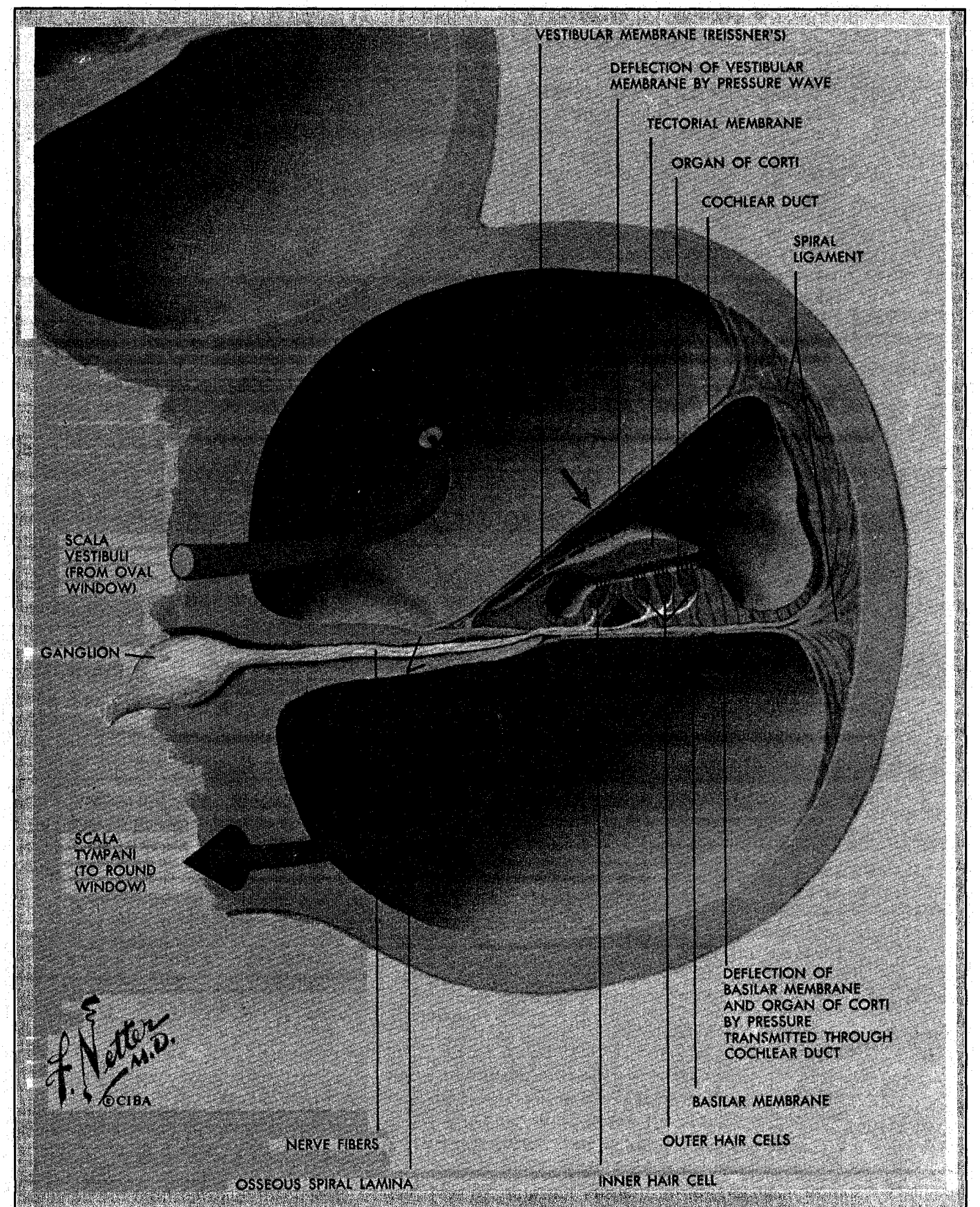

Figure 10. The cochlear duct moves in response to positive pressure in the scala vestibuli relative to scala tympani. The detail shown regarding the inner hair cells is not known to be an accurate representation. (© Copyrighted 1970 CIBA pharmaceuticals, CIBA GEIGY Corporation).

cochlear models nonlinear. Nonlinear aspects of cochlear theory will not be discussed here [17], [4].

Two-dimensional cochlear macromechanics

In this section I will try to give a bit of the flavor of the extended theories of cochlear mechanics in order that the reader may better appreciate how and why they represent an improvement on the transmission line theory.

The first step towards an improved theory was taken by Ranke in 1950 in what he called a "short wave" theory. This theory was intended to complement the transmission line theory, which was considered to be a "long wave" theory. Short wave theory is most accurate near the cutoff frequency while long wave theory is best for lower frequencies [44]. Ranke's attempts were historically significant, but never actually developed into a useful theory for many reasons. For example, it is not known how to optimally interface the long wave model wave to the short wave model, since some sort of matching procedure is necessary.

Then in 1972, M. Lesser and D. Berkley proposed a rectangular box model of the cochlea, where the scalae were straight, and the cochlea was symmetric about the basilar membrane. This geometry is shown in Fig. 13. The main point of their paper was to demonstrate the importance of extending the models to two dimensions because of the effect of this extension on the solutions. Their line of reasoning inspired research that kept people busy computing for at least ten years. As mentioned, via numerical methods, we have now moved beyond the two-dimensional formulation into the realm of three-dimensional models. More time is needed to fully evaluate the significance of these more complicated calculations and models, but it presently appears that they do *not* close the gap, as was originally hoped, between the mechanical model and neural measurements of cochlear frequency response. Thus the most important problem which remains unsolved in

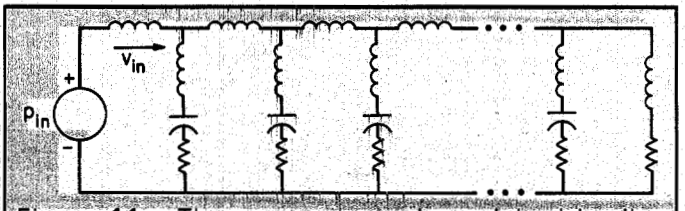

Figure 11. The most commonly exploited basilar membrane-cochlear model is the transmission line model. In this model the inductors represent masses of the cochlear fluid (series elements) and the basilar membrane (shunt inductors). The inductor values are frequently assumed to be independent of their position along the length of the cochlea. The stiffness of the basilar membrane is represented by the shunt capacitors. These elements are position-dependent, and are usually assumed to vary exponentially with position. They are most stiff (smallest capacitance) near the stapes. Thus the resonant frequencies of the shunt elements, taken in isolation, are largest at the stapes (base) and smallest near the helicotreme (apex). This model, called the transmission line model, or the one-dimensional model, has been popular since it was first introduced by J. J. Zwislocki in 1948. This model does not give as sharp a high frequency cutoff as two- and three-dimensional models. However, it does capture many of the essential features of the system in a qualitative way, such as the traveling wave observed by von Bekesy.

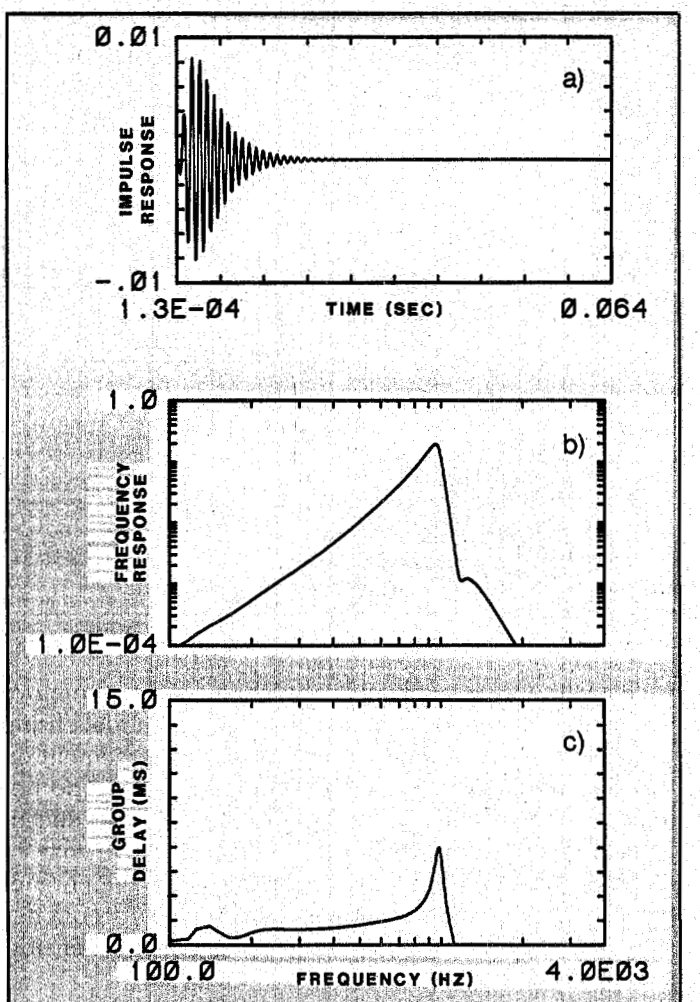

Figure 12. When the response of the model is viewed at a given place on the basilar membrane as a function of frequency, the response is found to be a band-pass filter. The slope on the high-frequency side of a real cochlea is place-dependent and varies from 50 dB/oct for the low frequency place, near the helicotrema, to more than 500 dB/oct for a high frequency place. This latter slope would give a 42 dB change in output for a semi-tone change in frequency (a change corresponding to going from C to C# on a musical scale). By comparison, the slope on the low frequency side of the model filters is quite shallow, and is between 6 to 12 dB/oct. Because this slope is so shallow, the cochlear filters on the basilar membrane are sometimes referred to as low-pass filters rather than band-pass. This is technically not correct, but practically speaking, is a reasonable characterization. If we were to show the pressure across the basilar membrane instead of the basilar membrane displacement, the transfer functions would be strictly low-pass. In the first panel (a) we see the impulse response of panel (b), the model transfer function magnitude, defined by the output displacement relative to the stapes input displacement. In the third panel (c) we show the corresponding group delay.

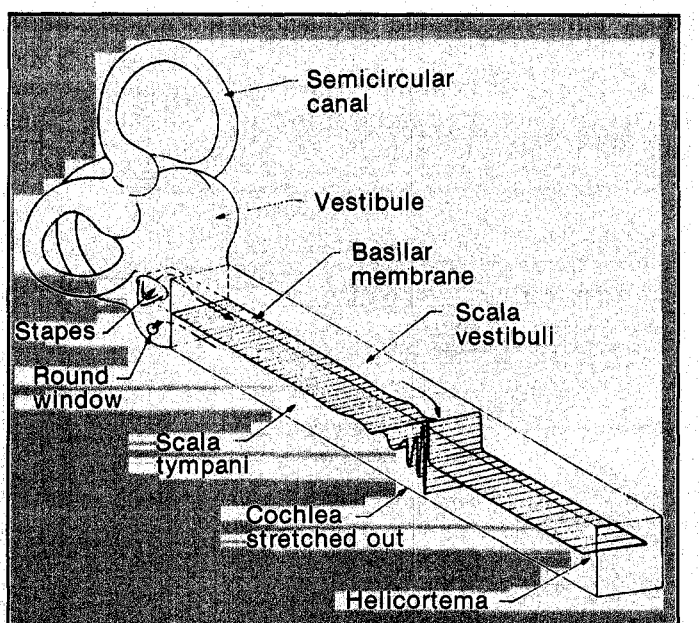

Figure 13. This figure which has been taken from Zweig et al. (1976), shows the traveling wave at one point in time on the basilar membrane. Because of the dispersive nature of the basilar membrane, a wake appears behind the main pulse. This pulse also becomes broader as it travels due to the loss of the higher frequency components. (Reprinted from Zweig et al., 1976).

cochlear theory is explaining the sharpness of tuning of the neural response. The two-dimensional models bring the neural data and model calculation into agreement on the high frequency side of the tuning curve, but do not improve the match on the low frequency side. Recent measurements indicate that a 20 dB difference may in fact experimentally exist [41] due to the basilar membrane-to-hair cell transduction process, on the low frequency side of the tuning curve. We shall discuss such a transformation in the next section.

Cochlear micromechanics

Micromechanics refers to the mechanics of the organ of Corti. This aspect of cochlea theory has a much shorter history than that of cochlear macromechanics.

The most commonly accepted description of the motion of the organ of Corti was proposed by ter Kile in 1900. His concept is shown in Figs. 10 and 14 where we see how he proposed that the displacement of the basilar membrane could drive the hair cells in a radial mode of excitation. In Fig. 15 we see a similar description of this mode of excitation from Allen (1980). A simple analysis of the model reveals that the vertical motion of the basilar membrane is linearly related to the shearing motion seen by the hairs of the inner hair cells, which are now known to be the transducers used to sense the motions of the basilar membrane. Thus the model of ter Kile is equivalent to a lever which linearly converts the vertical basilar membrane motion into radial shearing motions appropriate to the excitation of the afferent inner hair cells.

This description seemed adequate as a first step, but several important problems remained. First, there have been no direct observations to confirm the ter Kile model, nor are there likely to be any in the near future, due to the inherent difficulty in making observations of such small motions in such difficult places. Second, the response of the neural signals, as we have described, do not seem to be in agreement with basilar membrane measurements [41]. It was hoped that a simple modification to the ter Kile model might bring together the various theories and the experimental data. I shall argue this possibility here.

At this point it is necessary to remove ourselves from the models and look at some experimental data, in order to understand the magnitude of the discrepancy between

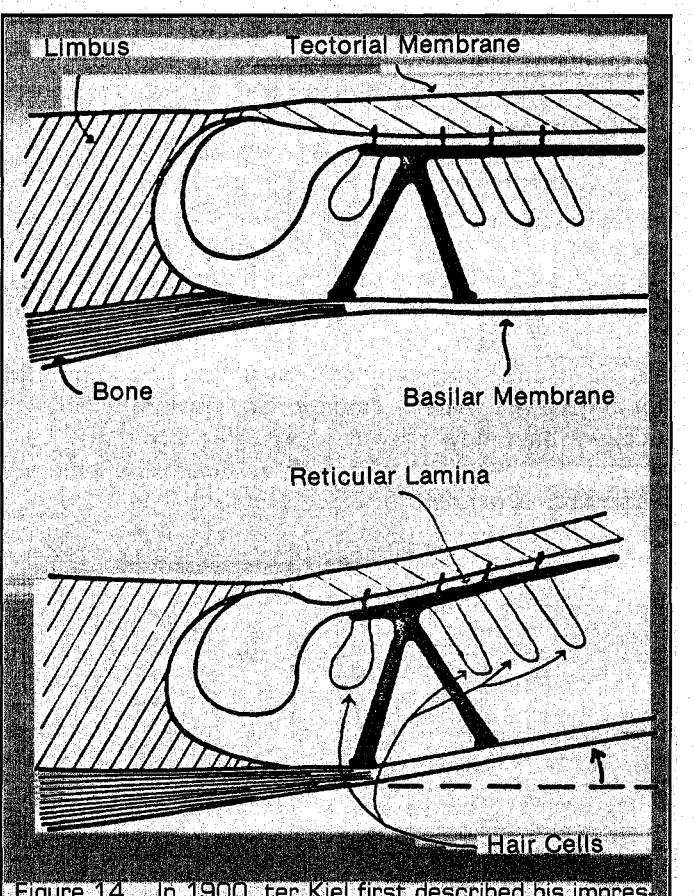

Figure 14. In 1900, ter Kiel first described his impressions of how the vertical displacement of the basilar membrane was transformed to a "radial" shearing sufficient to drive the hair cell cilia. At that time it was generally believed that the row of inner hair cells was connected to the tectorial membrane. It is now generally believed that inner hair cells are not driven directly by the tectorial membrane, but are dragged by the surrounding fluid. This would happen because the viscous "boundary layer" (a thin fluid layer where viscous forces dominate) is greater than the 6 micron distance between the tectorial membrane and the top surface of the hair cells. (This surface is called the "recticular lamina"). As a result, the relative shear of the two surfaces acts as a mechanical resistor, or "dashpot," as it is referred to in mechanical terms. The mechanical equivalent of the entire system is a lever; electrically, it is a transformer.

the mechanical tuning of the basilar membrane and the neural signal. In Fig. 16 we see tuning responses of the basilar membrane [52]. The measure which is being compared to the basilar membrane is the neural response.

The voltage in the hair cell has also been measured by Sellick et al. This voltage, called the receptor potential, is tuned like the neuron. It has been shown by Sellick et al. (1983) that the mechanical and receptor potential responses differ by at least 20 dB on the low frequency side of the response curve. Is this difference significant, or is it an artifact of the experimental technique? Unfortunately, we cannot be sure of the answer to this important question. To date no one has shown that basilar membrane responses are similar to neural, or equivalently, to hair cell receptor potential signals, in the mammalian cochlea. For this reason, the present micromechanical models are in a state of flux, and until the experimental data improve, and the many experimental questions are cleared up, this theoretical question will remain open.

I shall briefly mention two theories which are presently being considered which might explain the presently observed differences in frequency selectivity between model and experiment. The first is based on the idea that the tectorial membrane vibrates at its own resonant frequency near the resonant frequency of the basilar membrane. Two different versions of this idea were independently proposed by Zwislocki (1979), and by Allen (1980).

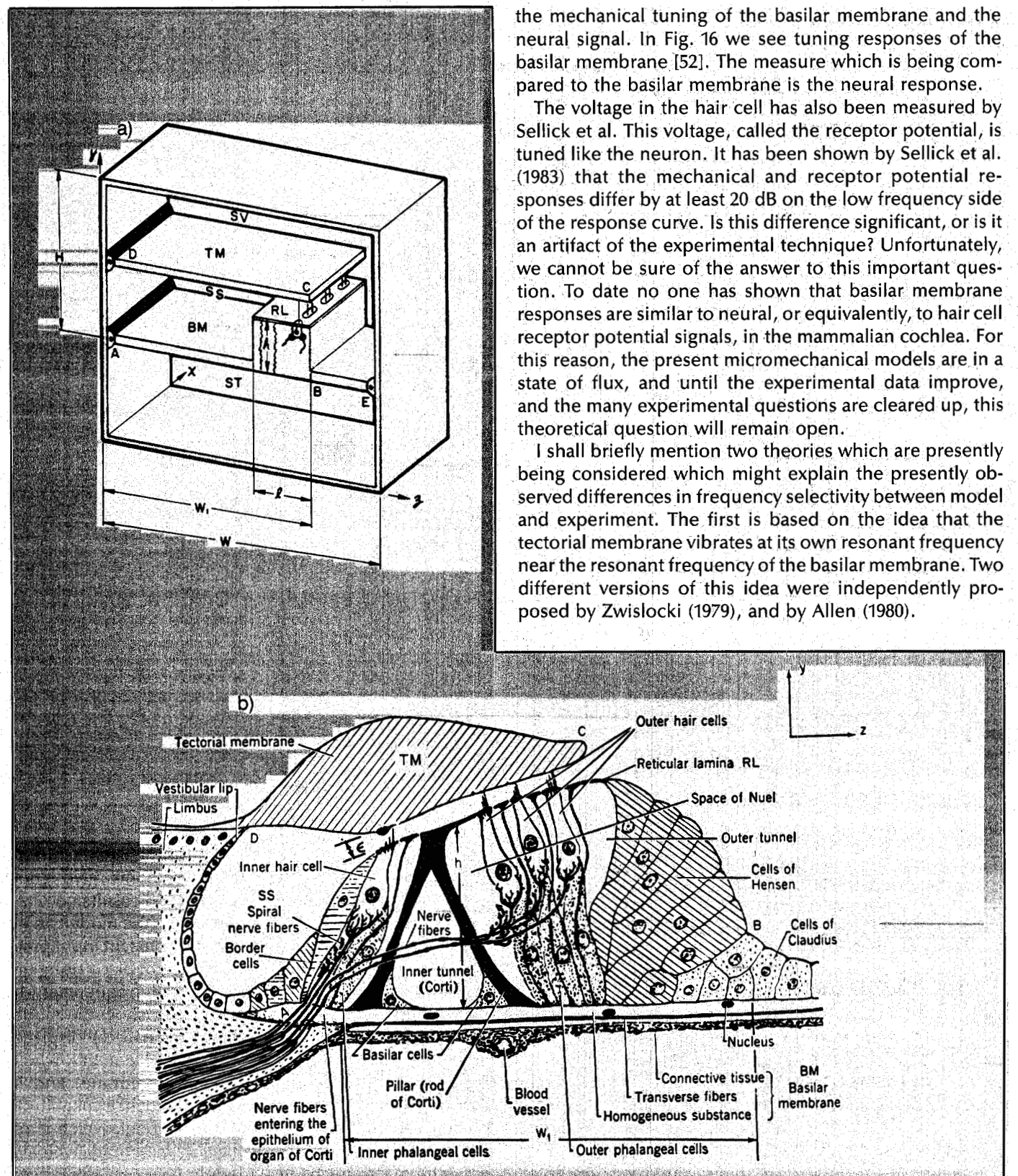

Figure 15. We show here a labeled three-dimensional representation of the previous figure (a) and compare it to a detailed labeled drawing of the organ of Corti. The dimensions labeled define the variables used in modeling the transduction system. The inner hair cells seen in (b) are the transducers that signal the central nervous system (CNS). The purpose of the outer hair cells is still unknown other than the obvious structural one. The cilia length of the outer hair cells define ϵ, the sub-tectorial space. The neurons connected to the outer hair cells are largely part of the efferent system. It has been shown that the CNS can modify, to some extent, the mechanical properties (e.g. the stiffness) of the outer hair cell cilia. The inner hair cells on the other hand appear to be passive displacement detectors. (Reprinted from Allen (1980).

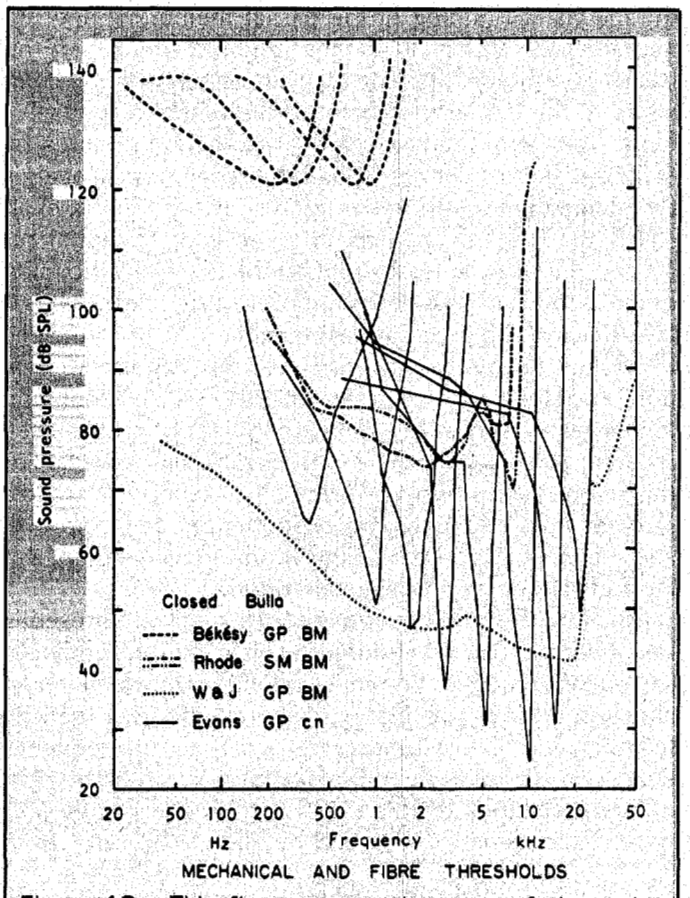

Figure 16. This figure summarizes one of the major problems in hearing research today, namely the observed difference between the measured basilar membrane frequency response and neurally measured frequency response. The problem is that most measurements of basilar membrane frequency response are not as sharply tuned as neural responses. In the figure the solid curves are guinea pig neural tuning curves. The other dashed curves are basilar membrane data from guinea pig and squirrel monkey. In the data of the figure we see the pressure at the input required to produce a fixed constant output. This way of plotting the data, for a linear system, gives the reciprocal of the usual transfer function. (Reprinted from Wilson, 1974).

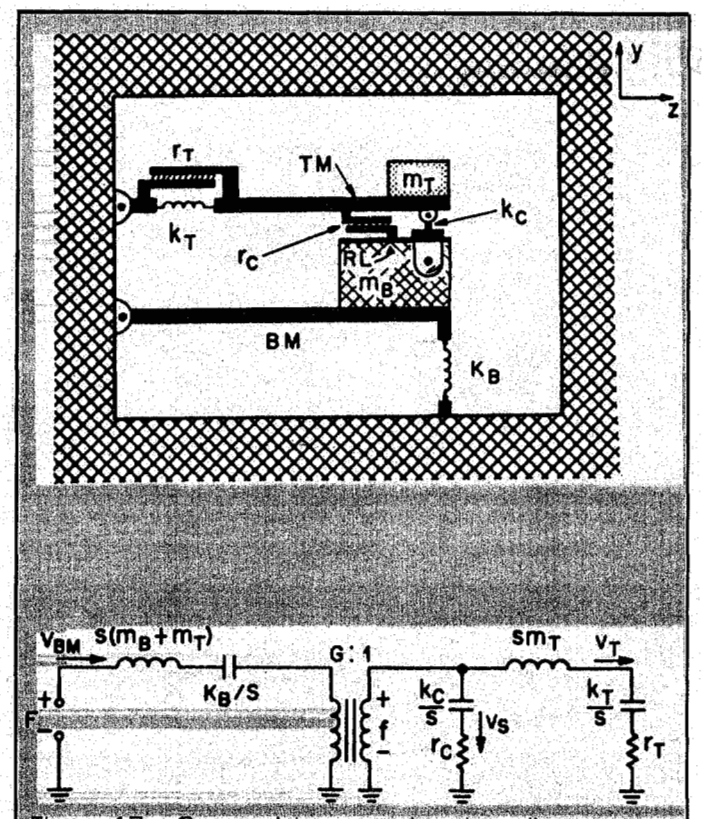

Figure 17. One model assumption that allows one to match neural tuning data, given the transmission line model, is to introduce a spring, or elastic element, in the tectorial membrane (element k_T of the figure). The addition of this element gives rise to poles and zeros in the transfer function between the basilar membrane displacement and the shear seen by the hair cells. This may be easily shown by studying the electrical equivalent circuit given in the lower part of the figure. (Reprinted from Allen, 1980).

In Fig. 17 we see a model extension of the ter Kile model where the tectorial membrane is given a new degree of freedom to vibrate in the radial direction [2], depicted here as the z-direction. By analyzing this mechanical circuit one may show that zeros must exist in the transfer function between basilar membrane motion (up-down motion) and the shear driving the hair cells (the space containing resistor r_c). We assume that this complex zero pair accounts for the low frequency difference observed between basilar membrane motion and neural response. In Fig. 18a we show cat neural tuning data for several neural units, and in Fig. 18b, a model result using the two-dimensional micromechanical model coupled to the resonant tectorial membrane model. In the model calculation we have held the model output response constant, and plotted the input pressure at the ear canal. Intermediate model results (not shown) for the cochlear input impedance and the cochlear microphonic give good agreement with experimentally observed results.

In Fig. 19 we show four measures from the model: In (a) we see the model neural output given constant input pressure in the model ear canal as a function of place along the basilar membrane for 6 different input frequencies. In (b) we show the neural phase. In (c) we show the model transfer function magnitude relating the hair cell displacement and the basilar membrane displacement which results from the resonant tectorial membrane, and in (d) we see the basilar membrane impedance magnitude for the resonant tectorial membrane model, which is required when calculating the basilar membrane velocity using the macromechanical model.

From Fig. 18 it is clear that the model does a good job of describing the neural data. Note that the resonant tectorial membrane transfer function (Fig. 19c) has a significant effect on the response for frequencies below the cutoff frequency.

A second alternative theory, designed to account for sharp neural tuning, has been proposed by Neely and Kim

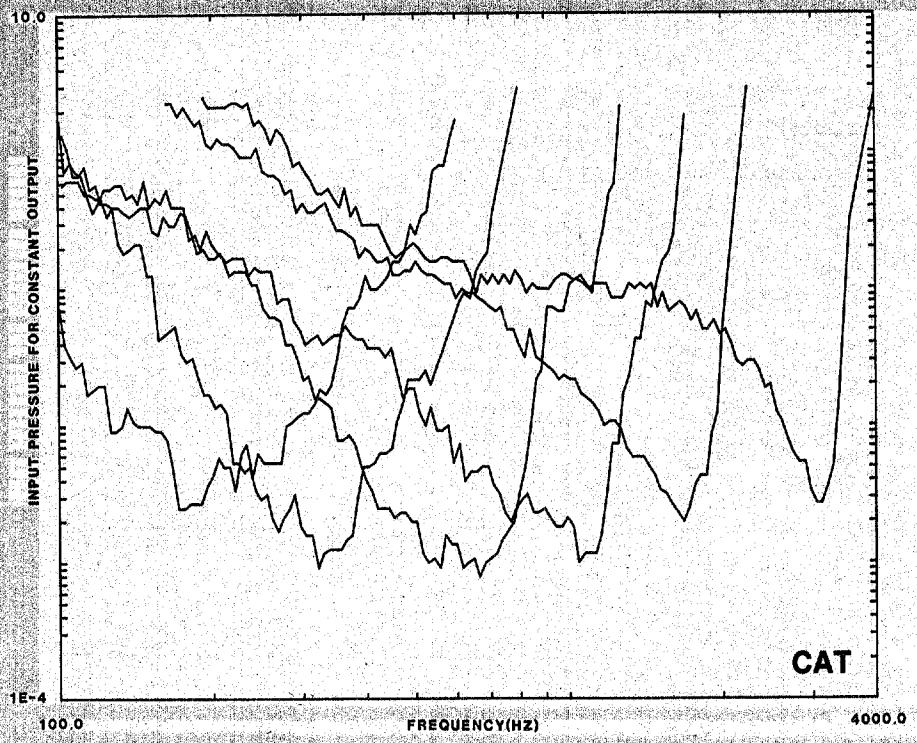

Figure 18a. We show here six cat low-threshold tuning curves which are equally spaced on the frequency axis. We only display units having characteristic frequencies between 100 Hz and 4 kHz because this is the important frequency range for speech communication. No similar data are available for humans. However, all known mammals give similar results. Note the amplitude range of the plot which covers 10^5, or 100 dB.

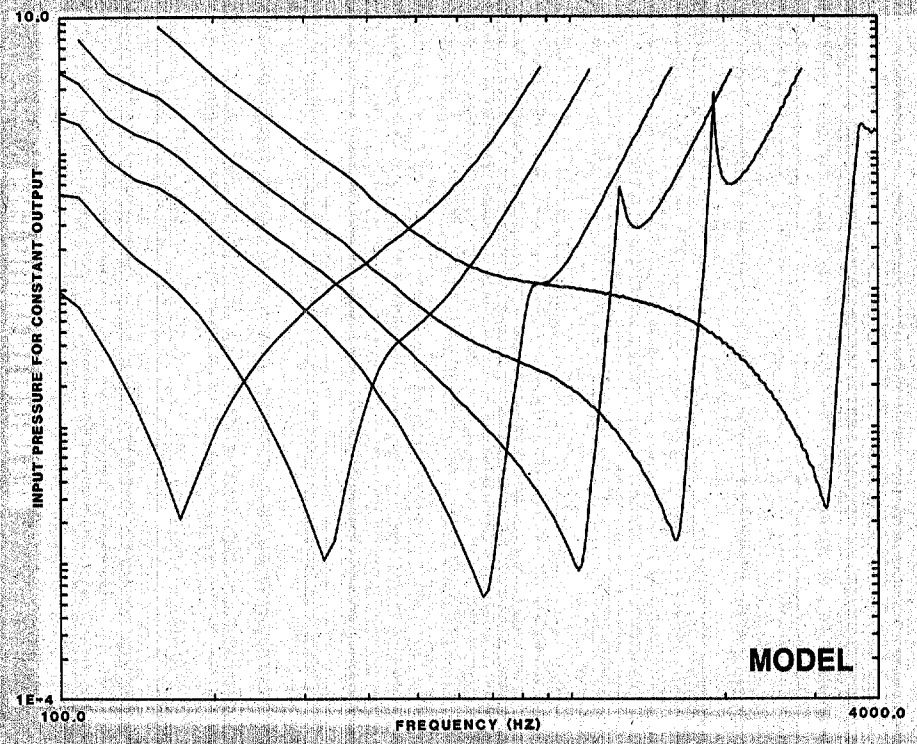

Figure 18b. The cochlear and middle ear models are used to simulate the ear canal pressure for a constant output, which here was assumed to be the shear displacement of the tectorial membrane-reticular lamina. The model calculation was done in the frequency domain with a linear two-dimensional cochlear model. The basilar membrane model is that defined in Fig. 17. The middle ear model is the same one described in Fig. 5. More details are given in the next figure.

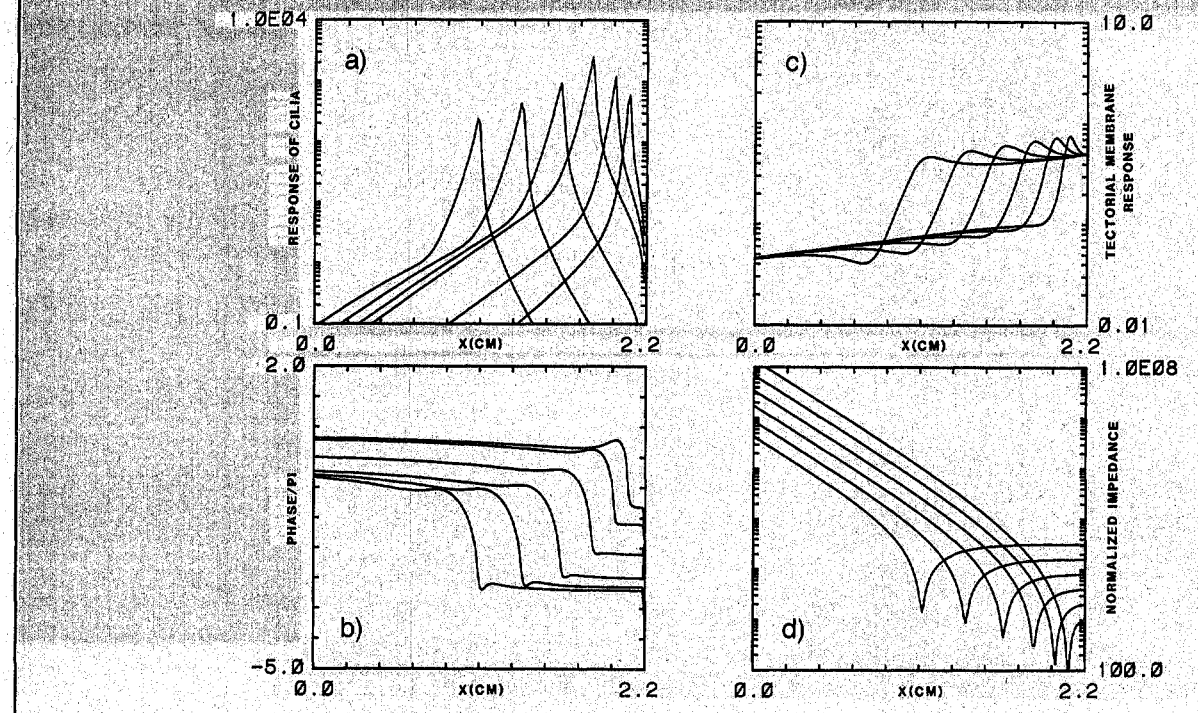

Figure 19. This figure is for a different set of model parameters, and shows how the responses vary as a function of position along the basilar membrane, for six different frequencies. Panel (a) gives the shear displacement; (b) shows the shear phase, (c) shows the basilar membrane to shear displacement transfer function magnitude, and (d) the assumed basilar membrane impedance function magnitude normalized by the basilar membrane mass. Note the large effect of the spectral zero in (c). The effect of this zero is to change the low-pass basilar membrane transfer function into a band-pass filter as seen in Fig. 18b. The fourth order impedance function (d) differs only slightly from the usually assumed second order impedance. The basilar membrane impedance is defined as the complex ratio of the pressure across the basilar membrane to its velocity. The shunt elements of the transmission line model of Fig. 11 represent that model's basilar membrane impedance.

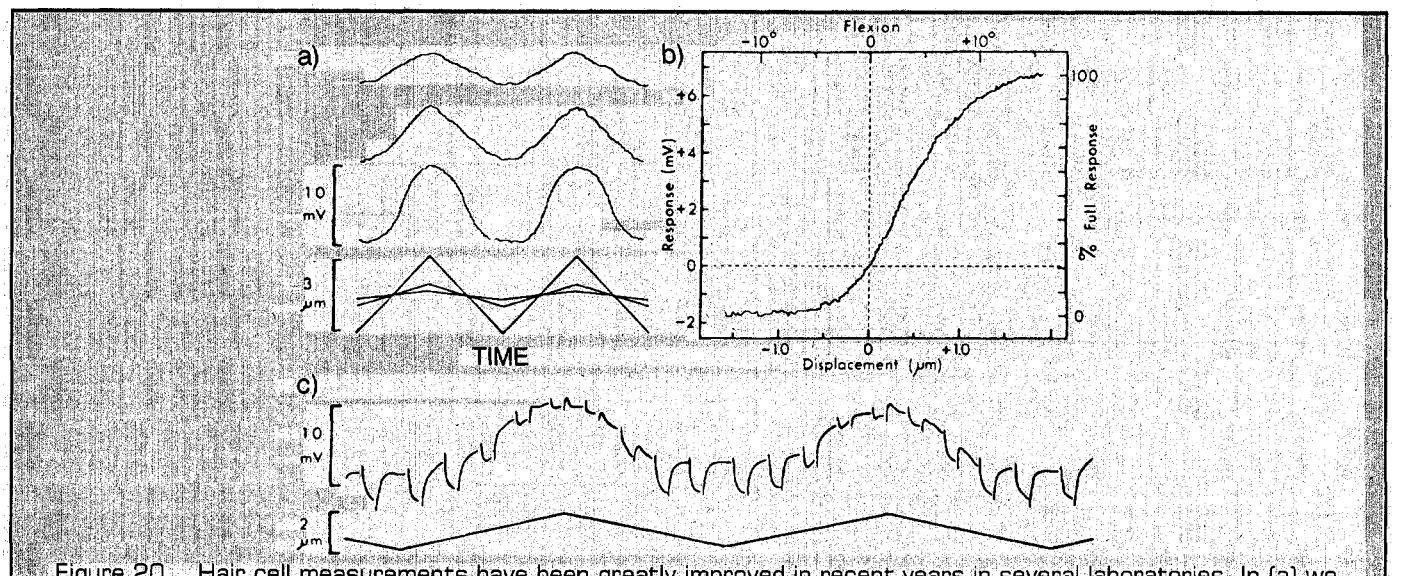

Figure 20. Hair cell measurements have been greatly improved in recent years in several laboratories. In (a) we see the internal voltage time wave form of a hair cell in response to three different displacement amplitudes. In (b) we see that an instantaneous relationship exists between the voltage and the input displacement. This cell is sensitive to less than one micron of displacement. The cell shown here is not a mammalian hair cell, and different hair cell types are likely to have different sensitivities. In (c) we see the receptor potential with a superimposed current injected by the measurement electrode. Since the superimposed square wave changes in magnitude, one may deduce that the electrical resistance of the cell is changing as a function of the input displacement. These data verify the model proposed in 1957 for hair cell transduction by H. Davis, shown in (a) of the next figure. (Reprinted from Hudspeth and Corey, 1977).

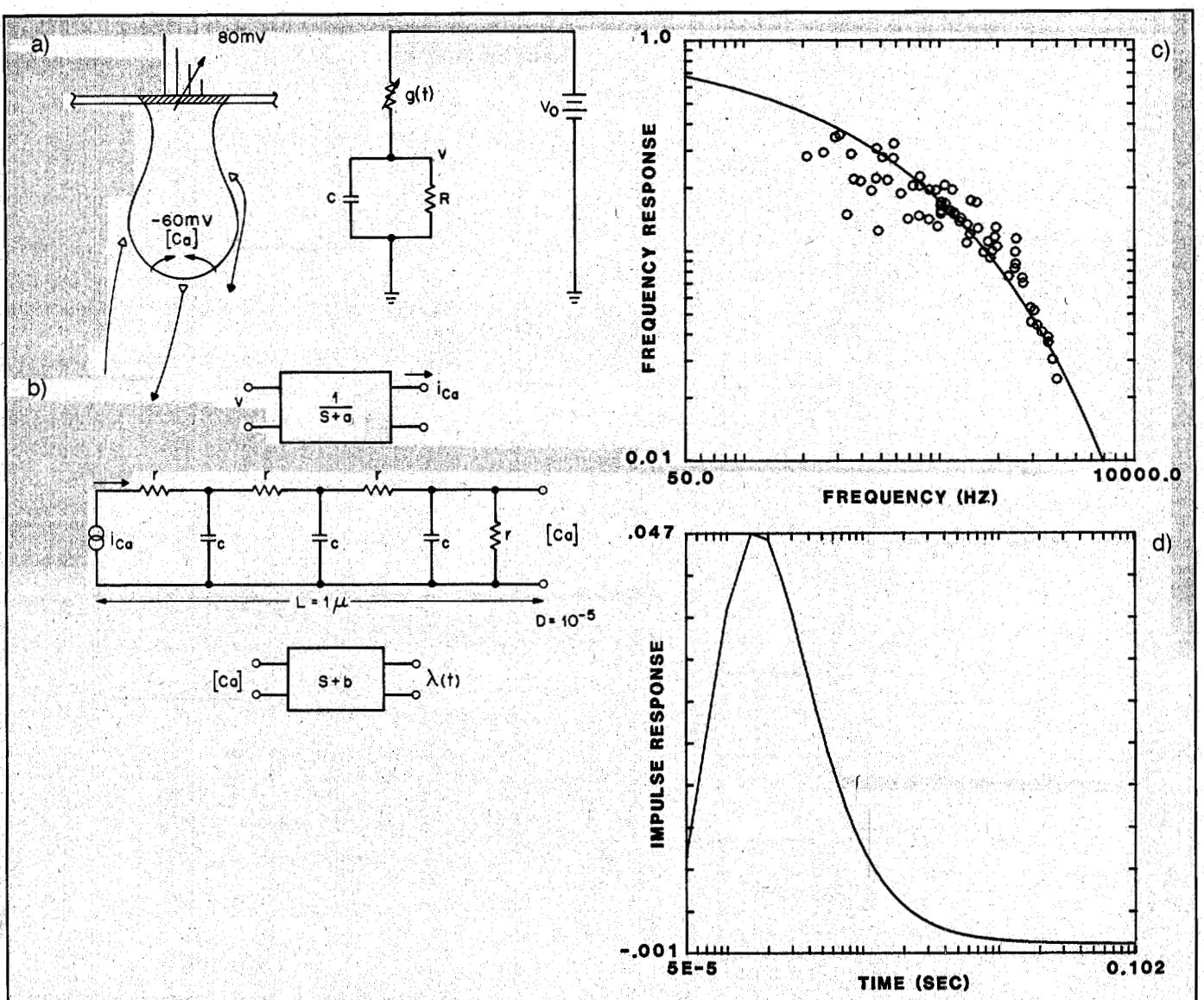

Figure 21. We present a model of a hair cell which gives a useful description of the hair cell response. In (a) is the classical "Davis" model, which has now been verified by direct experimental observation. The second part of the model has not been verified directly, but does allow one to model the neural signals. (b) We assume that the voltage in the cell opens channels giving rise to a calcium current, I_{Ca}. The leaky integrator shown ($s = \sigma + j\omega$) is assumed to have a fast time constant (≈.1 ms for example). The current then diffuses a distance equal to the diameter of the synapse, which is taken to be 1 micrometer, with a diffusion constant of calcium in solution. The diffusion process may be represented by an RC transmission line, which acts as a low-pass filter on the calcium concentration. We then represent the transformation from calcium to postsynaptic voltage, which drives the neuron, as a "leakly" differentiator. This last step is a phenomenological model, but seems to be useful from an input-output functional viewpoint. This last step might be viewed as depletion of vesicles (packets of neurotransmitter) around the presynapse as they are used up during a step of input signal. The frequency response of the RC diffusion line for the parameters described is shown as the solid line of (c). The open circles are a measure of the normalized phase locked component of the neural response as measured by Johnson (1980). In the lower panel (d) the impulse response of the diffusion line is shown, on a logarithmic time scale.

(1983), and has been worked out in some detail by Neely (1981) in his PhD thesis. This model calls upon the concept of negative damping, or resistance, in the basilar membrane. In this model the neural response and the basilar membrane response are identical, unlike the resonant tectorial model described above, and therefore the original ter Kiel model is used unmodified. In order to understand why this alternative proposal is an interesting one, it is necessary to describe some important recent experimental observations.

EVOKED ECHOS AND SPONTANEOUS EMISSIONS

Consider for a moment what one would observe if one were to pulse the vocal tract at the mouth, and then observe the response at the same point. Because of the

acoustic delay of the vocal tract, reflections would be seen over a time scale given by the travel time of the sound in the tube. Only if the tube were uniform, with a zero reflection coefficient at the lips, would no reflections be observed. The same situation must exist in our transmission line model of the cochlea: since the equivalent cochlear transmission line is not uniform, it must also give reflections. However because of the nature of the smooth variations of the line along its length, computed reflections from the line have been found to be unmeasurably small (Neely, personal communication).

In fact, when one does the experiment on a human ear, low level dispersive reflections *are* observed. The delay involved approximately corresponds to a round trip travel time along the basilar membrane, and the reflections are nonlinear in their behavior, since they grow at less than linear growth rate with input level [23]. Because the observations have a nonlinear character, it will not be easy to explain or model them until the nonlinear properties of the basilar membrane are better understood.

A second, and seemingly related phenomenon, observed much earlier, was that the threshold of hearing frequently has a low level microstructure which takes the form of a quasi periodic elevation in threshold as a function of frequency. Such threshold elevations would be characteristic of low level standing waves due to slight mismatches at different positions along the cochlear transmission line.

A third somewhat bizarre observation was the finding that narrow band tones emanate from the human cochlea. In animals similar tones have been correlated with damage to the cochlea, in cases where damage has been assessed.

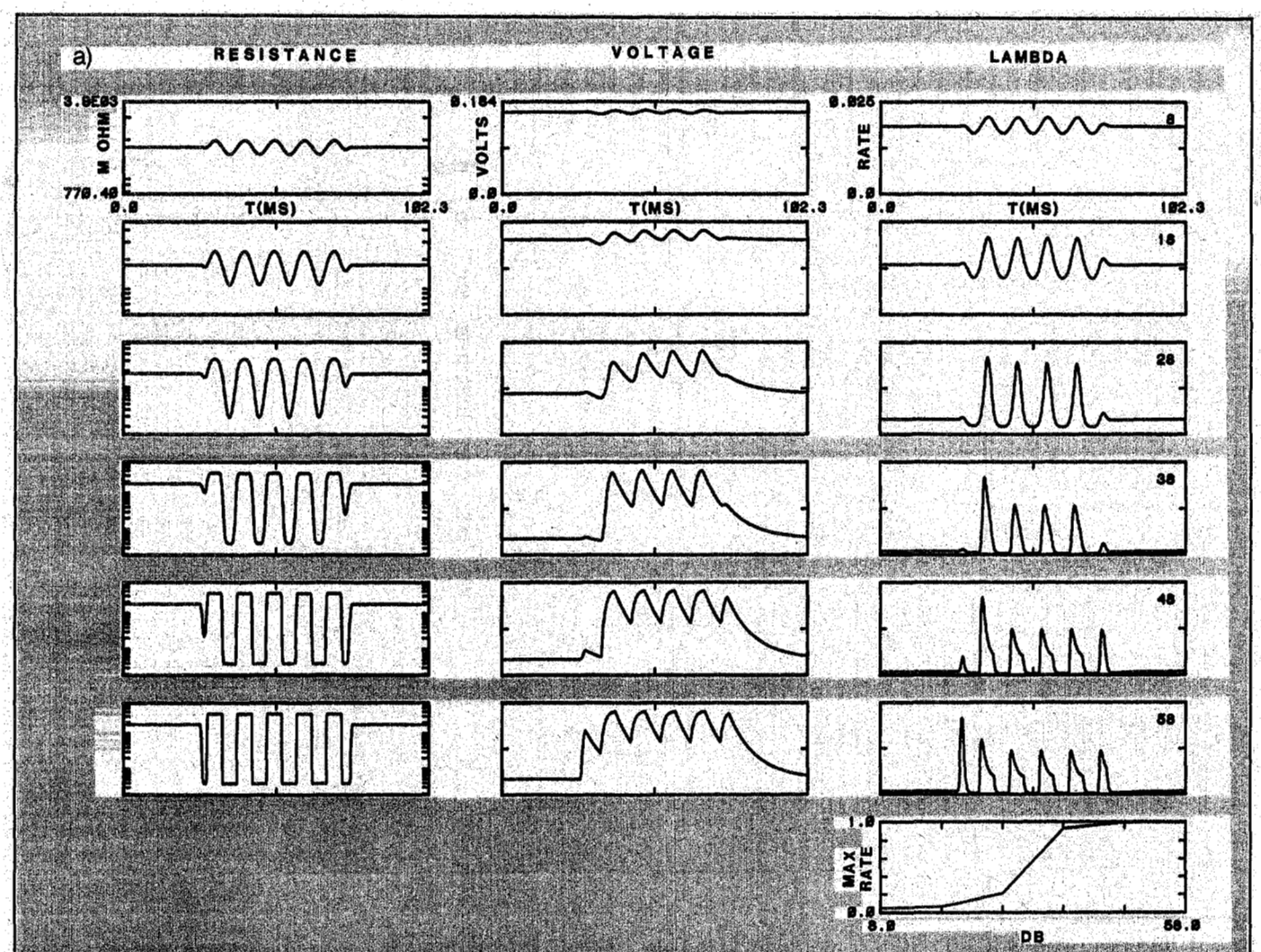

Figure 22. (a) The transduction model's response to a 110 Hz tone burst over a 50 dB level range gives an output which is quite similar to neural data (b) recorded under identical stimulus conditions. The first column of (a) is the electrical resistance change for the model. The second column is the model receptor potential. The rightmost column is the model output, which represents the instantaneous probability of finding a nerve impulse at each instant of time over the tone burst duration. The number in the rightmost corner is the signal level in dB. The bottom panel is the maximum value as a function of the input level. This last column should be compared to Fig. 22b, which is representative data collected from a cat neuron. (Allen, 1983a).

It has been suggested that these emissions are the result of very high Q resonances formed by standing waves on the basilar membrane due to reflections at a point of damage on the basilar membrane [24]. We note here that such a high Q resonance would be driven by any noise in the system giving rise to narrow band peaks in the ear canal power spectrum of the noise floor.

When these spontaneous emissions were first observed, many researchers were quick to conjecture that the cochlea was an active system which occasionally could become unstable. Hence models that incorporate negative damping, such as Neely's, are interesting. The use of negative damping in the model serves the function of increasing the Q of the basic resonance of the cochlear filter, and when the damping becomes too small, it is proposed that the system begins to oscillate. An apparent major failing of the "active" hypothesis is the form of the experimentally observed cochlear input impedance near an emission frequency. The acoustic impedance, as measured by Kemp (1979), shows increased damping near an emission, rather than a decreased damping. Also the reactive part of the impedance is consistent with the reflection model described here and by Kemp (1979). It is not presently clear what the impedance of the active model should be as a function of frequency since these active models have not yet been fully specified. For a presentation of the active model point of view, the reader should look at the papers [15, 12, 33, 25].

I do not think that any of these phenomena disturb the fundamental scenario of the cochlea performing its basic function as a linear filter bank. These low level acoustic effects are more likely an epiphenomenon, which while interesting to study, are not functionally important when taken in isolation. In any case, data on this question are rapidly accumulating.

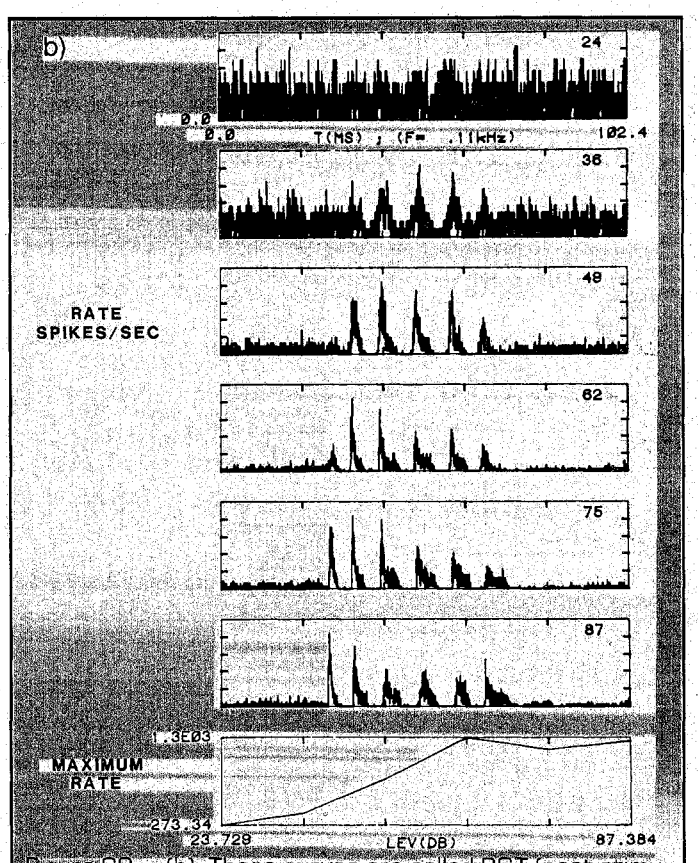

Figure 22. (b) These curves are called PST (post stimulus time) histograms. To form these data, each spike time determines the histogram bin to be incremented, computed as that time modulo the stimulus period. The sound pressure level, in dB SPL, is given in the upper right hand corner of each plot. The bottom panel is the maximum rate as a function of the input level. The model rate level function, and cat rate level functions are quite similar. Similar comparisons for tone data over the entire range of levels and frequencies between 100 Hz and 20 kHz produce similar agreement (Allen, 1983a).

COCHLEAR TRANSDUCTION

The third stage of processing of the signal is the transduction stage, which follows the micromechanics. We now believe that the mechanical to electrical transduction takes place at the inner hair cell site. This transduction is assumed to take place in the following way.

When the stereocilia of the inner hair cells are displaced, their electrical resistance is changed by the mechanical stimulus. The relationship between the mechanical input and the electrical response is that of a simple half wave rectifier, as is shown in Fig. 20. The data of Fig. 20 have been taken from the paper of Hudspeth and Corey (1977).

A potential of about 120 to 140 millivolts is biologically maintained across the cell. Thus, the current flow into the cell body is modulated by the mechanical displacements. This in turn modulates the voltage within the cell body. The equivalent electrical circuit model of a hair cell is shown in Fig. 21a. This model of transduction was first proposed by H. Davis in 1957, and has since been called the Davis model. A great deal of experimental data supports this model [10].

The accepted chain of events, beginning with the cell potential change, is as follows. The change in receptor potential modulates the opening of calcium channels at the bottom of the hair cell. When the channels open, calcium diffuses into the cell, and triggers the fusion of packets filled with neurotransmitter substance onto the cell wall. When these vesicles of neurotransmitter fuse with the haircell wall at the synapse, the transmitter substance is released into the synaptic junction, thereby changing the potential of the postsynaptic cell. This change in potential then leads to a modulation of the firing pattern of neural pulses on the nerve fiber leading from the synapse.

A recently proposed model which summarizes this latter sequence of events is shown in Fig. 21b [3]. First we model the calcium current by a leaky integration, and we assume that the calcium must diffuse the diameter of the synapse, which is about one micron. This diffusion process is represented by the RC transmission line of Fig. 21b. The vesicle production is modeled by the last box which represents a

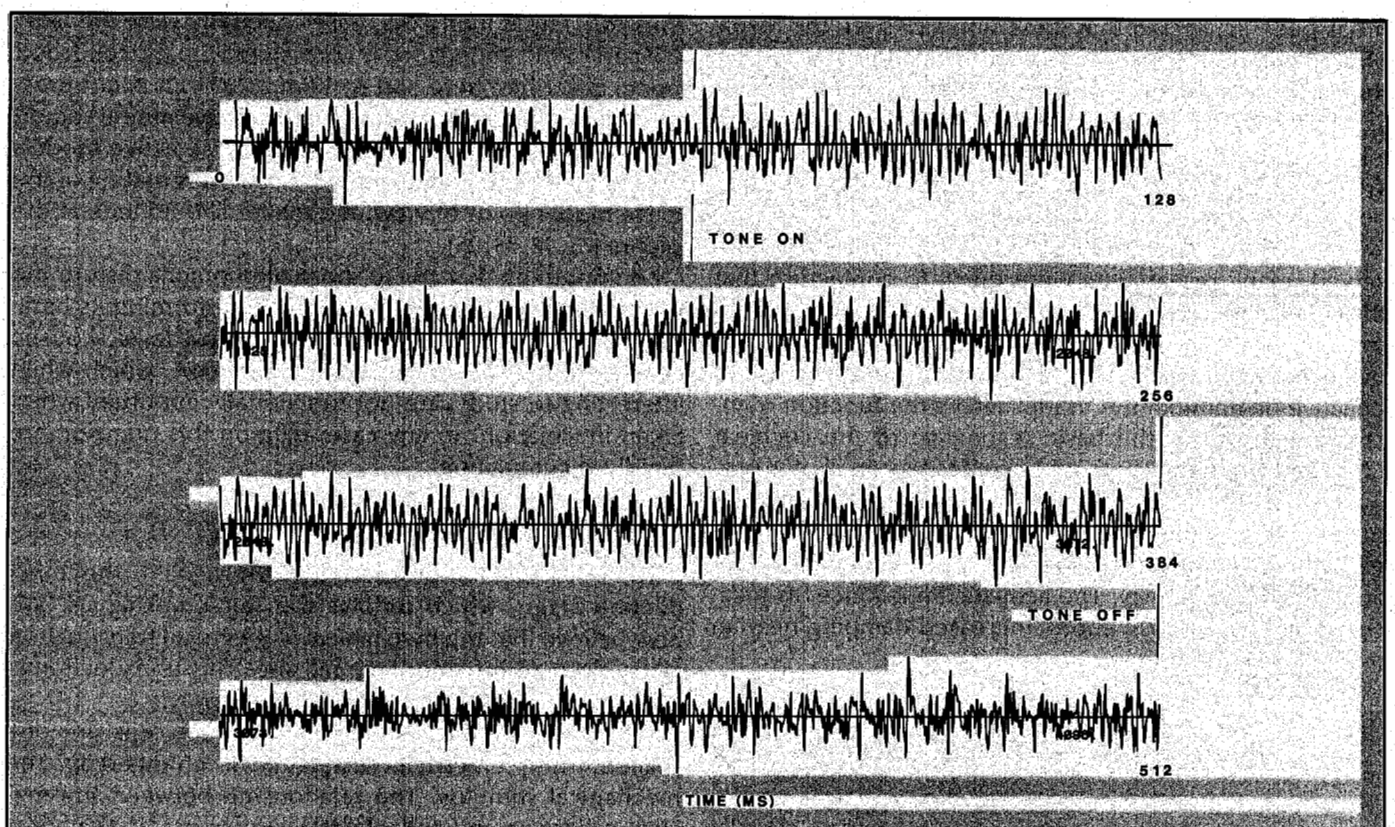

Figure 23. We begin here to describe the overall properties of the model by using a stimulus consisting of an 800 Hz tone burst imbedded in wide-band noise. The signal to noise ratio is 0 dB. Thus it is not easy to visualize the tone waveform.

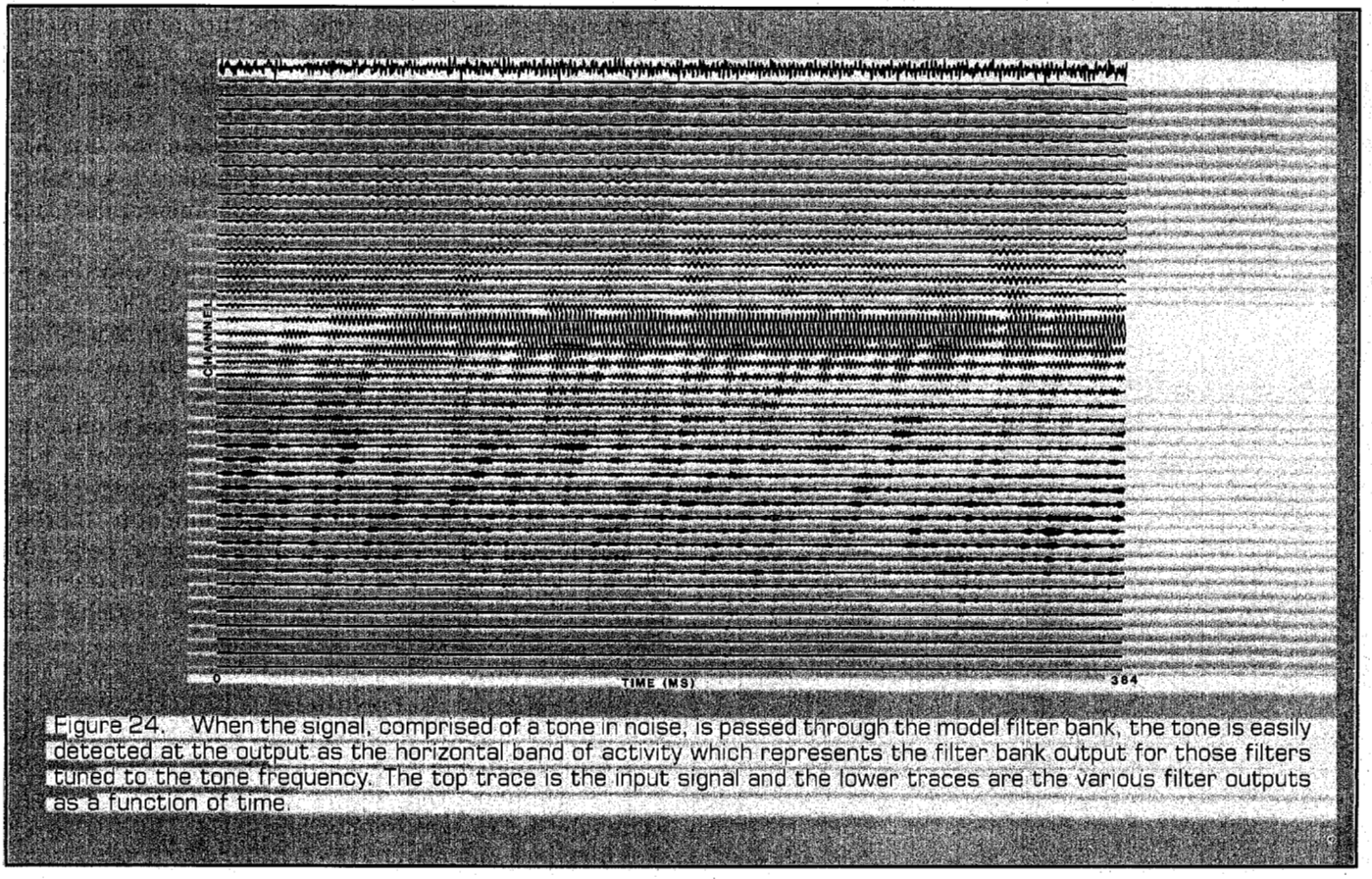

Figure 24. When the signal, comprised of a tone in noise, is passed through the model filter bank, the tone is easily detected at the output as the horizontal band of activity which represents the filter bank output for those filters tuned to the tone frequency. The top trace is the input signal and the lower traces are the various filter outputs as a function of time.

leaky differentiation. The output of the model is $\lambda(t)$ which is intended to represent the probability of finding a nerve spike at time t. The diffusion line acts as a low pass filter having the response shown in Fig. 21c, d, with the open circles showing Johnson's (1980) quantification of the magnitude of this low pass filter. The length of the model diffusion line was varied to best fit Johnson's data points, resulting in the one micron length estimate. The hair cell model's response to a tone burst is shown in Fig. 22a while experimental data from a cat [3] are shown for comparison in Fig. 22b.

Several other hair cell models may be found in the literature. The early model of Schroeder and Hall (1974) is important and has inspired others to pursue the problem. Also of interest is the work of Smith and Brackman (1982), who have studied several deficiencies of the haircell models experimentally.

The transduction model is important because it is a

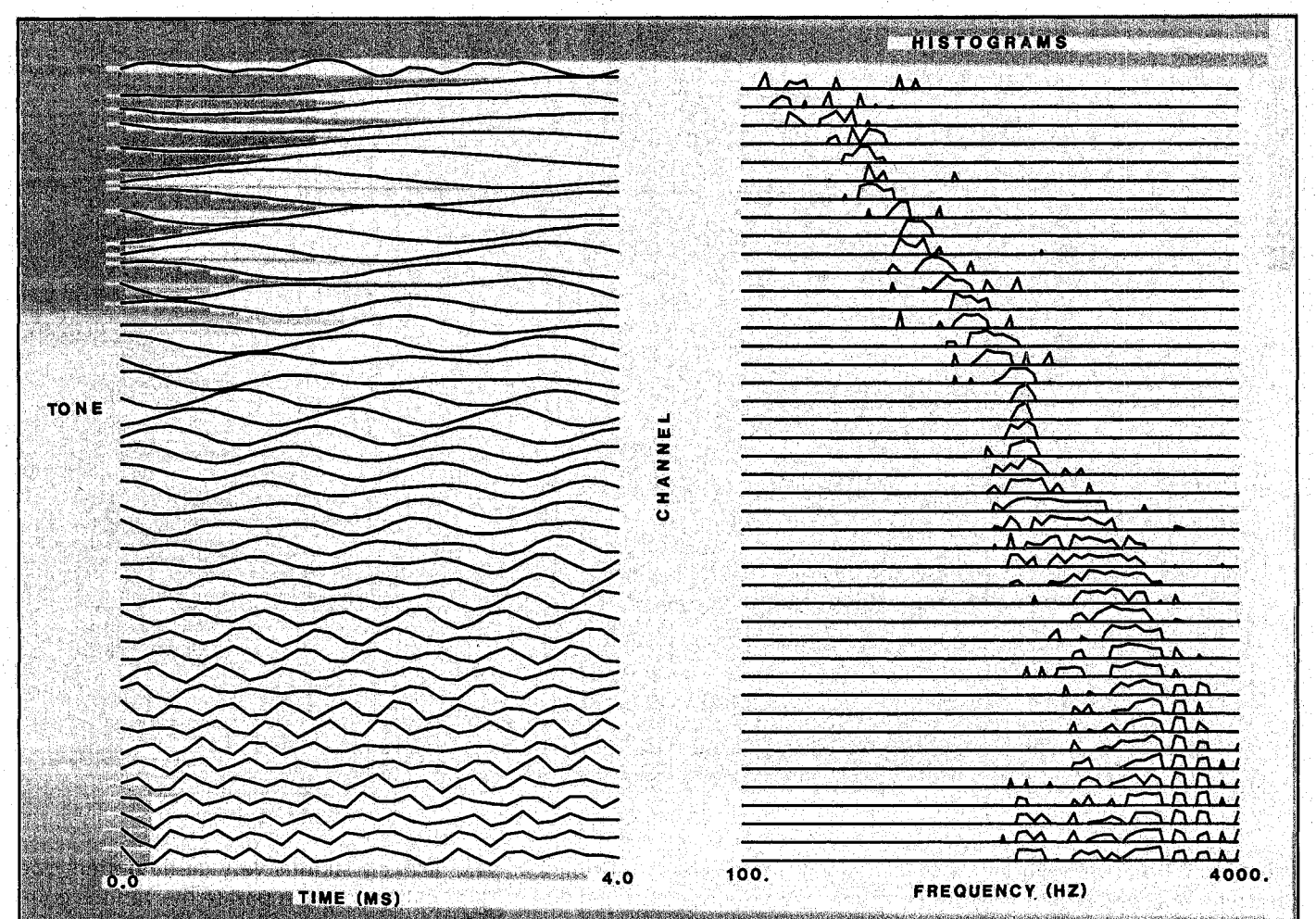

Figure 25. We now ask the question: "How can we display and enhance the features of the signal, given the representation found in the auditory nerve?" Based on the model of the hair cell, we assume here that the information is carried by the zero crossings of the multitudinous narrow-band signals. This is because the hair cell cilia appear to act as a switch, given moderate and high level signals, transforming the signals to peak-clipped signals. In an infinitely peak-clipped signal, the information is coded by the zero crossings. The method used here to extract this information from the zero crossings is to construct histograms of the intervals between zero crossings, as shown on the right of this figure. The left panel are the filter outputs, as in the previous figure. In those channels for which the tone dominates, the intervals are determined by the tone alone, and are therefore nearly identical. For those channels that are driven by noise, the interval histograms peak at a delay corresponding to the reciprocal of the characteristic frequency of each cochlear filter. In the right panel of this figure, we show histograms of the reciprocal periods, which results in a frequency scale rather than a time scale. The top trace shows the histogram for the lowest frequency channel, while the bottom trace is for the highest frequency filter. Note that in the middle traces, several channels have similar histograms. These are the channels responding to the tone burst. We may now enhance the tone component by taking the pointwise product of each trace with its neighbor, and then sum vertically over all the products, maintaining the abscissa as the independent variable. In this way we collapse the entire set of histograms to a display which is similar to one section of a spectogram. This procedure takes advantage of the orthogonal nature of uncorrelated neighboring channels. Only when two contiguous channels have a similar response, will their product be non-zero, as in the case of those channels responding to the tone.

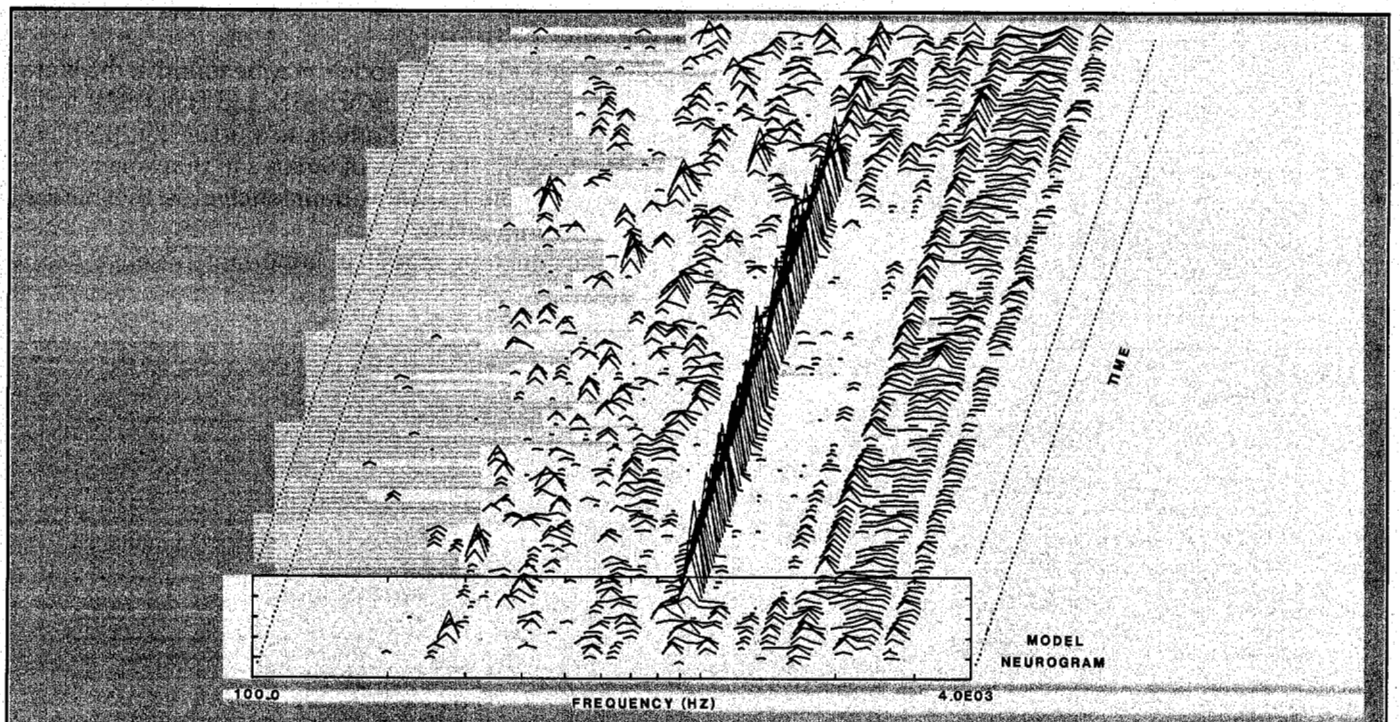

Figure 26. When we look at the "neurogram" generated by the method described in the Fig. 25 figure caption, the tone is clearly present in a low correlation background noise. Each horizontal trace corresponds to the sum-over-histogram products described in the previous figure caption. Recall that this display has been constructed from the zero crossings of the many narrow band filter outputs. An interesting property of the zero crossing representation of the neural signals is the so-called "FM capture" effect. When the tone dominates the zero crossing histogram, the effect of noise in that channel is suppressed. Such a capture effect has been previously observed in auditory nerve responses in many publications. Arthur et al. (1976) found such a suppression effect of one tone by another by looking at the Fourier transform of the period histogram (PST) of the two-tone complex, as the amplitude of one of the tones was varied. Recently Javel et al. (1981) characterized a similar effect, which they named "two-tone synchrony suppression." Sachs and Young (1980) and Giesler and Sinex (1982) also have described this effect given speech-like stimuli. In this figure we see the upward spread of masking of the noise background by the tone. In the model calculation, the masking is one-sided (it masks higher frequencies, but not lower) due to the asymmetrical shape of the filter frequency response. A tone can capture more filter outputs tuned higher in frequency than tuned lower in frequency because the filter slope is greater above the best frequency than below. Such a qualitative explanation of auditory masking has been accepted for many years [40, 47]. Thus the presence of masking in our model output might be expected, in retrospect. What came as a surprise to this author however was the significance of the effect in the model at low frequencies, with speech as the input. (See Fig. 29).

major nonlinear transformation on the signal which can improve our understanding of how the transduction mechanism functionally codes the information. The model described by Fig. 21a, b has been tested for tones and tone bursts as the stimulus. This model does not describe how the neural point process is generated, but instead describes the response of the neurons in terms of their underlying probability of firing $\lambda(t)$. One justification for adopting this description is that a large number of neurons leave each hair cell (typically 20 neurons leave the bottom of each inner hair cell); thus we have assumed that the information coded by the receptor potential is transferred via this ensemble of twenty neurons, rather than by the point process of an isolated neuron.

SOME EXAMPLES—A TONE IN NOISE

We may now follow signals through a cascade of the various component models to show how the system responds.

The linear part of the system is very easily understood since the cochlear filters simply separate the various components in a predictable (linear) manner. The response of the transduction stage however is more complicated and requires some explanation.

The first example is for a tone of 800 Hz imbedded in noise (Fig. 23). The outputs from the model are shown in Fig. 24. One interesting and important feature of the data is that because the filters separate the tone and noise, all of those filters which respond mainly to the tone carry the same waveform. Thus those filters tuned to the input tone frequency respond identically, while the other filters respond in random ways, since the noise is largely uncorrelated in bands.

The cochlear model, by its very nature, must be in some sense "bandwidth expanding," since the number of channels coming from the cochlea is so large. An important question that must be answered is: how can we code the output of the model for the special case of a speech input?

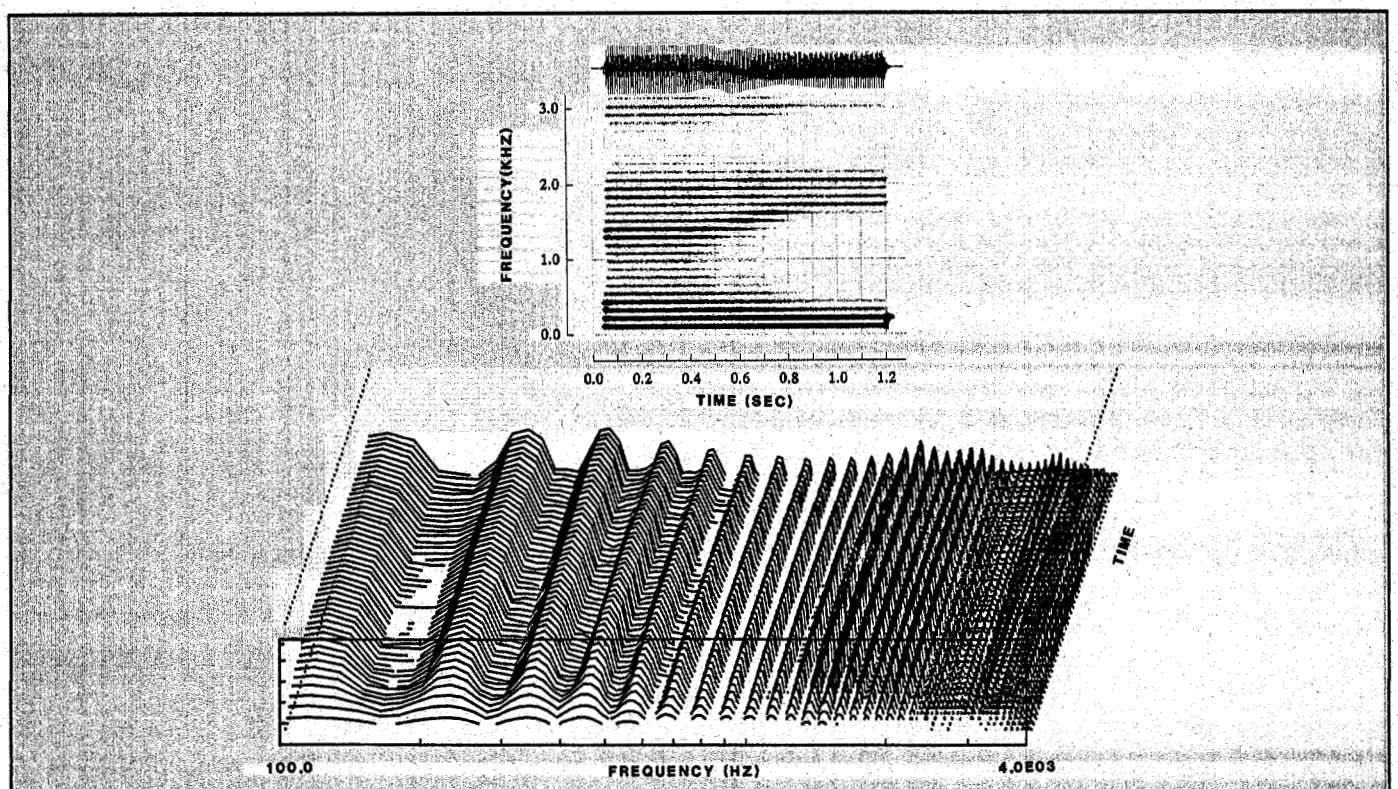

Figure 27. We see here a standard narrow-band speech spectrogram of a synthetically generated vowel pair. The pitch in this example is held constant at 105 Hz. The formant frequencies from 0.0 to 0.4 secs are 352, 1360, 1968 and 2960 Hz and from 0.8 to 1.2 secs are 216, 1736, 2052 and 3040 Hz.

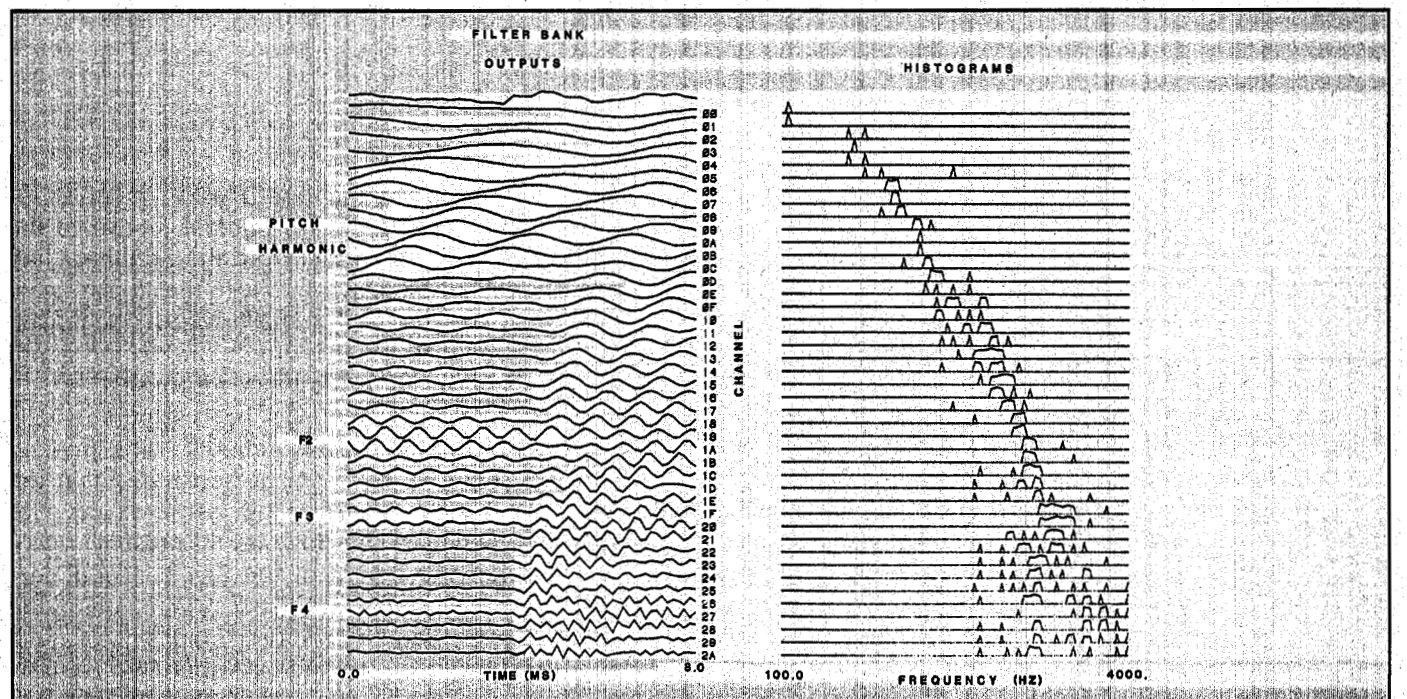

Figure 28. When the speech signal is passed through the model, the formant frequencies and pitch pulses are evident in the filter-bank output, as seen in the left panel. The formants are seen as ringing for those filters tuned to the formant frequencies. The higher the formant *Q*, the longer the duration of the ringing. For high-frequency channels, the pitch shows up as a periodic pulse excitation because many spectral components fall within one cochlear filter channel. At low frequencies, the cochlear filters resolve individual spectral pitch lines. This effect is not obvious in a single frame of frequency histograms, as seen on the right panel of this figure, but only becomes clear in the spectral display after taking products of neighboring histograms as described in the Fig. 25 caption.

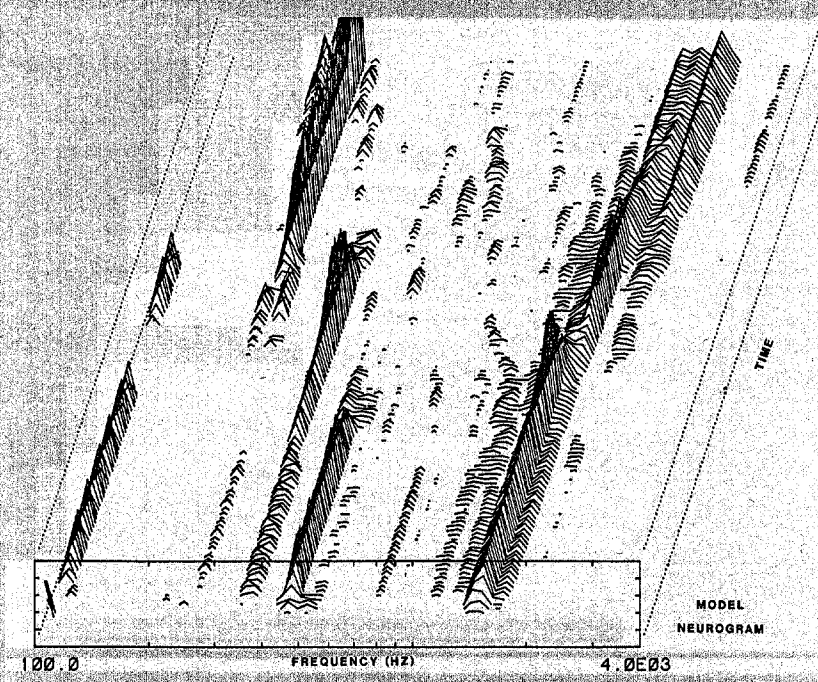

Figure 29. Many interesting effects may be seen in this model "neurogram" of the synthetic speech. First F_2, the second formant, is very clearly represented at 1360 Hz. However, the third formant at 1968 Hz seems to be masked because it is weaker than the second formant. A similar effect may be seen for the pitch harmonics at lower frequencies where the strongest pitch component dominates the output. For example, in the first steady state segment, the 4th harmonic at 420 Hz dominates all other components down to, but not including the first. In the later portion of the plot, the second harmonic dominates. Such masking can result from only 6 dB of emphasis by the formant. This analysis raises again an old question of perceptual interactions between pitch and the first formant. If masking effects in the auditory nerve are as significant as they are in this model, then the response in the nerve has been significantly simplified by the masking effect, with an enhancement of the stronger formants by the nonlinear effects of the hair cell.

For example, is it possible to classify the signals at the output of the model, thereby reducing the information rate? The number of possible ways of doing this is probably large, and in that sense the question is open ended. However one proposal that we have explored seems promising, and we shall describe it here.

Suppose we build time-varying interval histograms for each of the channels of the model. If the signals in two neighboring channels are identical, then the periods will be identical, and the histograms will peak at the same delay value as shown in Fig. 25. These similarities may be drawn out by computing the statistic of the sum over pairwise products of the histograms. These products will go to zero when the correlations are small, but stay large when the neighboring channels are similar. In Fig. 26 we show such a 'spectrogram' of correlations., Those channels responding exclusively to noise only show small correlations, and approach zero in the spectrogram. This measure is almost totally independent of the signal input level because of the nature of the hair cell model transducer which removes all of the level information above its threshold (since the model essentually codes zero crossings).

From Fig. 26 we see that the tone "masks" the noise for frequencies above the tone. This effect is similar to the "FM capture" effect and results from the nonlinear hair cell detectors which can respond only to the largest signal (the tone) at the expense of the smaller signal (the noise). The upward spread of this masking is due to the asymmetry in the cochlear filter frequency responses.

Such an upward spread of masking is experimentally observed in psychophysical experiments using masked tones. At this point we have not demonstrated that the masking observed in Fig. 26 is the same as psychophysically observed masking. However, its presence is not unexpected given the nature of the hair cell nonlinear transduction process which acts like a switch, coding the zero crossings of the band-pass filtered input signals. Such a capture effect has been studied by Arthur (1976) and by E. Javel (1981) and his colleagues, which they described as "synchrony suppression." Synchrony suppression should not be confused with two-tone rate suppression, which is a neurally observed nonlinear effect with quite different properties [4].

The masking effect described here is a direct result of the switch-like response of the hair cell cilia to multi-

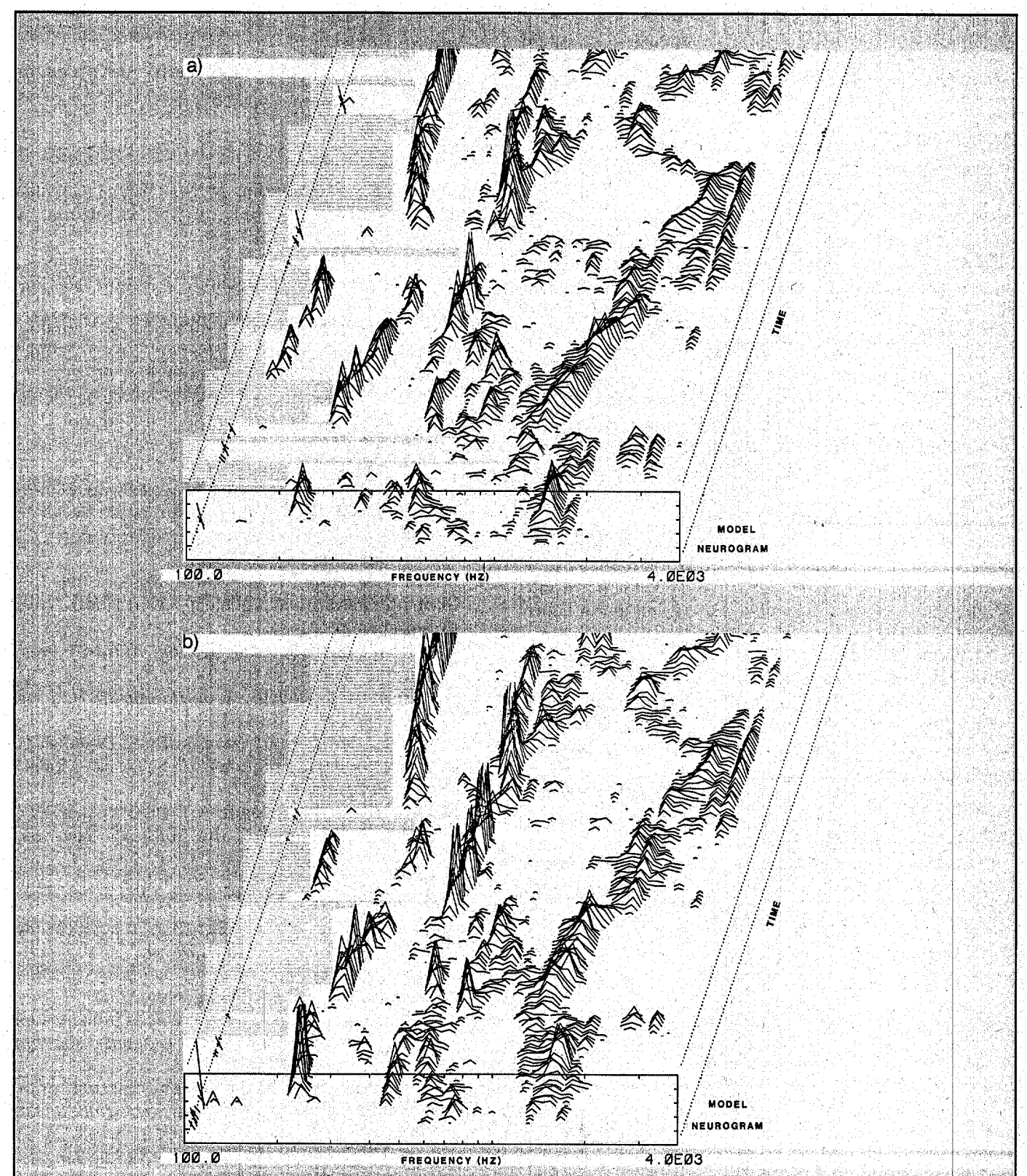

Figure 30. When model "neurograms" are computed using speech signals, the formants are enhanced as in the synthetic speech sample. The upper panel is a male speech sample (data from Fig. 4a) analyzed by the model. The lower panel is the same sample. However, the speech has been reverberated by the impulse response of a small room (data from Fig. 4b). Because of the reverberation, the features are smeared together in time when the speech energy is low, as in Fig. 4b. Although degradations are evident, many features are robustly represented. Are these robust features the perceptually significant features? Much more research is needed to answer this question.

component input signals. We shall show in the following examples that this masking effect appears to be quite significant in the speech coding process.

SPEECH INPUT

In Fig. 27 we show a traditional narrow-band spectrogram for a synthetic vowel-vowel pair /ai/. The time waveform is shown at the top. The two sets of formant frequencies are defined in the figure caption; the pitch is slightly greater than 100 Hz, and has been held constant. In the next figure, Fig. 28, we see the output of the model filter bank on the left, and the histograms of τ^{-1}, or inverse zero-crossing periods, on the right. This panel is similar to that of Fig. 25, and shows histograms, for each channel output, of the reciprocal of the zero crossing intervals. By looking at the zero crossings, we approximate the model haircell output for large input levels. Labels indicate the various formant frequencies and the pitch period epoch for this frame. The waveform on the top left is the input to the model.

When we form the model neurogram for the speech signal we find a great simplification, as seen in Fig. 29. Because of the masking of one pitch component by a stronger neighbor, only the greatest local harmonic component remains. More details are given in the figure caption.

Finally in Fig. 30 we show the model output for our reverberant speech sample of Fig. 4b. Although the reverberation has an effect on the details of the spectrum, many of the basic features are still clearly seen in the model spectrum such as the formant frequencies and the pitch structure at low frequencies. These examples demonstrate that a certain robustness could be maintained by the cochlea with respect to reverberation and noise, assuming the model neurogram is realistic.

Because of the peak clipping nature of the hair cell transducers, the model output is largely independent of level. In fact, one serious deficiency of the model in its current form is its inability to code or quantify loudness in any robust manner.

CONCLUSIONS

The cochlear model must stop here, because the function of the cochlea stops at the auditory nerve. Where do the signals go from there? We know very little about the signal processing that goes on at the higher centers of the auditory system. It is only possible to guess what is reasonable. The models described above (up through the hair cell model) are fairly successful at describing and predicting many phenomena that can be measured in the nerve fiber for a complex signal. However, other models for information processing and coding need to be proposed, modeled, and tested. For the moment, the models as quantified in this paper may be viewed as a stepping stone toward this future effort.

ACKNOWLEDGMENTS

I would like to thank Bill Seibert and David Berkley for extensive comments on an early version of the manuscript.

REFERENCES

[1] Allen, J. B., and Sondhi, M. M. (1979). "Cochlear macromechanics—Time domain solutions," *J. Acoust. Soc. Am.* **66**, 123–132.

[2] Allen, J. B. (1980). "Cochlear micromechanics—A physical model of transduction," *J. Acoust. Soc. Am.* **68**, 1660–1670.

[3] Allen, J. B. (1983a), "A Hair Cell Model of Neural Response," In *Mechanics of Hearing*, Ed's E. deBoer, M. A. Viergever, Delft University Press, 193–202.

[4] Allen, J. B., and Fahey, P. F. (1983). "Nonlinear Behavior at Threshold Determined in the Auditory Canal and on the Auditory Nerve," In *Hearing—Physiological Bases and Psychophysics*, Ed. by R. Klinke and R. Hartmann, Springer-Verlag, New York, pp. 128–133.

[5] Allen, J. B. (1983b). "Magnitude and phase frequency response to single tones in the auditory nerve," *J. Acoust. Soc. Am.* **73**, 2071–2092.

[6] Arthur, R. M. (1976). "Harmonic Analysis of Two-Tone Discharge Patterns in Cochlear Nerve Fibers," *Biol. Cybernetics* **22**, 21–31.

[7] Blauert, J. (1983). *Spactial Hearing*, The MIT Press, Cambridge, MA.

[8] Von Bekesy, G. (1960). *Experiments in Hearing*, McGraw-Hill, New York, N.Y.

[9] deBoer, E. (1981). "Short Waves in Three-Dimensional Cochlear Models: Solutions for a 'Block' Model," *Hearing Research* **4**, 53–77.

[10] Corey, D. P., and Hudspeth, A. J. (1983). "Analysis of microphonic potential of the Bullfrog's Sacculus," *J. Neuroscience* **3**, 942–961.

[11] Davis, H. (1957). "Biophysics physiology of the inner ear," *Physiol. Rev.* **37**, 1–49.

[12] Davis, H. (1983). "An Active process in cochlear mechanics," *Hearing Research* **9**, 79–90.

[13] Flanagan, J. L., *Speech Analysis, Synthesis, and Perception* (1972), Springer-Verlag, New York.

[14] Giesler and Sinex (1982). "Responses of primary auditory fibers to a brief tone burst," *J. Acoust. Soc. Am.* **72**, 781–794.

[15] Gold, T. (1948). Hearing II. "The physical basis of the action of the cochlea," *Proc. Royal Soc. Edinb. B*, **135**, 492–498.

[16] Guinan, J. J., and Peak, W. T. (1966). "Middle Ear Characteristics of Anesthetized Cats," *J. Acoust. Soc. Am.* **41**, 1237–1261.

[17] Hall, J. L. (1981). "Observations on a nonlinear model for motion of the basilar membrane," in *Hearing Research and Theory*, Vol. 1 (Academic, New York), 1–61.

[18] Helmholtz (1862). *On the Sensation of Tone*, Dover Publications (1954), New York, 406–410.

[19] Hudda, H. (1983). "Measurement of the eardrum impedance of human ears," *J. Acoust. Soc. Am.* **73**, 242–247.
[20] Hudspeth, J., and Corey, D. P. (1977). "Sensitivity, polarity, and conductance change in the response of vertebrate hair cells to controlled mechanical stimuli," *Proc. Nat'l Acad. Sci. USA,* **74**, 2407–2411.
[21] Javel, E. (1981). "Suppression of auditory nerve responses I: Temporal analysis, intensity effects and suppression contours," *J. Acoust. Soc. Am.* **69**, 1735–1745.
[22] Johnson, D. (1980). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," *J. Acoust. Soc. Am.* **68**, 1115–1122.
[23] Kemp, D. T. (1978). "Stimulated acoustic emissions from within the human auditory system," *J. Acoust. Soc. Am.* **64**, 1386–1391.
[24] Kemp, D. T. (1979). "Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea," *Archives of oto-Rhino-Laryngology* **224**, 37–45.
[25] Kemp, D. T., and Chum, R. (1980). "Properties of the generator of stimulated acoustic emissions," *Hear. Res.* **2**, 213–232.
[26] Knudsen, E. I. (1981). "The Hearing of the Barn Owl," *Scientific American,* December, 113–125.
[27] Kuile, E. ter. (1900). "Die Uebertragung der Energie von der Grundmanbran auf die Haazellen," *Pflueg. Arch. ges Physiol.* **79**, 146–157.
[28] Lesser, M., and Berkley, D. (1972). "Fluid mechanics of the cochlea, Part I," *J. Fluid Mech.* **51**, Part 3, 497–512.
[29] Lynch, T. J. Nedzelnitsky, V., and Peak, W. T. (1982). "Input impedance of the cochlea in cat," *J. Acoust. Soc. Am.* **72**, 108–130.
[30] Moller, A. R. (1983). *Auditory Physiology,* Academic Press, New York.
[31] Morse, P. M. (1948). *Vibration and Sound,* 2nd edition, McGraw-Hill, New York, p. 308.
[32] Neely, S. T. (1981). "Fourth-order partition dynamics of a two-dimensional model of the cochlea," Doctoral dissertation, Washington Univ., St. Louis, MO.
[33] Neely, S. T., and Kim, D. O. (1983). "An Active cochlea model showing sharp tuning and high sensitivity," *Hearing Research* **9**, 123–130.
[34] Neuweiler, G. (1980). "How bats detect flying insects," *Physics Today,* Aug., 34–40.
[35] Pickles, J. O. (1982). *An Introduction to the Physiology of Hearing,* Academic Press, London.
[36] Ranke, O. F. (1950). "Theory operation of the cochlea: A contribution to the hydrodynamics of the cochlea," *J. Acoust. Soc. Am.* **22**, 772–777.
[37] Sachs, M. B., and Young, E. D. (1980). "Effects of nonlinearities on speech encoding in the auditory nerve," *J. Acoust. Soc. Am.* **68**, 858–875.
[38] Schroeder, M. R., Hall, J. L. (1974), "Model for mechanical to neural transduction in the auditory receptor," *J. Acoust. Soc. Am.* **55**, 1055–1060.
[39] Schroeder, M. R. (1975). "Models of Hearing" *Proc. IEEE* **63**, 1332–1350.
[40] Schroeder, M. R., Atal, B. S., and Hall, J. L. (1979). "Optimizing digital speech coders by exploiting masking properties of human ear," *J. Acoust. Soc. Am.* **66**, 1647–1652.
[41] Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (1983). "Comparison between the tuning properties of inner haircells and basilar membrane motion," *Hearing Research* **10**, 93–100.
[42] Shaw, E. A. (1980). "The Acoustics of the External Ear," In *Acoustical Factors Affecting Hearing Aid Performance,* G. Studebaker and I. Hochberg Eds., pp. 109–125.
[43] Shaw, E. A. G., and Stinson, M. R. (1983). "The human External and Middle ear: Models and Concepts," in *Mechanics of Hearing,* Ed. by E. deBoer and M. A. Viergever, Delft Univ. Press, pp. 3–10.
[44] Siebert, W. M. (1974). "Ranke revisited—a simple short-wave cochlear model," *J. Acoust. Soc. Am.* **56**, 594–600.
[45] Smith, R. L., and Brachman, M. L. (1982). "Adaptation in Auditory-Nerve Fibers: A Revised Model," *Biological Cybernetics* **44**, 107–120.
[46] Stevens, S. S., and Davis, H. (1938). *Hearing, Its Psychology and Physiology,* John Wiley & Sons, Inc., New York, N.Y.
[47] Zwicker, E., and Scharf, B. (1965). "A Model of Loudness Summation," *Psychological Review* **72**, 3–26.
[48] Zweig, G., Lipes, R., and Pierce, J. R. (1976). "The cochlear compromise," *J. Acoust. Soc. Am.* **59**, 975–982.
[49] Zwislocki, J. J. (1948). "Theorie der Schneckenmechanik," *Acta Oto-Laryng. Suppl.* **72**.
[50] Zwislocki, J. J., and Kletsky, E. J. (1979). "Tectorial Membrane: A Possible Effect on Frequency Analysis in the Cochlea," *Science* **204**, 639–641.
[51] Zwislocki, J. J. (1980). "Five Decades of Research on Cochlear Mechanics," *J. Acoust. Soc. Am.* **67**, 1679–1685.
[52] Wilson, J. P., (1974). "Basilar Membrane Vibration Data and Their Relation to Theories of Frequency Analysis," in *Facts and Models in Hearing,* E. Zwicker and E. Terhart Eds., Springer-Verlog, 59.

Jont B. Allen was born in St. Charles, IL, on December 5, 1942. He received the B.S. degree in Electrical Engineering from the University of Illinois, Urbana-Champaign in 1966, and the M.S. and Ph.D. degrees from the University of Pennsylvania in 1968 and 1970 respectively. He then joined Bell Laboratories, Holmdel, NJ, in 1970 and transferred to the Acoustics Research Department in Murray Hill, NJ, in 1974. Dr. Allen is presently working in the areas of cochlear modeling, cochlear neurophysiology, digital communication theory and in digital signal processing applications. Some of his applied interests are in speech coding, room acoustics, dereverberation of speech signals and psychophysical modeling of room reverberation. His theoretical interests include short-time Fourier transform theory and cochlear modeling. Dr. Allen is a Fellow of the Acoustical Society of America and a Fellow of the IEEE. He is presently chairman of the Publication Board of the Acoustics Speech and Signal Processing Society and is a member of ADCOM of the same IEEE society. He is a past editor of the ASSP and has served on several committees in both the IEEE and the Acoustical Society. Dr. Allen is married with two children.