Excerpts from

HEARING: A 21ST CENTURY PARADIGM

including,

ELECTROCHEMISTRY OF THE NEURON

This material is excerpted from the full β-version of the text. The final printed version will be more concise due to further editing and economical constraints. A Table of Contents and an index are at the end of this paper.

James T. Fulton
HEARING CONCEPTS
1 (949) 759-0630
jtfulton@hearingresearch.net

This is a Draft

9 Overall Hearing Performance 4/7/12

Man lives in interaction with his physical environment. Actually, he lives in his perception of this environment. Smoorenburg, 1988

9.1 Overview of system performance, verification program & path to the summary

The previous chapters have described the components of the auditory system from a variety of perspectives at a detailed level. This chapter will try to assemble the material into a single brief and more cohesive scenario. In that sense, this chapter can be considered a summary of the previous material. However, the goal is to simultaneously provide a better understanding of the theory, mechanisms and interconnections that form the comprehensive model describing the overall auditory system. The material will include at least two facets not covered previously. First, the operational requirements placed on a given element of the system will frequently be discussed within the following sections. Second, additional methods of data collection will be utilized. Suggestions will also be made concerning new or improved methods of data collection.

The complexity of the auditory system makes its description awkward. Many writers have summarized the system by describing it as a time-dependent spectral analyzer. This description needs to be expanded in order to gain more precision. The spectral analyzer portion of the system is not time dependent but the portion following it is in one sense. The neural portion of the system is highly stimulus level dependent and responds by following various temporal profiles. It would be better to describe the neural system using a more conventional terminology. The auditory system is a state variable system, where its temporal course depends on both its initial or previous state and the stimulus.

The complexity of the system is highlighted in any discussion of the time constants of the system. There are at least three distinct and largely unrelated time constants that vary by species. First is the attack time constant associated with the leading edge of the response to a stimulus by the adaptation amplifier within the sensory neurons. This time constant is less than a few milliseconds under phonotopic excitation. Second is the decay time constant associated with the low pass filter at the output of the adaptation amplifier. This time constant is on the order of 6-10 seconds. Third is the second order time constant of the adaptation amplifier and its power supply following severe overload. This time constant, although seldom encountered in practice is easily measured in the laboratory. It is on the order of several minutes.

Describing the frequency limits of the human auditory system is difficult. The answer depends on the question. A typical audiogram will record an overall response without regard to the ability to discriminate between one frequency and another. However, frequency discrimination decreases to near

zero at frequencies above 15 kHz[1].

9.1.1 Gross organization of this chapter follows the organization of the hearing modality

[xxx add additional material re: Sections 9.1 to 9.8]
It is important to recognize the primary purpose of the hearing system. It is to extract information describing the external surroundings from the acoustic stimuli received. The requirement is not to relay that stimuli from the PNS to the CNS. As in the case of vision, there is no little green man within our brains that is listening to the signals delivered by the peripheral neural system. Information is extracted from the stimuli via two distinctly different functional tracts, the communications tract and the source location tract. Each of these tracts involves a variety of subsidiary tracts described here as paths.

Section 9.2 will present a variety of block diagrams describing the overall operation of these different tracts and paths. The remaining sections of the Chapter will summarize (primarily human) hearing performance based on different perspectives. **Section 9.3** will summarize a variety of individual elements of the system by reviewing some of the material of previous chapters. **Section 9.4** will describe hearing performance based on cumulative electrophysiological data while **Section 9.5** will describe the performance from a psychophysiological (psychophysical) perspective. **Sections 9.4** & **9.5** primarily review the normal operations of hearing. **Section 9.6** reviews technical data, such as adaptation and masking performance, acquired in a laboratory environment. **Section 9.7** reviews an additional group of data that frequently involves cognition or other higher level effects involving signal manipulation. **Section 9.8** will address a variety of situations related to both speech and musical communications. The uniquely specific performance requirements related to speech and music are reviewed in **Section 9.8.1**.

The bilateral organization of most animals leads to a binaural capability. The binaural capability supports additional performance related to both the communications and source location tracts. The enhancements to the communications function due to binaural hearing are not large. However, the source location tract is absolutely dependent on the binaural capability. **Section 9.9 xxx** will explore binaural source location in some detail. The contribution to communications of binaural hearing will only be touched upon in **Section 9.9. xxx**.

Verification of the model presented here is particularly important since it varies significantly from the conventional wisdom. To this end, a program is currently underway on a world-wide basis to involve PhD and post-Doc candidates in a series of manageable programs leading to the incremental verification of the overall model. In support of this approach, the steps in this program are documented under various headings in this chapter. They can be easily located by searching the document using the keyword, “verification” or “verifying” without the quotation marks. An index incorporating the word verification also appears at the end of the chapter.
- - - -
Complementing the operational summary described above, this chapter also presents a summary of the overall theory developed in this work with particular emphasis on the resulting top level circuit diagrams (**Section 9.x** xxx).

---

[1]Muraoka, T. Iwahara, M. & Yamada, Y. (1981) Examination of audio-bandwidth requirements for optimum sound signal transmission *J Audio eng Soc* vol. 29, pp 2-9

Describing the performance of the auditory system of an animal necessarily involves returning to some of the perspectives used earlier. The dual channel (narrowband tonal and wideband transient) nature of the PNS portion of that system, along with the closely related performance of the vestibular system, adds additional complexity to the discussion. The star-network configuration of the CNS makes describing discreet signal paths in this area even more difficult. Leaving the vestibular system to last, this chapter will address the cumulative performance measured at different nodes in the system. It will begin with a description of the cumulative performance associated with the first three stages of the acousto-mechanical system. The discussion will then bifurcate to discuss the electrophysiological performance of the wideband (IHC related) channels and the tonal (OHC related) channels separately. Performance measurements in these areas are typically obtained using drastic invasive techniques.

The discussion will also develop another dual characteristic of the hearing system, the presence of both fine and coarse channels within the overall tonal channel of human hearing. This duality has not been documented previously but its existence is well supported by the empirical database.

Converging again, the performance associated with various non-invasive electrophysiological experiments will then be discussed. The discussion of the auditory channel will conclude with a discussion of the psychophysical performance of the system.

Recognizing a clear distinction between the composite frequency responses related to basilar and tectorial membrane motion and individual frequency responses related to the tonal and broadband neural channels is critically important when exploring the performance of the auditory system.

The discussion related to the psychophysical performance of hearing cannot be comprehensive because of the vast literature related thereto. Hopefully, a framework will be defined that will allow other empirical experiments to be interpreted in detail when desired. This work will focus on only the most fundamental performance parameters.

The Theory of Hearing, the Electrolytic Theory of the Neuron and the model presented in this work resolve a wide variety of ambiguities and paradoxes found in the literature related to hearing. They provide a particularly valuable new insight into the binaural properties of the system.

The reader should take particular notice that much of the data in the literature assumes all of the sensory neurons of the auditory and vestibular systems share common characteristics. This work has shown this is not the case. Care should be taken to determine which domain the measurements in a given paper are concerned. This is sometimes difficult since the original author may not have recognized the scope of his own work. As discussed elsewhere, the introductory material in a paper is frequently designed to acquire credibility, rather than to provide an accurate scope to the paper. Statements such as "every normal auditory-nerve fiber shows an increase in discharge rate over some specific range of tone frequency[2]." is not helpful when the purpose and character of all of such fibers were not understood at the time of the paper.

The performance ranges measured for different parameters of members of a single species vary considerably. A range of 20 dB (100:1) is considered normal in the clinic for the sensitivity of the human auditory system. Many of these ranges are based on growth related phenomena and are therefore not described by Gaussian (normal) statistics. These parameters are described by *log-normal statistics*. Such statistics exhibit a narrow range on one side of the median value and an extended tail

---

[2]Johnson, D. (1980) The relationship between spike rate and csynchrony in responses of auditory-nerve fibers to single tones *J Acoust Soc Am* vol. 68(4), pp 11151122

on the other side. With respect to sensitivity, very few people exhibit a significantly better sensitivity than the median. However, losses in sensitivity among humans are measured in tens of dB.

Limited information is available in the literature concerning the mode of signaling used between elements and stages of the auditory system. This situation necessarily limits how experiments have been carried out in the past. This Chapter will provide an interpretation of how individual stages of the system works and provide information based on two sources; the literature and the theory provided here. Additional discussion will relate to cumulative performance at a given node of the system based on an initial acoustic stimulus. The interpretations will be supplanted with the best data available from the literature.

Most of the available data is cumulative performance data reporting on the measured signals at a given point based on an acoustic stimulus. Virtually no data is found in the literature based on electrical stimuli introduced later in the signaling path. What little data exists is based on the insertion of traumatic signals. Traumatic signals are signals that are not compatible with the signals expected by the subsequent circuitry. As an example, introducing a single square wave pulse at a potential of 100 volts amplitude at the dendrite of a stage 3 neuron is not compatible with the 100 mV or less dynamic range of the neural circuit that is expecting a pulse stream at a pulse rate centered around 200 Hz.

The vast majority of hearing data has been collected based on individual acoustic clicks or short bursts of single frequency tones. While there is some data based on stimuli consisting of chords (usually only two tones), tone sequences and verbal sentences, most of this material has not found its proper place in the framework of the literature. This chapter will attempt to place this data in an appropriate framework.

This chapter will also attempt to show where various information from clinical investigations fits in the same framework.

In the course of the discussion, suggestions will be made on improving a variety of clinical protocols, and exploratory research protocols (both invasive and non-invasive]

9.1.1 Overview of the performance of the auditory system OR IN 9.1

Looking at the system from a system analysts perspective, the hearing system, like the vision system and other biological systems is highly optimized. The system is optimized to the point of perfection for the performance requirements it is designed to meet.

The large dynamic range of the hearing system is not well understood. While the instantaneous dynamic range of the system seldom exceeds 200:1 (26 dB based on voltage), the long term dynamic range exceeds 1,000,000:1 (120 dB). This range is achieved through the use of a very high performance variable gain amplifier in the first neural stage of the system. Under optimum, quiet adaptation, conditions, the gain of this one amplifier can achieve 3500:1 (72 dB). At maximum gain, this amplifier allows the system to achieve the ideal condition, an external background limited noise performance.

From a noise perspective, the system operates at a level just above the noise level due to Brownian motion within the ear canal and the surrounding medium. On occasion, people complain of hearing noise routinely when the level is changed slightly to just below that of the background Brownian motion.

9.1.1.1 The variable gain characteristic of the normal hearing system

EMPTY

9.1.1.2 Noise performance of the system under normal and pathological conditions

When operating normally, the hearing system is near perfect from a noise perspective. This fact can be demonstrated by a number of simple tests.

The noise performance of the system can be examined from a variety of detailed perspectives. Its performance can be examined from the white noise perspective, the pathological perspective associated with the closed ear canal, and the methodology of signal extraction from below the background noise level.

Signal extraction from below the background level is a large subject area. The fundamental mechanism involves correlation designed to extract signals from the intrinsic random background noise level (white noise). This capability is the fundamental form of signal extraction usually associated with the concept of masking. Once this mechanism is understood relative to white noise, other forms of signal extraction in the presence of other types of random noise (generally described as colored noise) can be characterized. Kleppe has provided a figure illustrating colored noise[3]. Colored noise need not be random in character. It can be varied in the extreme to the case of a single tone of arbitrary frequency as a background. Intermediate between these two extremes are situations where the background is in the form of multiple simple tones (non-musical chords, interfering tones varying in frequency, and multiple correlated tones (generally described as musical chords). This range of interfering backgrounds makes the subject of masking very large and complex.

9.1.1.2.1 Impact of noise on electro-physical investigations

The high sensitivity of the auditory system under quiet-adapted conditions in the laboratory are profound. Except under the most controlled conditions, the Stage 3 signal projection neurons exhibit a continuous variation in action potential generation rate (the so-called spontaneous rate) due to the noise level sensed by the adaptation amplifiers of the sensory neurons. Thus, the spontaneous rate exhibited by a specific fiber is a function of the state of adaptation of its associated sensory neuron. The spontaneous rate of the Stage 3 signal projection circuit alone can be determined by suppressing any signal from the sensory neuron. This can be achieved by exposing the sensory neuron to a high signal level (greater than 70 dB SPL) for at least three *attack* time constants (xxx ms total) and making the determination in less than one *decay* time constant (xxx ms).

By performing the above procedure on a large ensemble of auditory nerve fibers, the true spontaneous rates for these projection neurons alone can be determined. Based on the character of the signals generated by the sensory neurons *en bloc*, it appears that most of the Stage 3 neurons will exhibit a true spontaneous rate near zero pulses per second.

The fact that the typical Stage 3 circuit exhibits a higher spontaneous rate due to encoding of a noise signal received from the sensory neuron can be confirmed by listening to (or viewing on an oscilloscope) the character of the signal after it has been decoded using an auxiliary decoder in the test equipment.

---

[3]Kleppe, J. (1982) Engineering Applications of Acoustics. NY: Artech House pp 164-166

9.1.1.2.2 Pathological noise conditions related to hearing

Looking at noise in the broad sense as any undesired signal input provides some interesting features of the auditory system. The simple act of closing off the ear canal can result in the observation of an enhanced noise level. As commonly observed in a quiet environment, closing off the ear canal can result in one's hearing the sound of blood rushing into the tissue around the ear canal in synchronism with the heart beat. Similarly, closing off the ear canal using a hard surface volume larger than the ear canal, such as a sea shell, can result in one's perceiving the random noise associated with the Brownian motion of the air molecules within that volume. Both of these simple acts suggest how close the normal auditory system is to background-level limited performance.

The large number of parallel signaling circuits found between Stages 1, 2 and 4 suggests the auditory system is not susceptible to single point failures in this region. On the other hand, at least one potential noise source is associated with each of these individual channels. The susceptibility of the auditory system to extraneous noise from pathological conditions in any of these channels is high. This situation may contribute to the high incidence of tinnitis in humans.

9.1.1.2.3 Tinnitis as a failure mode in hearing

The various forms of tinnitis can be associated with one of two types of failures in the auditory system. First, the failure can occur in the different types of signaling channels these individual parallel circuits support. Since the individual channels are designed to be so sensitive, an alternative failure mode is evident. It is common in man-made communications systems using phase-modulation to incorporate a "squelch circuit" to desensitize the receiver in the presence of noise alone in an individual communications channel. This circuit has the effect of eliminating the noise occurring between intervals of time containing useful signal information. It is possible that tinnitis is due to a failure in such a squelch circuit built into the Stage 4 of the auditory system. This condition will be discussed more fully in the next chapter.

9.1.1.3 Performance contributed by the higher cognitive centers EMPTY

Significant performance enhancement of the performance of the hearing system can be achieved using the cognitive centers of the CNS. Such performance is usually described using concepts such as mental concentration and mental attention.

9.1.2 Predominant signal insertion and extraction sites

Organizing the pertinent test data appearing in the literature into an appropriate framework is difficult. The best framework found to date is based on relating the stimulus location and recording location to the nodes defined in **Section 4A.1.2** xxx and following the format shown in **Figure 9.1.2-1**. The figure repeats an earlier figure from Chapter 4Axxx in the upper frame and expands on it in the lower frame. In conformance with the quasi-planes defined as interfaces earlier, node, B is defined to the left of the tympanic membrane in Stage B. Node C is defined to the right of the oval window. Node D is defined at the output of the piezoelectric element of each sensory neuron. Because of the limited importance of Stage 2 signal processing in the auditory system, node E is defined at the input to the Activa within the oscillatory connexus of the Stage 3 ganglion cell. Node F is defined at the initial segment of the axon of the Stage 3 ganglion cell. In many experiments, a later axon segment may be the data source (at theoutput of a Node of Ranvier). If so, the axon segment should be listed as F1, F2 etc. Node G refers to the input circuit of the stellate decoding circuit. This node is noise sensitive, as are all Node of Ranvier input circuits. Therefore the input circuits of Nodes of Ranvier are described

by counting antidromically using G1, G2, etc. to indicate the distance of the node from the associated. stellate neuron. Node H is defined as the output at the pedicle of the Stage 3 stellate decoding neuron. While this neuron is frequently labeled a stellate neuron in vision, it has been given many whimsical names in hearing, particularly within the cochlear nucleus (**Section 7.1 xxx**).

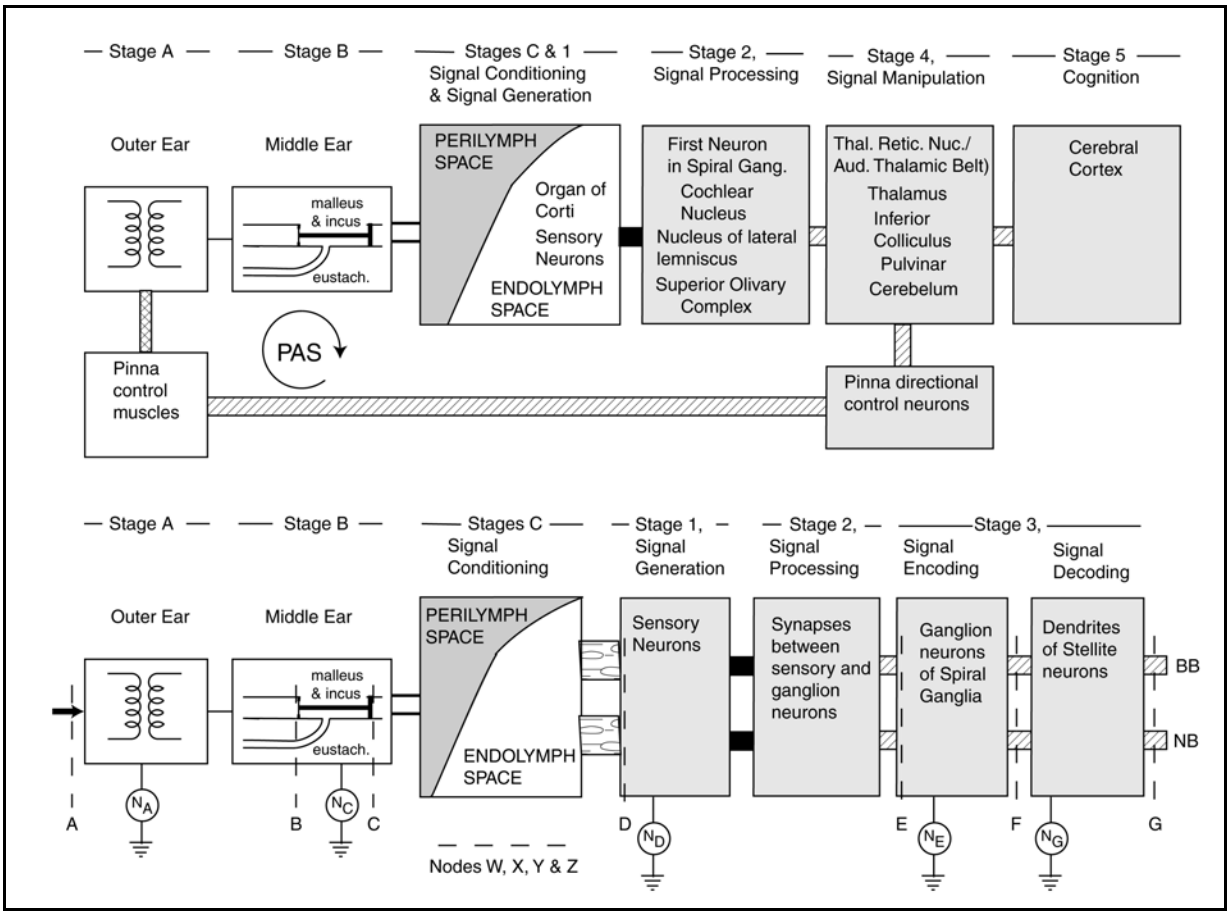


Figure 9.1.2-1 Node designations and potential noise sources within the initial stages of hearing. Top; block diagram developed in Chapter 4. Bottom; expanded block diagram defining the nodes and potential noise sources of hearing. BB; broadband signaling channels. NB; narrowband or tonal signaling channels. Nodes W, X, Y & Z are shown only conceptually. See text.

Data collected at the axons of the spiral ganglia based on far-field acoustic stimulation can be described as node A-to-F data. Similar data where the stimulus was applied to the tympanic membrane as directly as possible would be called B-to-F data. Data where the stimulus was applied to the oval window would be called C-to-F data

It is helpful to define a series of nodes describing point of data collection not related to normal auditory signaling. To accommodate the non-channel data collected relating to the vibration of the basilar and tectorial membranes, nodes designated Z and Y respectively will be used. To accommodate the data collected at the round window, node W will be used. Other evoked potential data (such as that collected at the surface of the skin) will be considered as collected at node V. When it is necessary to distinguish between the broadband and tonal channels, a subscript b or t will be applied to the appropriate node designation. Where the morphology of the ear has been changed significantly (e.g., cutting away of the pinna), the node designation will be prefixed by the term modified.

Note that the signal channels between nodes D and F are highly parallel. They are believed to consist of about 3,500 channels supporting the IHC neurons and a much smaller number supporting groups of OHC. The specific proportion of Stage 3 circuits can be determined from the population of auditory nerve fibers usually accessed in the laboratory at node F. As noted in Section xxx, this proportion has not been reported properly in th literature since most investigators have discarded or de-emphasized those neurons that did not respond to a narrowband (pure tone) stimulus.

9.1.2.1 The baseline circuit diagram for the first three stages of neural hearing

Figure 9.1.2-2 shows the detailed circuit diagram of the fundamental path supporting the first three stages of the afferent neural system related to hearing. [xxx show same nodal points as above.] The output of these circuits feeds into a complex mesh network containing all of the individual feature extraction engines of the CNS. The operation and interplay between these engines is poorly understood at present. **Section 9.4.4** xxx will discuss what is known about these portions of the system.

The operational details of each of the individual circuits shown in the figure have been discussed in previous chapters. This chapter will strive to define the performance achievable in hearing as a result of the cascading of these circuits and in conjunction with the previous non-neural stages discussed earlier. As noted earlier, it is critically important to recognize two major features of the system. First, the sensory neurons and the vast majority of the signal manipulation (Stage 4) and cognition (Stage 5) neurons are analog (electrotonic) neurons. They do not exhibit action potentials (the signaling feature most relied upon in prior research). Thus past research has largely been limited to tracking the action potentials of Stage 3, and their limited association with elements of Stage 4 & 5. Future research must pay more attention to the analog characteristics of these neurons. Second, the entire neural system is designed to operate in a constant maximum-signal amplitude regime. This regime has a dynamic range of typically 26-30 dB (200 – 300 to one). It is the responsibility of the adaptation amplifiers within the sensory neurons to adjust their performance to match the stimulus environment to the capability of the hearing modality. They do this by selecting an instantaneous range of about 26–30 dB from within the overall range of the external environment and translating that limited range into the range manageable by the hearing system. At no time does the *in-vivo* hearing system accommodate a total dynamic range of 80–120 dB. Much of the data reported in the hearing literature is the result of introducing very wide dynamic range stimuli over a period of time compatible with the time constants of the adaptation amplifier. The resulting data frequently suggests the system has a much wider dynamic range than it does (due to the dynamic changes within the adaptation amplifier largely unbeknown to the experimentalist).

Notice the Node of Ranvier and its associated axon segment are shown as being replicable in the figure. The total length of such a combined unit is generally less than two millimeters within the neural system. On long peripheral neurons, this unit may be replicated hundreds of times within one neuron. However, distances within the auditory system are very short and only lower orders of replication are normally found.

The nodes shown by callouts within circles in the above figures are those of primary interest to both the experimentalist and the analyst at this time. They will be referenced in the following discussions based on the nomenclature shown.

IN most neural paths associated with hearing, the path between node E and node E' consists of only

a synapse. However, in a few cases, a more complex Stage 2 circuit appears as shown in the alternate diagram. In this case, the signals from many individual sensory neurons (particularly OHC) are summed at this point. In some cases, the forward conductance of the synapse is bridged by a capacitor. This combination of a conductance and capacitance in parallel when combined with the conductance and capacitance of the input stage of the following connexus forms a lead-lag network of broad flexibility. It can be used to tailor the frequency response of the Stage 2 transfer function.

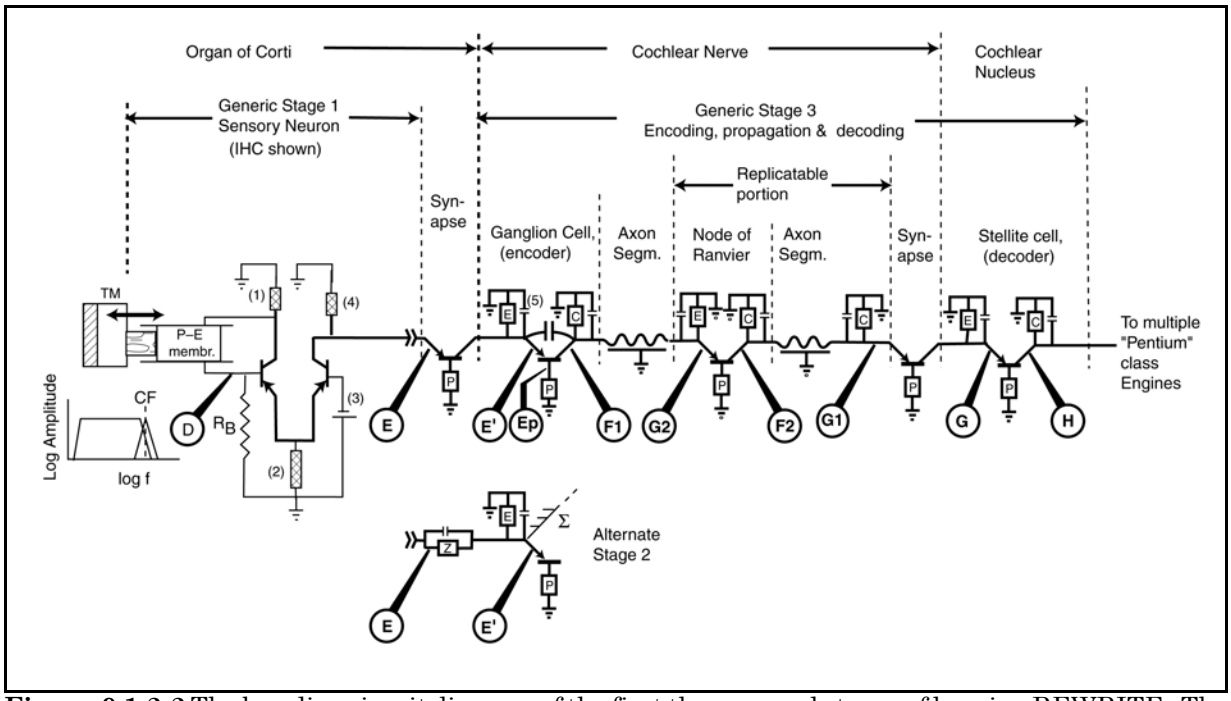


Figure 9.1.2-2 The baseline circuit diagram of the first three neural stages of hearing REWRITE. The load impedance (4) effectively limits the output bandwidth of the sensory neuron to 500–600 Hz. The insert at the bottom replaces the synapse and ganglion input circuit when a lead-lag network is called for. It adds a capacitance between the pedicle of the sensory neuron and the dendrite of the ganglion cell. See text.

9.1.2.2 The application of a stimuli to the system

The predominant mode of signal insertion in hearing research has been via the unmodified external ear. However, the application of the stimulus has involved three significantly different modes. The first involves application from a source remote from the pinna, and generally referred to as the open source mode. This type of source is defined as located at node A in this work. The second mode involves insertion of the acoustic energy directly into the ear canal. This mode generally involves closing off the ear canal from other acoustic stimulation and is described as the closed source mode. The act of closure can change the performance of the outer ear significantly. First, it eliminates the affect of the pinna on the frequency response. Second it can cause reverberation not found in the normal outer ear.

A third mode of stimulus application has involved significant surgical modification of the ear canal. This has involved removal of the pinna and part of the duct followed by the installation of an artificial ear canal containing a source. This modified closed source mode can have significant affect on the impedance of the outer ear.

Some investigations have attempted to apply an acoustic stimulus directly to the tympanic membrane. They have insured that the source was as close to the membrane as possible. This remains a form of open or closed source stimulation depending on how the entrance to the ear canal was treated. A few investigators have attempted to apply acoustic energy directly to the umbo at the center of the tympanic membrane. It is possible this method could be considered as applying a stimulus directly to node B of the system. Any method involving application of acoustic energy to the tympanic membrane, other than the open source method, is defined as occurring at a "modified node A" in this work. Hopefully future investigators will describe their use of modified node A in sufficient detail for the reader to comprehend the subtleties of the method.

To date, there has been little activity involving the application of non-acoustic stimuli to the auditory system in the research context. A growing exception to this statement has been the application of a strong electrical stimulus globally to the initial elements of the neural system. The goal has been to achieve hearing in those with a system failure prior to the sensory neurons. To date, these experiments have achieved a degree of success in bringing the totally deaf some form of auditory sensation in response to conventional speech input. However, much work remains in this area. The effect of the application of such gross signals to the environment of the sensory neurons is difficult to specify. It is possible they affect the input to the Stage 1 sensory neurons and can be considered a node D stimulus. It is more likely that they affect the input circuit of the Stage 3 ganglion neurons and should be considered a node E stimulus.

9.1.2.2.1 The problem of using a click generator

While a variety of investigators have used clicks to excite the auditory system, they have frequently taken short cuts due to the limitations on their instrumentation (recognized or not recognized). Frequently, the investigator has taken the waveform applied to his transducer as indicative of the output of that transducer. This introduces a large unknown in the experiments; what is the transfer function of the transducer?

Commercial crystal transducers, the projector elements used in earphones or microphones used as projectors, are not designed as pulse transformers. They are designed to respond reasonably uniformly over a wide frequency band. They rely largely upon the impedance and bandwidth of the driving electrical circuits to control the ultimate shape of any pulse created in the air. **Figure 9.1.2-3** illustrates a typical transducer output in response to a square pulse excitation[4]. This is the response measured in air in the open space between the transducer and a potential biological auditory system. The output shown in (**B**) should be compared to some of the electrophysiological signals recovered from the auditory system by recording the cochlear microphonic. An example from Kiang, et. al. is shown in (**C**). While the Kiang, et. al. paper compares their cochlear microphonic to the electrical stimulus at (**A**), and attributes any differences to the auditory system, it is more likely that the recording at (**C**) is a much more faithful reproduction of the actual acoustic stimulus at (**B**). Smith & Vernon specifically warn against this shortcut (pages 333 & 340) under procedures to adopt, "1. The only acceptable measure of stimulus magnitude is an acoustical measure—and this must be of the actual sound used to stimulate the ear."

[4]Smith, C.& Vernon, J. *eds*. (1976) Handbook of Auditory and Vestibular Research Methods. Springfield, Il: Charles C. Thomas. pg 335

Kiang sidestepped the above admonition by recording an electrical signal (N_1) at the round window (his figure 4.1). He also recorded a signal from an acoustic monitor in the outer ear (his figure 2.3). However, this monitor did not receive a signal with any resemblance to the click Kiang described using words. The signal from the acoustic monitor (AM) differed drastically from a square pulse with sloping leading and trailing edges. So did the electrical signal recorded at the round window (RW). It is clear that the signal recorded at axons of the auditory nerve cannot be considered to be due to the square pulse used to drive the acoustic transducer in the test set.

The stimulus must be described as Smith & Vernon required. The investigator must be sure he is using a monitor that provides a realistic copy of the stimulus. The differences between commercial condenser and dynamic microphones from the same manufacturer are significant.

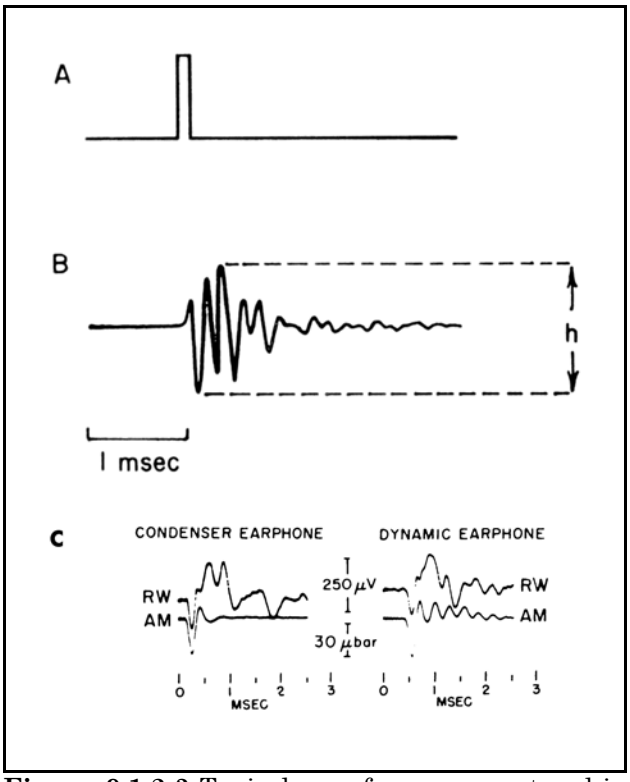


Figure 9.1.2-3 Typical waveforms encountered in acoustic click experiments. (A); a typical square pulse found in the electrical circuits of a test set. (B); the actual response of an acoustic monitor in the unrestricted airspace in front of the electrical to sound transducer of the test set. From Smith & Vernon, 1976. (C); examples of signals recorded acoustically by an acoustic monitor in the er canal and a probe at the round window. From Kiang, 1965.

9.1.2.3 The recovery of signals from the system

Exploratory research has been dominated by recovery of information about the system by psychophysical experiments (the subject reports his response through some form of action–normally speech or pressing a button). In the majority of the remaining experiments, the data has been recovered by probing the axons of the Stage 3 auditory nerve. This location is defined as node F in this work. In the absence of knowledge concerning the method of signal encoding at this location, the reduction of this data has been largely limited to statistical methods.

The remaining data collection efforts have focused on evoked potentials. These evoked potentials are characterized as non-signaling channel signals and are best addressed in the following section.

9.1.3 The auditory system physiology supporting performance evaluation

Most of the material in this Chapter is applicable to mammals. Smolders & Klinke have provided a brief discussion of the hearing related performance differences found among species[5]. The auditory systems of aquatic animals are very primitive. Hickman has suggested, "It is doubtful whether fish can hear in the ordinary sense of the word, although they are sensitive to vibrations through their

[5]Smolders, J. & Klinke, R. (1986) synchronized responses of primary auditory fibre-populations in *Caiman crocodilus* (L.) to single tones and clicks *Hear Res* vol. 24, pp 89-103

lateral line organs[6]." Significant differences are found in the hearing systems of amphibia because of their intermediate stage of evolutionary development between fish and terrestrial animals. Many exhibit no external ear and the tympanic membrane is exposed. The middle ear consists of a single rod like stapes that travels axially. The inner ear is much simplified and is filled with endolymph. No basilar membrane is present in the ear of the amphibian.

Avian and mammalian ears are functionally similar, except for the length and the degree of curvature in different species. The longer and more curved spirals are generally associated with wider frequency response.

9.1.3.1 Fundamental system characteristics important to performance evaluation

The auditory system employs two parallel signaling channels, a broadband channel (related to the IHC) and the tonal channel (related to OHC). The middle ear performs two-dimensional signal conditioning in order to support these separate channels (Chapter 4A, xxx). Neither the acoustic or the neural stages of the auditory system employ the concept of resonance in their operation. The overall system described so far has not employed any mechanically or electrically resonant structures to explain the performance of the auditory system. In addition, it has not exposed any external feedback mechanism involving the IHCs and OHCs that could generate a resonance condition. The only quasi-resonant condition uncovered is the phasing arrangement associated with the spacing of the IHC hairs along the tectorial membrane. This phasing can be interpreted as generating beams of acoustic energy aimed at specific OHCs. These OHC's exhibit a capture area that determines the range of acoustic frequencies that they are sensitive to.

9.1.3.2 Background

It is important to review several concepts before proceeding. A recent study of the vibratory motion of a Hensen's cell at the outer edge of the reticular lamina, claimed to be the response to AM modulated sound waves (at 100% modulation)[7]. The premise was stated at the beginning that "Natural sounds and speech contain amplitude-modulated waves." There may be a problem of interpretation here. These classes of sounds do not involve modulation in the engineering sense. They consist of linear mixtures of individual sounds, each of which may be transient or sinusoidal. There is a major difference between mixing multiple sounds by addition and modulation involving the multiplication of individual sounds in a nonlinear process. In the absence of a non-linear process, and in most applications a higher frequency "carrier," these signals do not represent an amplitude modulated (AM) signal. While the test signal used in the tests was definitely an AM modulated signal, and the data is very useful, it does not represent a typical natural situation.

In the same volume, Scherer, et. al. reported excellent data on the oscillatory properties of the tectorial membrane (TM).[8] They note, "It was confirmed that BM displacement was below the noise floor and, therefore, that the BM can be considered as clamped. This remains a controversial conclusion that is supported here.

---

[6]Hickman, C. (1970) Integrated Principles of Zoology, 4th Ed. St. Louis, Mo: C. V. Mosby pg 470
[7]Khanna, S. (2003) Response to amplitude modulated waves in the apical turn of the cochlea *In* Gummer, A. Biophysics of the Cochlea. Singapore: World Scientific pp 252-260
[8]Scherer, M. Nowotny, M. et. al. (2003) High-frequency vibration of the organ of Corti *in vitro In* Gummer, A. *ed*. Op. Cit. pp 271-284

The term amplification is frequently found in the literature of the cochlea. A change in energy level is frequently associated with a change in impedance under passive conditions (transformer action). This type of change in signal level is not normally associated with amplification in the engineering disciplines. Amplification in the engineering sense implies an active process that adds energy from a DC source to the overall output signal under the control of an input signal. Specifically, the ossicles of the inner ear are normally considered passive elements. Under this assumption, the ossicles do not introduce amplification. The removal of these elements may constitute a *loss* in signal level within the cochlea. Replacing these elements does not constitute amplification, merely the removal of a *lossy* condition. In another example, the outer ear collects energy over a solid angle that varies under a variety of conditions. The efficiency of this collection as a function of angle is frequently described in terms of a collection gain parameter. However, this is passive gain and it should not be associated with amplification, any more than the effectiveness of a megaphone should be considered an amplifier. It is a concentrator of available energy.

Another concern is with the terms synchronization and phase-locking. These terms are conventionally used to describe a second signal from an independent source being "aligned" with an initial source. It is not normally used to describe the output of an amplifier(s) in response to the stimulus applied to that amplifier. Thus, the output of an amplifier is expected to be aligned its own input. In acoustic research, the investigator often uses the terms synchronization or phase-locking to merely reflect the signal he measured had a direct relationship with the stimulus (frequently subject to time delays and electrical phase shifts due to the intermediate processes.

There is considerable ambiguity in the data found in the acoustic literature related to the early stages of the auditory system. Many authors have chosen to speak of their data as relating to the basilar membrane when it is action potential data recorded at the output of a neuron. This data involves undefined signal processing by undefined sensor neurons and possibly other neurons before reaching the point of recording.

[xxx summarize other concerns here, particularly the amplitude pressure problem]

Shepherd[9] has provided a figure that he assembled for pedagogical purposes based on the work of Hudspeth (1985), of Dallos (1985), of Crawford & Fettiplace (1985), and of Art, et. al (1984). **Figure 9.1.3-1** presents a similar figure but with the cytology of the sensory neuron replaced with material from this work. This change allows the performance of the overall circuit to be explained in much greater detail and does not rely upon alkali ion currents.

The data was collected under quasi-in-vivo conditions from a turtle, *pseudemys scripta elegans*. The complete head of the animal was isolated and one half of the head was removed to provide access to the remaining auditory system elements. Narrow pulse signals were applied to unspecified circuits of the superior olivary nucleus and signals were recorded from both auditory nerve fibers and from impaled sensory neurons. Art, et. al. indicated that they had no knowledge of the actual efferent signals applied at the synapse with the sensory neuron.

Frame B of the figure shows the response of an unspecified hair cell, with kinocilium, from the papilla of the turtle to transverse cilia motion. Note the symmetrical waveforms. There is no sign of rectification in the electrical response to a sinusoidal deflection.
[xxx describe the test configuration in some detail.

---

[9]Shepherd, G. (1988) Neurobiology. NY: Oxford Univ. Press pg 317

The figure has been redrawn using the electrolytic sensory neuron with the efferent pulse train applied to the dendroplasm of that neuron. Two output paths are shown connecting to the afferent signaling paths arbitrarily.

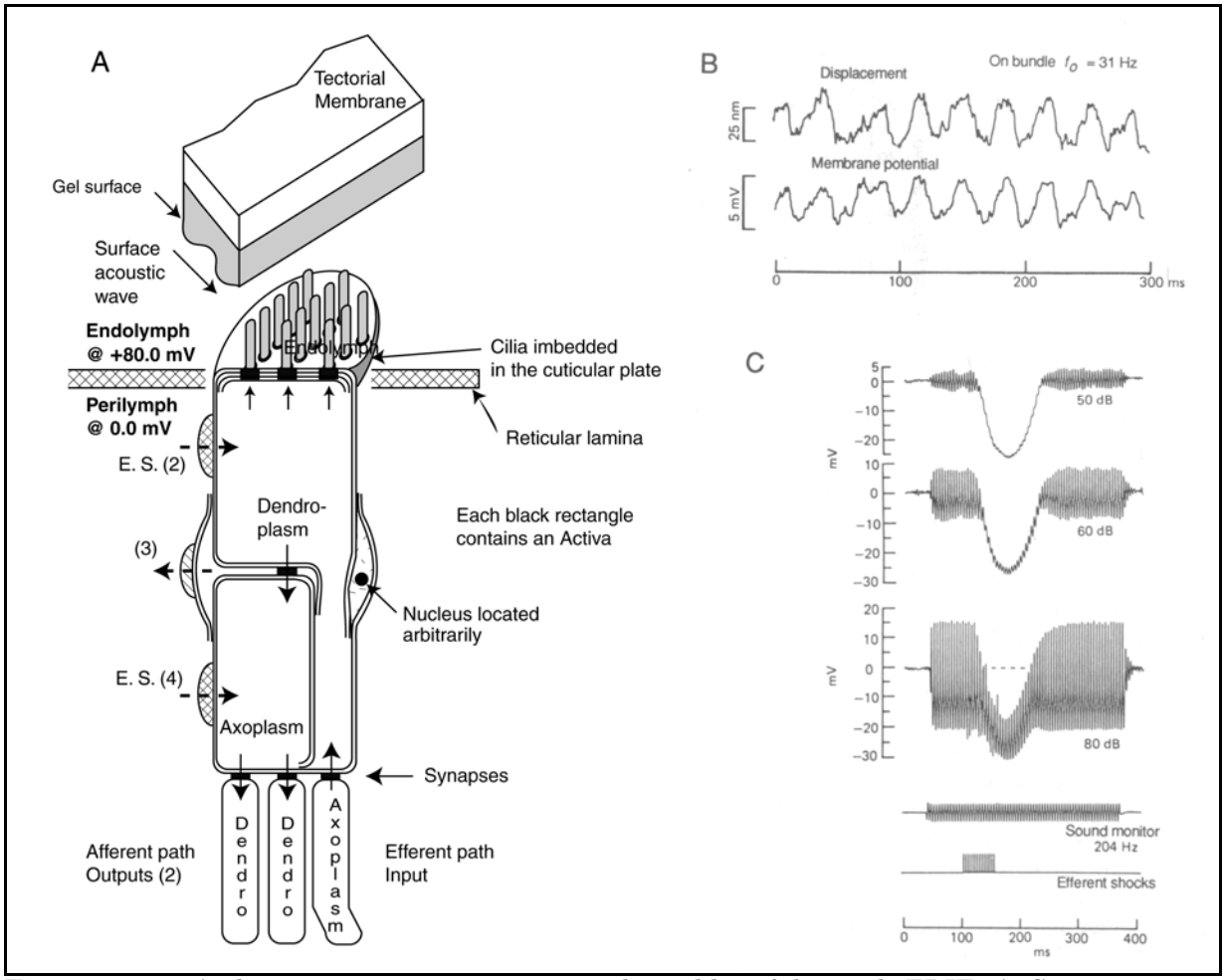


Figure 9.1.3-1 Auditory sensor measurements in the cochlea of the turtle EDIT. A; Generic sensory neuron with efferent input to the dendroplasm based on this work. Dashed arrows on the left represent DC electrostenolytic currents. No alkali metal ions are involved in the operation of this circuit. B: voltage versus pressure measurements in the turtle showing the fidelity of the voltage response under small signal conditions. No rectification is observed in these waveforms. Data from Crawford & Fettiplace, 1985. C; Amplitude versus intensity measurements in the presence of efferent desensitizing signals. See text. Data from Art & Fettiplace, 1984.

The positive action potential pulses arriving along the axon of the efferent neural path are integrated by the combination of the input synapse and the dendroplasm circuit. This results in a positive going pedestal potential in the voltage between the dendroplasm and the podaplasm of the sensory neuron. This positive excursion tends to turn the sensory neuron off. It caused the pedicle voltage to go 25 mV more positive than it otherwise would. The 204 Hz stimuli at 50 and 60 dB are completely cutoff by this mechanism. The 204 Hz stimulus at 80 dB is large enough that it is not cutoff by the efferent pulse train. However, the afferent signal is considerably distorted by the action of the efferent control

signal.

The effect of the 50 ms efferent pulse train is delayed by about 10 ms from the point of monitoring the efferent signal to the point of monitoring the afferent pulse train. The pedestal is seen to grow during the entire 50 ms efferent signal interval and then to decay with a similar time constant during an additional 50 ms. The overall impact on the output pulse stream approximates 100 ms.

The above papers are discussed in greater detail in **Section 5.4.6**. The many features of the sensory neuron demonstrated by these analyses suggest that the efferent neural system found in the cochlea of hearing is used for two purposes. First, signals from the motor neurons controlling the motions of the jaw appear to be sent also to the cochlea to provide a short term muting of the PNS during chewing. Second, it appears that occasional action potentials delivered to the sensory neurons over the efferent system can suppress any signal associated with noise sources found in the input circuitry of the sensory neurons. The presence of these pulses would be an effective deterrent to one form of tinnitus (Section 10.xxx). Since these pulses would also suppress hearing low level signals at the threshold of hearing, the pulses could be suppressed when a subject is concentrating on hearing such signals.

9.1.3.3 The use of Fourier Transforms in auditory research

[xxx address Fourier Transform versus Fourier sine series, Fourier cosine series etc.]
[xxx address Fourier Transform for repetitive signals versus non repetitive]

The concepts associated with the Fourier Transforms between the temporal and spectral (frequency) domain are not used with precision in many articles in the hearing literature. Considerable care must be taken to differentiate between four distinct types of signals in the temporal domain:

1. A true impulse function

2. A pulse function of finite characteristics

3. A continuous tone

4. A periodically interrupted tone.

The discussions of the Fourier Transform, and its specialized corollaries such as the Fourier Series and the *fast* Fourier Transform, found in texts on hearing are generally aimed at the undergraduate student without significant mathematical preparation[10]. Research activities require a more concentrated understanding of these important tools.

Contrary to the figure by Plomp reproduced by Green[11], every frequency component in the decomposition of a periodic pulse train consists of a continuous tone of constant amplitude for the duration of the train. A more appropriate rendition of Plomp's concept appears in Section xxx.

[xxx add words.

9.1.3.4 Understanding the frequency response of tonal and broadband channels

A major challenge during the last 50 years has been explaining the highly asymmetric shape of the frequency responses measured within the tonal channel of hearing. Most of these measurements have been based on motions of the basilar membrane. On the high frequency side, slopes of from 100 to 1000 dB per octave have been reported repeatedly based on many species[12,13,14]. Rhodes gave the precision of his estimate as 100 dB +/– 50 dB per octave, an astounding range from an engineering perspective. No estimate of the error, or even the physiological source, of the 1000 dB per octave value was given.

The **SAW-based Dual-Channel Electrolytic Theory of Hearing** provides a direct and simple

---

[10]Geisler, C. (1998) Op. Cit. Appendix A
[11]Green, D. (1975) Pitch Perception *In* Tower, D. *ed*. The Nervous System, Volume 3. NY: Raven Press pp 149-150
[12]Kiang, N. (1965) xxx
[13]Rhode, W. (1970) Observations of the vibration of the basilar membrane in squirrel monkeys using the Mossbauer technique *J Acoust Soc Am* vol. 49(4), pt 2, pp 1218-1231
[14]Evans, E. (1972) The frequency response and other properties of single fibres in the guinea-pig cochlear nerve *J Physiol* vol. 226, pp 263-287

explanation of these measurements. It centers on the operation of the dispersive properties of the two-dimensional SAW filter. See **Section 4.3.4**.

9.1.3.4.1 The simple theoretical case

There are four situations that can be compared with the frequency spectrum measurements found in the broadband and tonal channels (and by measuring nearby residual vibrations of the BM).

1. Multiple (isolated) resistor-capacitor circuits in series provide slopes related to small multiples of 6 dB per octave.

2. Inductor-capacitor circuits can achieve 12 dB per octave per stage.

3. More complex Chebishev and Butterworth passive filters can achieve 42 dB per octave.

4. Single stage tuned LC circuits with Q = 100 can achieve 60 dB per octave.

Asymmetrical filters can be formed by combining examples from 1 and 2 but their frequency responses will show the slopes indicated. Examples 3 and 4 always give symmetrical bandpass characteristics.

The asymmetry, and the steep slopes associated with the measured frequency responses quickly bring the appropriateness of the underlying protocols into question. **Figure 9.1.3-2** shows the actual versus the expected situation. Normally, when measuring a frequency response characteristic, the stimulus is designed to provide a constant intensity signal over a band of frequencies much wider than the circuit being evaluated. This has been the traditional assumption with respect to the above measurements. However, the model proposed in this work suggests this is not the actual situation. Frame A shows the frequency response of the stimulus as normally assumed and as it is based on the two-dimensional SAW-filter model of the Organ of Corti. The dispersive two-dimensional SAW filter removes the high frequency content from the signal as it moves down the gel-surface of the tectorial membrane.

Frame B shows the nominal frequency response of a bandpass filter. The filter may be formed of individual non-resonant filter sections, of a resonant filter section, or due to a window in conjunction with a spatial/frequency dispersion filter. The later configuration is frequently described as a monochrometer. The resonant and monochrometer forms are invariably symmetrical. The best frequency of the filter is described by its center frequency, BF.

Frame C shows the result of exciting the filter by the expected and actual stimulus and recording its output. Based on the two-dimensional SAW-based filter formed by the gel-surface of the tectorial membrane, energy at high frequencies is removed from the stimulus before it gets to the location of a particular IHC or OHC (or nearby accessible surface of the basilar or tectorial membranes). The result is the asymmetrical characteristic usually measured.

Because of the multiple filter sections of the dispersive filter occurring before the filter under test, the high frequency content of the stimulus is greatly attenuated. This is why the high frequency skirt of the overall response falls at such a high rate, typically greater than 100 dB per octave.

The measured BF is clearly an erroneous value with respect to the actual filter, and should be corrected in precision work. The correct BF is determined by the spatial/frequency capture area of the individual OHC or IHC relative to the gel-surface of the tectorial membrane. This capture area is

generally symmetrical about the location of the OHC or IHC.

9.1.3.4.2 Theoretical description of the frequency responses at the output of Stage C or 1 at high frequencies

The signaling channels of the neural system are designed to accommodate an instantaneous dynamic range of about 26-30 dB. This range may be limited by either background noise or limitations in the dynamic range of the amplifiers. However, it is common in audiometry to attempt to measure dynamic ranges of 40-120 dB. These wider dynamic ranges are normally not measured instantaneously, but over a period of time. This period of time is frequently long with respect to the adaptation time constants of hearing in the species under test. As a result, most of the test data in the literature must be described as "adaptation modified" data. Much of the spectral data acquired electro-physically is also cumulative data including the acoustic frequency response of one or more of Stages A, B and C, depending on how the stimulus was introduced.

These factors make the interpretation of data in the literature difficult without an adequate discussion of the acquisition protocol. **Figure 9.1.3-3** shows a modification of the previous figure using a logarithmic amplitude scale.

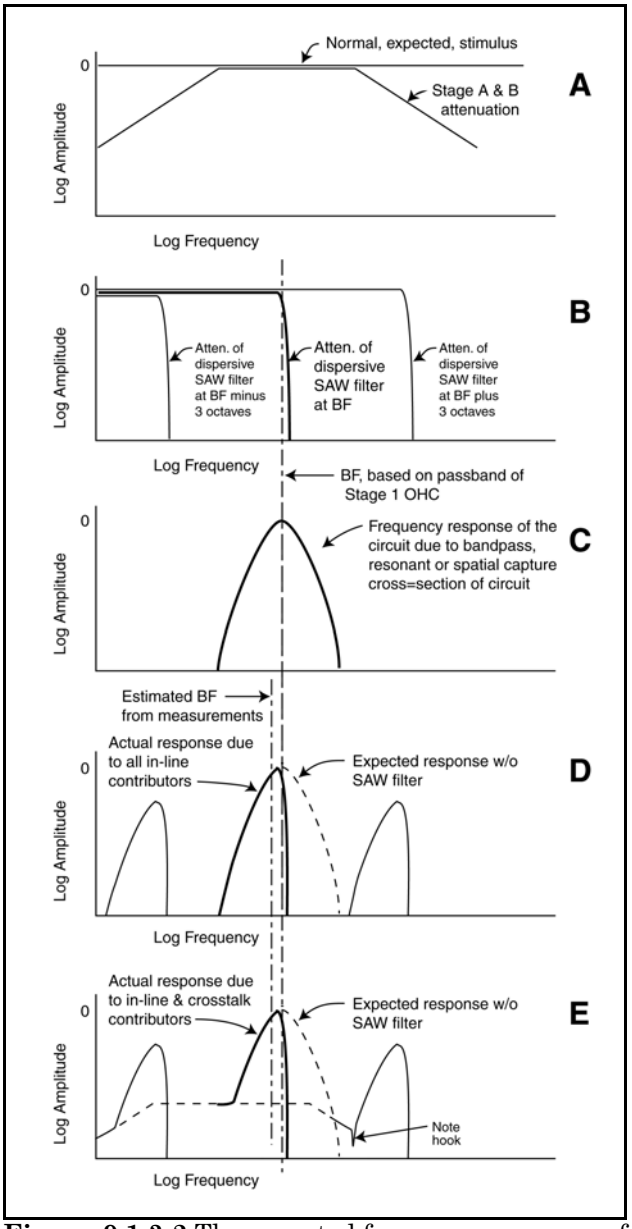


Figure 9.1.3-2 The expected frequency response of tonal and broadband channels in hearing. A; expected and actual stimuli spectra. B; frequency response of filter showing actual best frequency. C; measured frequency response of filter and stimuli showing error in estimated best frequency. See text.

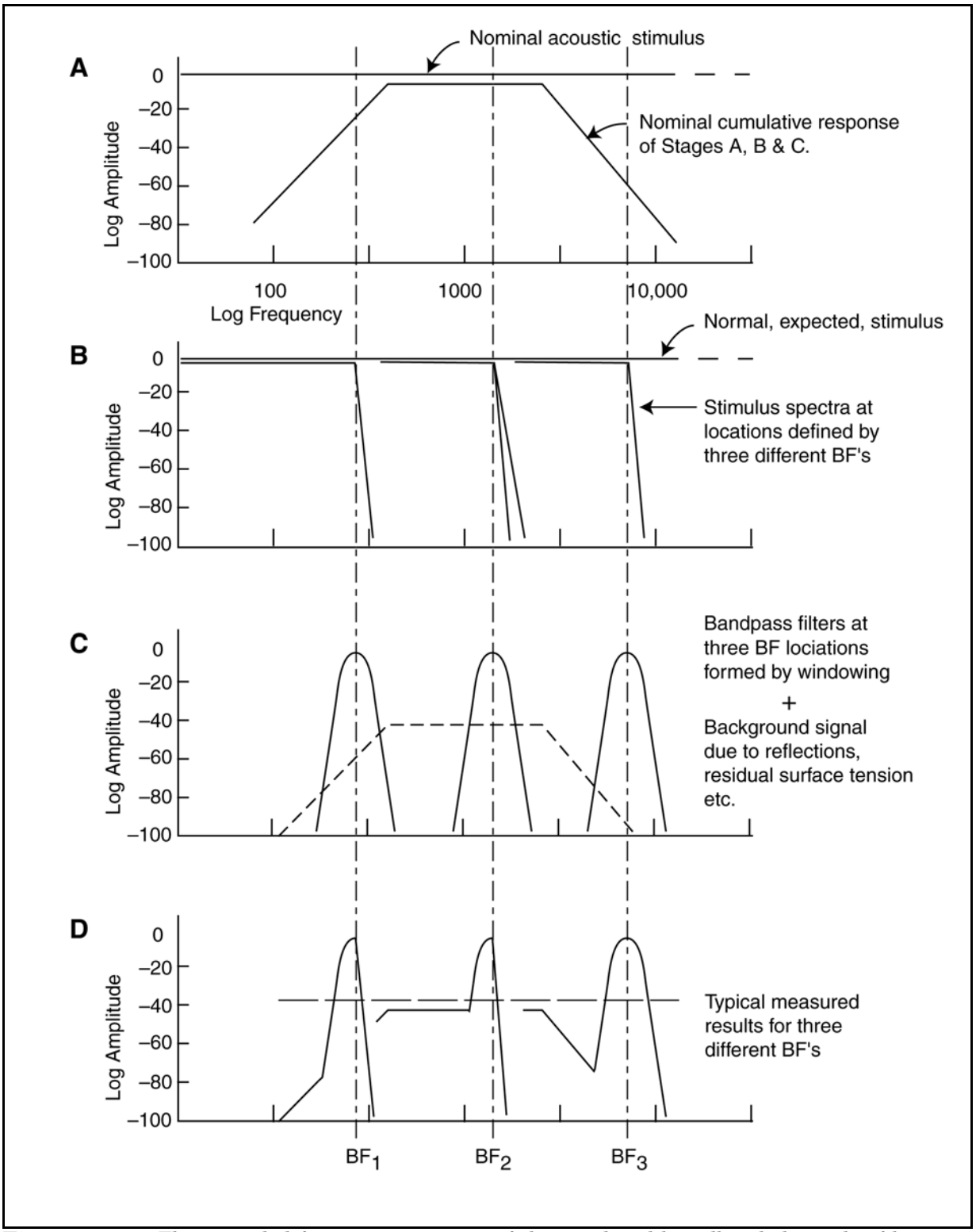


Figure 9.1.3-3 The extended frequency response of the tonal and broadband channels of hearing. Frame C shows the bandpass characteristics of the tonal portion of the gel-surface above the OHC. The bandpass of the broadband portion above the IHC extend farther to the left. Only portions of waveforms above the dashed line in frame D are used by the neural system. See text.

20 Biological Hearing

Frame A shows the spectrum of a flat stimulus and a nominal transfer function representing either the middle ear, Stage B, only or the cumulative performance of the outer and middle ears, Stages A & B. Frame B shows the transfer function of the acoustic portion of the inner ear under three conditions; with cutoff due to dispersion at the frequencies of BF_1, BF_2 or BF_3. Frame C is a composite showing two distinct features of the operation of the SAW filter on the gel-surface of the tectorial membrane. The overall energy spectrum traveling along Hensen's stripe in the absence of any dispersion is shown by the dashed line. All of the energy to the right of each BF is removed during the dispersion process. The remaining energy is sensed by the IHC. The energy associated with each BF is redirected in the direction of individual OHC. The energy spectrum associated with an individual OHC is shown by the three peaked responses. The width of these spectra are controlled by the dispersion performance of the filter and the capture cross-section of the OHC (the window associated with that OHC).

In measurements of the residual vibration of the basilar membrane, using either mossbauer or doppler laser interferometry techniques, spectra similar to those in Frame D are obtained. These spectra are the result of summing two acoustic vibration, the residual portion of the broadband spectra from the region of the IHC and the residual tonal energy spectra from the region of the OHC. The center frequency and relative amplitude of the individual combined spectra shown in Frame D depend on the location on the basilar membrane sensed by the instrumentation. The waveforms at BF1, BF_2 or BF_3 represent different locations along the length of the basilar membrane. The relative amplitude between the broadband and tonal components of each waveform depend on the lateral location of the sensed area along the basilar membrane. As the sensed area approaches the line of IHC, the narrowband contribution of the composite is reduced. Alternately, as the sensed area approaches the line of the OHC array, the narrowband contribution of the composite is reduced.

> This is the first Theory known to have described the origin of the notch to the left of the tonal response for BF_3 shown in Frame C.

The excellent agreement between the above waveforms predicted by this Theory and empirical measurements in the literature are excellent. [xxx refer results from both mossbauer and laser doppler] The data by Rhode (1970) using the Mossbauer technique is an example.

Empirical electrical measurements are available in two forms from the pedicles of the Organ of Corti. Most of these measurements relate to the OHC. [xxx provide references and discussion] They must be examined carefully. Many of them were acquired over periods of time that allow the adaptation amplifiers to change the electrical gain of the circuit while the test stimulus frequency is changed. Under these conditions, the waveforms are similar to those in Frame D. They consist of the primary tonal component, centered on the BF, plus a residual background from the broadband energy stream at a given OHC location. This background is introduced due to the greater than zero tension coefficient of the surface-gel and miscellaneous reflections of stray energy within the Organ of Corti.

9.1.3.4.2 Theoretical description of the frequency responses at the output of Stage C or 1 at low frequencies

[xxx belongs later in chapter. Only be the framework should be discussed here]
In 1996 Geisler & Cai provided a conceptual analysis of the apical portion of the auditory system based on the conventional unidirectional propagation of energy within the cochlear partition[15]. They introduce the term afferent auditory nerve (AN) fibers and ignore the terms IHC & OHC completely. The term IHC only occurs once in their entire paper, where it is equated to the discharge of auditory nerve fibers as a group (page 1554). The paper contains too many assumptions and interpretations, but the data appears useful. Some of the responses (ex. fig 3) appear to be related primarily to measurements of the motion of the basilar membrane. The three frames in Figure 5 utilize a variety of different scales without making the changes particularly evident. They introduce the term spatial –tuning curve which they describe as "produced, for any one chosen frequency, by plotting the threshold intensities, at that frequency, of all qualifying AN fibers in a particular animal, as a function of the fiber's location within the cochlea." They used the map of Liberman to relate frequency to the spatial location of their test (See **Section Figure 4.5.4.4**).

As a result, the "stylized drawings" in their figure 5 does not interpret the energy profile shown in their upper two frames at frequencies significantly below their frequency, f_2, the spectral contribution due to the IHC channels. Their analysis does not explain (or illustrate) the notch frequently observed adjacent to the peak associated with high frequency tonal channels (the right waveform in frame D above). If the material on pages 1550 through 1553 and the first paragraph of page 1554 of their article were extended to explain the role of the broadband (IHC-based) channels, and the bi-directional character of the SAW filter within the cochlear partition as a part of their analysis, it would be in excellent agreement with this work. The analysis would then offer a distinctly different explanation for the rapid falloff in energy at frequencies above CF =f_0. Such an analysis should overcome the need for most of their speculations in their discussion section.

Figure 9.1.3-4, reproducing their figure 2, shows the distinctive shape of their AN responses, which correspond to the OHC responses defined within this work. Note the reversal of the frequency scale in this figure.

The slope of the low frequency skirt of the 2.5 kHz OHC (dashed line on the right) corresponds to an attenuation of about 65 dB/oct. This slope appears to be shared with the low frequency skirts of all of the OHC in the figure.
.

[15]Geisler, C. & Cai, Y. (1996) Relationships between frequency-tuning and spatial-tuning curves in the mammalian cochlea *J Acoust Soc Am* vol. 99(3), pp 1550-1555

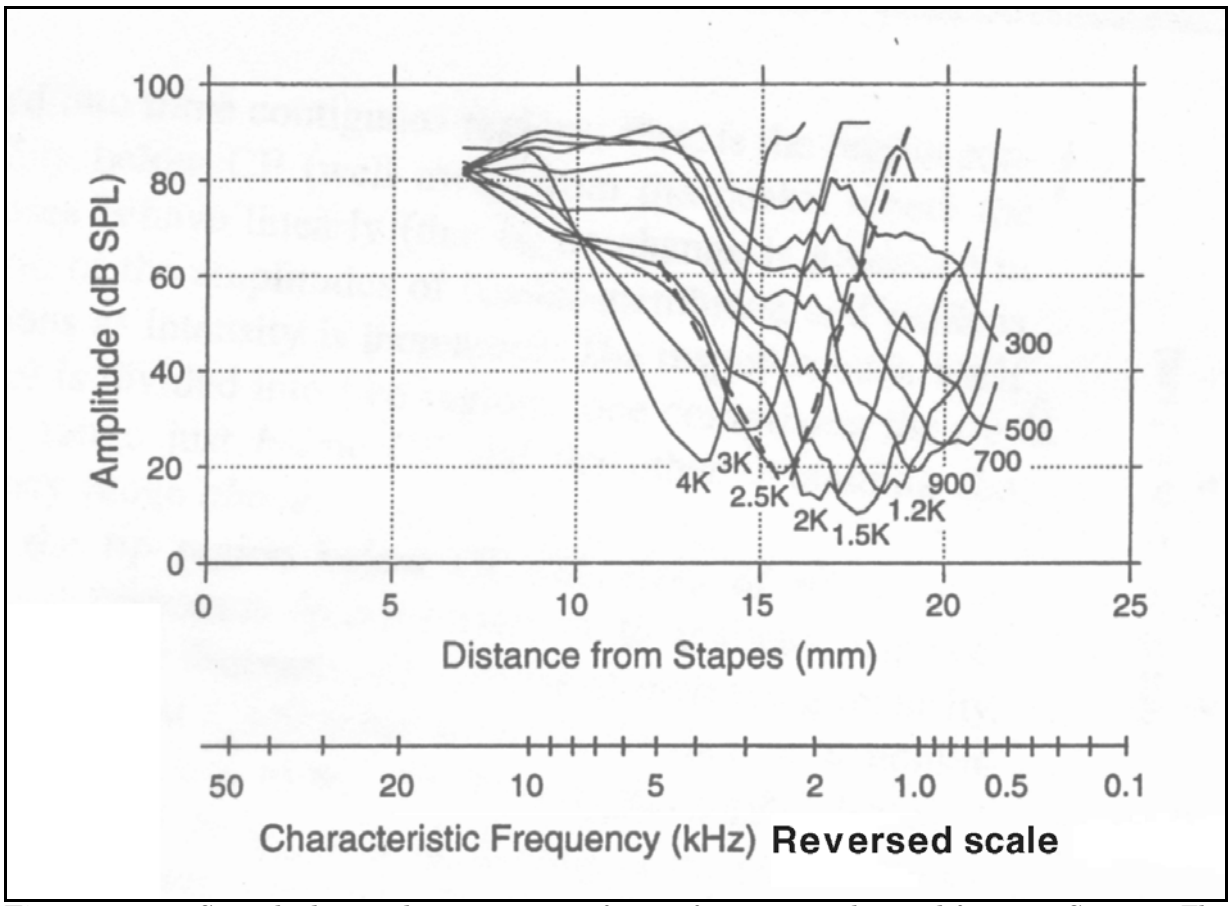


Figure 9.1.3-4 Smoothed spatial tuning curves for ten frequencies obtained from cat C93030. The first–order regression lines calculated for the rising and falling slopes of the 2.5 kHz curve are also shown (dotted lines). From Geisler & Cai, 1996.

9.1.3.5 The limited amplitude/frequency performance achievable in the acoustic environment

Investigators have frequently overlooked the performance of the IHC neurons while showing a preference for the OHC neurons. This has been due largely to their method of signal neuron selection. They have located sensory neurons that responded maximally to tones applied to the overall system. However, the OHC and IHC neurons are designed to process different signals. The OHC neurons are designed to process narrowband acoustic signals and are tailored to show maximum sensitivity, on a per cycle acoustic bandwidth basis, at a specific frequency. On the other hand, the IHC are designed to process broadband acoustic signals and are tailored to show maximum sensitivity to a broadband signal. As a result their sensitivity, on a per cycle acoustic bandwidth basis. is low for a specific frequency. However, it is high for a complex signal integrated over a band of frequencies. The quality of these neuronal circuits, often defined by their gain-bandwidth (or dynamic range-bandwidth) product, are similar for the OHC and IHC.

9.1.3.5.1 The frequency discrimination achievable by the OHC in the operational acoustic environment

A majority of the frequency responses found in the hearing literature exhibit a narrow response superimposed on a broader background response. These waveforms have found limited interpretation based on the previous paradigm. However, the current paradigm is based on the SAW-filter, acoustic antenna theory applied to the cochlear partition, the finite surface tension of the gel-surface of the tectorial membrane and signal summation at the output of some of the sensory neurons. This paradigm provides a more detailed interpretation of these frequency spectra. It is clear from acoustic radiation theory that it is difficult to achieve sidelobe levels greater than –20 dB relative to the peak in an acoustic beam[16]. It is also clear that the rapid adaptation of the sensory neurons allows the investigator to record the frequency response of an individual sensory neuron over a very large dynamic range if the stimulus application intervals are sufficiently long and not too closely spaced in time. If the test protocol is designed to prevent adaptation during the duration of a stimulus and during the interval between stimuli, considerably different frequency responses are obtained for a given sensory neuron. These responses are more indicative of the instantaneous performance of the auditory system upon which much of its operational performance is based. They typically exhibit an short term dynamic range of less than 23-26 dB. For the signals associated with the OHC neurons, the frequency response within 23-26 dB of peak response is that descriptive of the operational capability of those neurons.

9.1.3.5.2 The limited sensitivity demonstrated by the IHC in the operational acoustic environment

9.1.4 The use of PST Histograms in hearing research

The use of histograms in exploratory research can be very dangerous in the hands of inexperienced investigators. The histogram can manipulate the raw data in a variety of ways that effectively hide many of the characteristics of the original data. The problem is compounded by the ease of assigning a name to the histogram created by a particular investigator. These names are frequently non-descriptive at the detail-level and similar to those used by others under different circumstances. The problem has been and is compounded in auditory histograms where the relationship between the stimulus and the action potential stream usually recorded at node F is not understood.

By proper preparation of a histogram, a true replica of the signal waveform (the information content of the signal in the channel) can be recovered from the action potential pulse stream.

9.1.4.1 Background

[xxx expand to include background discussion of various types]
[xxx add period histograms in title and text]

9.1.4.2 The post-stimulus-time histogram

Kiang documented the use of the post-stimulus-time histogram (PST) in detail in his 1965 monogram[17]. In this monogram, he fervently called for a better theoretical model of the auditory system in the second paragraph. “Neurophysiology has at times been described as data-rich and theory-poor. The

---

[16]Koch, W. (1973) Radar, Sonar and Holography: An Introduction. NY: Academic Press pg 58
[17]Kiang, N. (1965) Discharge Patterns of Single Fibers in the Cat’s Auditory Nerve. Cambridge, MA: MIT Press research monograph #35

neurophysiology of hearing however may almost be described as speculation-rich and data-poor. There is a serious lack of systematic data at virtually every level of the auditory system." Unfortunately, Kiang did not change this situation and it remains true today. His concept of the PST based on the frequency of pulse streams failed to recognize the fundamentally different form of the modulation technique used in hearing. A new form of PST is developed and demonstrated in **Sections 7.3.4-5**). All of his voluminous PST data reveals new information if converted to the new PST format.

The following discussion is based on Kiang's original PSTs. It can be enhanced by introducing the transformation described in the referenced sections.

While Kiang has made many contributions in the field, the adoption of his label for the "primary auditory units" he focused on within the auditory nerve has not served the community well. It has obscured the actual operation of the auditory system. As shown in this work, the neurons found in the auditory nerve are actually stage 3 axons emanating from the spiral ganglia. The spiral ganglia obtained the initial signal information from the stage 1 sensory neurons by extending dendrites to them in a very complex arrangement that is indicative of the stage 2 signal processing accomplished prior to stage 3 encoding of the signals.

The summary provided by Kiang in his chapter 10 is supported by this work, with the possible exception of his item D. The transition from amplifier to integrator operation within the sensory neurons needs careful definition. In general, the transition can be defined in terms of the characteristic of the low pass filter causing the transition. In this case, the transition occurs between 500-600 Hz in a variety of animals. If one chooses to use the frequency at which the fundamental frequency component tends to disappear at node E (more than 20 dB below the amplitude at 500-600 Hz, then a frequency closer to 5 kHz is observed. This is the value Kiang relies upon (as demonstrated in his figure 4.6 discussed below). [xxx check this last sentence or two]

Kiang addresses the awkward term "inhibition" [his quotation marks] in the context of hearing on page 124. [xxx may not belong in this section.]

Kiang's explicit discussion of the PST form of data presentation was excellent. However, it suffered in three areas. Although he shows a finite pulse interval in an example in figure 3.3, his analysis was limited to the impulse situation. Second, it failed to note the folding of the original data that is inherent in the presentation. This folding obscures significant features in the data as discussed in detail by Smith & Brachman[18]. Third, as he noted using the term N_1 to describe the response recorded at the round window, "While the N_1 recordings are relatively stable from click to click, the time pattern of spike discharges fluctuates considerably for identical stimuli that are well above 'threshold'."

The problem with the PST is that it treats the data from both the amplifier regime and the integrator regimes of neurosensor operation the same. The nature of the PST is shown in chapter four of the monograph by Kiang. A temporal response is sampled using a much narrower sampling window, bin, where the bin number is determined from the peak in the response of the original stimulus, or in Kinag's case, the peak of the movement of the round window in response to the acoustic stimulus. Because of the velocity of the conductive acoustic wave through the cochlea, the difference between these two times is small relative to the signals generated in the neural system. The criteria for counting is whether the signal exceeds a specific threshold during the sampling interval, not whether

[18]Smith, R. & Brachman, M. (1980) Operating range and maximum response of single auditory nerve fibers *Brain Res* vol 184, pp 499-505

the response began during that sample interval. The protocol is to continue sampling the waveforms generated during a finite interval that includes multiple stimuli at generally at shorter fixed intervals. The bins, collecting data during the same interval after each stimulus are then summed.

For stimuli at frequencies below the neurosensor transition from amplification to integration (near 600 Hz in humans and possibly as high as 2 kilo-Hertz in cats), the resulting histogram displays the profile of the individual action potentials developed in response to the generator waveforms produced by the neurosensors. The first profile occurs following a total delay that includes the delay associated with the frequency discriminator, the initial neurosensor and any subsequent delays associated with stage 2 signal processing. The second profile generally occurs following a shorter delay. At very low stimuli repetition rates, the interval between the pulses in the PST is indicative of the interval between the action potentials generated to encode the intensity of the generator waveform. This interval is inversely proportional to the intensity of the stimuli.

If the interval between the stimuli becomes shorter, a situation is encountered where the histogram related to the second stimulus begins to overlap that of the first stimulus. Thus the interpretation of the PST in the absence of recognizing this overlap becomes difficult. In addition, the histogram related to only the first pulse may show some hysteresis due to mechanisms within the stage 2 and stage 3 processes. The stage 2 process may include a thresholding function and may also include pre-emphasis circuitry. Zwislocki in 1974[19] and Sokolich, et. al[20]. provide excellent examples of the effect of stage 2 pre-emphasis in the signals recorded at auditory neurons. The stage 3 process may exhibit a significant refractory interval under some conditions.

[xxx should this material be moved up into the background section?]
Figure 9.1.4-1 illustrates a variety of situations encountered in the PST histogram, including those described above. (**A**) shows the four basic types of signals that can be applied to the auditory system. The impulse is the basic function. It exhibits a finite area, defined by zero width and a finite height. The real pulse, exhibits a finite area also defined by a finite width and height. It need not be square. The third case shows a pulse of long duration, and includes a real rise time and fall time. The fourth is the sinusoidal excitation. These four types can be combined into any conceivable acoustic stimulus waveform using the rules of Fourier Analysis.

---

[19]Zwislocki, J. (1974) Cochlear waves: interaction between theory and experiments *J Acoust Soc Am* vol 55, pp 578-583, fig 6
[20]Sokolich, W. Hamernik, R. Zwislocki, J. & Schmiedt, R. (1976) Inferred response polarities of cochlear hair cells *J Acoust Soc Am* vol 59, no 4, pp 963-974, figs 4 & 5

(**B**) shows a typical impulse stimulus and the resulting generator waveforms at the output of the sensory neuron under five different conditions of stimulus amplitude. Note the latency is measured from the point where the response departs the horizontal axis. The latency is a function of the integrated area of the impulse. The amplitude of the generator response of the sensory neuron is also a function of the integrated area of the impulse. In most measured waveforms of this type, the latency includes a large component due to the finite velocity of the surface wave traveling along the tectorial membrane between the oval window and the location of the sensory neuron of interest. This component can have a value in multiple milliseconds.

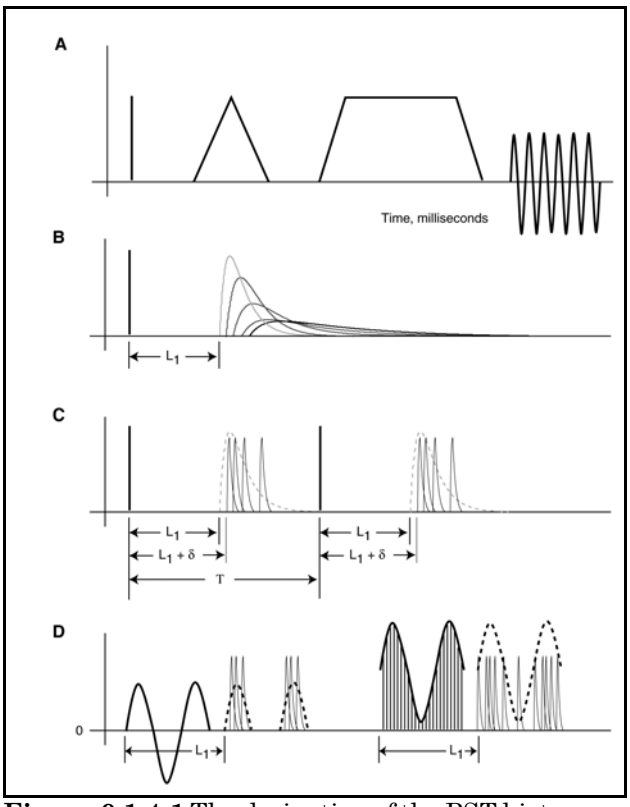


Figure 9.1.4-1 The derivation of the PST histogram and its peculiarities. A; four basic forms of stimuli. B; The responses of a sensory neuron to an impulse. C; the parameters of the responses of the auditory neurons to two simple acoustic impulses. D; the typical response of the sensory neuron and the following ganglion cell to a sine wave. By combining these forms, the generator waveforms and the action potential patterns associated with any stimulus can be described.

(**C**) shows the action potentials generated in response to the earliest generator waveform in (**B**). The first action potential is delayed by the travel time, δ, through the stage 2 signal processing. The following pulses are generated after delays proportional to the intensity of the initial generator waveform. They eventually stop when the remnant of the action potential no longer exceeds the threshold level of the stage 3 ganglion cell.

If a stream of acoustic impulses are presented to the ear, each impulse will generate a group of action potentials transmitted along an axon within the auditory nerve. If the time interval between the impulses is too short, the second group of action potentials will attempt to encroach on the time span of the first group. To the naive, this may cause the action potential stream to appear continuous. On the other hand, the ganglion neuron cannot generate action potentials below a minimum spacing determined by its refractory period. Thus, the overall action potential stream may show artifacts due to this refractory period.

When a low frequency sinusoidal waveform is presented as a stimulus as in (D), the sensory neuron generates an output waveform that tracks the positive excursion of the waveform based on the piezoelectric effect. The latency, L, has been shown expanded here for ease of interpretation. This generator waveform causes the stage 3 ganglion neuron to generate one or more action potentials during the positive half of the sinusoidal waveform. Laboratory results have shown that the sensor neuron does not generate an output during the negative excursion of the stimulus. In the absence of a response, the stage 3 action potential generator is also dormant.

For more complex stimuli, the generator waveform becomes more complex. It frequently exhibits a leading segment that reaches a higher potential than the sustained level for the remainder of the

waveform. This is characteristic of the stage 1 sensory neuron. The overall waveform also displays a tail as the pulse decays. In specific neural channels, the generator waveform of the stage 1 sensory neuron in response to a long pulse stimulus may be differentiated and summed with a copy of the original waveform in stage 2 in a process described as pre-emphasis. This process has the effect of advancing the initial action potential generated in stage 3 and preventing any action potentials being generated during the tail of the waveform.

All of the situations shown are for a stage 3 neural channel characterized by an axon with a negligible spontaneous rate of action potential generation.

By combining the illustrated waveforms, and recognizing the frequency limitations of the various mechanisms (particularly the limited bandwidth of the adaptation amplifier), the response of the first three stages of the auditory system to any stimulus can be described in detail. The response can also be expanded to include influence of the stage 0 physiological elements if desired.

When preparing a PST histogram, the complications related to the generator waveforms and action potential streams described above are frequently obscured by the binning process. Using (**D**) as an example, let a second sine burst be inserted at a time, T, following the first stimulus. As long as the interval T is sufficiently long relative to the time constant of the adaptation amplifier, the responses will be as shown in the figure following every time, $t_{n+1} = t_n + T$. By bining synchronously with the start of the first waveform using bins centered at sub-multiples of the interval, T, a totally readable PST histogram will be obtained. Using fixed interval bins can lead to complications if the intervals are not spaced at sub-multiples of the time T. The impact is minor but can be seen in many PST histograms. Investigators frequently note that repeated PST histograms show slight differences when all other parameters are maintained.

- - - - -

[xxx changes to a new subject here]

Figure 4.2 in Kiang shows a wide variety of PST histograms where the primary variable was the critical frequency, CF, of the axon examined. The implication of course is that all of the channels examined were spectral channels, even though their transient response was being documented. Up to 0.93 kHz, everything looks normal. [xxx continue analysis here.] The responses are a series of groups of bin counts separated by finite intervals. Note the latency of the first group in each set is inversely proportional to the CF of that set. The dominant factor in this latency is related to the distance of the relevant neuron from the base of the cochlea. This factor is approximately 0.15 to 0.2 ms/mm times the distance along the tectorial membrane to the neuron of interest. At 13.5 kHz, this latency factor has become less than one third of a millisecond while it is more than 4.5 milliseconds at 0.43 kHz.

The highest bin counts are found in the 1.0 and 1.2 kHz records, nominally the frequencies of peak auditory sensitivity. In the range of 0.43 kHz to 2.1 kHz, the time between each pair of groups is given by the reciprocal of the frequency, CF. This is a clear indication that the channel was amplifying the applied signal (or in the language of the hearing community, was phase-locking to the input stimulus). Beginning at 0.93 kHz, a new group begins to appear before the first major group. By 2.1 kHz, this new group is the group of dominant height. Beginning with 2.6kHz, only one group is recognizable following each stimulus at the scale of the provided histograms. Kiang noted that this inability to separate the pulses into discrete subgroups is intrinsic (page29). This initial group is providing a signal indicating the presence of energy at that frequency but the intensity of that energy is not high enough to support a second action potential after a period of 1-2 msec.

One of the procedures used by Kiang was to subtract the average SR of an axon from his data. While

this appears logical, it must be noted that the action potentials associated with high SR channels are not synchronous with his stimulus. These channels use a free-running oscillator to create the action potentials. The signaling data is encoded onto this free-running pulse stream by phase modulation. This method does not force the action potentials to start at a given time. Therefore, it is not surprising that his bin counts would not be repeatable as a function of time following the stimulus. Only in the case of axons with minimal SR is the action potential timing synchronous with the stimulus.

The practice of defining the latency with respect to the peak in the bin count of the histogram as used by Kiang (page 25), while common in hearing research, is not defendable on theoretical grounds. This interval includes at least four independent components. The first component is the acoustic delay within the outer ear–middle ear channel. The second delay is associated with the surface wave traveling along the tectorial membrane as mentioned above. The third is the finite delay associated with the generator waveform. The generator waveform exhibits an abrupt departure from the horizontal axis under noise free conditions. It is the time of this departure that defines the latency of the sensory neuron signal. The fourth is the rise time of the generator waveform after initiation, as given by the P/D equation. The rise time is a variable in itself. It is not directly related to the latency defined in the second component.

The fact that the delay associated with the velocity of the surface acoustic wave along the tectorial membrane is the dominant element in Kiang's latency is seen clearly if his figure 4.5 is replotted using the reciprocal of the CF. Note that his latency includes the rise time of the leading edge of the first cycle of the stimulus. The result is a curve that can be overlaid with the distance from the base of the cochlea versus the delay for neurons at that distance (**Section xxx**). His figure 4.6 is excellent evidence that the interval between the peaks in the PST histogram is the same as the period of the excitation at frequencies below about 5 kHz in cats. The relationship may extend to higher frequencies.

Histograms labeled period histograms, post-stimulus-tme histograms and peri-stimulus histograms (among others) will be reviewed here.

9.1.4.3 The use of histograms in analyzing action potential pulse streams EMPTY

[xxx period (refered to the source) and period referred to the nominal pulse interval of the pulse generators.]

9.1.5 The signal polarity resulting from acoustic stimulation

> The literature contains information concerning the polarity of the electrical signal(s) within the neural system resulting from a "unidirectional" acoustic stimulus. Caution is suggested in interpreting this data based on the compressibility of air. Within a short distance, the energy in an initially unidirectional stimulus will be observed to exhibit periodic regions of raised and lowered pressure typical of any propagating longitudinal acoustic wave.

It is useful to specify the polarity of the generator waveform relative to the acoustic stimulus. Smolder

& Klinke have provided data for the caiman, *Caiman crocodilus*[21]. Their data shows that action potentials at the output of Stage 3 ganglion cells are first observed following the rarefaction phase of a click. The minimum delay before a response is shown as approximately two milliseconds for a tonal channel sensory neuron with a BF of 578 Hz and 100dB SPL. This response generated many fewer impulses than the subsequent condensations and rarefactions. However, they discount the reduced count in their discussion. They note that this polarity is the same as in mammals. To achieve an action potential, a positive-going generator potential is required. Such a positive- going generator potential is the result of a negative-going potential at the base (dendritic) terminal of the corresponding sensory neuron.

9.1.6 Bandwidth limitations in the auditory system

The goal of the auditory system is the extraction of information concerning the location of acoustic sources and the interpretation of any cues associated with the temporal and frequency content of the stimulus from that source. This extraction is performed in a series of steps along one or more parallel feature extraction paths. These paths exhibit significantly different bandwidths, generally as a decreasing function of their distance from the ears.

The so-called conduction portion of the hearing system consists of the air passages and mechanical linkages of the hearing system, along with the ancillary sound conduction through the boney housing of the labyrinth. In the case of some marine animals with highly developed direction finding capabilities, this conduction is also supported by a large chamber filled with an oil of specific acoustic properties. These channels are broadband and when combined exhibit bandwidths up to about 20 kHz in humans. They can support bandwidths up to xxx kHz in bats and other species.
The mechanical portions of the cochlea are parts of this conduction system. However, the cochlea operates in two modes, a longitudinal mode and a lateral mode. The longitudinal mode exhibits a broadband frequency response compatible with the upper frequency achievable in a given species. The IHC arranged along this longitudinal structure all intercept a broadband range of frequencies, except their maximum frequency of interception decreases with distance along the cochlear partition. The lateral mode forms the frequency dispersive element of the auditory system. It disperses the acoustic energy based on frequency content (high frequency content first). The dispersed energy is intercepted by the OHC sensory neurons. As a result, different neurons are excited by a narrow range of acoustic frequencies. This mechanism is the origin of the frequency-place relationship of hearing. It also explains why the IHC farther along the cochlear partition only receive acoustic energy at ever decreasing frequencies.

The neural system exhibits a more complex arrangement of channels with ever decreasing bandwidths. The OHC and IHC neurons are able to faithfully convert acoustic signals into electrical signals up to a bandwidth of about 600 Hz in mammals. Above this frequency, the sensory neurons act as integrators (commonly but erroneously described as rectifiers). Above 600 Hz, they produce an output signal that follows the envelope of the acoustic signal applied to them. There are indications in the literature that there may be a subset of sensory neurons that have frequency responses up to about 1000 Hz and these neurons are involved in impulse based source location activities.

The neural system between the sensory neurons and the middle and upper brain operates within the frequency constraints imposed by the sensory neurons, generally below 600 Hz except for some possible neurons involved in source location.

---

[21]Smolders, J. & Klinke, R. (1986) synchronized responses of primary auditory fibre-populations in *Caiman crocodilus* (L.) to single tones and clicks *Hear Res* vol. 24, pp 89-103

The neural circuits of the middle and upper brain operate at lower bandwidths. Their output is characterized by a fusion frequency of about 20 Hz. This upper limit is analogous to the fusion frequency, sometimes called the flicker frequency limit of vision. While the upper limit of upper and middle brain activity is low, this activity involves a great many parallel paths. The resulting information handling capacity of the brain is much higher than suggested by the individual circuit bandwidth.

9.1.6.1 Difficulty in measuring the broadband slope above CF

Experimentalists have had a great deal of difficulty measuring and interpreting the slope of the broadband response above a nominal critical frequency. The reason has been two-fold, the lack of an adequate model of the auditory modality and the unique attenuation characteristic associated with the frequency discrimination function. **Figure 9.1.6-1** illustrates the problem encountered by only one investigator. The extracted parameter describing a deterministic mechanism had to be displayed as a scatter diagram.

With the appropriate filter roll-offs available from this work (**Section 4.3.4**), the appropriate data can be extracted from the responses more effectively.

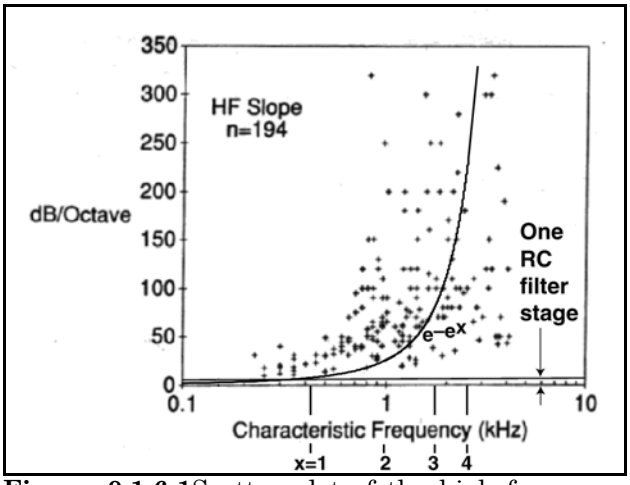


Figure 9.1.6-1 Scatter plot of the high frequency slopes of 194 neural tuning curves from *Tiliqua rugosa* (measured by taking the slope between the 3 and 23 dB points above CF) versus CF. From Manley, Yates & Koppl, 1988.

9.1.7 Re-defining the concepts of phase-locking, and synchronism in hearing

In the collection of signals at the Stage 3 axons of ganglion neurons, a lack of understanding of the fundamental nature of these signals has led to the introduction of the concept of synchronism between these signals and the stimuli (specifically the degree of synchronism between a group of action potentials and a particular stimulus). A similar concept describing the phase of the measured signals (which frequently involved a group of action potentials) relative to the stimuli.

With a clearer understanding of the nature of the Stage 3 axon signals, the concepts of phase-locking and synchronism can be largely eliminated from the hearing vocabularly. The signals at the Stage 3 ganglion axons can be described deterministically based on the acoustic signals applied to the mechano-acoustic portions of the ear, the analog signals generated and passed through the Stage 1 sensory neurons and the method of encoding used by the Stage 3 ganglion cells.

While there is a clear phase relationship between these circuit elements, this phase is not concerned with the term phase-locking as used in electrical engineering. In that common usage, the phase of an independent source is adjusted to be fixed (locked) to that of an initial source). Nor is the phase measured in hearing research in any way a result of a locking action between two signals. The phase of the resultant signal, anywhere in Stages 1 through 3 of the hearing system is a direct result of amplification and encoding of the original signal. Any shift in phase along the signal path is due primarily to the time delay resulting from the normal phase shift associated with analog filters, from

signal propagation delays and from signal regeneration delays.

The hearing system exhibits the unusual capability to sense signals at frequencies above the bandwidths of individual neural signaling channels. It does this by clipping the signal provided at Node D and integrating the resultant signal using a low pass filter prior to Node E. The resulting output remains synchronous with the envelope of the applied signal, but not necessarily with the fundamental frequency component.

9.1.7.1 Data reported concerning synchronism

Hind, et. al. have provided data on their synchronism coefficient based on a series of asynchronous two-tone tests[22]. The data appears to reflect the overall frequency response of the hearing system prior to the probe location in the auditory nerve. The "synchronization" is lost with respect to one tone and gained by the other tone as their relative amplitudes go through 1:1.

9.1.8 Using modulated signals in performance evaluation

Investigators have frequently employed amplitude modulation (AM), frequency modulation (FM) or other modulations to their basic acoustic signals to evaluate performance. Reviewing the literature, the usage appears to be more due to convenience than any clear technical objective.

Amplitude modulation of a test signal by a second frequency of less than 0.5 percent of the unmodulated carrier is analogous to natural acoustic signals, and is therefore easily, and efficiently, processed by the hearing modality. If the carrier frequency coincides with the best frequency of an OHC-channel, the signal at the pedicle of the OHC will display the modulating frequency of the AM signal. If the carrier frequency coincides with the sloping characteristic of the tuning curve of an OHC, the pedicle signal will be a distorted form of the modulating signal. If the AM signals is modulate by a second frequency significantly higher than 0.5% of the carrier, the situation becomes more complex. The signal will be spectrally separated into its components and individual components may stimulate separate OHC neurons. In this case, the output of a particular neuron will present the waveform of its particular modulation component. In both of the above cases, the IHC-channels will also be stimulated but in a less concise manner. The IHC channels are stimulated following high pass filtering. For AM signals with modest modulation frequencies (and carrier frequencies above 1000 Hz), the early IHC channels will each receive a full copy of the modulated signal and each will demodulate the signal completely. The pedicle output of each IHC channel will be a copy of the modulating signal. The same situation will apply to higher modulation frequencies and the earliest IHC channels. For more apical IHC channels, a point will be reached where the lowest modulation sideband will be filtered away leaving only a signal of the single sideband with carrier type. This signal will still be processed by the IHC subsystem but the output at the relevant IHC pedicle will be distorted in form. The appropriate stage 2 and stage 3 circuits are not designed to represent phase information associated with the above distortion. As a result, the distortions will frequently be overlooked in the signal propagating along the auditory nerve. If propagated, they will lower the performance of the stage 4 correlation and information extraction circuits.

The use of frequency modulated signals is a more complicated situation. Processing of these test signals is different in the OHC-based and IHC-based channels. Narrow band frequency modulated

---

[22]Hind, J. Anderson, D. Brugge, J. & Rose, J. (1967) Coding of information pertaining to paired low-frequency tones in single auditory nerve fibers of the squirrel monkey *J Neurosci* vol 30, pp 794-816

signals, where the deviation ratio is less than 0.5%, can be used to excite individual OHC channels. However, they are not demodulated by the OHC of an OHC channel unless the center frequency of the test signal coincides with the sloping side of the channels tuning curve. In this case, the FM signal is converted to a distorted AM signal and the apparent modulation waveform is reproduced at the pedicle of the OHC. For FM test signals with deviation ratios greater than a few percent, the situation becomes significantly more complex and depends on the frequency of the modulation. To save space, the reader is referred to texts on modulation theory such as Panter[23].

The conflicting perceptual data based on FM stimulation reported by Eggermont & Ponton can be rationalized based on the above discussion[24].

9.2 The top level circuit diagrams of Hearing

The literature contains a variety of simple block diagrams relating to one putative mechanism of the hearing system or another. It also includes a variety of mathematical models attempting to place various pieces of data in context but showing little relationship to the underlying physiology. One of the most recent of these mathematical models is by the Dau team[25],[26]. The mathematical model presented by the Dau team is only described relative to performance up to one kiloHertz. No effort was made to relate the model to the physiology of the hearing system of any animal. The filtering approach and sequence in Dau are not supported here. Other mathematical models, including Dau's appeared in a recent symposium record[27].

No material could be located that attempted to describe the circuitry of the hearing system in a comprehensive manner based on the physiology of the system. The following sections will present such a set of circuit diagrams. These diagrams are not meant to emulate or simulate the hearing system. ***They are meant to describe the actual circuits*** of hearing to the greatest extent possible using standard scientific and engineering symbols and nomenclature. The level of detail obviously becomes less as the description moves toward the signal manipulation and cognition stages of hearing. However, the circuits of Stages 1 through 3 are described in considerable detail.

9.2.1 The overall top level circuit diagram of hearing

]xxx consolidate with material in **Section 9.5.3.**]
As described in earlier sections of this work, the hearing system operates in a variety of modes in order to perform its mission of keeping the subject safe in, aware of and understanding of the important features related to its auditory environment. To do this, it employs a complex system of many subsystems. The system is highly modular. It uses one basic type of sensory neuron (a type that is shared with the visual, and presumably all, sensory neurons) in two distinctly different applications. It uses a signal processing subsystem, following sensing, that is strikingly similar to the visual equivalent. It uses a signal propagation system that is shared with all other sensory and motor neural

---

[23]Panter, P. (1965) Modulation, Noise, and Spectral Analysis. NY: McGraw-Hill Chapter 7
[24]Eggermont, J. & Ponton, C. (2002) The Neurophysiology of Auditory Perception: From Single Units to Evoked Potentials *Audiol Neurootol* vol 7, pp 90–91
[25]Dau, T. Puschel, D. & Kohlrausch, A. (1996) A quantitative model of the "effective" signal processing in the auditory system *J Acoust Soc Am* vol. 99(6) pp 3615-3631 (two papers)
[26]Dau, T. Kollmeier, B. & Kohlrausch, A. (1997) Modeling auditory processing of amplitude modulation *J Acoust Soc Am* vol. 102(5) Pt 1, pp 2892-2919
[27]Dau, T. Hohmann, V. & Kollmeier, B. (1999) Psychophysics, Physiology and Models of Hearing. Singapore: World Scientific

systems. The architecture of the auditory CNS appears to be the same as that of the visual and other sensory modalities, with the thalamus playing a primary role (rather than the cerebral cortex as generally assumed). The role of the thalamic reticular nucleus in the command and control of the hearing as well as other sensory modalities is critical to the operation of the system.

Section 1.3.1 provided a top level block diagram of the hearing modality. **Figure 2.4.2-1** provided a top level circuit schematic of hearing. **Figure 9.2.1-1** provides a more detailed top level circuit schematic of the hearing system. The many features of the hearing modality illustrated in the figure have been collected from the analyses and discussions in the previous chapters. Before going into detail, note that no probabilistic operators are used in the system, except potential random noise generators. The system is entirely deterministic. Also, except under pathological (or very unusual test) conditions, all signals below the dash-dot line are abstract. They exhibit no sinusoidal component(s). The signals consist of short term pedestal values (sometimes spoken of as DC values in the literature) associated with a specific location within the cochlear partition that can also be described as a neural address. These pedestal values will be called signatures. A signature may exhibit the temporal characteristics of the envelope of the stimulus, of the envelope of one frequency component of the stimulus, or of the envelope resulting from additional signal processing. The figure is an expansion of **Figure 4.3.2.2(B)** and continues to show the bilateral dispersive transversal filter described there. As a convenience, the high frequency inner hair cells are shown on the left and the lower frequency outer hair cells are shown explicitly on the right. In reality, these hair cell types are paired, with an incremental portion of the high frequency component of the traveling wave separated from the energy traveling along Hensen's stripe and redirected to the OHC. The residue is sensed by the corresponding IHC of the pair.

All of the sensory neurons of hearing exhibit the same neural circuit topology and essentially the same circuit element parameters. This topology, and the potential variation in parameters will be illustrated after this initial discussion.

Virtually all of the circuit pathways shown between the sensory neurons and the various engines of the CNS (in both nerves and commissure) are implemented by Stage 3 Signal Propagation circuits. These are not shown for simplicity. Most of the circuit pathways below the dash-dot line consist of multiple neurons and are appropriately designated nerves or commissure. Only a few paths have been made heavier to emphasize their multiple neuron nature.

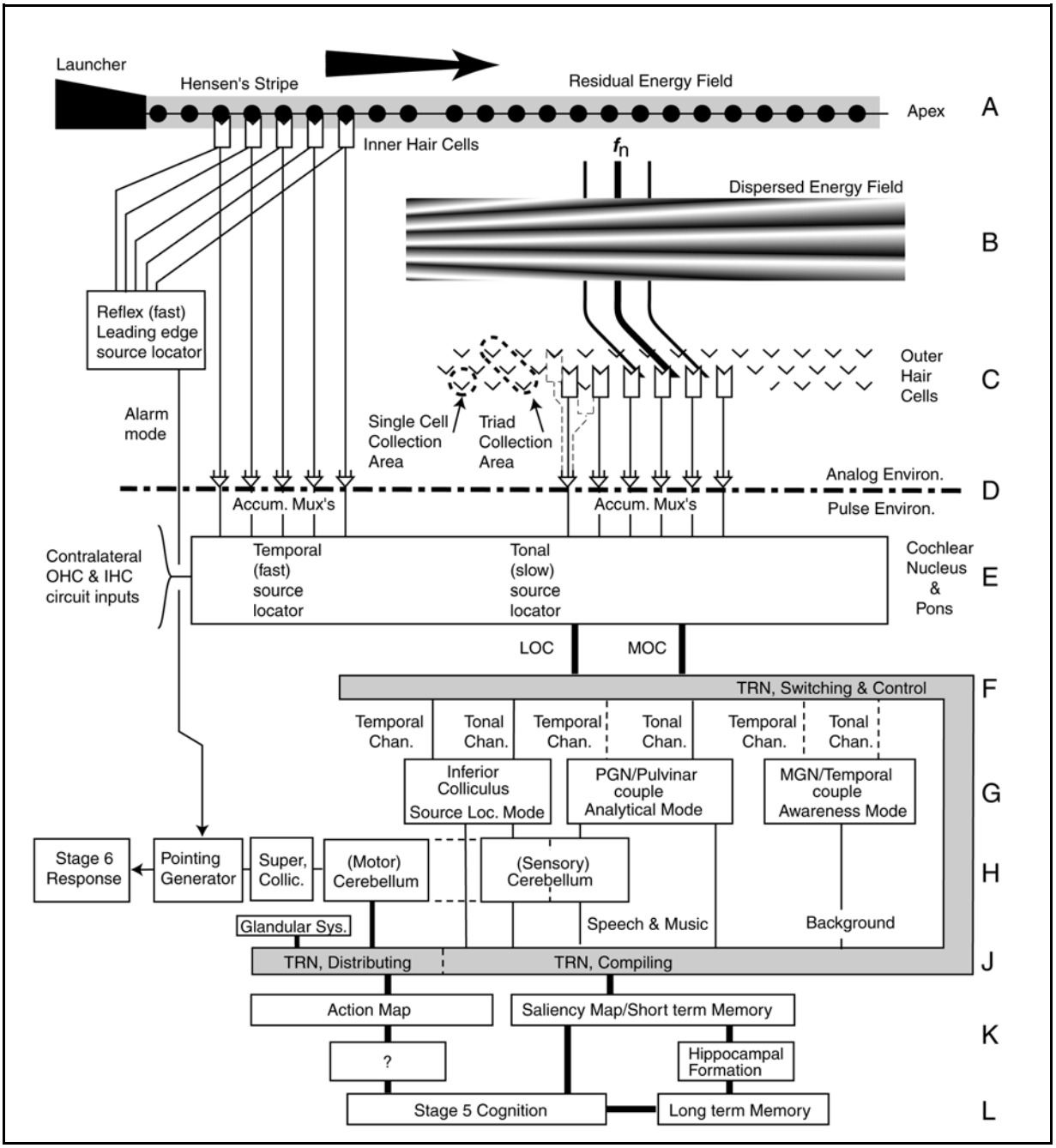


Figure 9.2.1-1 EDIT The top level circuit diagram of the hearing system applicable to humans and other advanced primates. All elements in the figure are deterministic. There are no probabilistic operators in the visual system (except that associated with random noise). At least five functionally distinct signaling pathways are illustrated using the labels A, B, C, D, E- - -. ADD WORDS

The figure starts at the upper left (of row **A**) with the launching of the wideband energy received from the oval window of the cochlea into Hensen's Stripe on the surface of the tectorial membrane facing the sensory neurons. The figure shows Hensen's stripe uncoiled for illustrative purposes only.

Hensen's stripe is not functional when uncoiled. This wideband energy progresses from left to right in a diminishing mode as the energy is dispersed into an energy field moving perpendicular to Hensen's Stripe. As the energy propagates along Hensen's Stripe, it encounters the stereocilia of the IHC sensory neurons.

The black dots are meant to represent the location of IHC cilia that sample the total residual energy field within Hensen's stripe. The rest of the IHC are only shown on the left for illustrative purposes. A key feature of the cochlear partition is the dispersal of the acoustic energy at **B** in accordance with the curvature of the coiled stripe. If mapped from above, the surface of the tectorial membrane would exhibit an amplitude profile similar to the rainbow shown (except it would be following the curvature of the stripe. The energy at a given frequency (represented by the hockey stick shaped lines to accommodate the uncoiling process) is intercepted by individual triads of OHC cilia (shown by the v-shaped symbols). Only a few OHC associated with the middle row of the hair cells are shown using solid lines.

The sensory neurons are only shown symbolically in rows **A** & **C**. Their circuitry will be discussed in detail in the next sub-section.

Row **D** highlights the multiplexers used to accumulate signals generated by the individual IHC and OHC. As discussed in Section 9.xxx, the degree and pattern of multiplexing is not well documented for either the IHC or OHC channels. The dashed lines between the OHC multiplexers and the OHC suggest the accumulation of all of the signals generated by one diagonal group of OHC as a minimum. The level of accumulation may be a function of distance from the base of the cochlear partition and may vary significantly between humans and other species. The output of the multiplexers are streams of pulses encoding the amplitude of the signatures described above. It is these stage 3 pulse streams that are projected to the next higher neurological level of hearing.

Row **E** represents all of the circuitry found at the neurological level of the Pons. This level is responsible for considerable signal conditioning prior to the projection of the signals to higher levels over the lateral olivary channel (LOC) and the medial olivary channel (MOC). In a general context, the LOC has been associated with wideband (temporal) signatures and the MOC has been associated with narrowband (tonal) signatures. However, many of the actual signals are more complex than these labels would suggest. The scope of the signals projected by the circuits within the Pons have not been fully investigated.

As shown, these channels also project signals resulting from binaural processing associated with source location processes. It is proposed that this initial source location processing employs more sophisticated circuits, and more extensive circuits, than proposed earlier by Jeffress. The proposed circuits generate pulse widths describing the time difference between the binaural signals generated from an individual stimulus. The signals generated by each ear describe the start time of both the overall energy envelope of the stimulus (using the IHC) as well as the start time of individual frequency components of the stimulus (using the OHC). This processing provides a wealth of information concerning not only the source of the stimulus but also reverberation within the local environment.

Row **F** illustrates the first of the many critical roles of the thalamic reticular nucleus, the control and switching of signals received from the PNS and the lower brainstem. As a general rule, both temporal and tonal channel signals are passed to each of the major elements of the mesencephalon. The binaural pulse width signals are known to be passed to the inferior colliculus. This work has discussed the tonal signals delivered to the multi-dimensional correlator of the PGN. It is highly likely that temporal signals are also delivered to this correlator. The character of the signals delivered to the

MGN remain poorly defined at this time. It is likely that the role of the MGN/temporal lobe couple is very similar to the well documented role of the lateral geniculate nucleus/occipital lobe couple in vision[28]. Both the PGN and MGN are shown closely associated with their lookup table counterparts, one half of the pulvinar and one of the temporal lobes.

The output of the IC, PGN/pulvinar, and MGN/temporal circuits are interps exhibiting minimal cochleo-topicity (although the nerves projecting those interps may exhibit receptive fields traceable back to multiple locations along the cochlear partition). The figure also describes the potential for additional information extraction by passing the interps to the sensory portion of the cerebellum for conversion into higher level percepts. All of these outputs are passed back to the TRN for compiling into a meaningful high level percept that can be placed into the saliency map (row **J**). As discussed in Chapter 8, the TRN plays a major role in controlling all of the major sensory modalities and most of the responses of the system.

The saliency map and action map of the parietal lobe, and the stage 5 functions of the frontal lobe (cognition) are shown symbolically at this time. Also shown symbolically are the stage 6 response tasks that are also controlled by the TRN. These include the actions of the glandular system as well as the skeleto-motor system. The motor system instructions, passed from the frontal lobes via the action map, are expanded into specific commands by the motor portion of the cerebellum (if necessary) before implementation by the superior colliculus and various signal generator nodes.

The multiple roles of the TRN and the cerebellum suggest their likely participation in many activities, including dreaming. The complete cerebellum is seen to occupy a unique position in the neurological system. It is used to learn new complex percepts (from all of the sensory systems) and to learn new motor response patterns. The complete TRN is seen to be in a position to control these learning processes and to control their latter expression. As an example, the TRN can allow the cerebellum to place percepts acquired at an earlier time into the saliency map. It can allow the cognitive functions of stage 5 to explore these percepts and to instruct the system to make specific responses. However, it can inhibit these responses. The result is a dream sequence where the cognitive system thinks something is or did happen and responds accordingly but the body does not actually move.

9.2.1.1 The computational models of the hearing modality–ear to cochlear nucleus

The material in earlier chapters has provided all of the foundation necessary to implement a mathematically sophisticated computational model of hearing. **Figure 9.2.1-2** presents a diagram indicating the nature of each of the major elements in the signaling chain from the external ear to the cochlear nucleus. Beyond the cochlear nucleus, individual paths must be discussed separately because the engines involved are mission specific.

The equations have all been written in a form that allows them to be concatenated into a single equation, requiring a computer to assemble and manipulate the resultant overall equation. However, once accomplished in order to demonstrate its realism with regard to the real hearing modality, the equations can be concatenated only as far as desired to develop the overall performance of the system up to any desired point.

To achieve the overall analysis and to establish the stability of the system as modeled, it is necessary

[28]Fulton, J. (2004) xxx book

to employ both s-plane and z-plane analysis from Control Theory of electrical engineering. These techniques are quite conventional, with the possible exception of modeling the output of the LOC (IHC) encoders into the z-plane. The method of encoding used introduces a variation in the pulse interval that is not normally encountered when a clocked sampler is employed in an analog to phasic conversion.

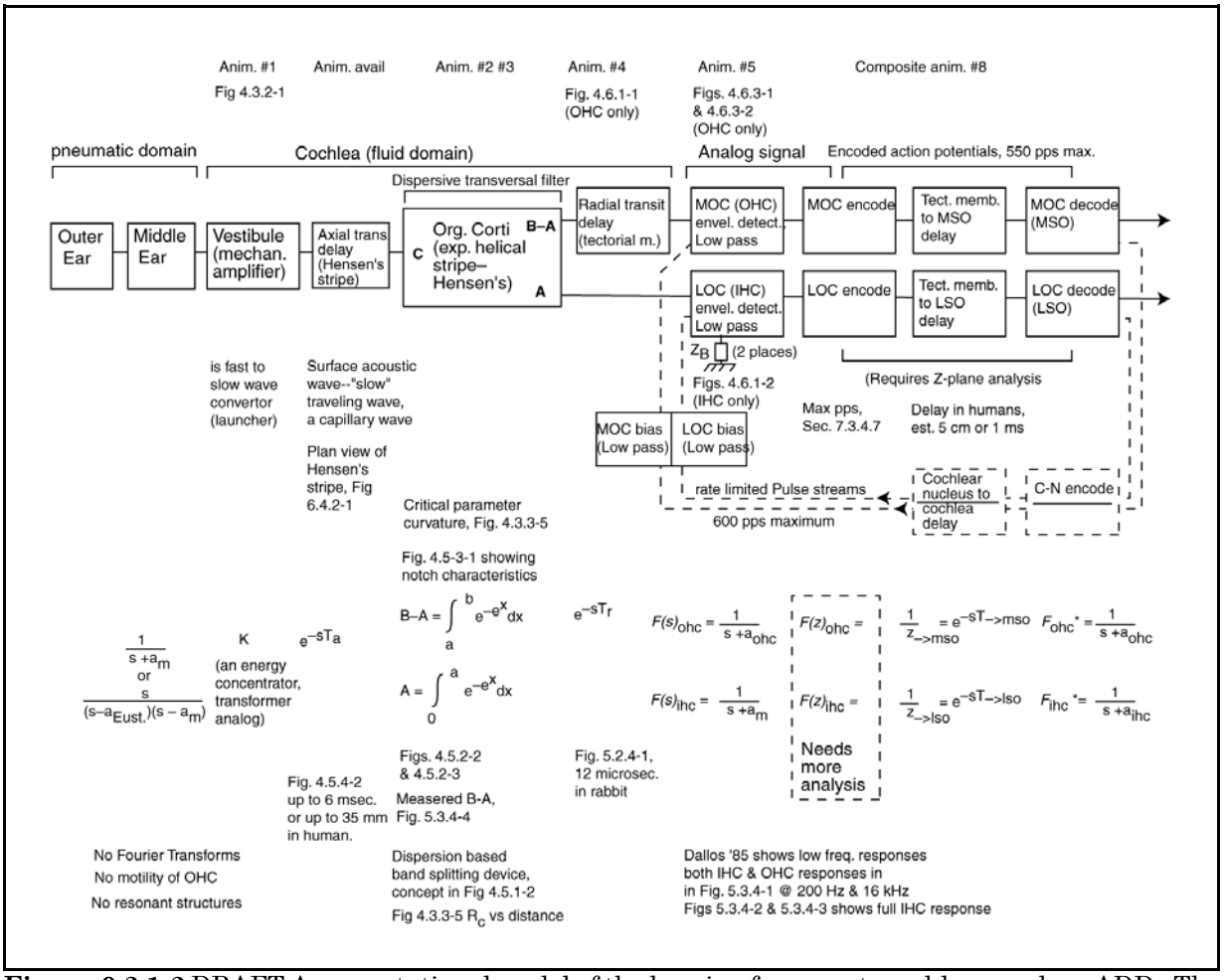


Figure 9.2.1-2 DRAFT A computational model of the hearing from ear to cochlear nucleus ADD. The dashed part of the figure can be removed in Illustrator. This figure should show all of the equations in a form that allows them to be concatenated.

As part of the energy concentration process, virtually no energy is delivered to the basilar membrane. The basilar membrane acts as a seismic block. As such, its motions are only as a result of the imperfect transfer of energy from the tectorial membrane to the sensory neurons (which are mounted to this seismic block)

.

The figure notes that the system employs no Fourier Transforms, no motility due to the sensory neurons, and no resonant structures at auditory frequencies (other than those of the outer ear and potential parasitic resonances within the structures of the middle ear). The concentration of energy within the vestibule, and the initial length of Hensen's stripe of the cochlea is the primary source of "amplification" prior to the electrolytic amplification following transduction within the sensory neurons. This concentration of energy is a passive process, not unlike that associated with a

transformer in electrical circuits (if the *voltages* of the transformer are equated to the *energies* of the slow surface acoustic wave launcher).

Note the equations begin in the pneumatic environment, provide a transition to the fluid environment of the inner ear, transition again to the analog neural environment and finally transition to the phasic environment before reconstituting the analog environment presented to the cochlear nucleus for the start of information extraction.

Figure 9.2.1-4 annotates the alpha characters shown on the Organ of Corti box. C is a wide band energy packet traveling along Hensen's stripe. B is an incrementally narrower energy packet after the high frequency energy packet, C–B has been stripped off into the OHC channel due to the curvature of Hensen's stripe in accordance with the Marcatili Effect. The procedure is repeated at B leading to the next incrementally narrower broadband packet at A with B–A being separated off into the OHC channel. Thus, the block labeled Org. of Corti in the above figure corresponds to the distance between adjacent IHC sensory channels of the curved bilateral dispersive transversal filter of **Figure 4.6.1-1**. This is the mechanism that converts the auditory modality from a single analog signal path for each ear into approximately 1000 individual IHC (LOC) and 1000 individual OHC (MOC) analog channels, all operating in parallel.

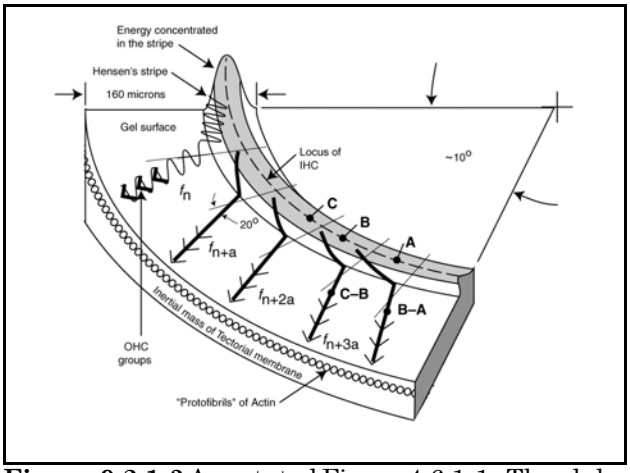


Figure 9.2.1-3 Annotated Figure 4.6.1-1. The alpha letters correspond to those along the Organ of Corti in the previous figure. C–B and B–A are narrowband energy packets separated from Hensen's stripe by the Marcatili Effect. A, B & C are sequentially wider broadband energy packets.

- - - -

Note also that the sensory neurons (a.k.a. hair cells) of both types operate as envelope detectors above their cutoff frequency and little or no output is presented at the carrier frequency with which they are associated. Their 3 dB bandwidth is on the order of 600 Hertz. The stage 3 propagation system samples the output of the sensory neurons at not over 550 pps, further limiting the unambiguous bandwidth presented to the cochlear nucleus to less than 275 Hertz.

The computational model confirms that the idea of a feedback signal from the sensory neurons through the cochlear nucleus and back to the sensory neurons involving a high auditory frequency is not plausible.

9.2.1.2 The top level circuit diagram of the sensory neurons of hearing

Figure 9.2.1-4 shows the top level circuit diagram of the generic sensory neurons of hearing. The figure can be compared with a variety of simpler block-type models of the sensory process. The circuit shown is generic. The more specific circuits for individual OHC and IHC neurons are developed in

Chapter 5. Each sensory neuron contains two Activa forming distinctly different circuits that are closely linked in a configuration known as a differential pair. The Activa on the left forms the adaptation amplifier and that on the right forms the distribution amplifier. The adaptation amplifier can be considered a distributed amplifier with a piece of it supporting each stereocilia of the cell. However, the total current through the equivalent emitter (the terminal with the arrowhead) flows into the differential circuit containing the impedance labeled (2). In a differential amplifier circuit, the two emitter currents shown operate to maintain a constant current through (2) to the best of their ability.

The OHC and IHC sensory neurons are called upon to accomplish a variety of tasks. The most important is to provide the necessary variable degree of amplification so that the remainder of the neural system can operate in a nominally fixed amplitude signal environment. That is to say, additional amplification is not found within the remainder of the neural system, except possibly indirectly. This requirement is the reason for the adaptation mechanism. At stimulus levels below 45 dB SPL, the sensory neurons operate at maximum amplification factor. For intensities between 45 dB and 100 dB, the amplification is continuously reduced in order to limit the maximum signal output level at the pedicles of the neurons.

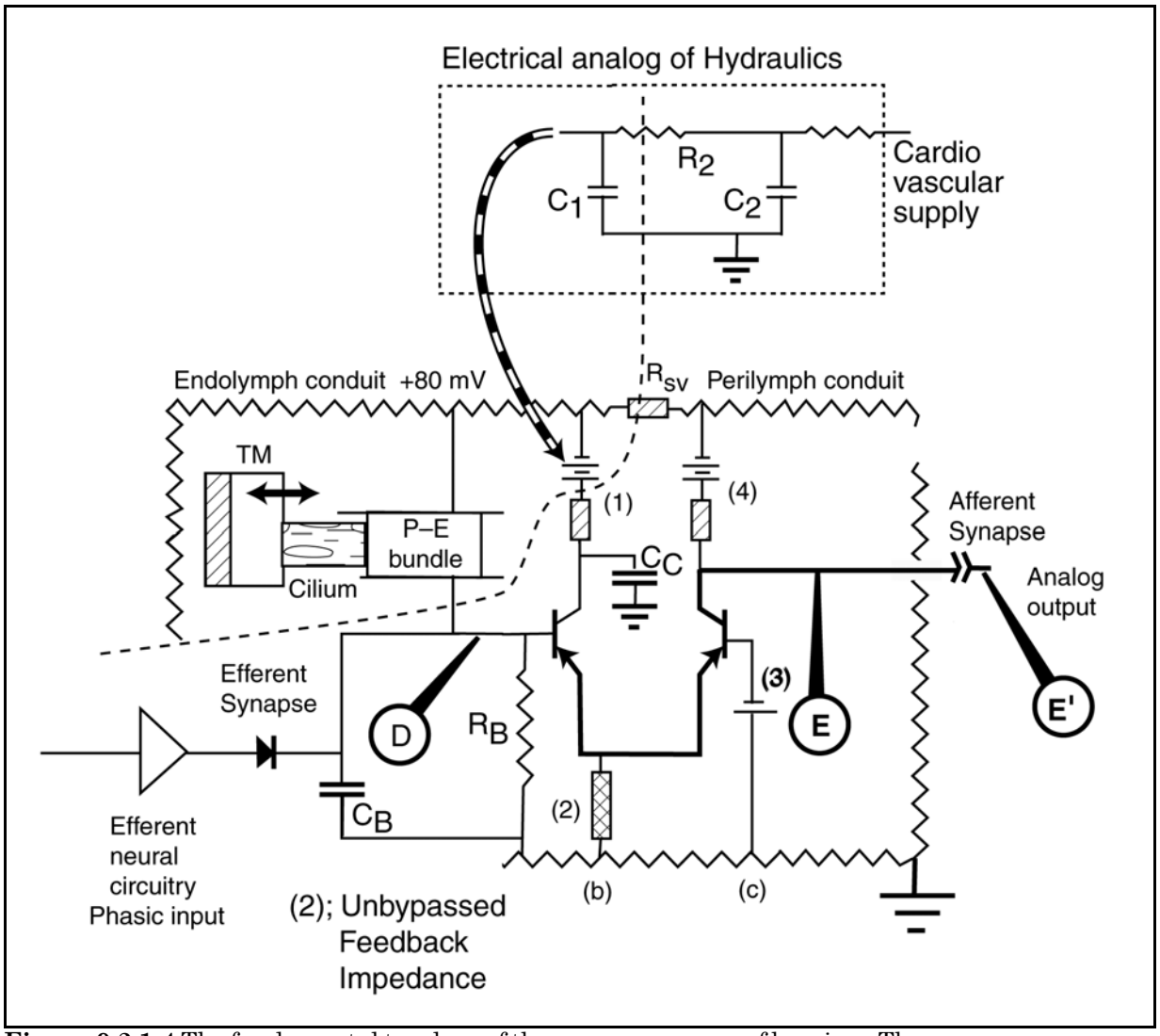


Figure 9.2.1-4 The fundamental topology of the sensory neurons of hearing. The sensory neurons are immersed in two electrically isolated fluid conduits. The dynamics of the piezo-electric bundle is not shown in this figure. The bulk of the tectorial and basilar membranes act as passive inertial masses in this circuit.

Variable amplification (gain) is achieved in the adaptation amplifier circuit by a variety of mechanisms. The dominant mechanism involves the continual replacement (not shown) of the cross-linking chemical bonds within the piezo-electric material in the base region of the cilia. If these bonds are broken faster than they are replaced, the average signal amplitude at node D will be reduced and the gain of the adaptation amplifier circuit is reduced. A more rapid mechanism involves the average potential between the base and the emitter of the adaptation amplifier. This potential can vary rapidly and change the average current through the adaptation amplifier Activa. This will result in a change in the average value of the potential at the distribution amplifier pedicle (node E).

Finally, the efferent neural circuitry connecting to the input circuit of the adaptation amplifier can also

modify the gain of that circuit. The amplification achieved in the adaptation amplifier is also controlled by the electrical bias at the point labeled (D) relative to the emitter terminal of the Activa. Normally, this bias is generated by the impedance, R_B. However, in some situations, this bias is further controlled by the efferent neural system as shown. The efferent neurons synapse with the junction between the base terminal of the Activa and the impedances, R_B and C_B. In this situation, the action pulses delivered to the synapse are rectified by the synapse and the resistor and capacitor network. This action changes the DC bias at (D) in proportion to the pulse rate of the action potential stream.

The electrical signal applied to the base terminal of the Activa is generated by the piezoelectric transducer associated with each stereocilia. This transducer generates a voltage in proportion to the stress placed on the stereocilia by the motion of the surface acoustic wave on the surface of the tectorial membrane (with the bulk of the tectorial membrane and the bulk of the basilar membrane acting as stabilizing ***and non-functional*** inertial masses).

The bandwidth of the base circuit of the adaptation amplifier is very wide and plays not role in the hearing system. However, the bandwidth of the collector circuit of the adaptation Activa plays a major role in hearing. This circuit has a nominal bandwidth of 600 Hz in most mammals. While configured as a low pass circuit. it is part of the negative internal feedback network formed by the impedance labeled (2). As a result, the overall adaptation circuit operates as an integrator as well as an amplifier. Furthermore, its feedback coefficient varies with signal intensity level because of the limited ability of the distribution amplifier to equalize the current flowing through the emitter terminal of the adaptation Activa. As a result, the adaptation amplifier exhibits a gamma (ratio of AC output to AC input signal level) of one at low signal levels and a gamma of nominally 0.3 at signal levels in the phonotopic regime of 45 to 100 dB SPL.

All of the above discussion related to the current through the adaptation and distribution amplifiers. However, the output of the distribution amplifier is an AC voltage. The conversion from a current to a voltage is performed by the impedance labeled (4) acting as a diode. This diode is formed by the type 2 BLM associated with the pedicle of the axon of the neuron. As a result, the output of the sensory neuron is an AC voltage representing the current change within the sensory circuits. At high DC current levels, the AC voltage is compressed due to the curvature of the diode characteristic.

Because the adaptation amplifier acts as both an amplifier and as a integrator, the output of the neuron consists of a pedestal (a slowly varying component roughly describing the envelope of the stimulus) and a true copy of the high frequency component of the stimulus waveform. As the frequency of the high frequency stimulus increases, the amplitude of the “true copy” decreases beginning in the vicinity of 600 Hz and tends to disappears at frequencies above 2,000 Hz. At high stimulus frequencies, it is the pedestal that constitutes the output signal of the individual sensory neuron.

The output of every sensory neuron is delivered to one or more afferent synapses. Each synapse is an active device formed by an Activa. They act effectively as electrical diodes of a nominally fixed impedance. This impedance does play a role in the Stage 2 signal processing following the sensory neurons. In some cases, the capacitance between the input (in this case the emitter) terminal and the output (collector) terminal is an important parameter of the synapse circuit.

The electrical analog at the top of the figure represents the power supply supporting the collector terminal of the adaptation amplifier. It plays a major role in the peristimulus and post-stimulus performance of the sensory neurons (**Section 5.x.xxx**).

9.2.2 Potential top level circuit diagram of the tonal channels of hearing

42 Biological Hearing

The potential performance of the human hearing modality depends on the architecture of the multiplexing plan shown in the top level circuit diagram. Determining this architecture is a very difficult task requiring the latest instrumentation. No significant investigation into this architecture has occurred in the last 50 years. This section will describe the optimum capability of the modality on the assumption that signals from each triad of OHC sensory neurons are summed in a separate stage 2 multiplexer. Later discussions in this chapter will consider higher levels of summation (involving multiple triads) that may be documented along the length of the cochlear partition. The level of summation need not be uniform with distance along the partition.

Figure 9.2.2-1 describes the circuit diagram proposed for the tonal channel of hearing. The OHC are shown along the left edge beginning with the nominal cell #1 at the apex and proceeding to cell 3456 (3500) at the base. Only one row of the OHC is illustrated. However, the accumulation multiplexers are shown combining the signal from the OHC in each of the three rows of OHC that exhibit the same nominal characteristic frequency. The signals from these multiplexers (the initial neurons associated with the spiral ganglia) are passed directly to the multidimensional correlator of the thalamic reticular nucelus (TRN) for correlation. The multiplexers are grouped in nominal sets of 432. Each set consists of the signals representing one octave of frequency and is represented by a distance of 4 mm on the nominal human cochlear partition. The first 432 signals are introduced along the bottom edge of the correlator as shown, beginning with the first accumulation multiplexer representing the characteristic frequency centered on 27 Hz (note A0 on the musical scale). Multiplexer 433 delivers its signal to the location corresponding to the note A1, multiplexer 865 delivers to location A2, etc.

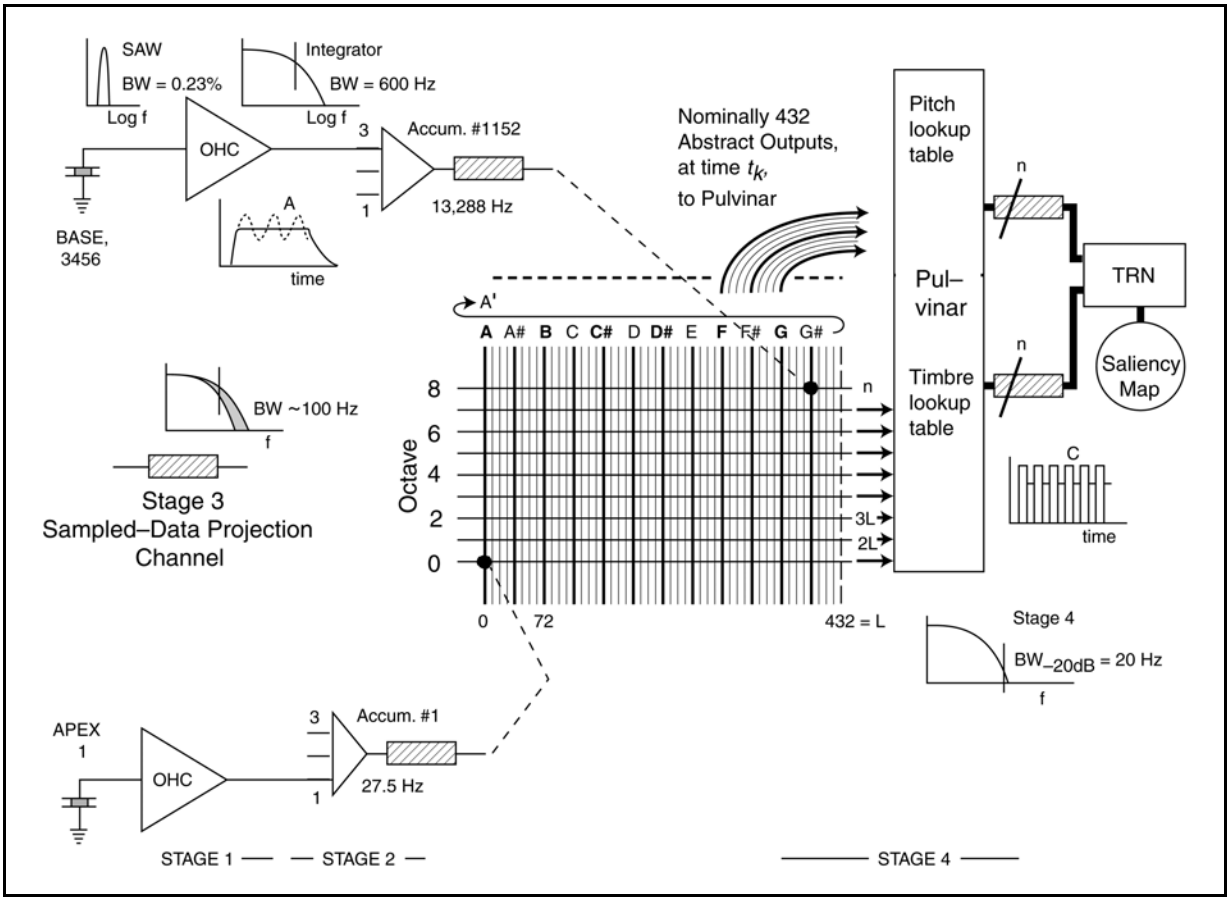


Figure 9.2.2-1 Candidate Circuit diagram of the tonal channels of hearing. This candidate would provide optimum frequency segregation by summing each individual row of OHC sensory neurons. For illustration purposes, the lowest row of sensory neurons is assigned a characteristic frequency of 27.5 Hz. The highest row is assigned a characteristic frequency of 13,288 Hz. The correlator is arranged in 9 rows of 432 columns. Each row extends over one octave. Column spacing is 0.23%.

The correlator employs sensing neurons arranged vertically. The dendrite of each sensing neuron synapses with the axon at each node within a specific column (nominally eight nodes). of the two dimensional correlator shown. There are nominally 432 sensing neurons. The axon of each of these sensing neurons (shown by the curving fan of lines)delivers its output to the lookup table provided by the auditory portion of the pulvinar. The pattern displayed by these 432 sensing axons at any given interval are compared with the stored data of the pulvinar to establish an initial interp describing the information contained within the original stimulus during the equivalent earlier interval (accounting for the time delay of the overall signal paths). As shown at the right, the number of signaling channels used to form the output of the lookup table is not known. The signals associated with these channels are delivered to the saliency map for cognition by the engines of Stage 5. In complex situations, the information may be processed further, in conjunction with the cerebellum, leading to a more sophisticated percept that is delivered to the saliency map.

The bandwidths of the channels leading from the dendrites of the OHC to the saliency map is continually decreasing. The OHC exhibit an output pass band with an integration frequency of 600 Hz, the Stage 3 circuits exhibit a nominal limit of 100 Hz, and the Stage 4 circuits exhibit a passband limited to about 20 Hz (the fusion frequency) for any individual circuit path.

The viability of this diagram is supported from several perspectives. It is compatible with the just noticeable frequency differences measured using highly trained subjects (**Section 9.6.2**) and it is compatible with the perceived pitch performance of similar subjects (**Section 9.7.4**).

By treating the signals collected by accumulation of the signals from the three OHC exhibiting the same characteristic frequency, the number of individual channels remains at approximately 3500. If these are distributed across the rows of the correlator in groups of 432, the correlator displays 8 octaves. Each octave consists of 12 semitones with the space between semitones divided into 36 finer tones. This very fine spacing agrees quite well with the just noticeable frequency difference achieved by highly trained individuals. This just noticeable difference is in the region of 1/64 to 1/72 of a semitone. Ward notes, "Under optimum conditions, changes of as little as 0.1% in frequency (Konig, 1957) can be detected 75 percent of the time by the best listeners[29]." The model and the literature suggest the just noticeable difference is a statistical value corresponding to about 1/72th of a semitone, with a 50/50 chance of being described as the 1/36th semitone on either side of it.

The ability of a human to report his perception of a stimulus is highly dependent on his level of musical training. In the case of an individual who is musically naive, his response to almost any stimulus is merely that it is relatively high or low and pleasant or unpleasant. In my case, I have never acquired any sense of a musical scale. Neither am I aware of any harmonic relationships although I am very sensitive to "sour notes" and certain discordant sounds when I encounter them. Highly trained individuals perceive a completely different acoustic environment.

The figure suggests that a trained human perceives pitch as related to a specific column of the two-dimensional correlator. The fineness of pitch determination depends greatly on the level of training, or experience, of the subject as well as the stimulus used. Pitch determination based on a single pure tone may be more limited than that based on the harmonic series of tones produced by an instrument.

The more tones causing amplitude values to be transferred to places within a single column of the correlator, the higher the amplitude of the output of the sensing neuron related to that column. The higher the output related to a single column, the stronger the perception of that column. If the subject is trained to relate the perceived sensation related to that column with a specific fundamental frequency, he will forever after describe that sensation by that fundamental frequency (unless he is retrained).

Note that the fundamental frequency does not need to be present among the places receiving a signal within a column. The column describes all of those places that are harmonically related in cochleotopic space. This fact plays a major role in music theory. It is compatible with the harmonic character of the tones produced by a single note on such an instrument and explains the significance of certain simple musical chords known to be pleasant to the ear.

9.2.3 The critical bandwidth of human hearing

Past experiments have frequently defined a ubiquitous but unexplained critical bandwidth associated with the hearing system when subjected to virtually any narrowband applied stimulus. This broad critical bandwidth appears to describe the susceptibility of the system to external interference while

[29]Ward, D. (1970) Musical Perception *In* Tobias, J. *ed*. Foundations of Modern Auditory Theory. NY: Academic Press Chapter 11

focused on the applied stimulus.

A unique feature of this critical bandwidth is that it is always centered on the frequency of the applied narrowband stimulus. Because of this feature, the critical bandwidth cannot be described in the context of a group of fixed bandwidth filters. The critical bandwidth feature appears to be associated with the concept of attention. Based on the above figure, the critical bandwidth appears to be related to the number of sensing neurons that are of interest to the pulvinar interp extraction mechanism at any given instant. This bandwidth will be discussed in greater detail in **Section 9.xxx**. It is typically described as on the order of ±8% of the stimulus frequency.

[**Figure 9.2.2-1**] suggests an interesting feature of the critical bandwidth that can be verified by experiment. The figure suggests that the critical bandwidth is a feature of the output side of the multi-dimensional correlator. When reflected to the input side, the susceptible frequency range is described at the cochlear partition by a Riemann Transform. This situation suggests the hearing modality is susceptible to interference from frequencies quite distant from the applied stimulus, but harmonically related to the critical bandwidth surrounding the stimulus frequency.

- - - -

The neural circuitry of the auditory system does not support signaling channels with bandwidths wider than a nominal 600 Hz. The primary bandwidth limiting elements of the system are the sensory neurons acting as integrators. Except at frequencies below or near 600 Hz, all signal manipulation within the auditory system is based on spatial place along the cochlear partition rather than actual temporal frequency of the stimulus.

The measured noise performance of the human auditory system under high S/N conditions has repeatedly shown a noise bandwidth that is typically one-third of an octave wide and centered on the dominant tone in the stimulus package (**Section 8.xxx** & **9.5.4 xxx**). These findings suggest the noise performance of the system is determined by the number of amplifiers involved in supporting the spatial integration range of a correlator at a given time. This number may be dynamically changeable in order to optimize the information extraction capability of the system.

The above description provides strong guidance that the tonal performance of the human auditory system does not involve frequency selective filters in the PNS (potentially in the cochlear nucleus) and does not involve frequency selective filters in the CNS. The guidance is that the tonal signal manipulation subsystem is spatially oriented, not frequency oriented.

Beginning in the 1930's with Fletcher and culminating in the 1980's with Zwicker, a large volume of work was reported attempting to describe the architecture and performance of the auditory system in the presence of noise (**Section 8.3**). This work did not result in a definitive description of any circuitry within the auditory system. Some investigators proposed frequency-selective circuitry within the PNS (apparently within the cochlear nucleus) while others proposed frequency selective circuitry within the CNS.

Because of its prominence in the technical literature, it is difficult to ignore the possibility of a multiple frequency band filter configuration associated with tonal channel hearing. Major efforts have been made to describe a coarse tonal analysis function following the discovery of the enigmatic critical band character of the hearing modality. Zwicker made monumental efforts to understand this characteristic from a frequency filtering perspective. However, the relationships he defined and his results have not been useful. While he attempted to describe multiple adjacent critical bands with definable edges located somewhere in the neural portion of the hearing modality, no such band edges have ever been

located with precision. Instead, the instantaneous bandwidth of a given band appears to be describable only as a percentage width of the center frequency chosen. This feature suggests the critical band phenomenon is not related to the frequency domain but to the spatial or place domain.

Moore and his associates have continued research into this psychophysical phenomenon using the label "equivalent-rectangular-band" (ERB)[30,31]. ERB's remain much wider than the minimum resolvable frequency of human, and animal, hearing.

Colomes, et. al. have attempted to use the Zwicker concept in the development of a signal compression codec for music[32].

The critical band concepts of Zwicker, Zwicker & Fastl[33], Moore, and Colomes et al., relying upon a series of fixed bandwidth discrete filters within the hearing modality, are not supported in this work.

The fact that the Stage 4 Signal Manipulation engines operate primarily on abstract analog signals devoid of sinusoidal content is of key importance in hearing. It means that conceptual models and analyses in the literature relying upon sinusoidal signals and Fourier Transforms are falsified. They should be dismissed when planning an applied research program. The method of information correlation within Stage 4 relies upon low bandwidth analog signatures associated with specific addresses, not tones.

9.2.4 Top level circuit diagrams of the source location system- - Path D EMPTY

None of the material reviewed in preparing this work provided a discussion of both impulse-related and tone-based source location. These two mechanisms are clearly distinct and serve different purposes in hearing. The impulse based source location mechanism provides a very fast estimate of the source of any stimulus (particularly high intensity sources) and is designed to support the alarm mode of hearing and the safety of the animal. Impulse based delays to the point of interpretation within the lower brainstem are on the order of xxx, including any delay within the acoustic channels of the outer, middle and inner ears. The tone based mechanism suffers an inherent delay due to the relatively slow transport velocity of sound along asignificant portion of the cochlear partition. It must then collect enough information to perform a differential time calculation between the tonal signal received by the two ears. This calculation requires the presence of a tonal component in the signal received from the cochlea and is therefore limited to frequencies below or near the nominal 600 Hz integration frequency (certainly below 1.5 kHz). While differential time measurements with a precision of 50 microseconds can be achieved in the tonal mechanism, this precision is only obtained after a significant time delay. This delay consists of a travel delay of about xxx ms plus a delay during the rise time of the first neural response at the site of time difference calculation. At 200 Hz, the rise time to the first peak in a sinusoidal stimulus is 1.25 ms. If three peaks are needed to confirm the calculations, a total delay of at least xxx ms is involved in the perception of a tone-based source location calculation at 200 Hz. A

---

[30]Glasberg, B. & Moore, B. (1990) Derivation of auditory filter shapes from notched-noise data *Hear Res* vol 47, pp 103-138

[31]Peters, R. Moore, B. & Baer, T. (1998) Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people *J Acoust Soc Am* vol 103(1), pp 577-587

[32]Colomes, C. Lever, M. et. al. (1995) A perceptual model applied to audio bit-rate reduction *J Audio Eng Soc* vol. 43(4), pp 233-239

[33]Zwicker, E. & Fastl, H. (1999) Psychoacoustics, Facts and Models, 2nd Ed. NY: Springer pp 158-164

physical response involves more delay.

Source location is discussed in summary in **Section 9.9 xxx** and in detail in **Section 8.xxx** or **6.xxx**.

9.2.5 Top level circuit diagrams of Paths C, E, etc EMPTY

Little can be said of the requirements on or the results of operation of the unique portions of paths C, E, etc. at this time. However, the arrival of fMRI techniques suggests that it will be soon possible to isolate the purpose of specific areas of the cortex and possibly develop specific complex stimuli that will cause responses in only small areas of the temporal lobe. This would allow much more tailored electrophysiological investigations that could define both the requirement and the performance associated with specific areas of these pathways.

9.2.5.1 The top block diagram of leading- edge-based source location EMPTY

9.2.5.2 The top block diagram of phase-based source location EMPTY

9.3 Describing individual stage performance in hearing (electrophysical data)

9.3.1 Stage 0 performance: The acoustical and mechanical elements of hearing

Stage 0, the non-neural portion, of the hearing modality is more complex than that of other modalities. As expected, this complexity introduces greater variation into the performance of the modality. **Figure 9.3.1-1** shows a circuit schematic of stage 0. The figure is expanded from a similar figure in Rosowski & Merchant[34] built on an earlier version by Kringlebotn[35]. These earlier versions provide notional descriptions of the content of the impedances shown within the boxes of this figure. However, these simple circuits are not complex enough to represent the actual performance of these elements.

> In Kringlebotn (1988), he did not address the role of the Eustachian tube. He determined the shunt legs of his circuit are negligible when evaluated against the available data. After citing Zwislocki (1975), he appears to support the idea that the input admittance of the cochlea is resistive (L_C & C_C are negligibel) , thereby simplifying his model in this area. If the cochlea employs a slow SAW, there would be no internal reflection within the time scale of admittance measurements in a healthy cochlea. In that case, it would always appear resistive in his model. Kringlebotn noted additional experimental data was needed in this area.

The figure provides many clarifications of how stage 0 operates. In this figure, the transformer symbol is used to represent either transitions from one type of environment to another (from the pneumatic to mechanical environments; (T-E)/U & S/V), to represent a mechanical lever (U/S) or an acoustic horn (F/T & F/MOU). The variable resistor symbol is used to represent muscular activity having an effect on stage 0 operation. Mus. ten. refers to the muscular tension of the muscles affecting the pressure in the eustachian tubes and tympanic cavity. Acu. Refl. refers to the acoustic reflex associated with the muscles supporting the ossicular bones. The capital symbols, P & U refer to pressures and volumetric velocities. The capital letters F & V refer to forces and translational velocities. Capital letters W & A refers to *transverse* velocities and amplitudes of the SAW wave at the tectorial membrane/endolymph interface. The subscripts F, T, MOU, CAV,U, S, V, C, TEC & RW refer to the free field, outer tympanic surface, aural cavity, tympanic cavity, umbo, stapes, vestibule, cochlear fluid, tectorial membrane & round window.

The outer ear is shown in an expanded form. An upper path supporting the application of acoustic energy to the outer surface of the tympanic membrane. The lower path indicates the application of acoustic energy to the inner surface of the tympanic membrane via the eustachian tube. The transformer (T–E)/U delivers energy to the umbo of the malleas bone that represents the *difference* between the contribution from the ear canal and the eustachian tube channel.

The inner ear is also shown in an expanded form. The upper path supporting the propagation of acoustic energy as a slow SAW along the tectorial membrane. The energy is selectively extracted from the SAW by the stage 1 neural circuits of the OHC and IHC (only the OHC paths shown). The lower path shows the flow of some acoustic energy traveling as a fast bulk acoustic wave (BAW) through the perilymph to the round window where it passes back into the tympanic cavity.

The upper path is also shown as the dispersion portion of a transversal filter. Multiple signals are

---

[34]Rosowski, J. & Merchant, S. (1995) Mechanical and acoustic analysis of middle ear reconstruction *Am J Otol* vol 16(4), pp 486-498

[35]Kringlebotn, xxx (1988) xxx *Scan Audiol* vol 17, pp 75-85

extracted from the tectorial membrane at different locations (characterized by different time delays and frequencies, beginning with the highest frequency, F_H) and passed up the signal chains to the stage 4 signal manipulation circuits of the CNS. It is there they are reassembled in time coincidence using computational anatomy techniques acting as the assembly portion of a transversal filter..

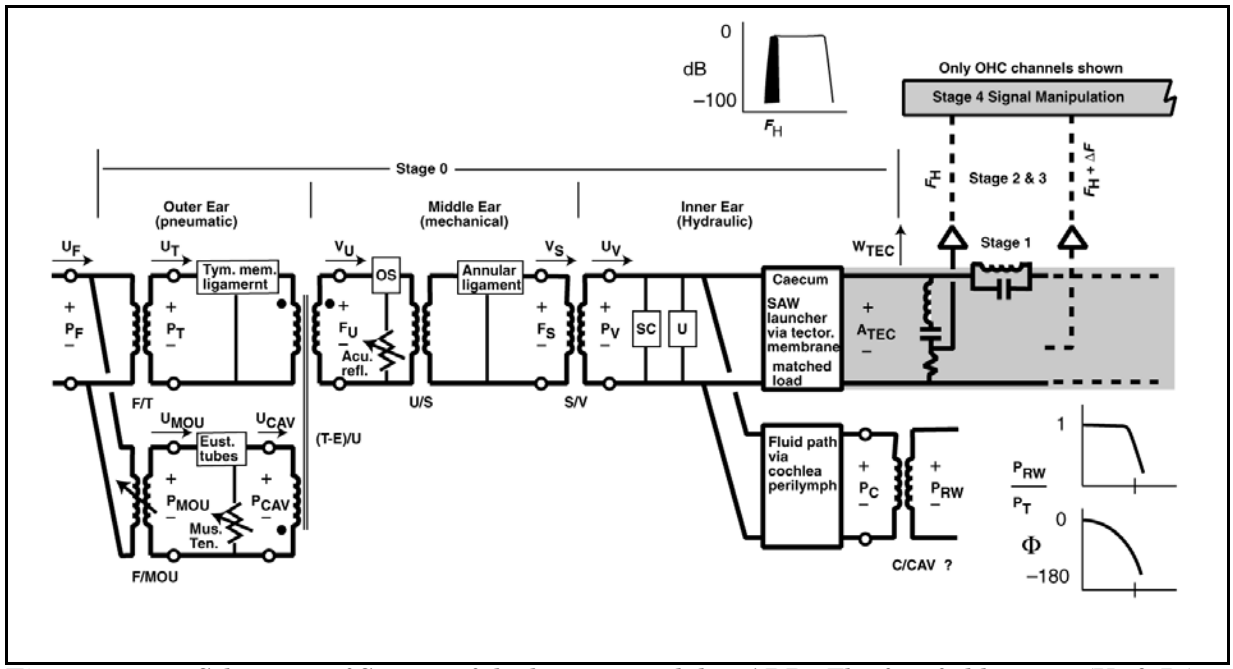


Figure 9.3.1-1 Schematic of Stage 0 of the hearing modality ADD. The free field energy (U_F & P_F) at upper left is applied to the tympanic membrane via two distinct paths (through the ear canal & through the eustachian tube) that are out of phase. Shaded areas represent input and output structures of a transversal filter configuration. The lower shaded structure represents the slow SAW traveling along the liquid crystalline face of the tectorial membrane. The upper shaded structure is entirely within the neural system. The lower right path and graph relate to the energy propagated through the perilymph as a compression wave. OS; ossicular bones. SC; spiral canals. U; utricle. See Text.

Many different signal paths through stage 0 have been studied, the easier ones to access being studied most. These studies have used different definitions of the outer, middle and inner ears. Many studies related to the middle ear during the 1960's and 1970's actually studied the signals at the round window, P_{RW} or F_{RW}, relative to the signals present at either the outer surface of the tympanic membrane, P_T, or at the stapes, F_S (frequently generated by energy introduced into the ear canal only.. As shown by the graphs at lower right, these measurements led to unexpected responses. Of particular interest is the small phase delay associated with these responses. They indicate the energy following this path left the cochlea before the SAW energy associated with even the earliest frequency, F_H, had reached the location of the cochlea where they were extracted from the main channel. This time difference indicates the energy introduced into the cochlea as a compression wave in the perilymph was not instrumental in the operation of the cochlea as an energy transducer. The limited bandwidth of the signal at the round window is caused by the long and small diameter fluid pathway from the vestibule to the main perilymph chambers.

A more appropriate study of the middle ear would compare the energy delivered to the vestibule, at the wetted surface of the oval window, P_V & U_V, to that supplied at the umbo, F_U & V_U. However, acoustic energy is frequently introduced directly into the ear canal, bypassing the contribution via the eustachian tube. This protocol does not provide a realistic representation of the middle ear transfer

function. Without the eustachian tube input, the middle ear acts as a low pass filter. With the eustachian tube contribution, the middle ear acts as a bandpass filter with a 6 dB/oct roll off at low frequencies. The results of these studies are reported below.

Using the data from Guinan & Peake shown below, it is noteworthy that the signal at the round window for 1000 Hz arrives after a phase shift of 100 degrees or about 0.3 msec. whereas the 1000 Hz signal arrives at the stage 1 sensory neuron via the SAW after about 4 msec (figure 4.5.4-2). ***The acoustic energy via the SAW arrives at the sensory neurons long after the acoustic energy associated with the bulk perilymph has left the cochlea altogether.*** The 4 msec delay associated with the 1000 Hz frequency is equivalent to a phase delay of 1440 degrees!

Puria & Allen have provided a more detailed look at the operation of the tympanic membrane as a flexible surface exhibiting time delay[36].

9.3.1.1 The overall sine wave performance of the outer ear

The overall sine wave response of the outer ear has been measured repeatedly under all possible conditions. Obviously the data associated with a group reflects the average of a great variation in external ear structures. To quantify these variations, a large variety of measurements have been made using standard "heads" and various shaped external ears(pinna). **Figure 9.3.1-2** shows the data of Wiener & Ross[37] They describe the stimulus as a progressive sound field. This would probably be described as a free-field today. [xxxon real ears.]

[xxx what happened to citation to Shaw/]
[xxx discuss head on separately from side angles.]
[xxx then discuss difference between left and right for a source on the left ala xxx who showed this data]

The variations in performance between individuals at a given frequency is significant due to variations in the dimensions of the ear canal. This variation appears due to differences in ethnicity, head size (and indirectly gender due to head size) of the test subjects.

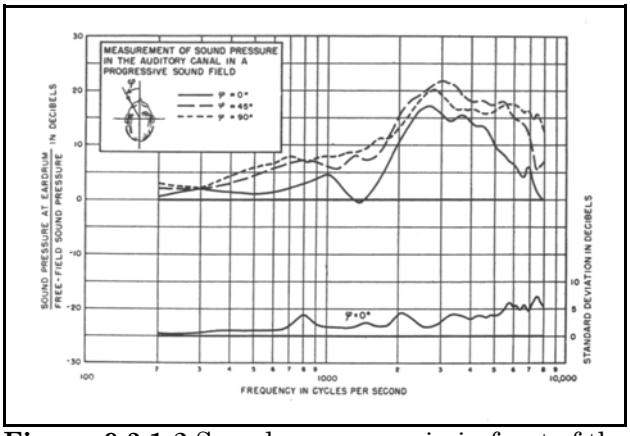


Figure 9.3.1-2 Sound-pressure gain in front of the tympanic membrane in a progressive sound field for three azimuthal angles as shown. From Wiener & Ross, 1988.

9.3.1.2 The overall sine wave performance of the middle ear

Limited data is available on the performance of the middle ear in humans. Some work has been done on cadavers, but it has been questioned consistently as to its relevance.

Dallos and others have suggested the middle

[36]Puria, S. & Allen, J. (1998) Measurements and model of the cat middle ear: Evidence of tympanic membrane acoustic delay *J Acoust Soc Am* vol 104(6) pp 3463–3481
[37]Wiener, F. & Ross, D. (1946) The pressure distribution in the auditory canal in a progressive sound field *J Acoust Soc Am* vol. 18, pp 401-408

ear operated as a low pass filter. **Figure 9.3.1-3** shows the data of Guinan & Peake dated 1967. This response obviously represents a low pass filter with additional peaking and notching. (presumably obtained by introducing acoustic energy directly into the ear canal of the cat). The slope of the high frequency area beyond the notch suggests the low pass filter contains two poles, one near 3000 Hz and one near 6000 Hz.

Guinan & Peake also provided averaged data for 25 cats that erased both the peaking and the notch from the composite curves.

Puria et al. provided more detailed information on the performance of the human middle ear in a useful form[38]. However, the ear tissue was removed at autopsy and subsequently placed in dry ice for storage. They did compare their results from frozen and non-frozen samples to validate their results. The stimulation protocol was limited to the ear canal. However, their figure 4(A) suggests the source was removed far enough in case 3 to allow participation of the eustachian tube in the stimulation process. The lower frequencies were attenuated as much as 40 dB in this case.

When free field acoustic energy us used to stimulate the middle ear via both the ear canal and the eustachian tube, the data available suggests the middle ear operates as a bandpass filter due to the presence of the eustachian tube. This tube operates as an effective pressure bypass around the tympanic membrane at low frequencies.

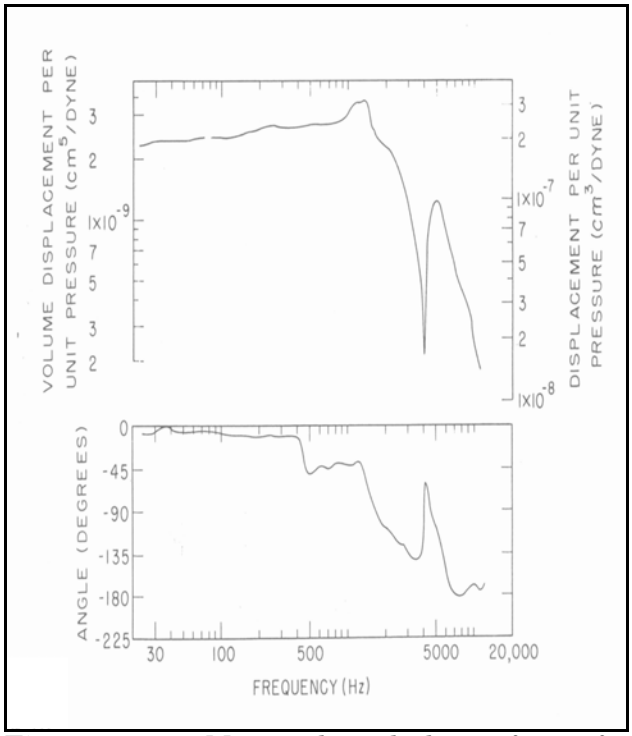


Figure 9.3.1-3 Magnitude and phase of transfer function without eustachian tube participation in one cat. Intact ear, bulla closed, and septum in place. Performance is more complex than a simple low pass filter. From Guinan & Peake, 1967.

Relkin describes the middle ear function as peaking at 40% efficiency near 1.0 kHz. He gives the efficiency as approximately 10% at 100 Hz and 0.6% at 4.0 kHz[39].

[38]Puria, S. Peake, W. & Rosowski, J. (1997) Sound-pressure measurements in the cochlear vestibule of human cadavers. *J Acoust Soc Am* vol 101, pp 2745-2770

[39]Relkin, E. (1988) Introduction to the analysis of the middle-ear function *In* Jahn, A. & Santos-Sacchi, J. *ed.* Physiology of the Ear NY: Raven Press pg 121.

More recently, Rosowski & Relkin provided data on a middle ear under free field stimulation. **Figure 9.3.1-4** shows their data describing the transfer impedance between the pressure at the wetted surface of the oval window at the vestibule and the surface of the tympanic membrane facing the ear canal in human cadavers when the tympanic pathway was also stimulated.. Their data shows a peak efficiency of about 20% at 800-2000 Hz compared to an ideal hydraulic ram of the same dimensions. The line labeled anatomical transformer ratio represents their calculated middle ear performance calculated from the dimensions of the middle ear operating as a simple hydraulic ram. It should be noted that this ratio and the "efficiency" quoted above from Relkin may not be the same as the efficiency calculated from the output energy divided by the input energy of the middle ear.

A dashed line has been added for discussion purposes. Conventional wisdom is that the middle ear is a low pass filter. However, the presence of the Eustachian tubes provides a pressure relief mechanism around the tympanic membrane. As a result, the middle ear should show a 6 dB/oct roll off at low frequencies when the Eustachian tubes are performing nominally. This is the nominal roll off (dashed red line) in the data from Rosowski & Relkin. The 3 dB point for this roll-off is 400 Hz. The deviation of the data from the dashed line at low frequencies suggests the Eustachian tube alone is also represented by a bandpass filter (with a passband in the 100 to 200 Hz region.

Interestingly, the high frequency roll-off is not that of a simple low pass filter. As seen more obviously in the phase plot, the high frequency roll-off involves two break points, the first at about 4000 Hz and the second at about 7000 Hz.

[xxx confirm this]
It is clear that the condition of the acoustic channel provided by the Eustachian tubes and the oral cavity are important in determining the low frequency performance of the middle ear.

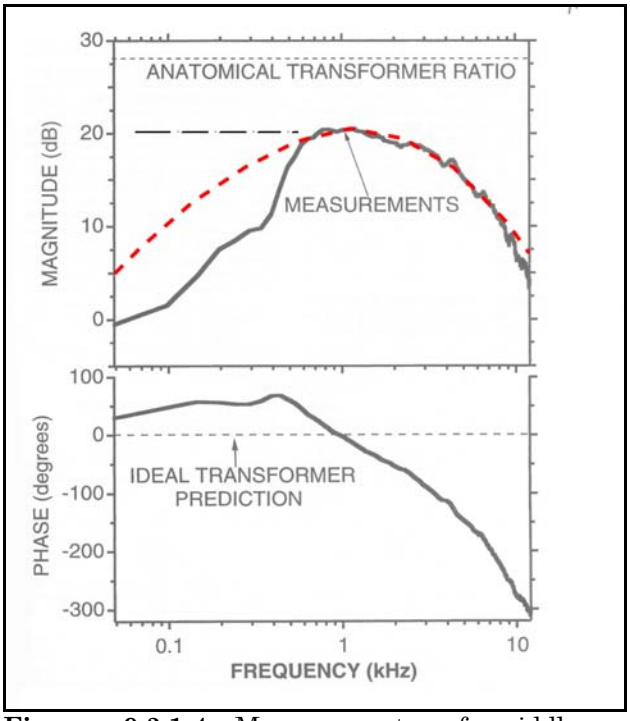


Figure 9.3.1-4 Measurements of middle-ear pressure gain ADD in human cadaver ears. The dashed lines represent the mechanical advantage of an ideal hydraulic transformer. Dashed line at 21 dB added for discussion. Dashed red curve is three pole bandpass filter overlay Modified from Rosowski & Relkin, 2001.

Aibara et al[40]. have provided similar and contemporaneous bandpass measurements to those of Rosowski & Relkin. They also provided data on the input impedance of the inner ear measured at the oval window. It was 21 gigaohms (21 x 10^9 ohms) with a phase angle of zero between 0.1 and 5.0 Hertz (based on a sample of 12 ears). If correct, this resistive impedance is indicative of no reflection from the cochlea, as expected for a slow SAW being dispersed by the Marcatili Effect. However, they gave

[40]Aibara, R. Welsh, J. Puria, S. & Goode, R. (2001) Human middle-ear sound transfer function and cochlear input impedance *Hear Res* vol 152 pp 100–109

no explicit definition of their Ohm.

The Ohm of Aibara et al. appears to have units of pressure/volume velocity or pressure/(area x velocity) or dynes•sec/cm^5 when expressed in CGS units. This is the "acoustic impedance" or acoustic Ohm defined in **Section 1.2.2.2**. Aibara et al. appear to be calculating the *input impedance of the vestibule at the oval window* by calculating the ratio of the vestibular pressure/tympanic pressure (a numeric value) and dividing it by the volume velocity of the oval window divided by the tympanic pressure. This calculation is equivalent to the ratio of the vestibule pressure divided by the volume velocity of the oval window, $P_V/(V_S \bullet A_{OVAL})$. **Figure 9.3.1-5** shows their results.

The very small change in phase angle over the audio spectrum is quite interesting; as noted above, it suggests no reflection from the cochlea of healthy ears.

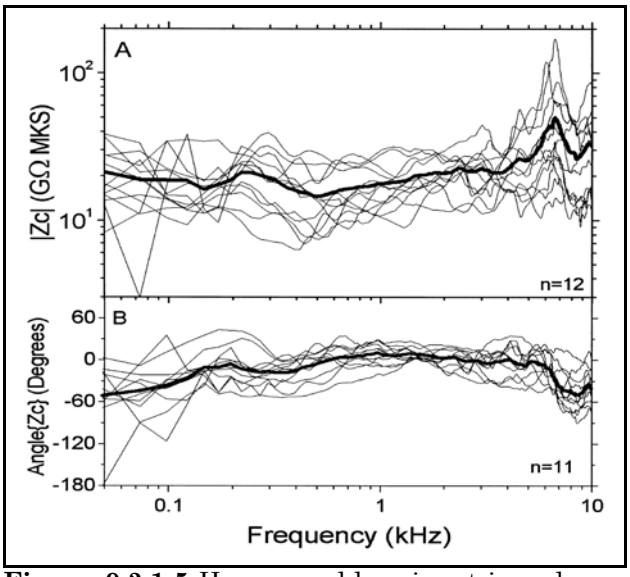


Figure 9.3.1-5 Human cochlear input impedance, Zc, in 12 human temporal bone ears. (A) shows the magnitude and (B) shows the phase angle. The curve for each ear represents the mean of three consecutive measurements of vestibule pressure and stapes velocity made within 1 h after hydrophone insertion. The mean magnitude and phase (thick line) of all ears is also shown. The phase from one ear was excluded because of a phase measurement error. From Aibara et al., 2001.

9.3.1.2.1 The acoustic reflex of the middle ear

Rabinowitz has provided good data on the acoustic reflex of the human ear[41]. The data is summarized by Rosowski & Relkin in **Figure 9.3.1-6**.

9.3.1.2.2 Effect of tympanic cavity pressure on hearing

Kringlebotn has provided data on the effect of positive and negative pressure in the tympanic cavity on the performance of the tympanic membrane in cattle[42]. As one might expect, the performance degraded in both instances, particularly at frequencies below 1000 Hz..

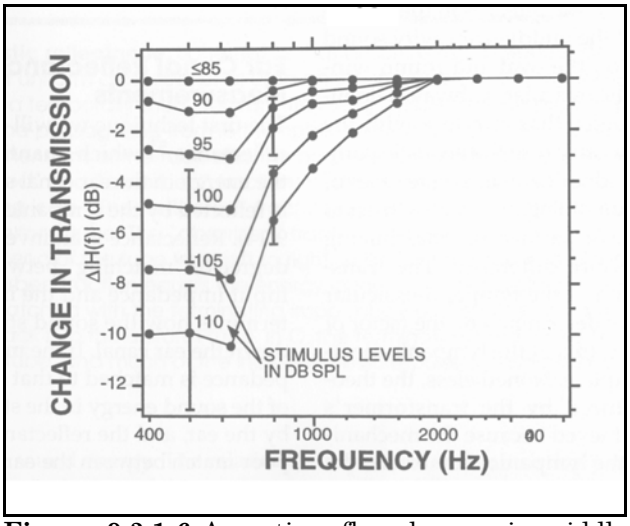


Figure 9.3.1-6 Acoustic reflex changes in middle ear transfer function if humans. Produced by bursts of band-pass noise (2–4 kHz of varied stimulus level in the contraleateral ear. Note the size of the error bars. From Rabinowitz, 1977.

[41]Rabinowitz, W. (1977) acoustic–reflex effects on the input admittance and transfer characteristic of the human middle ear *Dissertation* MA: MIT

[42]Kringlebotn, M. (2000) Frequency characteristics of sound transmission in middle ears from Norwegian cattle, . . *J Acoust Soc Am* vol 107(3), pp 1442-1450

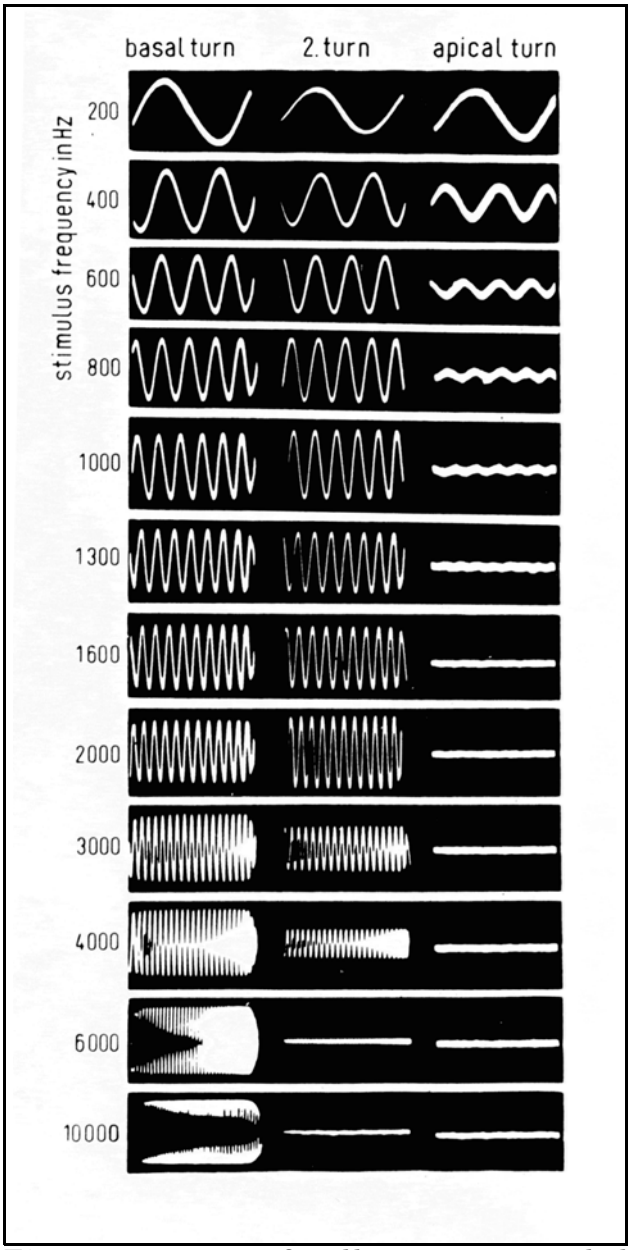


Figure 9.3.1-7 Oscillograms recorded simultaneously from the 1st, 2nd and 4th turns of the cochlea in the guinea-pig using differential recording techniques. The amplification was adjusted such that at the lowest stimulus frequency the amplitudes of the cochlear microphonics were equal in magnitude for the three locations of recording. The stimulus intensity was adjusted such that at all frequencies the amplitude of the cochlear microphonic was of equal magnitude in the basal turn. From Tasaki et al., 1952.

9.3.1.3 The overall sine wave performance of mechanical portion of the inner ear

Tasaki et al. provided the sine wave performance of the inner ear in 1952[43]. **Figure 9.3.1-7** shows their oscillograph data based on the simultaneous recording of responses at three points along the cochlea of the guinea-pig using a differential measurement between the scala vestibuli and the scala tympani at each location. They describe this differential measurement as a "cochlear microphonic," in contrast to the conventional single ended measurement made at the round window.

Tasaki et al. did not provide any physical description of the cochlear partition or circuit diagram including the sensory neurons as part of their investigation. In a previous paper, they did provide a gross physical cross-section of the

[43]Tasaki, I. Davis, H. & Legouix, J. (1952) The space-time pattern of the cochlear microphonics (Guinea Pig), as recorded by differential electrodes. J Acoust Soc Am vol. 24(5), pp 502- 519

cochlear partition[44]. Based on **Section 5.2.3** of this work, these signals were generated extraneous to the fundamental operation of the sensory neurons. This fact they surmise on page 586. Davis et al. describe a "summation signal" or "summating potential" that it is now clear was a misnomer. The actual waveforms do not involve a summation but correspond closely to the pedestal described in detail in **Section 5.3.4**. The voltage signal in the scala vestibuli was derived from scala media voltage that was in turn derived from the current flowing in the collector circuit of the adaptation amplifier of the sensory neurons. The voltage signal in the scala tympani was derived from the current flowing in the common emitter circuit of the sensory neurons. The summation potential is the pedestal generated by the low pass filter associated with the single-ended adaptation amplifier within each sensory neuron. Both the sinusoidal component of the signal and the pedestal component are represented in these two extraneous signals.

Probes of 10 micron diameter were inserted into the fluids of these two scala but their precise distances from the sensory neurons were unknown.

The signals appear to be dominated by currents flowing in the inner hair cells. This is particularly true where signals at 750 Hz were picked up from all along the cochlear partition. However, the potential for a component from the outer hair cells should not be discounted.

One must be careful in interpreting the calculated velocities of propagation in their figure 17 based on the recorded electrical signals. They show no time lag for 750 Hz signals traveling along the first 8 mm of the cochlear partition and then major delays associated with the next 10 mm. Tasaki et al. did not account for the electrical delay between the piezoelectric signal created at the cilia of the hair cells and the point of signal recording. This relatively constant delay must be subtracted from the recorded delays. Neither did they account for any phase shift associated with the high frequency rolloff associated with their signals. The resulting velocities within the cochlea do not vary nearly as much as indicated in their figures 14 through 16.

The uniformity of the amplitude of the signal for frequencies below the best frequency suggest virtually no attenuation as a function of frequency along Hensen's stripe. This lack of attenuation implies little velocity dispersion with frequency below the dispersion frequency.

- - - - -

[xxx rewrite this paragraph to match the figure and find it a home]
Figure 9.3.1-8 provides a template describing the frequency response of the xxx ear. The literature generally assigns two specific frequencies to this bandpass characteristic, a low pass parameter near 500 Hz and a high frequency parameter near 5000 Hz. This section will attempt to justify more precise values.
[xxx be sure to address the phase-locking interpretation, see section opening the topology section]

Greenwood has provided data based on masking experiments that appears to bear on the high and low frequency parameters of human hearing[45].

---

[44]Davis, H. Fernandez, C. & McAuliffe, D. (1950) The excitatory process in the cochlea *PNAS* vol 36, pp 580-587
[45]Greenwood, D. (1961) Critical bandwidth and the frequency coordinates of the basilar membrane *J Acoust Soc Am* vol. 33(10), pp 1344-1356

9.3.1.4 The overall sine wave performance of mechanical portion the human auditory system

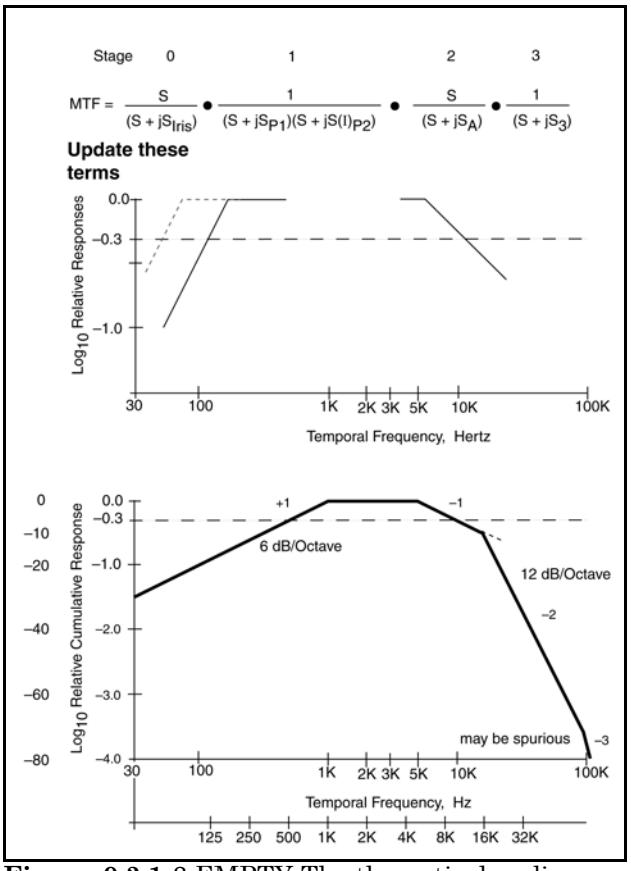


Figure 9.3.1-8 EMPTY The theoretical audiogram for the human based on several simplifying assumptions. The lower frame is similar to that of Masterton et al. in style but shows two high frequency roll-off segments.

[xxx split into outer, middle and inner ears]
The auditory system is frequently described globally using what is called simply an audiogram[46]. Masterton et al. show the basic form. Note the frequency scale is frequently plotted as logarithmic to the base 2 on audiograms. Such a plot does not introduce any distortion when re-scaled to a different logarithmic base. The Masterton et al. paper was a very early attempt to parameterize hearing among the phylogeny of mostly mammals. Unfortunately they had very little data to work with and it originated with a variety of other investigators. Their statistical weighting left much to be desired. As shown in Masterton, et. al., the area of best frequency frequently varies widely among members of the same species. This variation appears to be due to variations in the form and cleanliness of the outer ear. While citing one bottlenose dolphin as part of their statistics, they fail to define its lower frequency limit (it generally does not reach 1 kHz) and mis-represent the high frequency limit of dolphin hearing; and its "maximum binaural time disparity" by assuming it ears are on opposite sides of its head behind the eyes.

Geisler has provided a very descriptive figure showing the performance of the human auditory system in a variety of situations[47]. It also compares its high frequency response with that of the cat. However, it does not describe the automatic adaptation (loudness control mechanism of the system. [xxx probably include this figure]

---

[46]Masterton, B. Heffner, H. & Ravizza, R. (1968) The evolution of human hearing JASA vol 45(4), pp 966-985, fig 4
[47]Geisler, C. (1998) Op. Cit. pg 21

While many audiograms describing various conceptions of the fundamental performance of the auditory system have been drawn based on empirical evidence, this section will attempt to provide a traceable audiogram based on specific measurements related to the individual mechanisms involved. **Figure 9.3.1-9** shows a theoretical audiogram based on this approach.

It is difficult to correlate this figure with the work of other individual investigators because of the empirical nature of their work. They frequently describe their results semantically rather than graphically.

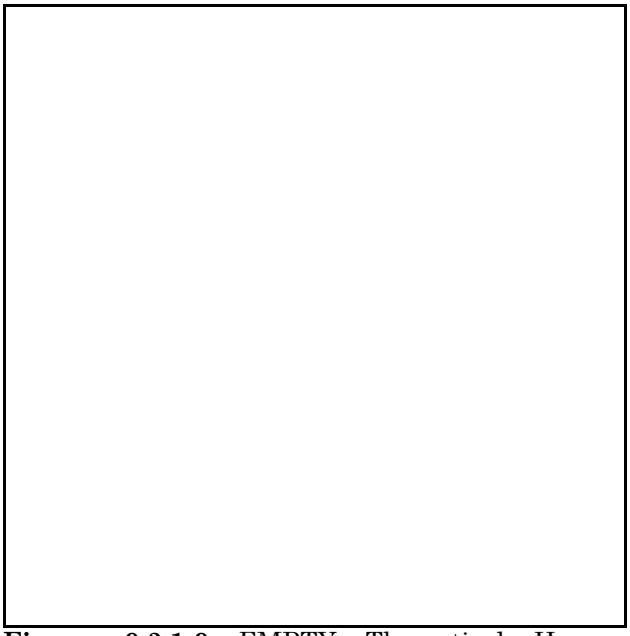


Figure 9.3.1-9 EMPTY Theoretical Human audiogram for a source in the mesotopic region compiled from the following individual sine wave responses.

Figure 9.3.1-10 displays the overall performance of the auditory system as adopted earlier by the ISO as ISO 226:1987 but now replaced by ISO226:2003 and presented below. The original curves were compiled by Fletcher & Munson from their extensive measurement program[48]. This family of curves includes the signal gain associated with the external ear as well as the physiological aspects of the auditory system. The data was collected by asking people to judge when pure tones of two different frequencies were the same loudness. As noted in **Section 9.3.2.4.1** and born out in laboratory practice, equal loudness measurements involve a very difficult judgement on the part of the subject. The curves are the average results from many subjects, so they should be considered general indicators rather than a prescription as to what a single individual might hear.

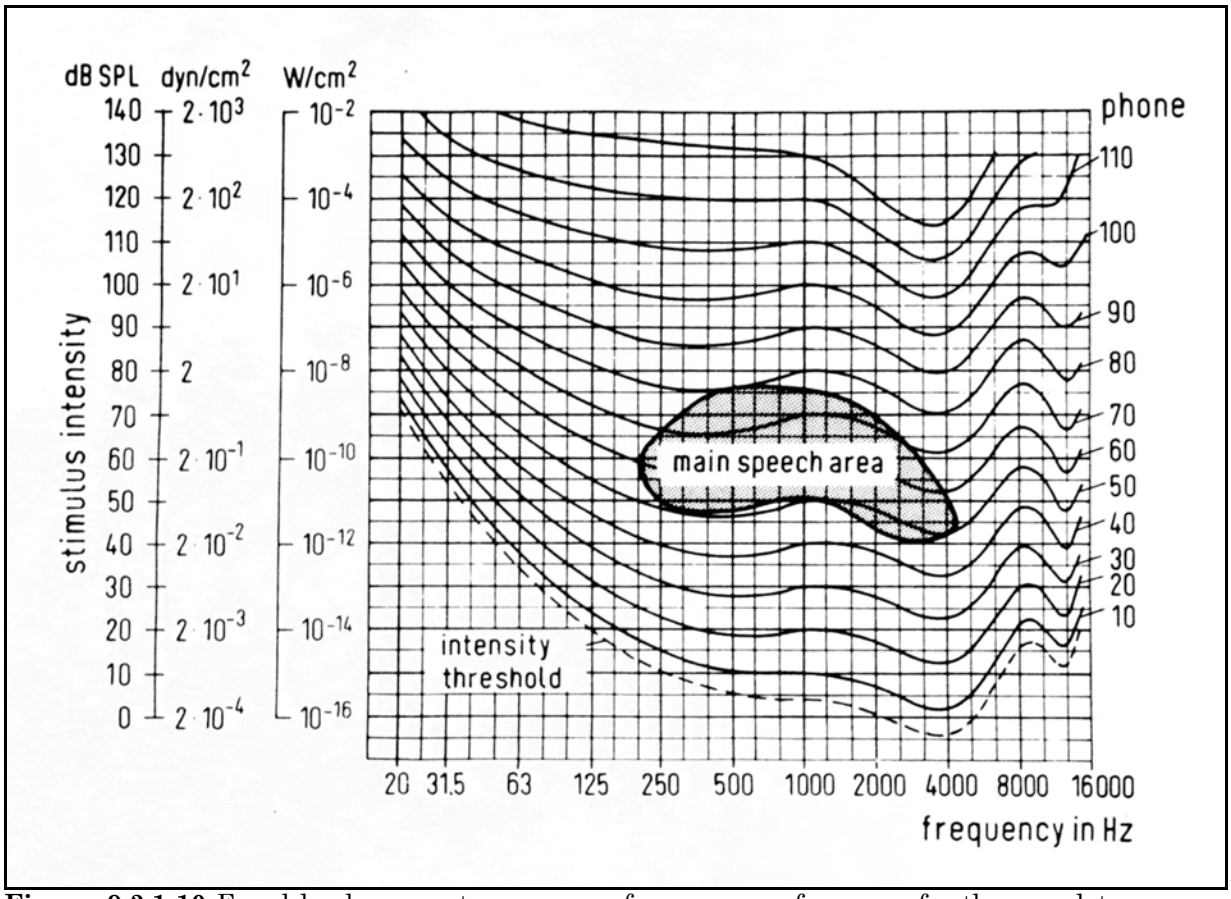


Figure 9.3.1-10 Equal loudness contours versus frequency performance for the complete average human ear. DIN Standard 45630, Sheet 2. It reflects the original data of Fletcher & Munson, 1933 as redetermined by Robinson & Dadson in 1956. The Fletcher & Munson curves stopped at about 12,000 Hertz and differed materially from these curves.

Below 63 Hertz, an attenuation of about 25 dB/oct is associated with the intensity threshold. Most models in the literature do not reflect this rate of roll-off and the measured roll-off is frequently much steeper than 25 dB/oct in laboratory animals (Geisler, 1996)

.

[xxx In the '30s, researchers Fletcher and Munson first accurately measured and published a set of

[48]Fletcher, H. & Munson, W. (1933) xxx *J Acoust Soc Am* vol. 6, pp 59+

curves showing the human's ear's sensitivity to *pure tone* loudness verses frequency ("Loudness, its Definition Measurement and Calculation," *J. Acoust. Soc. Am*., vol. 5, p 82, Oct. 1933). They conclusively demonstrated that human hearing is extremely dependent upon loudness. The curves show the ear most sensitive to pure tones in the 3 kHz to 4 kHz area. This means sounds above and below 3-4 kHz must be louder in order to be heard *just as loud*. For this reason, the Fletcher-Munson curves are referred to as "equal loudness contours." They represent a family of curves from "just heard," (0 dB SPL) all the way to "harmfully loud" (130 dB SPL), usually plotted in 10 dB loudness increments.

D. W. Robinson and R. S. Dadson revised the curves in their paper, "A Redetermination of the Equal-Loudness Relations for Pure Tones," *Brit. J. Appl. Phys.*, vol. 7, pp. 156-181, May 1956. *These curves supersede the original Fletcher-Munson curves for all modern work with pure tones.* Robinson & Dadson curves are the basis for ISO: "Normal Equal-Loudness Level Contours," *ISO 226:1987* -- the standard until recently replace by ISO226:2003.

The threshold curve of the ISO226:1987 standard does not exhibit a very high rate of attenuation as predicted by application of the Marcatili Effect.

Users of either of these curves must clearly understand that they are valid *only* for pure tones in a free field, as discussed in the following by Holman & Kampmann. This specifically means they do NOT apply to noise band analysis or diffused random noise for instance, i.e., they have little relevance to the real audio world. A good overview is T. Holman and F. Kampmann, "Loudness Compensation: Use and Abuse," *J. Audio Eng. Soc.*, vol. 26, no. 7/8, pp. 526-536, July/August 1978. xxx]

The performance of the outer ear alone, the pinna, has also been measured. **Figure 9.3.1-11** shows this data. The curves show the signal gain or loss given by the human external ear to a continuous tone with a sound pressure level (in decibels), as the source azimuth in the horizontal plane was varied in 15 degree increments, and as a function of frequency. Zero angle corresponds to a source directly in front of and on the midline of the head. Higher performance is actually achieved for one ear at an angle of 60 degrees from the midline. The negative values correspond to the source being on the other side of the midline with the ear being shadowed by the head.

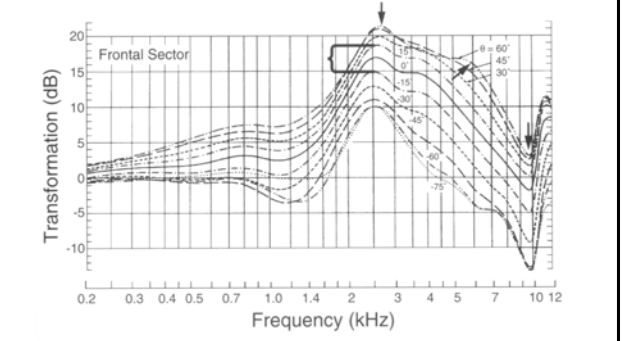


Figure 9.3.1-11 The acoustic performance of the standard external ear. Arrows and brace indicate important features. See text. From Shaw, 1974.

Geisler says in an example, "the common peak near 2.5 kHz is due to the "quarter-wave-length resonance" of the ear canal,(see Appendix B), which affects equally sound coming from all directions. . . ."(page 31) He did not explain why he put the expression in quotation marks. The situation is more complicated than that, and the Appendix contains an unfortunately placed typographical or computational error. While the negative response (anti-resonance) at 9.6 kHz can be related to a 1/4λ distance of nine mm in air (or more likely related to a 3/4λ distance of 27 mm), the variations at lower frequencies would involve longer distances. The destructive interference at 1.37kHz would involve a total path length difference (1/2λ) of about 124 mm. The numbers suggest the overall response is due more to variations in collection efficiency with frequency (due to both the shape and texture of the pinna) than with interference phenomena. As an example, the notch in the 5–7 kHz region (slanted arrow) is due to the interference originating in the central depression of the pinna, also known as the concha.

The use of the term resonance with the term quarter-wave-length in quotations suggests an underlying problem. The condition is actually based on the phenomenon of interference, not resonance. As discussed briefly in his short appendix, the human ear canal is about 30 mm long and the impedance of the elements associated with the inner ear is different from the impedance of air, both in the ear canal and in open space. The result according to propagation theory is a reflection of some of the energy at each, even gradual, impedance change. These reflections cause a reinforcement or degradation in the total sound pressure at a given location in space along the path. A positive reinforcement is called constructive interference (or resonance in the vernacular). A negative reinforcement is called destructive interference (or anti-resonance in the vernacular).

The above discussion surfaces another fact. The effects of interference observed when measuring the sound response of the outer ear using continuous sine wave tones would not be observed if the response was obtained by taking the Fourier Transform of the impulse response of the same ear. In the time domain, the signature of the two impedance mismatches associated with the ear canal would be different. It would be a reverberation in the recorded signal (essentially a delayed echo at specific frequency ranges signaling the impedance mismatches in the system. Interferences would occur at delays near odd multiples of 0.2 milliseconds for a simple ear canal of about 30 mm length. Negative interferences (enhancement) would occur at delays near

multiples of 0.4 milliseconds.

By removing the impact of the pinna, the net physiological affect of the inner ear and the neural portions of the auditory system can be described. **Figure 9.3.1-12** shows this reconstruction. Three major features can be readily seen. First, the perceived loudness curves vary in a very consistent pattern over significant changes in input sound pressure. These curves are measured using a differential technique where the sound level is established for a period of time and then the minimum threshold is measured as a function of frequency. A major portion of this variation in sensitivity (about 70 dB in audio pressure), over the middle range of frequencies is due to the adaptation function built into the ARC.

As in the visual system, performance as a function of stimulus intensity can be described using several distinctive regions, although they have not been explored as fully in the audio system. These are a normal middle range (corresponding to the photopic range in vision) where the perceived sound level is largely independent of the actual stimulus level. This is the range in which the adaptation process is active. Within this range, differential signal levels about a mean level (regardless of frequency) are reproduced faithfully and the subject does not report any change in the timbre (distribution of frequencies within the perceived stimulus) of the stimulus with intensity changes. At sound levels above this range, the circuitry within the auditory system begins to saturate. Performance deteriorates with stimulus level, and the timbre associated with the source is quickly lost. The stimulus just sounds loud, whether it was considered musical at a lower level or not. At the other extreme, the adaptation amplifier operates at maximum gain. In this case, a decrease in stimulus level is perceived as a direct decrease in sound. This mesotopic region, is critically important in the design of a concert hall. A subject at the rear of the hall will perceive variations in long term stimulus level unrecognized by a subject in the middle of the hall. At still lower levels, the subject will take action to increase the stimulus level arriving at his ears in order to achieve an adequate signal-to-noise level for understanding (largely without concern for tonal quality).

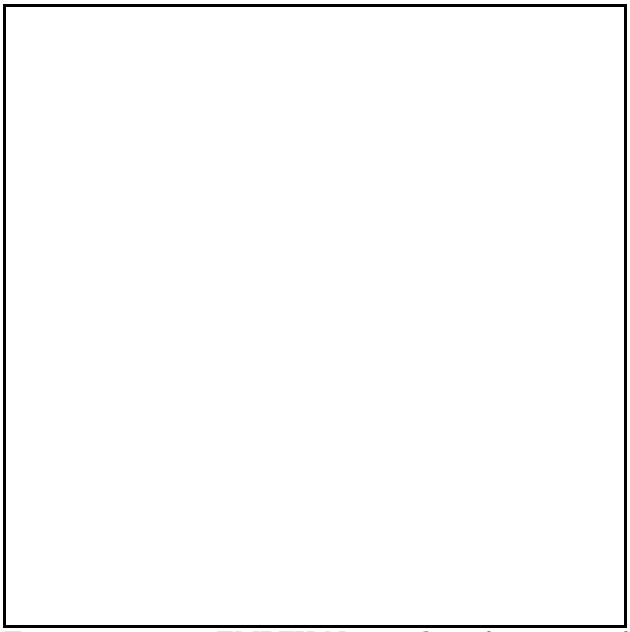


Figure 9.3.1-12 EMPTY Nominal performance of the inner (?) ear and physiological portion of the human auditory system (excluding the outer ear and ear canal).

Second, the rolloff of the low frequency response is at 18 dB per octave. This corresponds to three simple low pass filters operating in series. Since the initial rolloff in the region of 150 to 500 Hz is prominent, it suggests that one of the filters has a half-power point near 250 Hz. The other two appear to have similar half-power points near 100 Hz. One of these filters is almost surely that associated with the adaptation process itself. The other two are

probably associated with the operation of the ossicles of the inner ear and the surface wave launcher associated with the stapes and round window. Third, the high frequency performance of the human ear varies dramatically from individual to individual at frequencies above 10 kHz. Much of this variation relates to the intricate shape of the pinna. Some of it is due to age. These variations account for the ragged readings at the highest frequencies reported by Fletcher & Munson and by Shaw. The characteristics in this area are probably not statistically relevant.

9.3.1.5 The sine wave response of the cochlea

Developing a definitive characteristic for the sine wave response of the cochlea is currently difficult because of a lack of empirical data on the low frequency performance of the cochlea. The high frequency performance is well characterized by the operation of the Marcatili Effect.

The difficulty in deriving the theoretical response of the cochlea at .low frequencies is primarily one of geometry. The modified Hankel function (see Section 4.3.3.2) describing the curvature of Hensen's stripe at low frequencies currently contains an arbitrary term $(1– (u/22)^4)$ in order to fit the observed data from only a few researchers. It is the precise shape of the curl of the function near the apex of the cochlea that determines the precise dispersion of low frequency energy. At the same time, the spatial arrangement of the inner and outer hairs nearest the apex of the cochlea describes the ability of the neural system to capture the dispersed energy. Data related to these two mechanisms has not been found in the experimental literature. It can be assumed the sensory neurons are arranged as in Figure 5.2.4-1, but confirmation is necessary.

Figure 4.5.4-2 suggests the modified Hankel function leads to a significant reduction in the length of Hensen's stripe dedicated to a given octave of the auditory spectrum, about 3%/oct compared to 15%/oct in the middle frequency region.

9.3.1.6 The sine wave response of the composite stage 0 human auditory system

Figure 9.3.1-13 shows the contributions of the outer, middle and inner ears to the composite audiogram of human hearing. The middle ear is shown as a bandpass filter due to the pressure relief provided at low frequencies by the Eustachian tube.

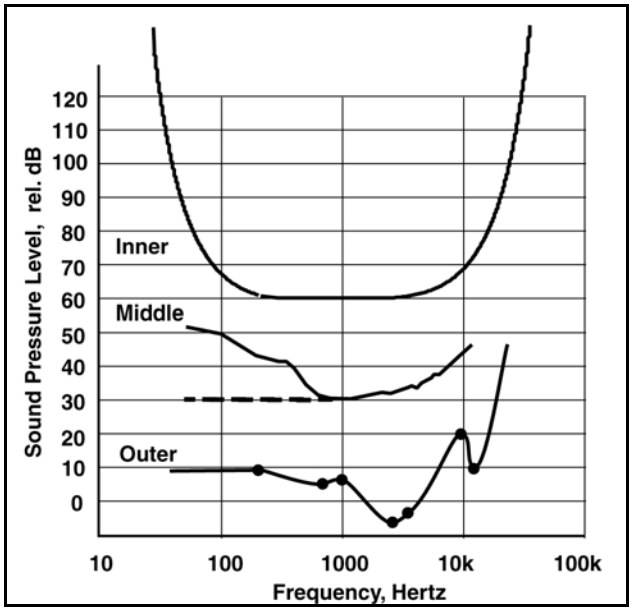


Figure 9.3.1-13 Contributions of outer, middle and inner ears to the audiogram of humans for free field acoustic stimulation from the front. Middle ear shown with Eustachian tube functional (solid line).

9.3.1.7 The most recent ISO 226:2003 equal-loudness-level contours

Figure 9.3.1-14 documents the most recent ISO 226:2003 Standard for human equal-loudness-levels[49]. Even this 2003 Standard has been in revision since 2005. The discussion accompanying this Standard includes the observed difference in human performance when experiments employ tight fitting headphones versus one audio projector mounted centrally in front of the subject and two or more speakers emulating a natural environment.

The red lines in the figure represent the new Standard. They appear to reflect the impact of the outer ear as described by Shaw but do not document the actual high frequency cutoff predicted by the Marcatili Effect within the cochlea. The high frequency limit appears to be dominated by the high frequency limit associated with the middle ear. Similarly, the low frequency roll-off does not appear to follow any recognized theoretical or intuitive function. The low frequency regions of the curves do show compression at low frequencies which can be expected as a result of adaptation at higher phon levels.

[49]http://www.aist.go.jp/aist_e/latest_research/2003/20031114/20031114.html

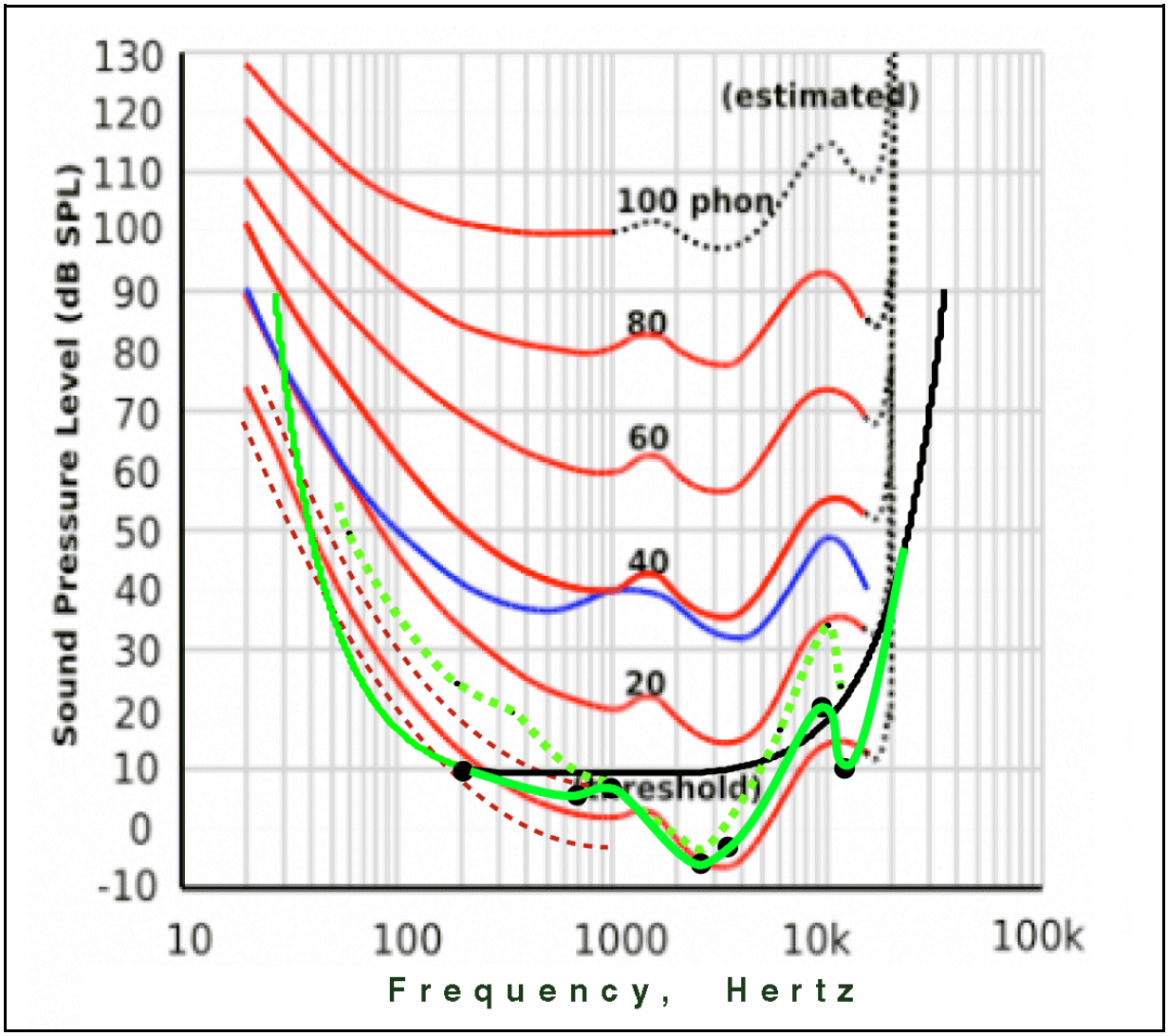


Figure 9.3.1-14 The latest ISO 226:2003 standard versus theory. Solid red lines; the Standard. Dashed red lines; one standard deviation from the mean of the Standard. Blue lines; the previous (1987) standard at 40 Phons. Black line; the transfer function of the cochlea (inner ear) based on the Marcatili Effect. Solid green line; transfer function of the combined inner ear (cochlea) and outer ear, based on data of Shaw, 1974. Dashed green line, transfer function of complete ear based on outer ear data of Shaw and middle ear data from Puria et al. See text.

The blue line suggests the change in the ISO standard between the 1987 and 2003 variants at the 40 Phon level. The dashed red lines have been added to indicate the one standard deviation about the mean of the threshold standard (from ISO 28961 draft supporting ISO 226:2003).

The black line represents the envelope of the cochlea transfer function of the MOC channels based on the Marcatili Effect. The low frequency roll-off of the cochlea exhibits the very rapid increase in attenuation with falling frequency associated with individual OHC in the laboratory. The high frequency roll-off of the cochlea also exhibits the very rapid increase in attenuation with increasing frequency associated with individual OHC in the laboratory.

With the introduction of the outer ear transfer function reported by Shaw (**Section 9.3.1.4**), the overall

predicted audio spectrum of the human ear is given by the green line. The green line assumes a flat transfer function for the middle ear out to at least 15,000 Hz (no impact of the eustachian tube at low audio frequencies).

If the in-vivo middle ear transfer function is that suggested by the findings of Puria et al., the dotted green line represents the best estimate of the theoretical threshold equal-loudness-level contour.

The resulting theoretical audio spectra are a very good match to the psychophysically–based spectrum of the ISO 226:2003 Standard. With a flat middle ear response below 15,000 Hz, the predicted curve (solid green line) is generally within one standard deviation of the Standard at short wavelengths and always within two standard deviations above 30 Hertz. With the inclusion of the middle ear transfer function reported by Puria et al., the predicted curve (dotted green line) is on the opposite side of the Standard by up to two standard deviations at low frequencies and shows greater deviation from the Standard at high frequencies than the curve without Puria's input.

> The ISO 226:2003 Standard was released by Technical Committee 43 and applies to binaural hearing from a free field source directly in front of the subject. Subjects were restricted to 18-25 year olds with "otologically normal hearing." Draft ISO 28961of 14 June 2010 provides the standard deviations at frequencies below 10,000 Hertz supporting the main standard. The values are all in the ±5 dB range. Standard deviations for frequencies above 10,000 were not available from that draft.

The nominal 5 dB standard deviation associated with the empirical Standard and the variation between the mean and the theoretical audiogram without a significant middle ear contribution (solid green line) appear to be dominated by two conditions. The variation near 8-10 kHz, is probably due to variations in the absolute dimensions of the outer ear canal. Part of this variation may be due to differences in ethnicity, gender and size of the test subjects. The variation below 200 Hz is probably due primarily to the size and condition of the eustachian tube of the subjects and whether they had their mouth open or closed (and their nasal passages functional).

> A blockage of the eustachian tubes generally results in the perception that low frequency hearing has been accentuated compared to the higher frequencies, in accordance with the green lines of the figure.

The solid and dashed green lines straddle the Standard at low frequencies, suggesting the precise condition of the eustachian tube and the state of sound conduction through the mouth and nasal passages play a significant role in determining the operating audiogram of an individual in this region. The state of sound conduction through the mouth and nasal passages should be defined as part of the next version of the ISO 226 Standard.

The differences in the ripples in the threshold curve at high frequencies between the Standard and both the solid and dashed green curves shows the need for more documentation concerning the dimensions of the outer ear canal, and the high frequency transfer function of the *in-vivo* middle ear.

9.3.2 Stage 1 circuit parameters common to all signal paths EMPTY

9.3.3 Stage 2 circuit parameters common to all signal paths EMPTY

9.3.4 Stage 3 circuit parameters common to all signal paths

[xxx summarize encoding from Chapter 7]
Stage 3 signal propagation plays a unique role in the neural system. Attempting to transmit an analog signal a distance of more than one millimeter is a challenge in the neural system. The capacitance associated with an extended plasma membrane is very large. As a result, the impedances involved force either the use of much larger Activa driving more powerful connexuses within individual neurons, or much larger diameter axons (even if they are myelinated). An attractive alternative is to convert the signal to a phasic signal that can be transmitted more efficiently. This is the role of the Stage 3 neurons, to encode the analog signals at the output of a signal processing circuit, transmit the information in a coded phasic form, and recover the analog information for further processing at a remote location.

The ability of the Stage 3 neurons to transmit signal information efficiently, has caused them to be used wherever it is necessary to transmit neural signals over one millimeter or more. Thus, Stage 3 neurons are found throughout both the peripheral nervous system of virtually all animals and the central nervous system of larger animals. They interconnect virtually every pair of engines in the neural system, whether they be located within Stages 2 through 6 of the system.

Stage 3 signal propagation in the auditory system is identical in signaling parameters to that used in the other sensory subsystems of the neural system. The highlights of the operation of the elements of the Stage 3 have been discussed in Chapter xxx and more extensive material is available in PBV, Chapter 13 [xxx check]. The methods of encoding the information generally takes two forms, a method of delta modulation to transmit monopolar information (usually associated with intensity related information) and a method of phase modulation to transmit bipolar information. Phase modulation is similar to frequency modulation but has special features that makes it the preferred choice in neural systems. Both of these methods employ the Activa in either a a driven or a free-running relaxation oscillator circuit. These circuits do not generate continuous sinewaves. They generate individual pulses intermittently, like the neurons driving the heart muscle, and are described as relaxation oscillators.

The phasic signals generated in Stage 3 circuits can still only be transmitted over distances of one or two millimeters. However, they enjoy the ability to be regenerated nearly perfectly by simple regenerative oscillators. These circuits are also relaxation oscillators. They are placed at intervals along the axon and are known morphologically as Nodes of Ranvier. While very energy efficient, Nodes of Ranvier as well as the other relaxation oscillators introduce significant time delays into the signaling channel. These time delays have made it difficult to speak of the propagation velocity of phasic neural signals in the literature.

All biological relaxation oscillators use the same circuit configuration within the connexus. The primary difference between them relates to the emitter-to-base bias voltage (approximately equivalent to the dendrite-to-podite voltage).

As shown in figure xxx, the velocity of phasic signal propagation along a myelinated axon (measured over a distance not including a Node of Ranvier) is quite high. However, the frequent introduction of Nodes of Ranvier cause the average velocity of the neural signal (measured over distances larger than two millimeters to be considerably lower. While the instantaneous velocity of neural signals is on the order of 4400 m/sec, the average velocity (or speed) is nearer to 44 m/sec or less. [xxx check values]

9.3.4.1.1The method of delta modulation used to encode monopolar analog signals EMPTY

[xxx driven oscillators]

9.3.4.1.2 The method of phase modulation used to encode bipolar analog signals EMPTY

[xxx free running oscillators]

9.3.4.1.3 The Nodes of Ranvier and signal regeneration EMPTY

[xxx a question of bias.

9.3.4.1.4 The circuit used to recover both delta and phase modulated signals EMPTY

9.4 Describing cumulative performance at intermediate nodes in hearing (electrophysical data)

9.4.1 Techniques and protocols in electrophysiology

9.4.1.1 Performance evaluation using non-invasive techniques

9.4.1.1.1 The Auditory Evoked Potential and Auditory Brainstem Responses

The location of the neural circuits of the auditory system make it all but impossible to record individual neural signals without major surgery on the subject. As a result, investigators have sought to remotely measure the signals associated with the individual physical sections of the auditory system. These have involved a variety of auditory evoked potentials (AEP) and other auditory brainstem measurement techniques. Rose, et. al. defined several of these techniques and described the utility of the results in 1971[50]. Naunton & Fernandez provided a broader review of the subject in 1978[51].

Jacobson has discussed the difference between the conventional electroencephalogram (EEG) and the auditory evoked potential[52]. "In stimulus-related response measured, the spontaneous EEG voltage far exceeds that of the EP." As a result, it is necessary to employ sophisticated signal recovery techniques to isolate the AEP responses. These generally use algebraic addition of multiple recording intervals synchronized to a repetitive stimulus waveform. Bandpass filtering is also used generally to limit the undesirable noise in the recordings. As he also noted, "In contrast to receptor potential generation, neurogenic potentials originate from the acoustic nerve and other neural sites with the auditiory CNS. They comprise action potentials . . . ." What he did not say was each feature in the waveform was due to the sum of many action potentials and all of these action potentials were positive

---

[50]Rose, d. Keating, L. et. al. (1971) Aspects of acoustically evoked responses *Arch Otolaryng* vol. 94, pp 347+
[51]Naunton, R. & Fernandez, C. eds. (1978) Evoked Electrical Activity in the Auditory Nervous System. NY: Academic Press
[52]Jacobson, J. ed. (1985) The Auditory Brainstem Response. San Diego, CA: College-Hill Press

going during their initial stage. These summations are sensitive to the amplitude of the stimulus and the level of adaptation of the subject (due to prior and possibly continuing background stimulation). In the case of binaural stimulation, the composite also reflects the source location determining properties of the system.

Because of their physiological importance, as well as their ease of measurement, signals have also been recorded from the "minibrain" associated with the brainstem. These auditory brainstem responses, ABR, are usually obtained non-invasively in humans. However, in other species, it is common to use subdermal needle electrodes. Brittan-Powell & Dooling have recently summarized the work in birds, much of which is based on a small altricial Australian parrot, the budgerigars or *Melopsittacus undulatus*[53]. It is more difficult to make precise measurements and analyses in smaller animals. The transit delays between individual waveforms along a neural path are much smaller causing the waveforms to overlay each other.

9.4.1.1.2 Interpreting AEP & ABR recordings

Many clinical investigators have attempted to define a set of features and standard measurements related to AEP and ABR recordings. Suggested standard waveforms have frequently been proposed. However, these attempts have all been empirical. They have suffered from a lack of knowledge concerning the source of the waveforms the investigators desire to characterize. The empirical results have also suffered frequently from the limitations (and the obsolescence) of the test instrumentation. Even a recent (2002) paper used a test configuration that is already obsolete from the manufacturers perspective.

As noted in the earlier chapters, the signals recorded as AEP or ABR waveforms are usually obtained by repeated integration of very noisy raw waveforms where the initial signal level is on the order of 1/20 the amplitude of the (frequently structured) noise. Extraction of this signal requires integration of many copies of the initial signal. If the initial signal is free of structured noise, and is synchronous with the sampling interval, a good representation of the sought waveform can be obtained. It is then necessary to interpret the waveform obtained.

Clinicians have generally attempted to define features in the recovered recordings based on the peaks and valleys in the waveform relative to the time of occurrence of the stimulus (or the time between such peaks). They have also attempted to calibrate acoustic click peak amplitudes against an equivalent sine wave amplitude. This procedure obscures the difference in energy (power) between these quite different stimuli.

Eggermont and Ponton summarized the value of AEP signals in 2002[54]. "At the current sophistication level of recording and analysis, evoked responses remain in the realm of extremely sensitive objective indicators of stimulus change or stimulus differences. As such, they are signs of perceptual activity, but not comprehensive representations thereof." After describing their interpretation of the analog auditory-evoked potential, they also noted, "the auditory-evoked N_1 response does not correspond to auditory speech perception in a one-to-one fashion."

---

[53]Brittan-Powell, E. Dooling, R & Gleich, O. (2002) Auditory brainstem responses in adult budgerigars (Melopsittacus undulatus) *J Acoust Soc Am* vol. 112(3) pt 1, pp 999-1008

[54]Eggermont, J. & Ponton, C. (2002) The Neurophysiology of Auditory Perception: From Single Units to Evoked Potentials *Audiol Neurootol* vol 7, pp 71–99

The waveforms related to the AEP and ABR are generated by a highly parallel DC coupled neural system consisting of both tonic (analog) and phasic (pulse) components. The parallel paths need not be of equal length or involve constant delays per unit length. The AEP and ABR signals are acquired at a location that is a variable distance from the different sources described above. Therefore, the form of the recovered signals are frequently a function of the location of recording, the nature of the stimulus and many other factors. The character of a recording obtained under binaural stimulation may be significantly, and fundamentally, different from the similar recording obtained under monaural conditions.

The signals obtained by nearly all of the test instrumentation available employ AC coupled head stages (or preamplifiers). The use of an AC amplifier to record a fundamentally DC signal always results in distortion of the information. Sometimes the distortion is considerable. As an example, the first cycle of the recorded signal will be at a significantly different voltage level than will subsequent cycles due to the charging of the isolation capacitance in the head stage. To explore these differences is tedious but necessary. Interpretation of the resultant AC waveforms also varies considerably between the clinical and exploratory research environments. The clinician tends to speak of latency extending to the peak of a response, rather than to its point of initiation. This obscures the relationship between transport delay prior to the beginning of the response and the rise time of that response.

Figure 9.4.1-1 provides a variety of waveforms that can be discussed individually. The ground rule introduced by Brittan-Powell & Dooling will be imposed initially (although it will be reworded); only the first two positive peaks in the composite waveform as a function of time, resulting from monaural stimulation, designated by sequential arabic numerals, will be described. Later, this restriction will be extended to examine subsequent features and the binaural case.

Frame **A** shows the signals generated by an impulse stimulus, (a short xxx going acoustic pulse with a temporal duration that is short relative to the bandwidth of the auditory system. This means the pulse has a duration that is less than the reciprocal of the maximum frequency response of the system under evaluation. Under this condition, the risetime of the initial analog waveform generated by the sensory neurons is largely independent of the duration of (but not the energy in) the stimulus. The waveforms apply to the simplest neural signal path from an IHC sensory neuron near the oval window through its associated non-differencing spiral ganglion cell within the spiral ganglia.

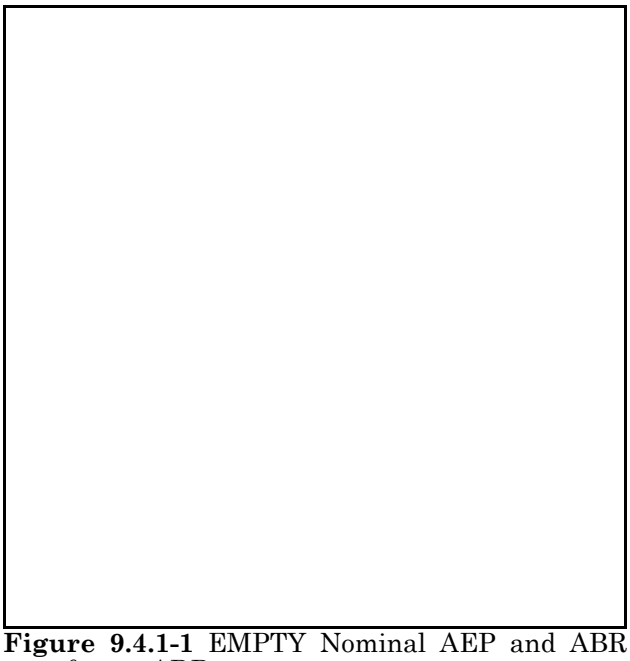


Figure 9.4.1-1 EMPTY Nominal AEP and ABR waveforms. ADD

The impulse is shown beginning at time zero. For an appropriate impulse, its width is negligible relative to the fineness of the temporal scale. Therefore, time can be considered to start at the midpoint of the pulse for convenience.

The acoustic energy requires a finite time to traverse the outer and middle ear. Air is a linear medium and it can propagate a unidirectional pulse of infinite extent. However, air is also a fluid. It will tend to flow

so as to neutralize a local pressure disturbance. At the detail level, the pulse will acquire a rounded form, with an undershoot, as it propagates in a real environment. It also requires a finite time to reach the first IHC of the inner ear. The effects of reflection and reverberation related to the outer and middle ear will be ignored for the present.

Under the above conditions, the first non-acoustic signal is generated by the piezoelectric transducer as shown. For a xxx going acoustic pulse, the polarity of the piezoelectric signal will be xxx-going. Depending on the size of the piezoelectric element, and the intensity of the stimulus at the sensory neuron, the waveform generated may take on any of the shapes shown, as predicted by the P/D Equation. Because of this variation, the only feature of this waveform that is the latency term in the P/D Equation which is separte from the transient response term. This is the true latency associated with piezoelectric transduction. It is shown as the point where the waveform departs from the baseline. The peak of the transient response occurs at a time given by the sum of the latency and the risetime of the leading edge of the transient. This sum is what is frequently described erroneously as the latency. ***This sum is stimulus amplitude, and generally test configuration, sensitive***.

This signal will cause the Activa of the adaptation amplifier to begin conducting as shown.

The complexity of the resulting waveforms is illustrated by the standardized waveform prepared by Jewett, Romano & Williston[55]. **Figure 9.4.1-2** also highlights that the signals, which are measured in millivolts when measured directly, are measured in the microvolt range from the exterior of the head. Such an attenuation of 1000:1 is suggestive of the amount of crosstalk and extraneous noise that can be accumulated into the ultimate waveform observed at the exterior of the skin. Brittan-Powell, et. al. have proposed a significantly different nominal ABR response (page 1000).

The technique used to obtain these waveforms is less than technically sophisticated. Schwartz & Berry illustrate the nominal locations used for the electrodes applied to the scalp of the head[56]. The locations exhibit little correlation with the electrical circuitry internal to the head or the various relatively insulating membranes separating large regions of the cortex. There is an equal problem with the sources of the stimulus. Their figures 5-8 and 5-9 show the variation in the recorded waveforms based on the type of source used.

Jacobson has provided data on an adaptation of the AEP known as the auditory brainstem response (ABR). It is the most recent addition to the clinical test battery in audiology.

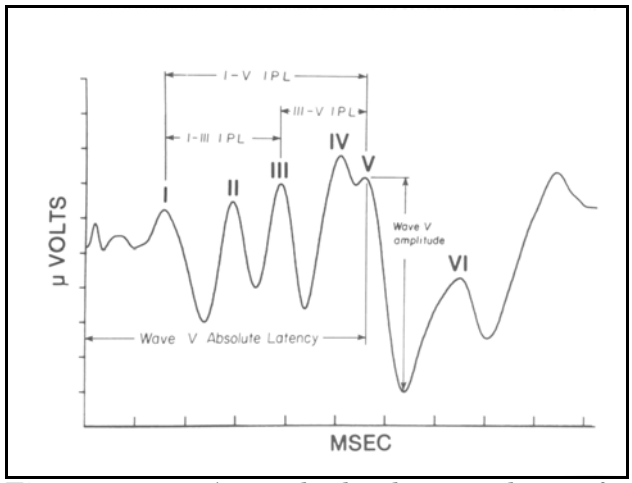


Figure 9.4.1-2 A standardized nomenclature for auditory waveforms measured externally. The absolute latencies of the individual features of the waveform are measured from the root of the feature. From Jewett, Romano & Williston, 1970.

[55]Jewett, D. Romano, M. & Williston, J. (1970) Human auditory evoked potentials: possible brainstem components deteced on the scalp *Science* vol. 167, pp 1517-1518

[56]Schwartz, D. & Berry, G. (1985) Normative aspects of the ABR In Jacobson, J. ed. The Auditory Brainstem Response. San Diego, CA: College-Hill Press Chapter 5

Figure 9.4.1-3 shows the interpretation of the source of the features in the waveform of Jewett & xxx by Moller & Jannetta (in Jacobson). The data is usually obtained in response to a tone contained within a rectangular envelope. An example is given in Jacobson of a 2000 Hz tone with a duration of 5 ms. The recorded pattern shows no sign of the fundamental frequency over the 20 ms recording interval.

A critically important point is that feature I, II and III are of opposite polarity in the Moller and Jannetta figure (relative to feature V) from that in the Jewett & xxx standard. Apparently two separate standard waveforms are required for the AEP and ABR techniques.

Jacobson describes the repeatability of the responses as good for clinical purposes. This appears to be based on the large numbers of stage 3 action potentials that are averaged to generate each of the features in the waveform. However, it is clear that adhering to a specific protocol is important. The form of the recording will change with test frequency (if the tone goes below the 500-600 HZ transition) and may change with any binaural stimuli.

Schwartz & Berry (also writing in Jacobson) have provided two waveforms obtained by using different acoustic sources. The ABR waveforms are grossly different at the detailed level.

Schwartz & Berry have provided statistical data on the normative values of latencies for the individual features as reported by 10 laboratories. The tests were performed at a nominal level of 60-70 SL. They give the average latency across these laboratories for feature I as 1.7 ms.

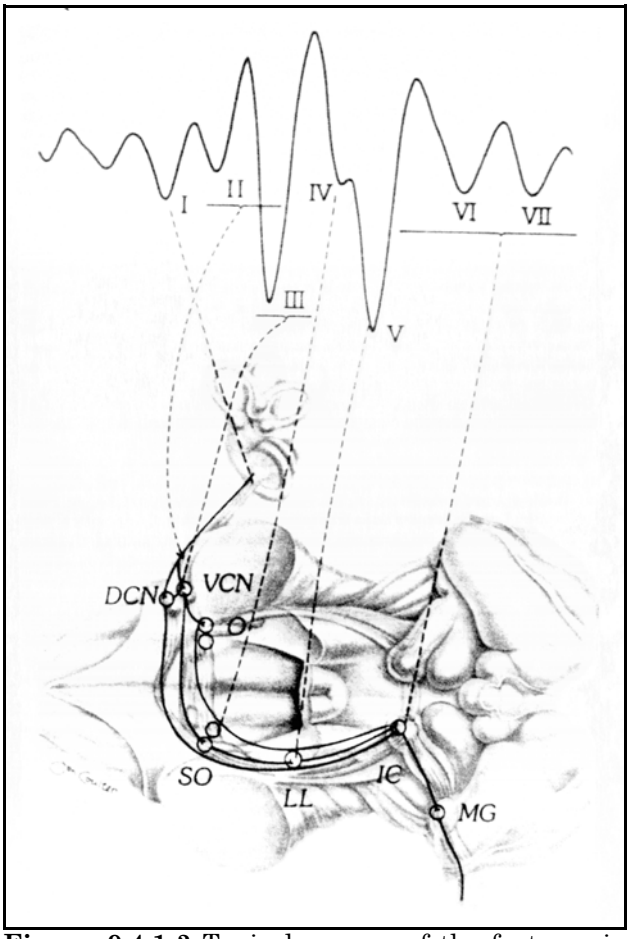


Figure 9.4.1-3 Typical sources of the features in ABR responses of man. From Jacobson, 19xxx.

Brittan-Powell, et. al. noted the general case, based on their definitions, "Latency decreased monotonically as a function of increasing intensity for waves 1 and 2 while the interwave interval remained relatively constant across intensities." Rewording this conclusion, the latency and possibly the risetime of wave 1 decreased with stimulus intensity. However the latency between waves 1 and 2 remained constant. Their figure showing the cochlear microphonic, the CCM (ne; compound action potential) and the ABR recorded simultaneously may be important. However, the reported waveforms were the result of considerable post collection manipulation that must be understood. The manipulation is described only briefly in the paper.

9.4.1.1.3 The human MLR type auditory evoked potential response

Bell, et. al. have reported on attempts to improve interpretation of electrophysical recordings of the auditory evoked potentials measured at the vertex of the human head[57]. The goal was to evaluate the state of anesthesia in surgical patients. The challenge was to recover the desired signal in the presence of significant noise (+20 dB above the desired signal).

9.4.1.1.4 Performance from MRI and fMRI techniques EMPTY

9.4.1.2 Invasive performance measurements EMPTY

Cochlear microphonic data

9.4.1.2.1 Cochlear microphonic and CCM data EMPTY

9.4.1.2.2 Two-tone performance in squirrels

Hind, et. al. provided interesting information in 1967[58]. The problem was the data was collected using non-synchronous tone stimuli and presented in histogram form. It also used histograms with seldom used names. The same team presented another paper in 1969 overcoming this shortcoming[59]. **Figure 9.4.1-4** is modified from figure xxx of the Brugge paper. The original figure assumed the OHC sensory neuron operated as a rectifier. However, the same data can be matched well assuming the OHC was operating as an integrator.

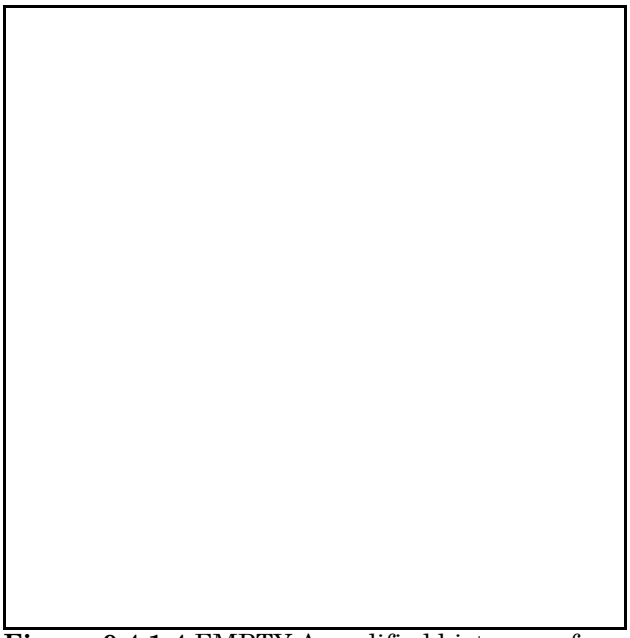


Figure 9.4.1-4 EMPTY A modified histogram from a two-tone test where the tones were harmonics of a common fundamental. Xxx ADD Modified from Brugge, 1969.

[57]Bell, S. Allen, R. & Lutman, M. (2002) Optimizing the acquisition time of the middle latency response using maximum length sequences and chirps *J Acoust Soc Am* vol. 112(5), pt 1, pp 2065-2073
[58]Hind, J. Anderson, D. Brugge, J. & Rose, J. (1967) Coding of information pertaining to paired low-frequency tones in single auditory nerve fiberes of the squirrel monkey *J Neurosci* vol. 30, pp 794-816
[59]Brugge, J. xxx (1967) xxx

9.4.2 Important uses of non-signaling channel data in hearing research

The long held assumption that the basilar membrane was a major functional element in the auditory system has left a large legacy of technical data, and putative theories. Much of the data remains pertinent and a place for it should be found in the framework of hearing theory.

There are at least three forms of data collection in hearing research that do not represent actual signals occurring within the normal signaling channels of hearing. These include signals related to the physical motion of the basilar and tectorial membranes, the electrical signal acquired at the round window of the cochlea and the evoked potentials acquired at a variety of remote locations, typically on the dermis of the subject. This type of data must be examined very carefully to determine its relevance and utility. It has frequently been misinterpreted as relating directly to "mainstream" auditory operations.

9.4.2.1 Mechanical motion of the cochlear membranes

9.4.2.1.1 Test equipment used to measure small motions

[xxx may duplicate somewhere else. Check occurrances of Mossbauer in indeces.]
Harrison has provided a discussion of the capacitive probe, Mossbauer and laser interferometric methods of measuring the motions of the basilar, tectorial and Reissner's membranes[60]. These techniques exhibit remarkable sensitivity to axial motion relative to the sensor. Sellick, et. al. note the impact of the size of the auxiliary source used in the Mossbauer experiments[61]. This can also be a problem in laser interferometry if an auxiliary mirror is used to increase the reflectivity of the surface being examined. This technique has been largely superseded by the laser interferometric technique. The major problem with the use of these techniques is the fact they record peak motions under high acoustic stimulus conditions that are far below the sensitivity range of the auditory sensory neurons. They routinely demonstrate that the *maximum* transverse amplitude of vibrations of the basilar membrane and the outer surface of the tectorial membrane are in the 10^{-2} micron range or less. Khanna & Leonard measured maximum motions of only 3-4 Angstrom, using both the Mossbaur and laser interferometer techniques, for a stimulus of 20 dB above the background level[62]. Many of their measurements were at amplitudes below the diameter of an atom The sensory neurons generally exhibit maximum signal output for displacements in the 1-2 micron range. [xxx check the numbers for the sensory range.] Measurements using these techniques tend to falsify the conventional wisdom that the basilar membrane is the primary mode of energy transfer within the cochlear partition than they support that proposition. Khanna & Leonard hint at this in their closing paragraph. "The present observations suggest that the mechanical elements that determine the primary response of the cochlea in the CF region are related to the OHCs. The most likely candidates are the combination of the stereociila bundles and the tectorial membrane. These may form an independent resonant system from that of the basilar membrane." Amen. Their abstract concludes, "These observations lead to the conclusion that the tuning properties in the CF region are predominantly determined by the

---

[60]Harrison, R. (1988) The Biology of Hearing and Deafness. Springfield, Il: Charles C. Thomas pp 44- 50
[61]Sellick, P. Yates, G. & Patuzzi, R. (1983) The influence of Mossbauer source size and position on phase and amplitude measurements of the guinea pig basilar membrane *Hear Res* vol. 10, pp 101-108
[62]Khanna, S. & Leonard, D. (1986) Relationship between basilar membrane tuning and hair cell condition *Hear Res* vol. 23, pp 55-70

mechanical properties of the OHC and not the basilar membrane."

9.4.2.1.2 Description of the motions of the basilar and tectorial membranes

[xxx node Z and node Y]

9.4.2.1.3 Description of the motions of Reissner's membrane

[xxx Rhode 1978] Rhode has made measurements on the motions of Reissner's membrane in response to a range of acoustic stimulations applied at Node A[63]. To quantify and calibrate the stimulus, he measured the motion of the malleus at Node B. The spectral data, plotted as the ratio of Reissner's membrane amplitude to malleus amplitude, appears similar to the data recorded from the tectorial and basilar membranes. However, there are differences in the range of 10 to 28 dB. For purposes of this work, Reissner's membrane will be assumed to mimic the motions of the tectorial (not the basilar) membrane.

[63]Rhode, W. (1978) Some observations on cochlear mechanics *J Acoust Soc Am* vol. 64(1), pp 158-176

9.4.2.2 Extraneous electrical (microphonic) signal data

9.4.2.2.1 Description of the cochlear microphonic (and CAP) at node W

Researchers have found it useful to record an *electrical* signal obtained at the round window of the cochlea. This signal has come to be known as the cochlear microphonic, CM. This signal has been known to be a composite for a long time. Davis & Saul noted in 1931 and 1932 that the response could follow the stimulation at frequencies as high as 100 times the highest observed action potential rate[64]. The origin of this composite waveform is illustrated in **Figure 9.4.2-1**. It is the electrical signal generated in response to an acoustic stimulation applied to the oval window of the cochlea. The collection point will be defined as node W. The originating node varies with the investigator. The signals indicated by the arrows originate at different times and at both the IHC and OHC neurons, based on the propagation velocity of the acoustic energy in the surface acoustic wave Hence, the resulting composite waveform can be quite complex if the acoustic stimulus is complex. However, the stimulus is usually restricted to a single tone. Under this restriction, a large number of the IHC are stimulated as the acoustic energy moves along the gel-surface of the tectorial membrane. However, these signals are dispersed in time due to the propagation velocity of the acoustic energy within Hensen's Stripe. Only a limited number of OHC are stimulated in a local area of the Organ of Corti are stimulated significantly, and their outputs are nearly in phase. Fortunately, the sensitivity of the IHC on a amplitude per Hz basis is much lower than the sensitivity of the OHC. However, it is usually necessary to perform considerable filtering to remove the background contributed by the IHC neurons and obtain a useful CM.

[64]Carterette, E. & Friedman, M. *eds*. (1978) Handbook of Perception, Volume IV: Hearing. NY: Academic Press pp 17-18

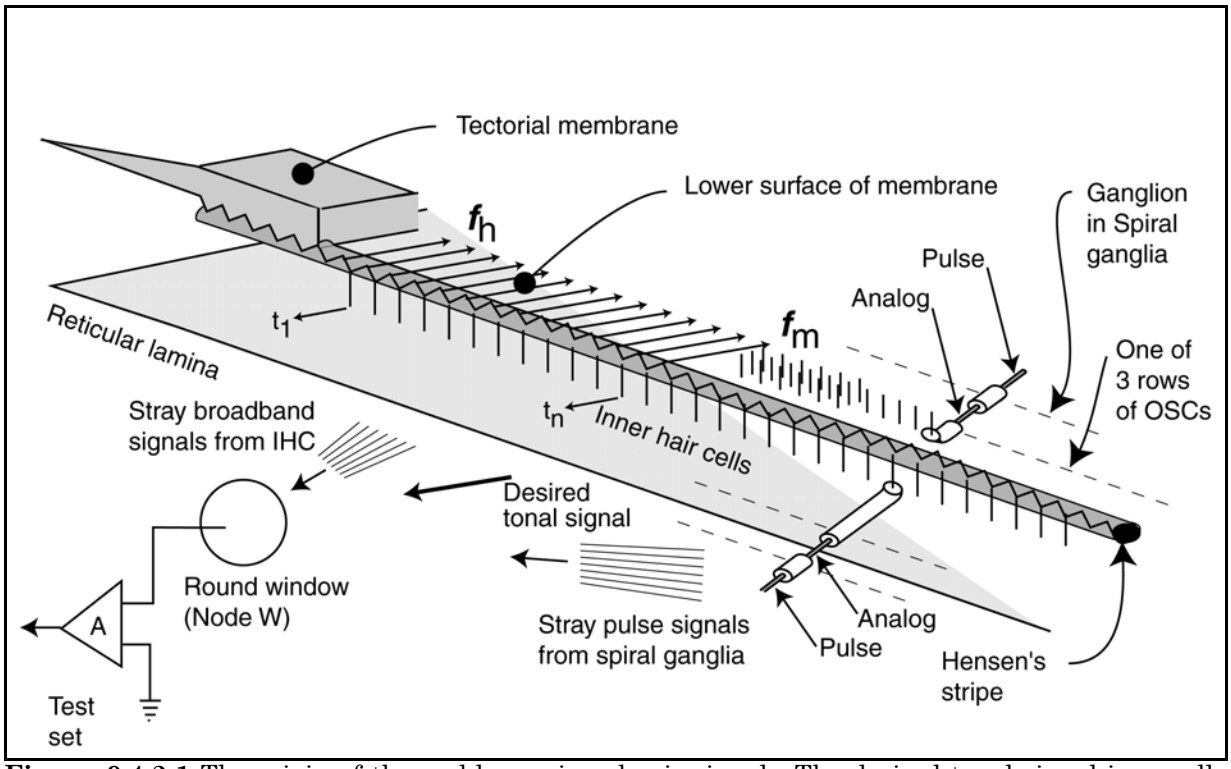


Figure 9.4.2-1 The origin of the cochlear microphonic signal. The desired tonal signal is usually buried in the noise contributed by both unwanted broadband and unwanted pulse signals. The test set ground location is not always specified precisely. The location of this lead can affect the relative amplitude of the component waveforms in the test set output.

9.4.2.XX2.1.1 The cochlear microphonic

The term cochlear microphonic is singular. It refers to an AC receptor potential measured (typically) at the round window. Recording from other locations will be discussed in the following section. Like any other evoked potential, it is a single potential measured at a single point due to voltages arising at a large number of distributed remote points. Obtaining the waveform requires surgical intervention.

Antoli-Candela & Kiang provide a good discussion of the cochlear microphonic response to 100 microsecond pulses[65]. They demonstrate the initial polarity of the response relative to the character of the stimulus. The majority of the subsequent response tends to be unipolar. While they describe their responses as compound action potentials, the primary responses represent the generator waveforms of the sensory neurons. They are clearly a function of stimulus amplitude (and adaptation) whereas action potentials exhibit constant amplitude regardless of stimulus level. The initial responses also exhibit a variable time delay as a function of stimulus level. This is another clear indication the waveforms are generator responses resulting from the phonoexcitation/de-excitation

[65]Antoli-Candela, F & Kiang, N. (1978) Unit activity underlying the N_1 potential *In* Naunton, R. & Fernandez, C. *eds*. (1978) Evoked Electrical Activity in the Auditory Nervous System. NY: Academic Press pg 165-191

mechanism. The low amplitude of their signals is characteristic of cochlear microphonic signals.

Gulick, et. al. have provided a discussion and several figures related to the cochlear microphonic[66] (pp 148-153). The figure shows the large dynamic range associated with the measured voltage and the fact that saturation (due primarily to compression at the sensory neuron pedicle) occurs at different frequencies. Gulick et. al. assert that if the CM is amplified and fed to a speaker, the investigator is able to recognize the signal as related to the input stimulus to the subject under evaluation. The fact that the measurement is a summation of the activity in virtually every sensory neuron is demonstrated by this observation. Although not stated, it appears their figure was obtained in response to short tone bursts as it shows no adaptation. The dashed lines in their figure suggest the response of the inner ear is quite linear. The envelope formed by replotting the sound pressured required to achieve a 30 microvolts output against frequency appears to be very close to the transfer function associated with the middle ear of the cat (maximum sensitivity in the 3,000 to 7,000 Hz range).

Gulick, et. al. have also addressed what is frequently called the compound action potential, another evoked potential sensing both a signal similar to the cochlear microphonic and similar to N_1 (page 162-166). As demonstrated in the next paragraph, the N_1 waveform is clearly an analog signal unrelated to any phasic action potential. The response is more properly labeled a compound cochlear microphonic (CCM). Sometimes a signal similar to N_2 is also recorded. The delays between these pairs of signals are functions of where the probe is physically placed. It is also a function of the frequency of the stimulus (above or below 500-600 Hz). These differences reflect the electrolytic propagation time of the neural electrical signals between these sources (very slow) as well as the electrical propagation time of these sources and the probe location (usually negligible). [xxx give reference to PBV–]

9.2XXX.2.1.2 The cochlear microphonic recorded under various conditions

The research value of the cochlear microphonic is unquestioned. However, its repeatability is poor since the precise location of the probe is frequently note identifiable and is generally not reproducible.

Figure 9.4.2-2 reproduces the cochlear microphonic recorded at the first turn of the cochlea in guinea pig[67]. [xxx add words from Rasmussen pg 25] The responses are clearly not distorted significantly over a sound pressure level ranging from 65 dB to 115 dB, although the output signal level intercepted from the pedicles does rollover. It is also clear that this output has not been rectified in any sense, although it may be sitting on the top of a pedestal that is not shown.

---

[66]Gulick, W. Gescheider, G. & Frisina, R. (1989) Hearing. NY: Oxford University Press

[67]Davis, H. & eldredge, D. (1959) An interpretation of the mechanical detector action of the cochlea *Ann Otol Rhin* vol. 68, pp 665-674

After commenting that the observed rollover is not normally seen in mechanical system, Davis makes an observation that is completely supported by this work. "This relationship is most unusual for a mechanical system, and it is even more extraordinary that the limitation and finally the reduction in the amount of electrical output is accomplished without distortion of the wave form. The ear seems to have a true 'automatic gain control'."

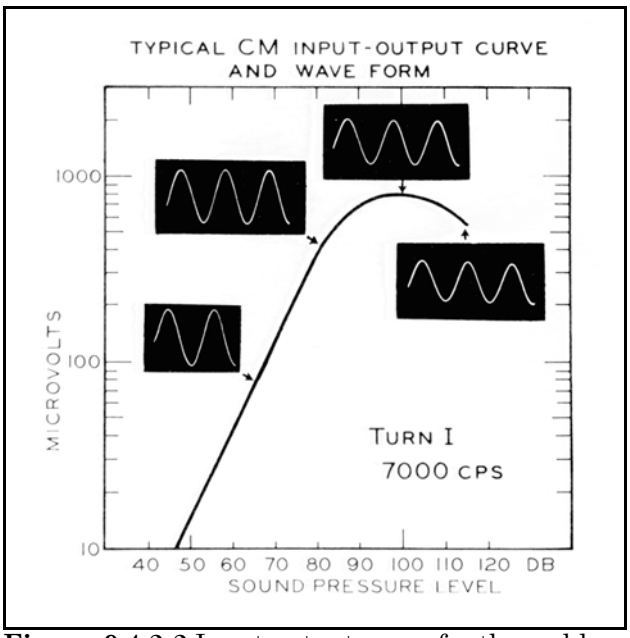


Figure 9.4.2-2 Input-output curve for the cochlear microphonic response of the first turn of the guinea pig cochlea to 7000 Hz tone burst. Note the absence of peak limiting at the highest sound intensity, or rectification at any level (although any pedestal may not be shown explicitly.. From Davis & Eldredge, 1959.

The cochlear microphonics shown in **Figure 9.4.2-3** from Tasaki are particularly important. They show the waveforms recorded at three different locations and a multitude of frequencies. They clearly demonstrate that the propagation medium within the cochlear partition does not exhibit a significant attenuation (and therefore relative phase shift) as a function of either frequency or distance along the cochlea. Similarly, the energy associated with a specific frequency is removed from the transmission channel. abruptly at or near a specific distance. Note the gradual reduction in the frequency response in the column labeled the second turn up to about 2000 Hz. This reduction is followed by a drastic reduction in the response within less than an octave (3000 to 6000 Hz). A similar situation is observed looking at the rows of data. There is very little attenuation along the cochlea until the energy at a specific frequency suddenly disappears. Below the critical frequency, the energy is propagated. Above the critical frequency, it is not. These are not the characteristics of a resonance phenomenon. They are the characteristics of a frequency dispersion phenomenon.

Davis has provided a similar set of waveforms[68]. His data is interesting in that he used a set of "differential electrodes" and provided calibration waveforms. It is also the rare paper that describes the electrical configuration of the test in detail. The calibration waveform was apparently carefully chosen as it shows no phase shift with location along the cochlear partition. The other waveforms do show a phase shift with position (below the cutoff frequency) as expected for any traveling wave.

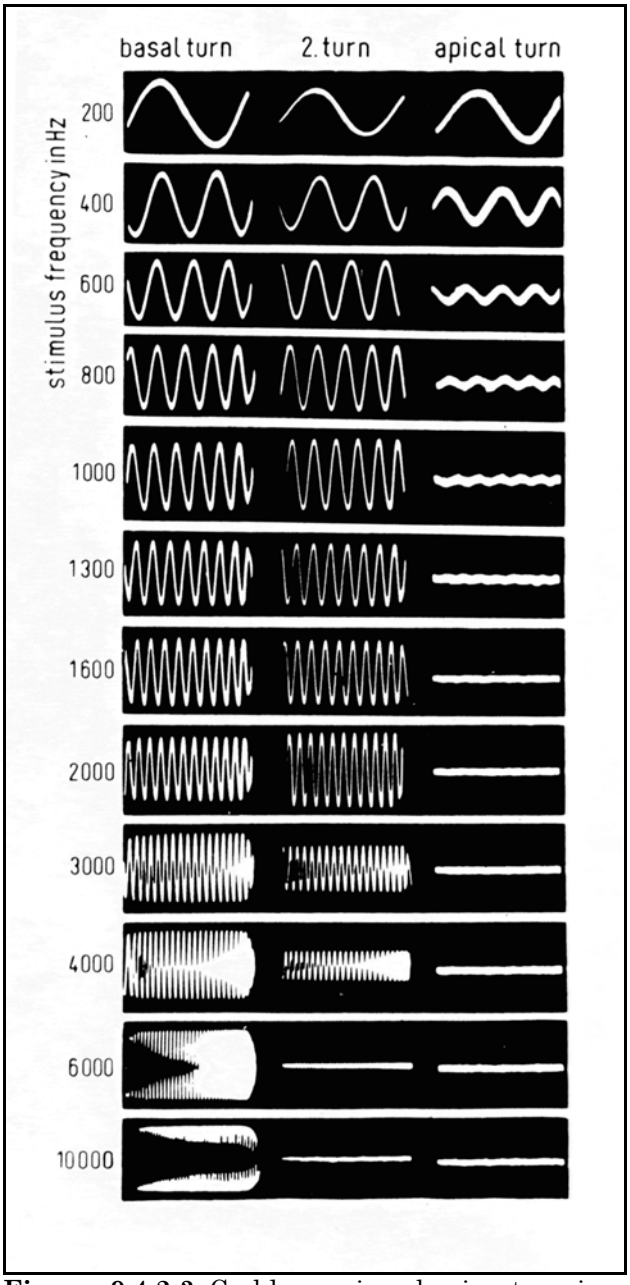


Figure 9.4.2-3 Cochlear microphonic at various locations and frequencies within the guinea pig cochlea. From Tasaki, xxxx.

[68]Davis, H. (1965) Mechanism of excitation of auditory nerve impulses *In* Rasmussen, G. & Windle, W. eds Neural Mechanisms of the Auditory and Vestibular Systems. Springfield, Il: Charles C. Thomas pg 24

9.2XXX.2.1.3 The cochlear microphonic after efferent neuron stimulation

Widerhold has provided good data on the nominal cochlear microphonic following stimulation of the efferent neurons of the cochlear nucleus. While the stimulation was not focussed on a specific afferent signaling channel, the data is important. It shows clearly that under some circumstances, the afferent cochlear microphonic exhibits a variation in N_1 amplitude as a function of such neurological stimulation. "The efferent fibers that cross the brainstem (crossed olivocochlear bundle, or COCB) were stimulated with a train of electrical shocks, delivered through bipolar electrodes in the COCB decussation, at the floor of the fourth ventricle. Summed responses of the cochlear hair cells and auditory nerve fibers were recorded . . . ." As noted by Wiederhold, this experiment has been performed many times with similar results. The efferent excitation is clearly inhibitory toward the afferent signal channel.

As noted by Wiederhold, the efferent channel must be stimulated at high shock rates (at least at their point of stimulation). In response to such rates, the effect on N1 builds up slowly and decays slowly with a time constant of about 100 ms. **Figure 9.4.2-4** provides a calibrated graph of the CCM describing this inhibitory action from Wiederhold. The curve in the absence of efferent stimulation shows the normal phonotopic operating range from about –60 to –35 dB. In the absence of adaptation, the normal response would follow the dash-dot line.

[xxx add words]

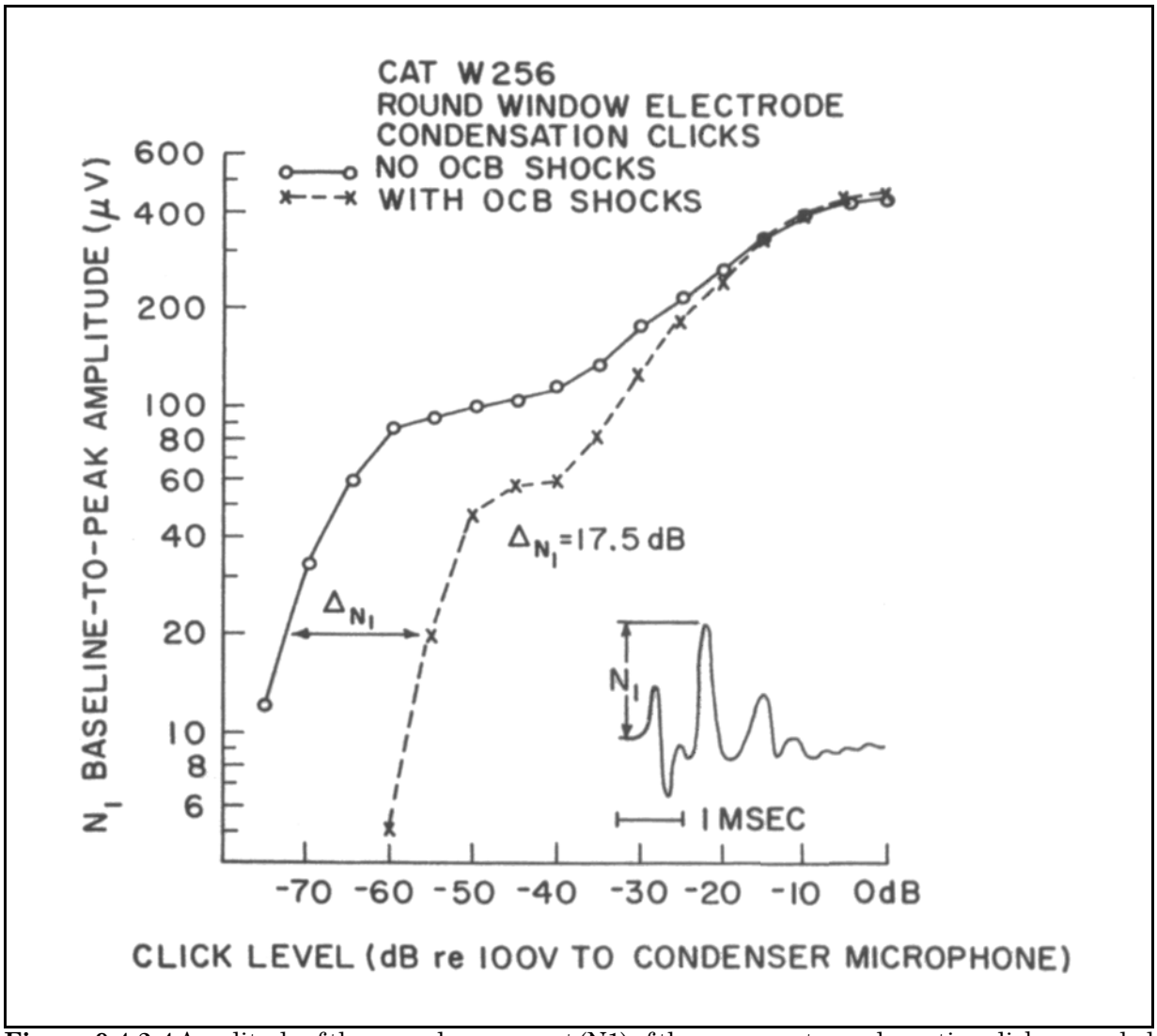


Figure 9.4.2-4 Amplitude of the neural component (N1) of the response to condensation clicks recorded with a gross electrode on the cochlea. Solid line; response to click alone. Dashed line; clicks preceded by 32 shocks to decussation of COCB at 400/sec. Delay from last shock to click is 10 msec. Inset; a typical response to illustrate baseline-to-peak measure of N1 amplitude. From Wiederhold, 1986.

9.4XXX.2.2 Description of the evoked potential signals at node V

The acquisition of evoked potentials as a non-invasive method has attracted some attention in the clinical environment. The method involves recording the voltage difference between two locations on the scalp, frequently the vertex of the head and the nape of the neck. These evoked potentials consist of signals collected remotely and simultaneously from a variety of poorly defined potential source(s). Considerable signal processing is usually required to eliminate extraneous signals not related to the auditory system in any way. The collected data is of little value to applied research in hearing.

9.4.2.3 Other data summarizing extraneous auditory performance between

nodes A and Z

9.4.2.3.1 Mechanical measurements related to the basilar membrane

[xxx incorporate the material concerning de Boer & Nuttall, 1997 here? It may be in 4 now.]
There is extensive data available concerning the motions of the basilar membrane in response to acoustic signals applied to the cochlear duct (frequently following modifications to Stage A). This data was collected in attempts to support the unidirectional basilar membrane vibration theories, whether membrane resonance or traveling wave based. The data shows the composite signal measured by a test target attached to the basilar membrane and tracked using either Mossbauer or Doppler laser techniques.

The data is largely irrelevant to the operation of the auditory system. However, it does suggest the basilar membrane is not a perfect inertial mass relative to the operating portions of the Organ of Corti (the gel-surface of the tectorial membrane, the cilia, and cuticular plate of the sensory neurons).

The data consistently shows the recorded waveforms due to basilar vibration are composite waveforms. They consist of a residual component of motion related to the energy passing the IHC and a residual component of the motion related to the energy passing the OHC near the target. In one case, the test target moved during the experiment and the change in the character of the composite waveform was noted. Rhode provided excellent data on these residual vibrations in relation to the motions of the malleus of the middle ear[69]. He also provided excellent data showing the motion of the basilar membrane and the adjoining boney limbus. His data demonstrates that the motion of the basilar membrane closely matches that of the boney limbus up to at least four kilohertz. The motion of the basilar membrane was 16 dB higher. However, the match gives further justification to considering the basilar membrane a non-functional inertial mass in the context of hearing. The amplitude of the residual vibration of the basilar membrane was reported as in the 0.01 to 0.06 Angstrom range near 7 kHz. This range is far below the vibrational range of the cilia within the Organ of Corti.

9.4.2.3.2 Electrical measurements related to the round window

Different communities have defined and measured both a cochlear microphonic and a complete cochlear microphonic (previously known as a "compound action potential"). Both are electrical signals measured with a probe placed near the round window, but similar evoked responses can be measured at other locations. **Figure 9.4.2-5** shows a typical waveform often used as a quasi-standard.

---

[69]Rhode, W. (1970) Observations of the vibration of the basilar membrane in squirrel monkeys using the Mossbauer technique *J Acoust Soc Am* vol. 49(4), pt 2, pp 1218-1231

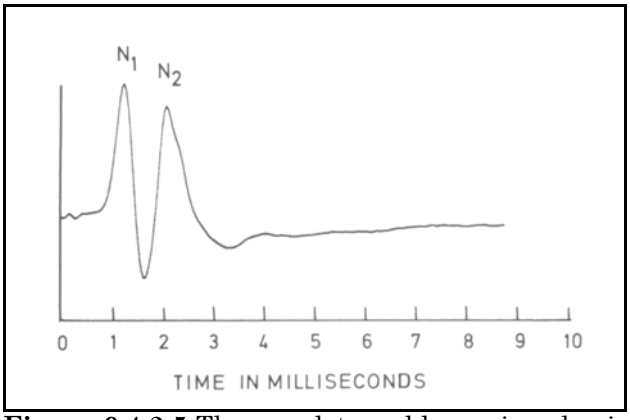


Figure 9.4.2-5 The complete cochlear microphonic (CCM) in response to a click sound recorded from the round window of a rat. Up is negative. N_1 is due to a summation of analog generator potentials. N_2 is due to a summation of subsequent Stage 2 signals. From Moller, 1983.

Moller has noted the confusion concerning the character of this waveform. Xxx add interpretation. [xxx see Section 4.1.3.]

Figure 9.4.2-6

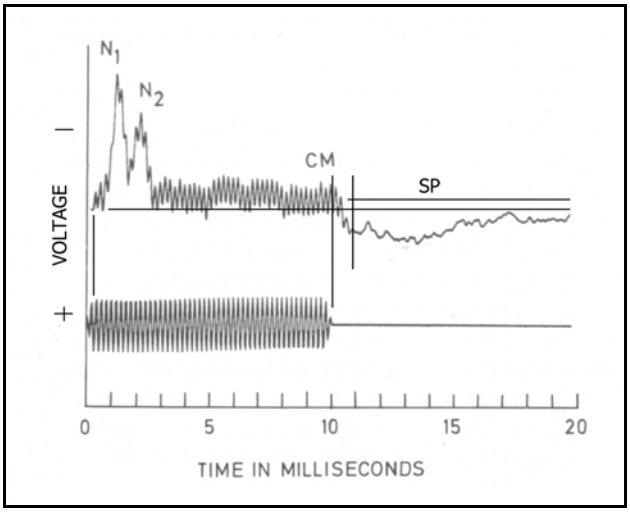


Figure 9.4.2-6 An unfiltered recording from the round window of a rat to show the CM, CCM (N1 & N2) and SP at the same time (A). CM; cochlear microphonic. CAP; compound action potential. SP; "summating potential." Note the delay due to the finite travel velocity of the tectorial membrane before the beginning of N1. B; tone burst at 5-kHz. See text. Data, without construction lines from Moller, 1983.

9.4.3 Cumulative Stage 0 performance, the physiological (non-electrolytic) acoustics ?CONDUCTIVE

9.4.3.1 Outer ear

The outer ear of chordates can be described as a curved horn waveguide terminating in a rigid barrier, the tympanic membrane. The tympanic membrane is described as rigid because the amplitude of its motions relative to the wavelengths of the incident acoustic energy are negligible. It is only with respect to the middle ear that the motions of the tympanic membrane are significant. With regard to the outer ear, the tympanic membrane appears as a lossy impedance containing a resistive element.

Under some circumstances, the tympanic membrane must also be considered a mechanical oscillator exciting the air in the waveguide. These circumstances relate to the otoemission mechanism and some reports of echoes originating within the middle or inner ear. However, the efficiency of this oscillator in exciting the air within the waveguide cannot be considered high. The amplitude of the physical motions of the tympanic membrane associated with these mechanisms is quite small (relative to the acoustic wavelengths). Specialized instrumentation is generally required to detect these emissions.

As noted frequently, the outer ear can be considered a transformer in electrical analogs of the auditory system. It collects energy, existing in the form of low amplitude pressure waves covering a large area, and concentrates that energy, in the form of a higher amplitude pressure wave covering a much smaller area, the effective area of the tympanic membrane. No active amplification is involved in this process.

The curve horn waveguide can be closed by some species under muscular control, either through steering of the pinna or actually collapsing the cross section of the horn at a point within the auditory duct under muscular control. These methods of controlling the amplitude of the signal reaching the tympanic membrane have largely atrophied in humans during evolution.

9.4.3.1.1 Location detection by time difference computation MOVE

The elements of the auditory system are largely symmetrical subsequent to the pinna in the auditory system. Because of this symmetry, the arrival time of acoustic signals at the two tympanic membranes can play an important role in source location. It is only necessary to compute the difference in travel time for the signals arriving at the two membranes and use a lookup table to perceive the angles (generally in the inertial space of the saliency map, not head-oriented space) to the source. The lookup table is needed because of the variation in the surface physiology of the pinna of a specimen, the instantaneous pointing of the pinna, and the interference to the signal paths contributed by other bodily structures, including the head.

The accuracy requirement associated with the location calculation is similar to that within the visual system. Source location by the auditory system is usually used to steer the visual system. Thus, the goal is to compute source coordinates within the acceptance range of the foveola of the visual system (typically +/– 0.6 degrees in azimuth and elevation. [xxx what level is typically achieved, provide reference] Once the source is located within this accuracy, the visual system is usually used to localize the source to a greater accuracy.

9.4.3.1.1 The steerable properties of the outer ear

The curved horn waveguide is steerable in most chordates (humans being a major exception). Casual observation of pets suggests the ears can generally be steered independently or in coordination. The steering of each external ear is performed by three pairs of muscles analogous to the muscles controlling each ocular globe of the visual system. The achievable pointing angle range in some species is quite large but has not been explored in this work.

9.4.3.2 Middle ear

The middle ear of a given species is designed to accommodate the environment of the animal and the required frequency range of the animal, and to perform its function faithfully over the life of the animal. The latter requirement places a variety of constraints on the middle ear and leads to a variety of pathological conditions when it fails to satisfy these constraints. Calcification of the ligatures connecting the small boney structures involved is a common failure.

The design of the middle ear differs markedly between terrestrial and aquatic animals. For aquatic animals, the impedance of the fluids in the outer and inner ear are similar. In these situations, the middle ear may be rudimentary. In reptiles, it may consist of a rigid column connecting the tympanic membrane to the oval window of the inner ear. For terrestrial animals, its primary role is to act as an impedance transformer (of sufficient bandwidth) between the low impedance of air in the auditory duct and the impedance of the liquids within the inner ear. This transformation is accomplished by connecting the tympanic membrane to the oval window of the cochlea by a mechanical lever. While much of the literature discusses this transformation from an amplitude perspective, it is the transfer of total energy that is of critical importance. It is more appropriate to speak of power in hearing because time and frequency play major roles in the mechanisms of hearing. Recent experiments have shown that the tympanic membrane acts primarily as a uniaxial piston in delivering power to the umbo at the tip of the malleus. This energy is transferred to the oval window of the inner ear by the lever system of the middle ear. The purpose is to drive the oval window acting also as a piston. The fact that it operates as a piston is highlighted by the fact the stapes interfaces with the oval window at two points of relatively large diameter. The oval window has an effective diameter. When it moves, it transfers power to the fluid within the inner ear. This power is not well characterized by the pressure associated with it alone since the impedances involved are poorly documented. The power is more easily described by the pressure exerted per unit area of the window multiplied by the effective area of the window and the distance the piston moves in a given interval. By rearranging these terms, as

discussed in Section xxx, it is possible to describe the work in terms of the sinusoidal velocity of the oval window (or attached stapes) for a given stimulus intensity. For sinusoidal stimulation, the velocity of the response is easily found from the amplitude of the response and the frequency. If one desires to describe the amplitude of the motion of the oval window, it is absolutely necessary that he also describes the effective area of the window or the impedance of the fluid medium being driven. The latter is difficult. The impedance changes continuously with the change in the shape of the ducts within the inner ear.

- - - -

Describing the energy transfer function for the middle ear requires close attention to the extent of the middle ear as discussed in **Section 2.3.3**. The stage is generally considered to begin at the outer surface of the tympanic membrane. Its ending is more difficult to define. It can be defined as the surface of the oval window in contact with the perilymph of the labyrinth. Alternately, it can be defined as the surface of the caecum membrane in contact with the perilymph. This latter location insures the transfer function only describes the acoustic energy that is actually effective in exciting the orthodromic tectorial membrane. In either case, the measurement requires a probe to be introduced into a specific location within the vestibule of the labyrinth.

Gan et al. have recently provided a three-dimensional (3D) finite element (FE) model of human ear with accurate structural geometry of the external ear canal, tympanic membrane (TM), ossicles, middle ear suspensory ligaments, and middle ear cavity[70]. They describe their model as a two cavity model and cite their earlier work on two one cavity models.. They quantify its accuracy by comparing it with the data for the outer ear of Shaw (1974) discussed previously in **Section 9.3.1.4.**

It is very difficult to find data on the transfer function of the middle ear in the literature for several reasons. It is extremely difficult to affect satisfactory probes and other instrumentation necessary to measure this transfer function *in-vivo* for any animal. Because the system is so tightly coupled, the experimental protocol and probes must maintain the relevant impedance levels between the outer ear and the subsequent inner ear.

Most experimenters have been satisfied to measure the input admittance of the middle ear by stimulating the tympanic membrane[71]. However, without a detailed model of the remainder of the mechanical system, this admittance cannot be converted into data describing the middle and inner ear separately. Dallos has presented some very old data describing the transfer function of the middle ear of the cat[72]. He describes the middle ear as a low pass filter with a very sharp roll-off at about 5 kHz. He models the middle ear as an electrical filter half-section containing a parallel tuned circuit in the series path and a parallel tuned circuit in the shunt path. [xxx reproduce his fig. probab. 3.18] While probably suggestive of the human middle ear, he notes the very species specific character of the middle ear transfer function.

Kringlebotn and Gundersen have provided a transfer function for the “middle ear” of humans by comparing the volume displacement at the round window to the sound pressure at the tympanic

---

[70]Gan, R. Suna, Q. Feng, B. & Wood, M. (2006) Acoustic–structural coupled finite element analysis for sound transmission in human ear—Pressure distributions *Med Eng Phys* vol 28 pp 395–404

[71]Huang, G. Rosowski, J. Puria, S. & Peake, W. (2000) A noninvasive method for estimating acoustic admittance at the tympanic membrane *J Acoust Soc.Am* vol 108 (3), Pt. 1, pp 1128-1146

[72]Dallos, P. (1973) xxx The Auditory Periphery fig 3.18

membrane external surface of the tympanic membrane[73]. **Figure 9.4.3-1** [xxx replace figure]shows their data from 68 humans for this combination of middle ear and fluid path through the cochlea to the round window. This figure should not be taken as representative of the transfer function of the human middle ear.

The figure shows the nominally 15 dB/oct roll-off beginning at 800 Hz. This roll-off is probably associated with the very long and narrow perilymphatic duct in humans noted by Kringle in his 2000 paper[74]. This perilymphatic duct is not involved in the excitation of the caecum membrane (beginning section of the tectorial membrane) of hearing. His 2000 paper includes good data on the effect of static over and under pressures on the performance of the inner ear cavity.

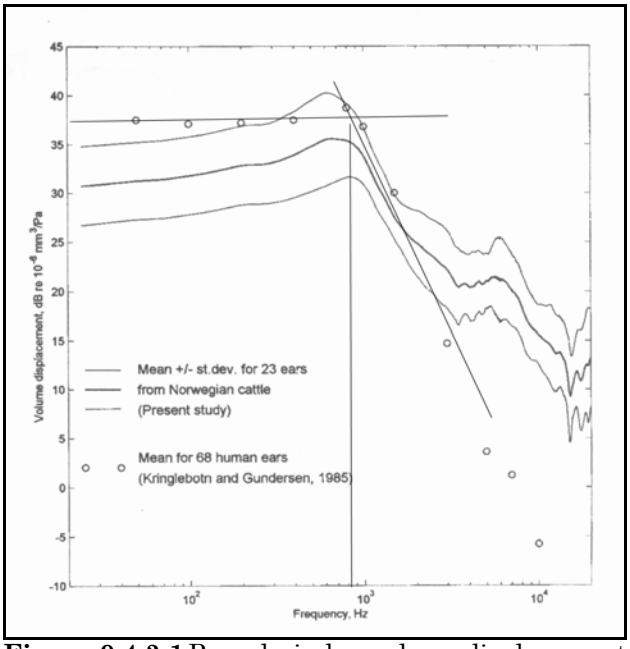


Figure 9.4.3-1 Round window volume displacement relative to sound pressure at the eardrum This figure is not indicative of the transfer function of the middle ear alone. REPLACE with figure from Kringlebotn & Gundersen, 1985.

Rosowski and Merchant have provided a very detailed paper on the human middle ear with good supporting data[75]. The paper was designed to support reconstructive surgery involving the middle ear.

Their figure 1a fails to illustrate the location of the caecum membrane within the vestibule; however, the arrow associated with the stapes footplate (U_S) in figure 1b points to its location. They present a hydraulic and electrical model of the middle ear that are both too simple for use here. Neither model can adequately reproduce the data of Shaw for the transfer function of the outer ear. They treat the middle ear as beginning at the outer surface of the tympanic membrane and ending at the dry surface of the oval window. They treat the oval window itself (defining the impedance of the oval window by that of the annular ligament supporting it) as part of the inner ear.. The cochlea is shown as a simple inductance/resistance (LR) shunt circuit connected to the annular ligament (modeled as a series resistance/capacitance (RC) circuit. No provision is made for loss of energy within the fluid portion of the vestibule. Interestingly, they equate the pressure within the vestibule (P_C) with the pressure at some location within the cochlea.

- - - -

[xxx this insert forms a lead-in to the next section. It may belong in the next section]

---

[73]Kringlebotn, M. & Gundersen, T. (1985) Frequency characteristics of the middle ear *JASA* vol 77(1) pp 159-164

[74]Kringle, M. (2000) Frequency characteristics of sound transmission in middle ears from Norwegian cattle, and the effect of static pressure differences across the tympanic membrane and the footplate *JASA* vol 107(3), pp 1442-1450

[75]Rosowski, J. & Merchant, S. (1995) Mechanical and acoustic analysis of middle ear reconstruction Am J Otology vol 16(4) pp 486-497

The paper by Nuttal & Dolan is critically important to the above discussion[76]. By making simultaneous measurements on the stapes attached to the oval window as well as discrete points on the surface of the basilar membrane, they showed that the local velocity of the transverse movement of the basilar membrane near an OHC was up to 101/1.5 = 67 times the velocity of the movement of the stapes. These were the values they tabulated at the best frequency (associated with the underlying area of interaction between Hensen's Stripe and the OHC) and 20 dB SPL stimulation. They describe this as a gain of near 40 dB (~38 dB) in connection with their figure 2.

The 40 dB figure corresponds to an impedance gain (analogous to that achieved in a transformer with properly matched source and load impedances). There is no reason to suspect an active amplifier was providing this gain. It is due to a change in the channel impedance propagating the energy (power).

In discussing their data, Nuttall & Dolan commented, "Among all of the previous published works on BM motion evoked by steady-state tones, it is difficult to find complete information relevant to the basic issues of active and passive mechanics." They then added a footnote describing their search. Their paper is discussed in greater detail in **Section 4.xxx**. The result of this impedance change will be discussed further in the following section.

The middle ear frequently contains several minor muscles that are able to modify the properties of the lever (the pivot point of the lever is not fixed). Changes in these muscles can vary the efficiency of power transfer and may also change the frequency response associated with the transfer function of the middle ear.

9.4.3.3 Inner ear

The precise operation of the inner ear *in toto* has largely eluded the exploratory research community. The combination of requirements placed on the combined acousto-mechanical and mechanico-electrolytic (neural) portions of the element are quite unique. This section will only address the acousto-mechanical portions of the inner ear. The requirement on the acousto-mechanical portion of this element is complex. Its primary requirement is to accept acoustic power and manipulate it in such a way as to provide two distinct and complex signaling outputs. One output is focused on analyzing the frequency content of the acoustic stimuli received. The other output is focused on the the impulse (or temporal) characteristics of the stimuli.

The most difficult requirement (by far) placed on the inner ear is to perform the above tasks within a volume (that of the cochlea) that is much smaller than the wavelength of the acoustic energy involved with respect to the fluids of the inner ear. It accomplishes this feat by quickly transferring the power from the fluid medium to a unique gel-like material. The material consists of a liquid crystal that is largely structure free at the particle size associated with audible frequencies. In the process of this transfer, the method of signal propagation is converted from the conduction mode to a special mode first investigated by Lord Rayleigh, the surface wave mode of power transmission. This mode provides a means of power transmission at a velocity of less than one percent of that associated with the

[76]Nuttall, A. & Dolan, D. (1996) Steady-state sinusoidal velocity responses of the basilar membrane in guinea pig *J Acoust Soc Am* vol. 99(3), pp 1556-1565

conduction of acoustic power through solids or liquids.

The liquid crystalline substrate described above is found within the cochlear partition but it is not associated with the basilar membrane. It is found on the under side of the tectorial membrane facing the reticular membrane and the remainder of the cochlear partition (including the basilar membrane). It coats the surface of the tectorial membrane extending between the two classes of sensory neurons (the IHC and multiple rows of OHC) and slightly beyond in two-dimensions.

The operation of the inner ear cannot be understood without recognizing the critical application of "Rayleigh" surface wave transmission, in two-dimensions, to the liquid crystalline substrate within the tectorial membrane of the cochlea.

The motions of the oval window discussed in the previous section constitute the input stimulation to the inner ear. In normal operation, the stimulus is a broadband signal with many features of white noise. The information content of the stimulus can be associated with how it differs from white noise.

The requirement on the inner ear is that it subdivide both the temporal (where the variations with time constitute information) and frequency spectra associated with the potential stimulus into a large group of specific signaling channels in preparation for extracting important ques from these signals. It is these ques that are processed further into interps and finally precepts that can be perceived by the cognitive elements of the neural system. A secondary requirement is that it perform this segregation in a highly reliable manner compatible with the life expectance of the subject. To this end, a degree of redundancy is called for.

To achieve the desired performance, the inner ear first slows the acoustic stimulus to a propagation velocity compatible with the dimensions of the cochlea. It then separates the energy of the stimulus, both the transient and frequency components of the stimulus (in parallel), into multiple individual channels.

With the stimulus energy propagating at a low velocity within the liquid crystalline substrate of the tectorial membrane, this is achieved by distributing the energy as a function of frequency, not with tuned circuits but by dispersion (as in an optical prism). The dispersed energy, and that continuing along Hensen's Stripe is then applied to the sensory neurons as shown in figure xxx and discussed further in Section xxx. [xxx below, repeat figure xxx from Chapter 4.]

9.4.3.3.1 The transfer of power to the liquid crystalline substrate

As discussed in Section xxx, little is known about the impedance environment associated with the acoustic energy transferred into the vestibule shared between the vestibular system and the auditory system. Two basic principles can be assumed; most of the energy turns in the direction of the fluid environment of the cochlea and the acoustic energy cannot propagate within the perilymph of the vestibular scala in the presence of a lower density (endolymph) within the cochlear duct. Most of the energy associated with the stimulus will migrate quickly to the endolymph. However, the endolymph is in intimate contact with an even lower density material, the liquid crystalline substrate of the tectorial membrane. This is where the energy finally migrates. In the process, it is converted into a surface acoustic wave moving at a forward velocity of about three meters/second (0.1 percent of the velocity of the original acoustic energy propagating by bulk conduction wave).

> The velocity of the surface acoustic wave within this substrate causes the 35 mm length of the human cochlear partition to appear to be about 35 meters long relative to the acoustic wavelengths in water. Conversion to a surface acoustic wave is an absolutely critical mechanism

in the operation of the auditory system.

Even within this substrate, geometry plays a major role, the energy accumulates within the region known morphologically as Hensen's Stripe. This very narrow region, about xxx microns by xxx microns supports the propagation of all of the energy associated with the original stimulus. Fortunately, the stimulus frequently looks like "colored noise" as described above. The individual frequency components add in quadrature and are poorly correlated in phase. These features keep the RMS amplitude of the surface acoustic wave at a manageable level within the Organ of Corti.

9.4.3.3.2 The dispersion of the stimulus energy into discrete channels

While traveling along Hensen's Stripe, the energy encounters the individual inner hair cells. The inner hair cells exhibit a unique arrangement of cilia protruding into the liquid crystalline region of Hensen's Stripe. Each inner hair cell projects up to 100 protein based cilia into this space, along with one longer, plasma filled cilia (the kinocilia). This single kinocilia actually contains a simple array of plasma filled microtubules. The protein based cilia sense the motion of the liquid crystal due to the propagation of the surface acoustic wave. This motion is transferred to the cuticular plate of the sensory neuron as discussed in Section xxx. Simultaneously, the kinocilia, which penetrates deeper into Hensen's Stripe, in conjunction with the kinocilia on adjacent IHC forms a dispersion device that redirects some of the energy in the stream toward the OHC located perpendicular to the IHC. This redirection is frequency selective. The result is the concentration of the energy in the original energy stream in the vicinity of different OHC on a frequency selective basis. [xxx repeat figure from xxx]

The frequency selective redirection of the available energy, combined with the spatial focusing effect of the dispersion mechanism can be quite effective in creating a high amplitude wave at the location of the OHC. It is this high amplitude wave that is sensed by the stereocilia of the OHC. It is this high amplitude wave that is the source of the signal measured by Nuttall & Dolan near the site of the OHC.

9.4.3.4 Verifying the model of the inner ear

It is critically important that the performance of the Organ of Corti and cochlea presented here is verified to the satisfaction of the hearing community. To this end, the author is working with a number of PhD and post-Doc students around the world, and the on-going Mechanics of Hearing (MoH) Workshop, to develop and implement such demonstrations and verifications.

The figures of **Section 9.2.1** and those referenced there in **Chapter 4** are the subject of this section.

Michael Rapson is currently involved in finite element modeling (FEM) of a generic cochlea[77]. His thoughts are presented here (20 August, 2011) only as part of an ongoing discussion.
.

I think that implementing your model could be broken into 3 stages, each of which would give valuable insights into the system:

1) Use the wave guide equations from your 2008 MoH paper to model the tectorial

[77]Rapson, M. (2011) personal communications

membrane. Input to the tectorial membrane is collected from the fluid / tectorial membrane boundary in the vicinity of the 'launcher' (details available in chapter 4, Sections 4.3-4.6, http://neuronresearch.net/hearing/pdf/4Physiology.pdf and the animations at http://neuronresearch.net/hearing/files/animation.htm) and output from the tectorial membrane is coupled into a set of radial segment models representing the micromechanics of the organ of Corti complex, where the coupling is primarily one directional. The organ of Corti complex interacts with the neighboring fluid. You obviously disagree with the parameters generally selected for the organ of Corti complex tuning, so we would tune it according to parameters you suggest. It would be interesting to see whether this model shows realistic basilar membrane tuning and possibly a 'traveling wave' as an epiphenomenon, because the basilar membrane tuning is one of the key physiological measurement that can be made.

2) From model (1) replace the analytic tectorial membrane model with a numerical model using FEM and fixed boundary conditions that support the required waves. The input to and from the tectorial membrane model would be as before.

3) Incorporate two way coupling between all the elements in model (2) this is essentially the final model you are interested in.

Jumping straight to (3) would require a long project and be fairly risky. Breaking it into these three steps offers a process that allows us to move towards the full model with intermediate (hopefully publishable) results.

9.4.3X Cumulative performance between nodes including Stages B, C and 1 NEW

[xxx add other later data]
In 1970, Dallos & Cheatham presented good data on the delays in the Organ of Corti of the guinea pig[78]. They also provided an interpretation of the earlier work in the literature. This discussion summarized the prior conflicting measurements. They did not provide a model nor did they discriminate between the acoustic delays and the electrical delays within the Cochlea. After exciting the tympanic membrane, they measured the CM at the round window[79]. [xxx check window location] To insure clean a CM, it is important to use a single-frequency stimulus and a low frequency electrical filter in the test configuration.

Their conclusion is directly applicable to the model presented here if basilar membrane is replaced by the gel-surface of the tectorial membrane. ". . .within the range of frequencies where CM at least approximately reflects the vibration pattern of the basilar membrane, measurements of microphonics indicate that the cochlea can be treated as a nondispersive medium, that is, one in which different frequency components propagate with the same velocity." The middle clause of this quotation asserts that the phase velocity is independent of frequency and is numerically equal to the resulting group velocity of the signal energy.

---

[78]Dallos, P. & Cheatham, M. (1971) Travel time in the cochlea and its determination from ccochlear-microphonic data *J Acoust Soc Am* vol. 49(4), pt 2, pp 1140-1143
[79]Dallos, P. Schoeny, Z. & Cheatham, M. (1971) On the limitations of cochlear microphonic measurements J Acoust Soc Am vol. 49(4), pt 2, pp 1144-1154

9.4.4 Data summarizing the auditory performance between nodes A and E

Palmer & Russell have provided one of the rare papers describing the generator waveforms at the pedicle of the sensory neurons in response to auditory stimulation. They also provided a summary of the relevant literature. Their goal was to understand the role of phase-locking in hearing. In this paper, they "addressed the question of whether the time constant of the hair-cell membrane is indeed the limiting factor in phase-locking."

9.4.4.X Stage 1 & 2 (signal generation and processing within the cochlea)

Stage 1 is defined as the first stage of the neural portion of the auditory system (as similarly defined in the visual system). It involves the conversion of the sensed mechanical energy into an equivalent electrical signal followed by the tailoring the characteristics of the sensed signal to insure compatibility with the remainder of the neural system. This tailoring involves both dynamic adjustment of the amplitude of the signal, adaptation, and shifting the DC amplitude of the signal and its impedance in a distribution amplifier.

In the framework of this work, little analog Stage 2 signal processing occurs in the auditory system of humans and most other chordates. Therefore, this subject will be discussed below as part of this section.

The details of Stage 1 signal generation activity has been discussed in Chapter xxx. The generation of the neural signal is intimately associated with the stressing of the cuticular plate of the sensory neuron due to the axial motion of the stereocilia. The microtubule structures associated with this plate exhibit piezoelectricity. They generate a voltage in response to a mechanical stress. It is this piezoelectricity that constitutes the initial electrical signal within each sensory neuron of the auditory system. Each sensory neuron also contains two distinct electrolytic semiconductor amplifiers (working in tandem) to accomplish the tailoring of the signal discussed above. [xxx consider a duplicate figure here]

The analog electrical signal that appears at the pedicle of each sensory neuron is called a generator waveform. It can be described in detail using the phonoexcitation/de-excitation mechanism described in Section xxx. This description, and the above mechanisms underlying it, are completely deterministic. No probabilistic mechanisms are involved or required to explain the creation of the generator waveform (as suggested widely in the prior literature).

As noted above, reliability plays a major role in the requirement placed on the Stage 1 portion of the neural system. To this end, the signal processing accomplished in Stage 2 centers around increasing the reliability of both the frequency and temporal portions of the signaling system. This is accomplished in the frequency channels by summing the signals from multiple rows of the OHC (arranged in echelon and possible arranged side by side, Section xxx).

> It is interesting to note that the individual OHC in the first of the multiple rows closest to the IHC exhibits a narrower frequency spectra, based on its energy capture cross-section, than usually measured at the output of a single stage 3 neuron in the auditory nerve. This fact, and the unevenness in the broad peak of each stage 3 neurons frequency response, is due largely to this mechanism for providing redundancy.

Stage 2 signal processing (and Stage 4 signal manipulation to be discussed later) is carried out

using neurons containing an Activa in a connexus (individual circuit) arranged as shown in Figure xxx. This configuration can also be converted to an oscillator for use in Stage 3 neural circuits.

Stage 2 signal processing is also carried out in the input circuitry of the ganglion cells who are primarily responsible for the encoding of analog signals into phasic signals as part of Stage 3. This signal processing usually takes one of three forms;

1. the summing of signals generated in separate Stage 1 or Stage 2 neurons,
2. the introduction of a lead-lag network (a type of frequency selective bandpass filter), or

3. the introduction of a threshold level to aid in the suppression of internal noise.

These capabilities are discussed more fully in the accompanying reference (PBV, Chapter xxx).

All of the Stage 1 signal generation and Stage 2 signal processing is accomplished in the analog signal domain.

9.4.4.1 Electrical measurements comparing measurements at node E and node W

Kiang has provided an interesting figure comparing several CCMs recorded at node W with PST histograms recorded at node E for units with various CFs[80].

9.4.5 Data summarizing the auditory performance between nodes A and F

Evans has provided excellent and extensive data on the performance of the guinea pig ear between a modified node A and node F[81]. He removed a large part of the pinna prior to his experiments. He defined bandwidths in the tonal channels much narrower than those predicted by the basilar membrane resonance and basilar membrane traveling wave theories of hearing. Evans also noted extremely high slopes in his observed pulse counts at node F that he attributed entirely to the frequency response of the circuit (values as high as 1000 dB per octave). This chapter will present a far different interpretation of the phenomenon he observed, related to both the frequency response of the acoustic monochrometer channels, the frequency response of the sensory neurons (under some conditions) and the potential thresholding associated with the input circuit biasing of stage 3 neurons. The data he attributes to latency can be divided into latency related to acoustic signal propagation (primarily within the Organ of Corti) and latency associated with neural circuit delays (potentially including fixed delays associated with one or more Nodes of Ranvier). As noted in **Chapter 1**, Evan's asserted his data falsified theories of hearing based on basilar membrane resonance or basilar membrane traveling waves. His remarks only considered such theories based on unidirectional energy propagation.

9.4.5.1 Math models summarizing the performance between nodes B-to-F

---

[80]Kiang, N. (1965) Discharge Patterns of Single Fibers in the Cat's Auditory Nerve. Cambridge, MA: The MIT Press Monograph #35 pg 31

[81]Evans, E. (1972) The frequency response and other properties of single fibres in the guinea-pig cochlear nerve *J Physiol* vol. 226, pp 263-287

Choi, et. al. have explored a simple mathematical model of the middle and inner ears[82]. It employs a third order equation to describe a nonlinear stage following a simple linear bandpass stage. It does not include any delay terms. Lacking a model of the underlying functional elements, Choi, et. al. have arbitrarily divided their model, which extends from node B to node F, into only a single linear and a single third order stages. No significant results were obtained.

9.4.5.2 Other functional models summarizing the performance between nodes B-to-F

While not defining the scope of their analysis clearly, the analysis of Geisler & Cai is a very important contribution[83]. It does not contribute any new empirical data but assembles concepts and data from many previous investigations, particularly those of Rhode[84].

9.4.5.3 An improved functional model summarizing the performance between nodes B-to-F

9.4.5.4 Performance of the adaptation amplifier derived from cumulative data

This section will address only adaptation at the circuit level. Overall adaptation performance will be addressed in **Section 9.6.4**.

Rhode provided excellent data on the performance of the adaptation amplifier in his 1978 paper.

9.4.5.5 The data of Rose & Weiss (cumulative bandpass)

[xxx expand on their work section considerably]
Rose & Weiss have assembled a large amount of data regarding the cumulative performance of the auditory system and presented it in three papers published simultaneously[85,86,87]. The data applies between the generator potential and the action potential pulse stream in a variety of species without definitive determination of the signal paths involved. They have also accumulated the available data across several species and drawn conclusions.

They have used some terminology that is unconventional, speaking of the cochlea of a lizard, as opposed to its papilla, as an example.

Interpretation of their work can be greatly improved if two modifications are made to the

---

[82]Choi, C-H, Chertoff, M. & Yi, X. (2002) Characterizing cochlear mechanoelectric transduction with a nonlinear system identification technique *J Acoust Soc Am* vol. 112(6), pp 2898-2909
[83]Geisler, C. & Cai, Y. (1996) Relationships between frequency-tuning and spatial-tuning curves in the mammalian cochlea *J Acoust Soc Am* vol. 99(3), pp 1550-1555
[84]Rhode, W. (1978) Some observations on cochlear mechanics *J Acoust Soc Am* vol. 64(1), pp 158-176
[85]Rose, C. & Weiss, T. (1988) Frequency dependence of synchronization of cochlear nerve fibers in the alligator lizard *Hear Res* vol 33, pp 151-166
[86]Weiss, T. & Rose, C. (1988) Stages of degradation of timing information in the cochlea *Hear Res* vol 33, pp 167-174
[87]Weiss, T. & Rose, C. (1988) A comparison of synchronization filters in different auditory receptor organs *Hear Res* vol 33, pp 175-179

fundamental framework they used. First, their work relies upon the term synchronization that is not accepted in this work, because it suggests a non-existent relationship (**Section 9.1.7**). While empiricists observe what they call synchronism between two measured functions, the typical case in hearing is that the second function observed is actually driven by the first function. Any phase error between the two functions is caused by a deterministic process, and not the error associated with synchronization introduced by a servo mechanism. Rose & Weiss recognize on page 152 that their synchronization filter is in fact a low pass filter speaking phenomenologically. Second, their data makes the common assumption that the action potential pulse streams are frequency modulated. As a result, their histograms are distorted (Section xxx). By replotting their histograms as suggested in the referenced section, their results portray a more accurate relationship between the action potential pulse streams and the generator potential generating them.

While lacking a specific model, they have presented considerable valuable data and drawn many important conclusions. Additional conclusions can be drawn by adopting the model of this work. In some cases, the histograms need to be replotted (**Sections 7.1.3**) to bring the precise meaning of the data into clearer perspective.

Figure 9.4.5-1from the first (?mistitled) paper by Rose & Weiss shows the signal analysis framework they adopted. It should be clear that the lower frame is generated by circuitry forming a direct path between the stimulus and the point of data recording. No synchronization of a second signal source is involved. This figure does not show the time delay between stimulation and response as clearly as it appears in the figure below. They defined a time interval T_{ss} that omitted the transient response within the more comprehensive interval, T_{on}. Thus, much of their analysis overlooks the significance of the transient elements in their data (or treats it awkwardly). They did recognize all three of the above features in their text associated with the figure. One feature of this figure is that a transient overshoot was recorded at a stimulus level of only 80 dB SPL. The sinusoidal form of the stimulus is well represented in the histogram if it is replotted using the scale defined in **Section 7.1.3**.

Figures 1 and 2 of their paper show the statistical variation to be expected in their data.

They did not spend much time relating the signals from their hair cells to the morphological location of those cells. The discussion accompanying their figure 8 sheds light on the difference between what they describe as the tectorial hair cells and the free standing hair cells in this animal. The definition was not morphological. It was performance related. Their definition (page 156) hinged on whether their rate based determination of the characteristic frequency was above or below 0.9 kHz. Hair cells were defined as tectorial if their rate-based

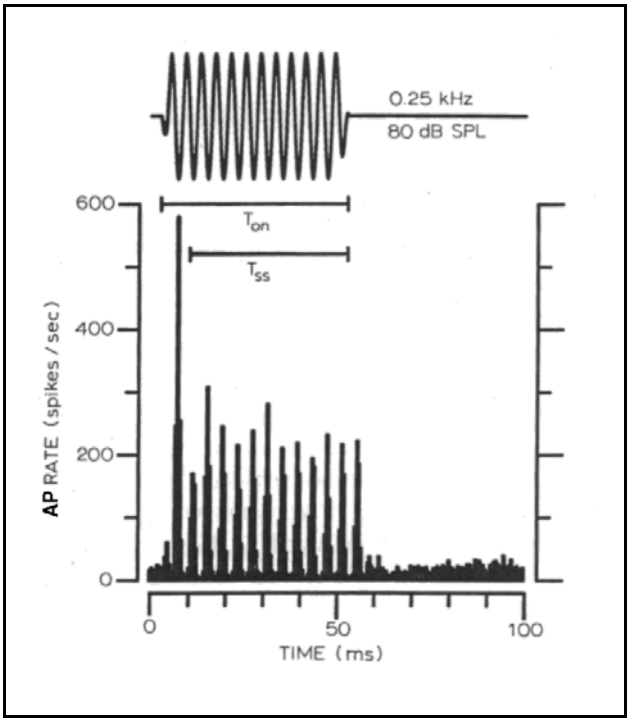


Figure 9.4.5-1 Instantaneous discharge rate of a free-standing cochlear nerve fiber in response to the specified tone burst. From Rose & Weiss, 1988.

characteristic frequency was below 0.9 kHz. All other hair cells were defined as free-standing. These definitions have not been encountered in the literature previously by this investigator.

Figure 9.4.5-2 shows a more comprehensive set of data from their second paper. Many features of the auditory system can be discussed based on this figure. The generator potentials at the hair cells clearly show the limited high frequency response of the signaling channel up to that point and the direct coupled character of the signaling channel. Note both a clear AC component and a significant DC component.

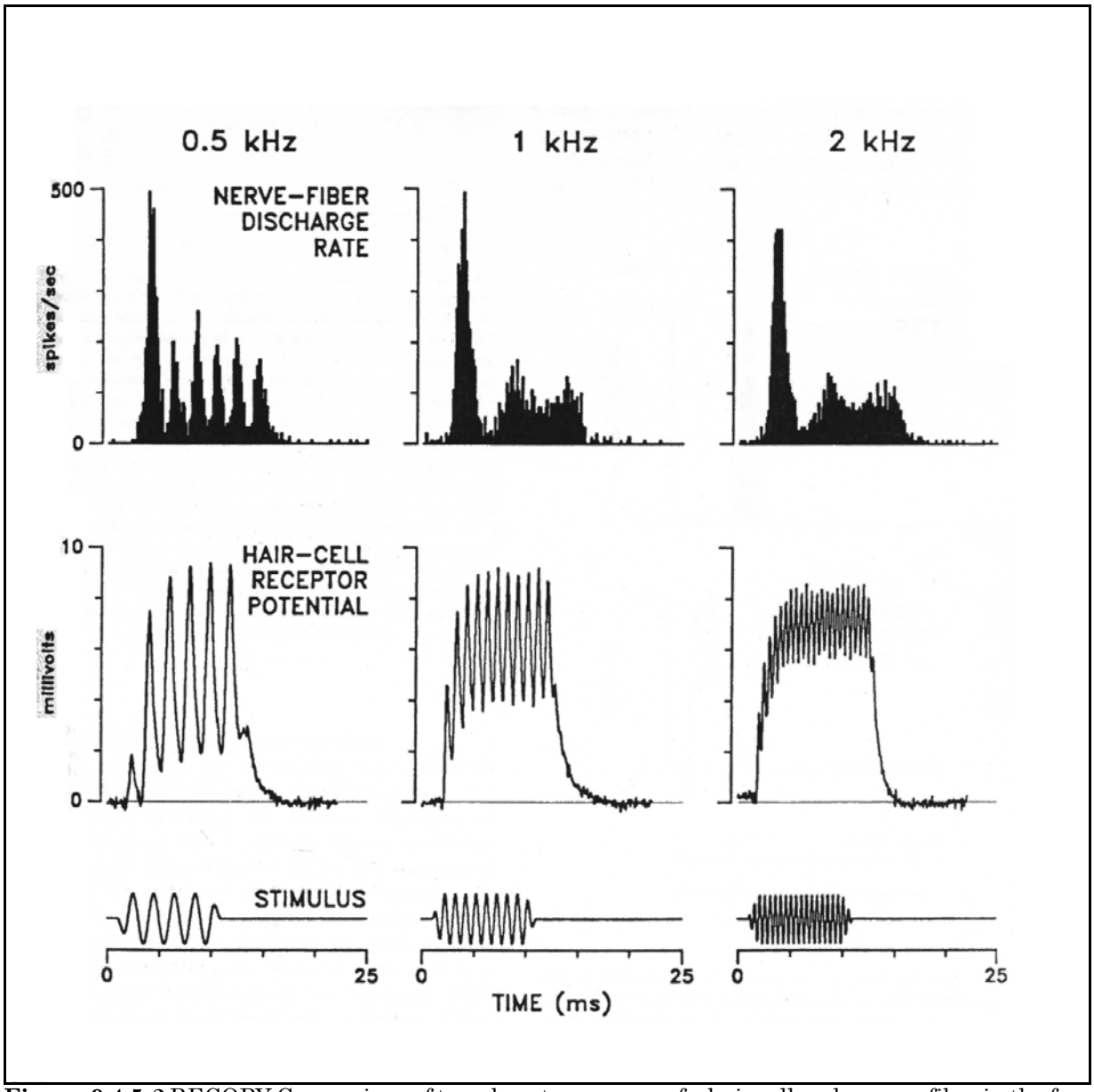


Figure 9.4.5-2 RECOPY Comparison of tone-burst responses of a hair cell and a nerve fiber in the free standing region of the alligator lizard cochlea. From Weiss & Rose, 1988.

The figure also shows that the action potential pulse trains are formed following a filter exhibiting a lead-lag characteristic. This filter introduces a significant boost in the initial response at the output of the ganglion cells. The upper left frame also shows that the pulse discharges are caused

by the individual peaks in the generator potential of the hair cells. This phenomenon does not involve a synchronizing mechanism, it is a simple response to excitation of the ganglion cell modulators to the generator potential (with a lead-lag network interposed between them). The upper frames of the figure also show that the overall bandpass of the circuits between the hair cells and the ganglion cells exhibits a low pass characteristic that they measured repeatedly as having a roll-off near 0.35 kHz. The result has been discussed elsewhere. The signals within the neural system representing frequencies significantly above 0.35 kHz (for this species) are abstract representations that do not contain any signature at the stimulus frequency. Weiss & Rose discussed the roll-off frequencies in different species in their second (? mistitled) paper. They documented that the roll-off frequency is highly temperature dependent (although the mathematical relationship they used to express that relationship is highly questionable).

Weiss & Rose (page 169) also note the nominal delay of two milliseconds between the generator potentials and the onset of the nerve discharges in their lizards (of undefined size). This delay would suggest a physical distance of not over 0.5 mm between the hair cells and the ganglion cells in these animals.

By treating the pulse streams as generated by pulse interval modulation of the ganglion cells in response to the DC level generated by the hair cells, the operation of this portion of the auditory system can be more easily understood. In this case, at frequencies considerably above the transition frequency the pulse streams only represent the intensity of the stimulus. The specific neural path identifies the frequency of the stimulus. At frequencies below or near the transition frequency, the pulse modulation combines the intensity of the stimulus with a signature frequency of the stimulus. It appears this signature is used in the source location portion of the auditory modality but not in the information processing portion of the higher brainstem and cortex. Rose & Weiss briefly assert this usage in the source location process on page 163. Their description of the usage of the tectorial and free-standing hair cells (page 163) is not sufficiently developed to consider in the framework of this work.

Figure 9.4.5-3 from Baden-Kristensen & Weiss (reproduced in Weiss & Rose, page 171) provides additional useful information. Unfortunately, the two upper frames were not obtained from the same signaling path. The ringing shown in the hair cell receptor potential is usually associated with the mechanical elements of the peripheral ear. The ringing is not resolved in the ganglion cell output. The delay between the ganglion cell output and the hair cell output is due to both the physical distance between the hair cells and the ganglion cells but also in this case due to the different location of the signal sources along Hensen's stripe.

Weiss & Rose attempt to invoke the chemical theory of the synapse in describing their data but note that little definitive information is available about this concept (page 173). They necessarily truncate their analysis in this area. Their conclusion on page 174 are necessarily brief. Without a complete model of the signaling system, it is not reasonable to make assertions about degradation of various signals.

The final Weiss & Rose paper in this set make several useful observations concerning the rapid roll-off of the frequency spectrum of various auditory channels, including the effect of temperature on this characteristic. They fit a regression line to the composite characteristic obtained by overlaying the characteristics from a number of species. The result was a slope of -106 dB/decade at a point near 10% response. They did not note that this curve was continuing to increase in

steepness with frequency, as shown in this work. Based on this and other similar values they obtained, they predict the order of the filter-functions they describe is between four and six. This is obviously based on simple RC filter theory and does not reflect the actual "double exponential" attenuation characteristic found in the frequency discriminator of biological hearing. They do note the significant change in transition frequency as a function of temperature in exothermic animals.

9.4.5.5 The data of Henry & Lewis (differencing channels)

Henry & Lewis obtained data at the cochlear nerve of anesthetized gerbils[88]. Their results did not conform to their postulated linear operation of the auditory system between the acoustic stimulus and the recorded pulse trains when using a trapezoidal pulse stimulus.

Their results clearly indicate that some of their observed pulse trains were related to differencing circuits within the spiral ganglia. When the stimuli tone was above the CF of the neuron under test, their PSTH recordings showed frequent transient responses related to the envelope of their stimulus but no sustained change during the stimulus (figure 2) or complex changes in the pulse rate during the sustained portion of the stimulus (figures 3 and 4). The output pulse rate in the PSTH recordings in their figure 5 decreased significantly with the lengthening of their pulse envelop and showed a longer post pulse recovery period.

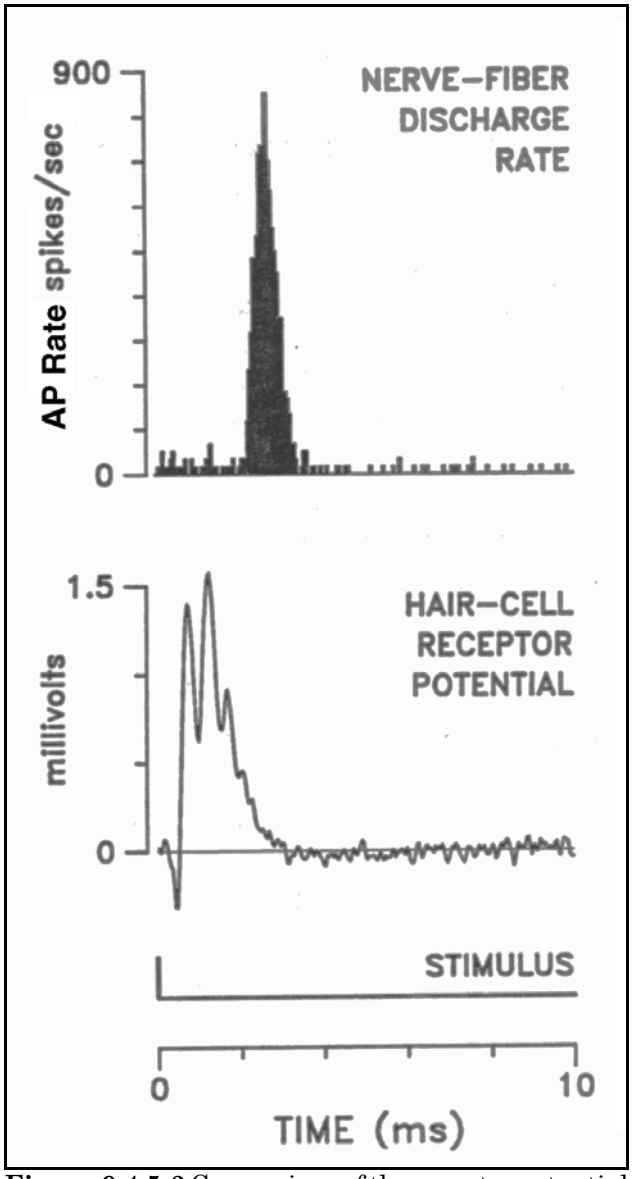


Figure 9.4.5-3 Comparison of the receptor potential and the discharge profile in response to an impulse. The CF was 2.2 kHz for the hair cell and 1.9 kHz for the nerve fiber. The levels of the two stimuli were comparable (about 6 μPa-s). From Baden-Kristensen & Weiss, 1983.

Their initial finding was that the neural system was not operating linearly over the amplitude range and onset and offset parameters they explored. They did not draw more substantial conclusions. Their data suggests significant Stage 2 signal processing prior to the encoding of the signals they recorded by the ganglion neurons of the spiral ganglia. In many cases, it appears the circuits accessed may be related to the alarm mode signaling channels (providing

[88]Henry, K. & Lewis, E. (1986) Cochlear nonlinearities implied by the differences between transient onsets and offsets to a tone burst *In* Wilson, J. & Kemp, D. *ed*. Cochlear Mechanisms. NY: Plenum Press pp 251-257

initial responses related to a change in the environment) rather than the tonal channels they assumed. A re-plot of their histograms with a revised vertical scale might provide more definitive data.

9.4.5.6 The data of Pfeiffer & Kim(two different populations) BRIEF

[xxx in measurements over 5 years and 1258 response patterns from 907 fibers, they developed two distinct classes of fibers in the cochlear nerve of cat based on clicks. All fibers of the first class had characteristic frequencies below 390 Hz. All fibers of the second class had characteristic frequencies above 380 Hz.] Other than that, their primary characteristics were related to compound histograms. Their class 2 fibers have histograms suggestive of signal processing within the spiral ganglia. Probably better answers if they used the new ordinate scale in their histograms.

9.4.6 Cumulative performance at individual nodes within the BROADBAND channels of Stages 1 through 4 NEW NEEDS RESTRUCTURING

Palmer & Russell have provided valuable performance data concerning the broadband (IHC related) channels in a variety of species[89]. However, careful study of their introductory remarks are warranted. In attempting to establish continuity between their work and the previous literature, they have adopted the paradigm of their time and introduced potentially misleading information into this section. As an example, in the first paragraph they reference multiple authors as supporting the contention "Because action potentials are elicited by unidirectional movements of the basilar membrane." The proper interpretation of their figure 9 falsifies this contention. The generator potential at the pedicle of both the IHC and OHC show a clearly bidirectional movement of the acoustic stimulus normally associated with the basilar membrane. In fact, this bidirectional stimulus originates with the gel-surface of the tectorial membrane. The clearly bidirectional movements of the basilar membrane are actually residual with respect to the actual mechanisms of hearing. On the other hand, the action potential data presented did not recognize the information encoding technique used transmit the signals using action potentials. Their results only confirm that all action potentials are positive going at the point of origination.

The recorded IHC generator waveforms in Palmer & Russell (1986) exhibit significant rise and fall times that they do not discuss. It appears likely that they used the same acoustic stimulus source as Smolders & Klinke (1986). They describe their stimulus as trapezoidal with rise and fall times of five milliseconds.

9.4.6.1 The importance of transients

[xxx page 174 in Moore, 1977]

9.4.6.2 The characteristics of the Stage 1 IHC generator potential

[xxx Palmer & Russell]

9.4.6.3 Phase-locking at low signal frequencies among IHC neurons

---

[89]Palmer, A. & Russell, I. (1986) Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential in inner hair-cells *Hear Res* vol. 24, pp 1-15

Figure 9.4.6-1 describes the electrical performance of Stages 1, 2 & 3 of the broadband channel of hearing.

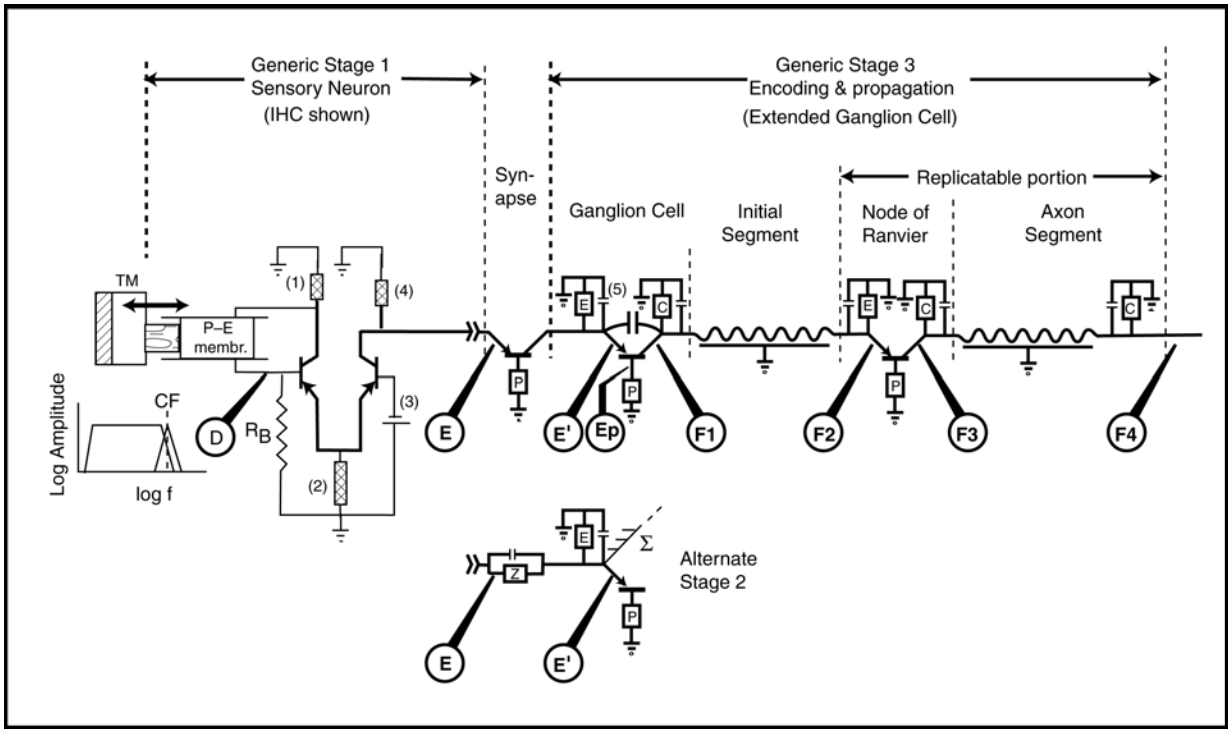


Figure 9.4.6-1 The operation of neural Stages 1, 2 & 3 broadband circuit of hearing, omitting the decoding function, shown schematically. The operation of acoustic Stages B & C, and inner ear are shown parametrically.

9.4.6.4 Cumulative time delays within the broadband channels

[xxx Smolders & Klinke, 1986]
See material in chap 5 (5.2.5.2) on P/D delay and data in 4 on delay due to SAW velocity.

9.4.7 Cumulative performance at individual nodes within the TONAL channels of Stages 1 through 4 to formants NEW

[xxx what channel does the following apply to]
[xxx Sachs provided useful histograms on the response of individual neurons within the auditory nerve to more complex, specific, sound formants (more complex stimuli)[90]. However, he did not discuss the purpose of the individual signals reported. He does provide a variety of references that have been active in the debate over the functional characteristics of the circuits involved.]

9.4.8 Overview of cumulative performance ending within Stage 4

[90]Sachs, M. (1984) Speech encoding in the auditory nerve *In* Berlin, C. Hearing Science. Philadelphia, Pa: Taylor & Francis Chapter 8, pp 263-307

9.4.8.1 Stage 4 (information signal manipulation)

Stage 4 of the auditory system is distributed more widely within the neural system than are the other sensory subsystems. This feature is introduced primarily to support the **Alarm Mode** designed to protect the animal. The signal manipulation related to the precision Acoustic Servomechanism will be discussed in Section xxx [xxx below]. The remainder of Stage 4 is designed to begin the feature extraction activity associated with the information carrying channels of the system. This is accomplished using the two information channels labeled LOC and MOC in the figures of **Sections**
3.xxx and 4.xxx.

A variety of feature extraction engines have been defined based on anatomy and morphology. Some have been investigated electro-physiologically, but primarily to determine "traffic flow" rather than information content.

It is likely, but so far unconfirmed in the literature, that the percepts and higher level interps are forwarded (by stage 3 circuitry) over multiple parallel neural fibers within various commissure of the CNS. This makes isolation and investigation of these signals particularly difficult. No data has been uncovered in the literature that describes these signals electrophysiologically.

9.4.8.2 Signals in the cochlear nucleus of Stage 4

[xxx note when the stimulus is amplitude modulated, the sensory neurons are acting as clamp circuits (Millman) which result in envelope detector operation Pantner pg 212.
[xxx note that the direction sensitive signal paths can detect the various inflected tones of speech very efficiently.]
Mollar performed a comprehensive series of exploratory experiments reported during 1969-75 and summarized in his 1983 text[91,92]. While well documented, they did not precisely define the functional location of the neurons accessed during the experiments. Most of the data refers to signals exhibiting more complexity than found within the auditory nerve, but potentially less complex than many signals recorded within other elements of the trapezoidal body. Moller noted a number of neuron responses that were sensitive to the direction of frequency change. These are suggestive of the type needed to recognize the change in frequency at the leading edge of many vocal expressions (although the functional purpose at this location may be quite different).

His figure 2.37 of 1983 is reproduced from his 1972 article. That article makes a number of assumptions concerning the origin and manipulation of the data that is questionable. A major part of his program was to explore the result of an AM modulated stimulus. Of particular concern is his assertion that the modulation index of the signal is increased by passing through the neural system. Such a mechanism has not been identified in the man-made circuits processing AM modulated signals. Moller adopts a logarithmic notation to describe the modulation index which is awkward. The 1972 paper may have incorporated some typographical errors also. The conventional mathematics and notation of AM modulation uses a simple linear percentage. Moller also introduces the use of a Bode plot. However, he did not interpret the meaning of his Bode plot.

---

[91]Moller, A. (1983) Auditory Physiology. NY: Academic Press pp 166-176
[92]Moller, A. (1974) Responses of Units in the cochlear nucleus to sinusoidally amplitude-modulated tones *Exper Neurology* vol. 45, pp 104-117

An alternate interpretation can be applied to many of these data based on the explicit model available from this work. In the case of the AM modulated stimulus, this work suggests the OHC sensory neurons are operating as integrators (in their normal mode). This integrating mode completely obliterates any sinusoidal component of the carrier frequency at 30 kHz in favor of a pedestal component. However, the bandwidth of the integrating filter is larger than, or compatible with, the modulation frequency. As a result, the OHC circuit operates as a clamp circuit[93] that is effectively a conventional AM demodulator[94]. Under this interpretation, **Figure 9.4.8-1**shows the modulation amplitude in the recovered (clamped and therefore asymmetrical) waveform, 2A, is twice as large as in the symmetrical waveform. However, the average height of the carrier,2B, is also twice that shown in the symmetrical version. As a result, the ratio of A/B remains the same as in the original stimulus. This analysis explains the apparent increase in modulation asserted by Moller. In at least most of his experiments, it is proposed that the modulation index did not change. It should be noted that in PST histograms for basic tone frequencies significantly above 600 Hz, the histogram represents the pedestal height reflecting the amplitude of the tone. It does not exhibit any sinusoidal content, unless the tone was also amplitude modulated.

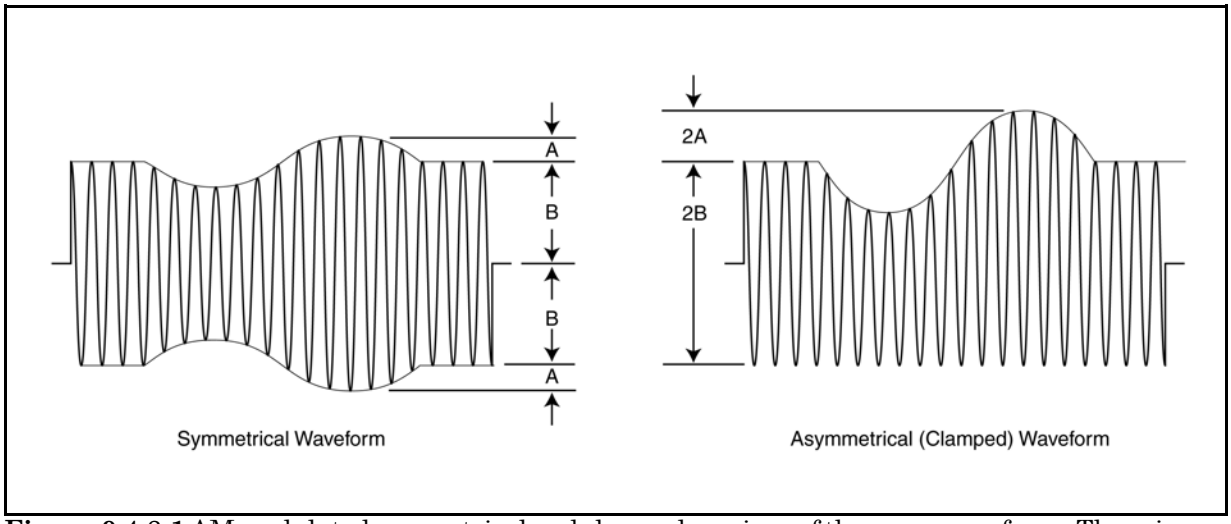


Figure 9.4.8-1 AM modulated symmetrical and clamped versions of the same waveform. The primary difference between these waveforms is the phase between the two modulating (Fourier) components. The amplitude of the Fourier components are the same. A correctly plotted cyclic (or period) electrophysiological histogram will reflect the clamped waveform correctly.

The Bode plot referenced by Moller is a powerful tool in the hands of an experienced circuit analyst. From the measured data, the form, stability and performance of the circuit under test can be readily determined[95]. The large phase shift with frequency in the circuit makes it obvious that one or more significant time delay elements are present. This time delay can be described by a phase delay that is inversely proportional to the modulation frequency. Subtracting out this delay relative to the peak frequency of the amplitude function in the Bode plot, (making the relative

[93]Millman, J. & Halkias, C. (1972) Integrated Electronics: Analog and Digital Circuits and Systems. NY: McGraw-Hill pp 114-115
[94]Pantner, J. (1965) Modulation, Noise and Spectral Analysis. NY: McGraw-Hill pg 212-218
[95]Truxal, J. ((1955) Automatic Feedback Control System Synthesis. NY: McGraw-Hill pp 549-554

delay equal to zero at peak frequency) does two things. First, it helps define the propagation delay between the stimulus location and the OHC neuron location of interest. Second, it highlights the residual phase shift associated with the integrating filter and any other electrical filters in the system. This residual delay can be compared to the rate of rise and fall of the amplitude functions in the Bode plot. The theoretical relationship is fixed. Therefore any deviation between the measured and theoretical relationships can be used to evaluate the precision of the experiment or the presence of unknown factors in the experiment.

Figure 9.4.8-2(A) reproduces Moller's figure 2.37 with a theoretical overlay as discussed here[96]. Frame **B** of the figure shows a generic block diagram of the test configuration based on this work. It includes several elements introducing a fixed delay at the acoustic frequency of 30 kHz as well as the potential location of several filter elements. The primary delays are the mechanical delay in the surface acoustic wave of the cochlear partition and the neurological delay associated with the signal propagation in Stage 3. The major filter elements are a high pass filter in the middle ear of Stage 0, the low pass filter (the integrator element) of the adaptation amplifier of Stage 1 and the low pass filter of either the Stage 3 or Stage 4.

The inset at upper right for 300 Hz has been annotated using the same symbols as Moller but showing the asymmetrical form of the stimulus waveform as it appears in a histogram. The modulation index, A/B, is the same in the histogram as in the stimulus (where B is measured from the spontaneous pulse rate). No value for the spontaneous action potential rate, SR, was given in the paper. See **Section xxx** for a further refinement in the ordinate of the histogram if precise results are desired. It was shown there that the ordinate of the conventional PST histogram is not linear. It is quite compressed at low pulse rate values. This compression distorts the height of the PST histogram relative to the baseline value of zero. The distortion can be considerable at lower pulse rates. The Bode plot allows rapid estimation of the circuit underlying the responses measured. Two dash-dot lines are shown in the figure. One slops up to the right at 20 dB/octave. This slope is the asymptotic slope for a single stage high-pass filter. The other slopes down at 40 dB/octave. This slope is the asymptote for s two-stage low-pass filter. These curves approximate the response measured by Moller for only one specimen quite well (particularly if the peak of the response is modified as shown by the dash-small-dot curve). It is safe to say the filters involved in his experiment consisted of one high-pass and two low-pass stages.

[96]Moller, A. (1972) Coding of amplitude and frequency modulated sounds in the cochlear nucleus of the rat *Acta Physiol Scand* vol. 86, pp 223-238

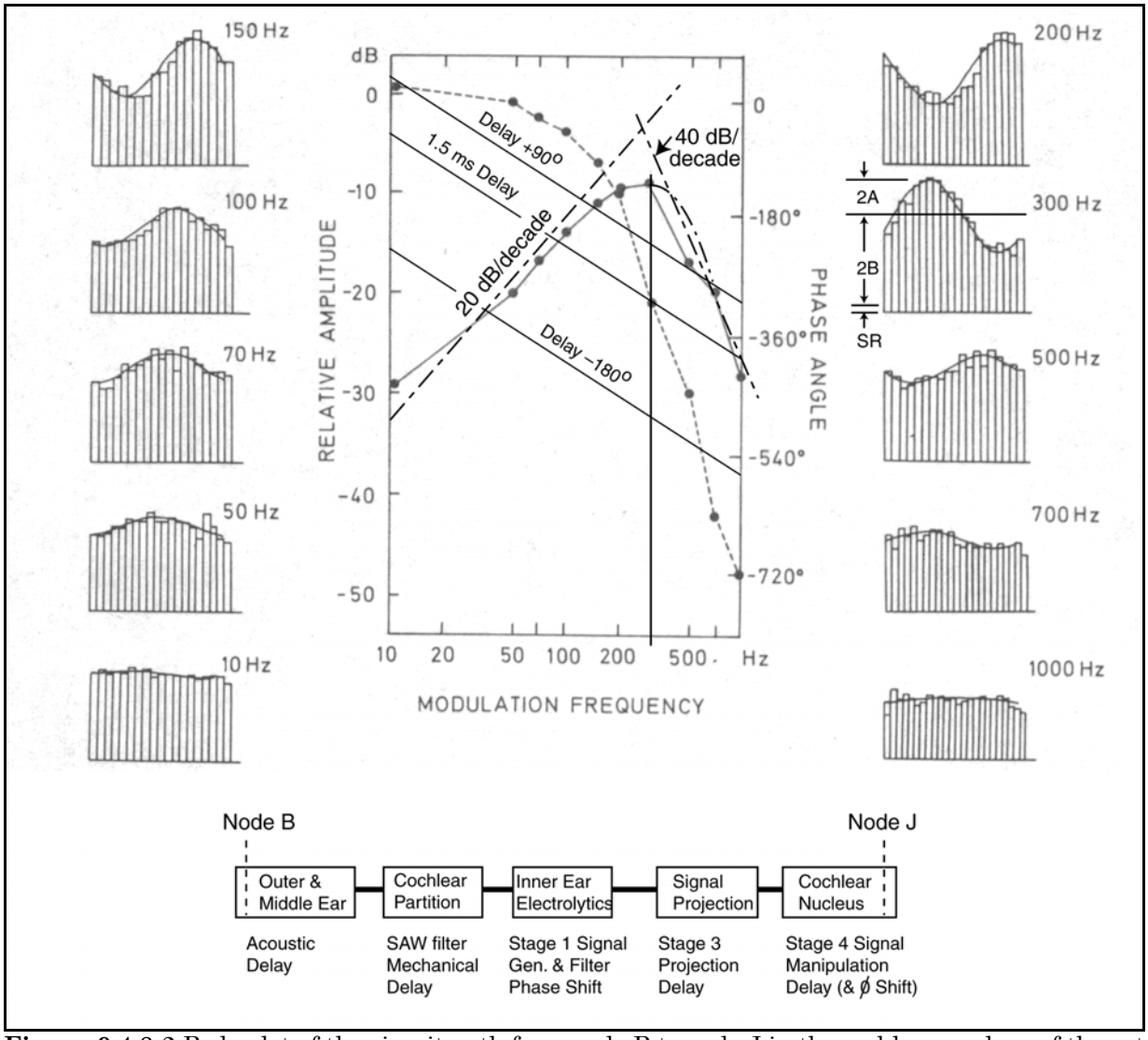


Figure 9.4.8-2 Bode plot of the circuit path from node B to node J in the cochlear nucleus of the rat based on a set of cyclic histograms and an AM modulated stimulus. A; Histograms and Bode plot from Moller overlaid with amplitude and phase shift asymptotes. See text. B; block diagram of the hearing circuit under test. Stimulus was applied at node B in the ear canal. Signals were recorded at node J in the cochlear nucleus. Modified from Moller, 1972.

Also shown in the figure are three diagonal lines going down to the right. The middle line represents a circuit delay of 1.5 ms as a nominal value for the total delay between nodes B and J in the network. This delay consists of the mechanical delay of the SAW filter plus the projection delay of Stage 3. This delay is drawn through the measured phase shift at an angle directly below the peak response. It becomes a new nominal 0 degree relative phase shift. Based on the assumed filters within the circuit, the phase response of the system should approach the nominal delay +90 phase shift line at low frequency. Similarly, the assumed filters should approach a nominal delay –180 degrees phase shift line at high frequencies.

It is clear that the relative phase shifts are too large for the filters assumed and the delay assumed. By raising the delay to about 1.8 ms, the diagonal lines are rotated clockwise and a better fit can be

obtained. However the curvature in the measured phase response at low frequency suggests more data points are needed to confirm the precision of the experiment. Based on this theory, it can be said the filter found between the outer ear canal and the cochlear nucleus of the mouse at the point probed is of the type assumed. The low pass has two poles near 600 Hz each and the high pass filter has a pole near 100 Hz. The total delay is on the order of 1.5-1.8 ms in this experiment consisting of a significant delay associated with the Stage 3 projection circuits, and possibly a small delay in the Stage 4 signal manipulation circuits.

The time delay of 1.5-1.8 ms agrees with a similar value in Kiang[97] and is in agreement with a later paper by Moller[98]. Kiang was working in the auditory nerve itself. As a consequence, it is suggested that the 1.5-1.8 ms was obtained for either the axons of Stage 3 neurons at their entry point into the cochlear nucleus or very close thereto. In the later paper, Moller noted the difference in delay in what he called primary and secondary neurons of the cochlear nucleus. The primary neurons appear to be at the input to the cochlear nucleus with the secondaries proceeding along the various signal paths of the engine toward the output area. His figure 5 shows these delays bunched around various delays extending to over 8 ms. He also indicates the shortest delays are associated with the posterior ventral cochlear nucleus (PVCN). Those with the longest delays were associated with the dorsal cochlear nucleus (DCN).

The presence of the proposed high pass filter appears consistent with the paper by Moller exploring adaptation within the system using an AM modulated stimulus[99]. This paper also addresses the effect of adaptation on the responses to AM modulated stimuli at the cochlear nucleus. The major finding is that the system provides a linear response at the cochlear nucleus for AM modulated signals of moderate modulation index in the mesotopic range (up to 40 dB above threshold). Besides supporting an integrator frequency near 600 Hz as suggested by the above Bode plot analysis, it also supports the start of internal feedback near 40 Hz (resulting in the phonotopic region above 40 dB in this case).

Finally, Moller presented a paper addressing the question of coding as well as linearity of signals within the cochlear nucleus[100]. The term coding was used in the generic sense. No attempt was made to determine the underlying coding algorithm. It would have been useful to apply Bode plot analyses to the data in these latter papers.

The author would be happy to communicate with any investigator wishing to replicate or extend this experiment and analysis or employ the Bode plot more fruitfully. While AM modulation appears very helpful in confirming the character of the integration filter and the internal feedback loop of the adaptation amplifier, this type of modulation is not normally found in speech and other natural sounds. Caution is required in expanding the results achieved to the underlying system. As an example, the total delay always contains the cochlear partition delay associated with the carrier frequency, not the modulation frequency. The propagation delay is also associated with the Stage 3 channel identified with the carrier frequency, not the modulation frequency.

---

[97]Kiang, N. Watanabe, T. Thomas, E. & Clark, L. (1965) Discharge Patterns of Single Fibers in the Cat's Auditory Nerve. Cambridge, MA: MIT Press, Research Monograph #35

[98]Moller, A. (1975) Latency of unit responses in cochlear nucleus determined in two different ways *J Neurophysiol* vol. 38, pp 812-821

[99]Moller, A. (1974) Responses of units in the cochlear nucleus to sinusoidally amplitude-modulated tones *Exp Neurology* vol 45, pp 104-117

[100]Moller, A. (1974) Coding of amplitude and frequency mmodulated sounds in the cochlear nucleus *Acustica* vol. 31, pg 292-299

9.4.8.3 Signals within other elements of the trapezoidal body of Stage 4 EMPTY

9.4.8.4 Signals in the inferior colliculus of Stage 4

Langner & Schreiner have provided a rich mine of exploratory data concerning the action potential pulse streams they located in the inferior colliculus, IC[101,102]. However, it is not clear from their text whether these apply to the input or output of the IC or whether they apply to the source location or other functions performed within the IC.

Figure 9.4.8-3 reproduces a modified figure 8 from Langner & Schreiner. For purposes of discussion, it will be assumed this waveform was recorded near the point where the signal stream first reached the IC. The stimulus is only shown symbolically. A 35 kHz sine wave will not reproduce at the scale of the figure. The estimated stimulus in this case is an analog waveform free of any sinusoidal content. The action potentials represented by the dots in the graph show no "synchronism" to the analog waveform. The estimated waveform is drawn showing a total delay from the acoustic stimulus to the point of action potential recording of about 6.5 ms. This includes the physiological delay associated with the middle ear, the delay within the cochlea and the delays within the neural system up to this point.

---

[101]Langner, G. & Schreiner, C. (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms *J Neurophysiol* vol. 60(6) pp 1799-1822

[102]Schreiner, C. & Langner, G. (1988) Periodicity coding in the inferior colliculus of the cat. II. Topographic Organization *J Neurophysiol* vol. 60(6) pp 1823-1840

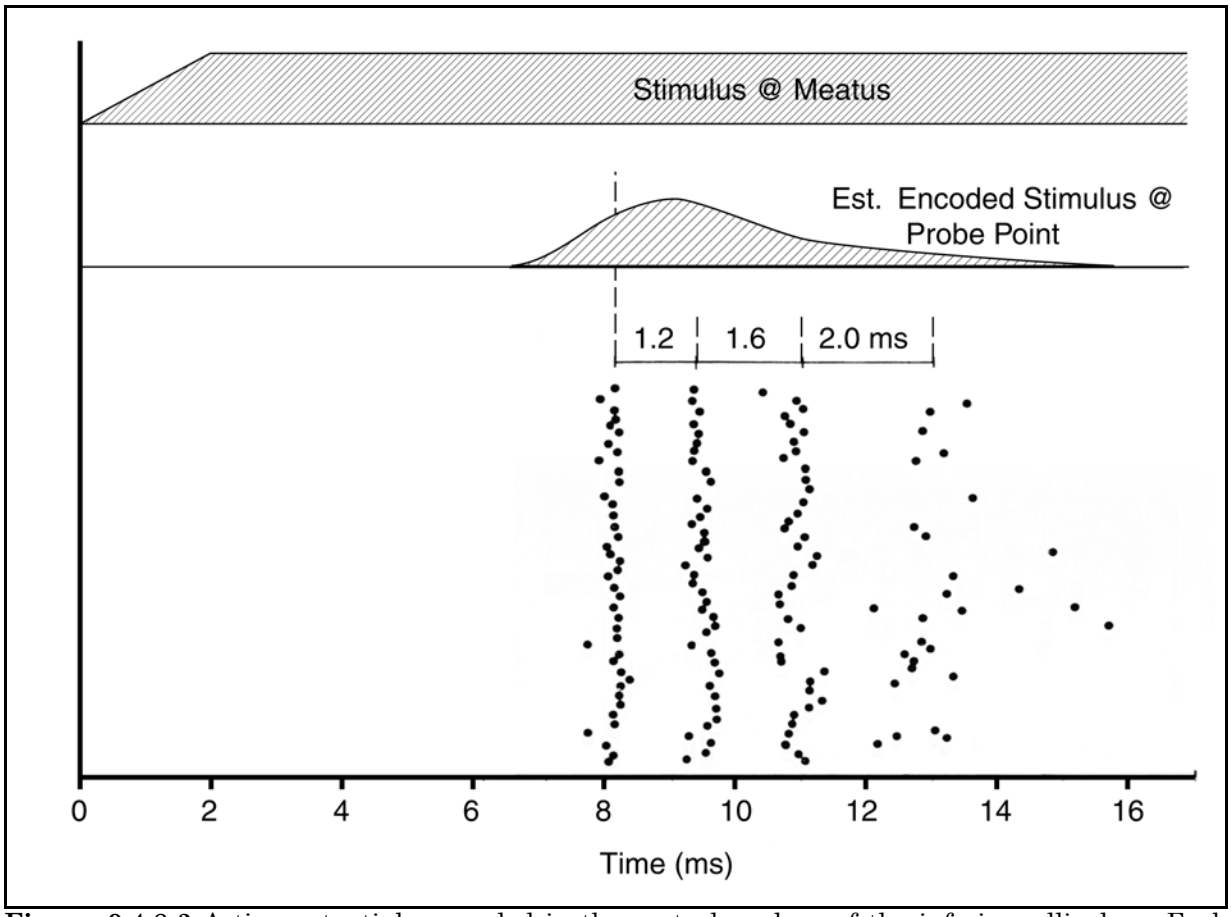


Figure 9.4.8-3 Action potentials recorded in the central nucleus of the inferior colliculus. Each horizontal row of dots is a separate recording session in response to an acoustic stimulus at 35 kHz introduced into a closed auditory canal. The pulse had a 2 ms rise to full amplitude and a 200 ms duration. See text. Data from Langner & Schreiner, 1988.

Based on the defined test stimulus and the resulting action potential streams at the central nucleus of the IC, the form of the signal (representing the envelope of the tone burst encoded at the output of the sensory neurons) can be estimated as shown in the figure. The first vertical row of dots represent the initial crossing of the ganglion threshold required to generate an action potential. The second vertical row represents the fact that the signal stayed above threshold. The third row of dots is delayed more than the second row, suggesting the signal intensity at the encoder had fallen slightly. The fourth row of dots is quite ragged. It suggests the signal that was encoded was decreasing, at a rather shallow rate, back to the baseline. As a result, the fourth pulse in each set varies considerably in time following the third pulse. Whether any other signal processing was involved in shaping the neural signal prior to encoding cannot be determined.

Note that the dots in each row that starts early, all occur early. This is suggestive of a variation in the timing of the stimulus, possibly due to the pulse gate generator keying. Note also that the fourth dot is missing in some groups, suggesting the signal at the ganglion encoder went below threshold prior to the expiration of the refractory period of the encoder Activa circuit.

In a set of experiments where the stimulus was AM modulated at 100%, distinct patterns were

obtained that showed the signals at the IC were in synchronism with the envelope of the stimulus (their figure 2). This is to be expected. The responses are in fact driven (deterministically) by the envelope of the stimulus. The fact that the recorded signals might reproduce the entire sinewave of the envelope of the stimulus is obscured because of the inappropriate vertical scale used in their histograms. The histograms as presented compressed the low pulse count portion of the histograms to the point of invisibility. Replotting them in accordance with the guidance of **Section xxx** should show a symmetrical histogram reflecting the envelope of the stimulus, as filtered by the mechanisms between the stimulus point and the recording point, properly.

Their figure 3, showing the characteristic frequency (CF) of a group of their signals, would also be more illustrative of their actual spectral bandwidth if replotted according to the guidance in **Section xxx**.

Several systematic problems are seen immediately in the graphs that should be explained before any detailed analysis is performed. The location of the first recorded pulse group varies with respect to the nominal starting time of the stimulus in an inconsistent manner suggesting the performance of the gating circuitry was not documented sufficiently (figures 2a, 2b and 2c). No pulse or time hack was presented in these figures to show the actual start or phase of the modulation as it related to the start of the tone burst or time zero.

It is clear that the first pulse group in each recording is due to the transient associated with the start of the tone burst as it is modified by the adaptation characteristic of the sensory neurons. This initial transient is not synchronized to the subsequent modulation. The height of the initial group with respect to the subsequent pulse groups, is indicative of the lack of synchronism between the start of the tone burst and the modulation.

The brief analysis presented above suggests several findings. The signals recorded by Langner & Schreiner in their figures 2 and 8 were generally recorded at the input to the IC. It also suggests that these signals were not significantly modified by any Stage 2 signal processing prior to their arrival at the IC. Finally, it confirms that the overall bandwidth of the channels recorded for single neurons within the IC varied. The variation appears to be uncorrelated to the carrier frequency of the channel but to exhibit some structure (their figure 4). The variation in bandwidth of a channel varied by over an order of magnitude at a given CF. This structure suggests that many of the signals were recorded at later points in the IC after additional signal manipulation had been performed. This appears to be confirmed by figure 6 where the total time delay recorded for a channel with a given CF varied widely (by as much as 15 ms).

The papers of Langner & Schreiner provide additional information beyond that presented here, including comments concerning periodicity pitch (**Section 9.7.5** xxx).

Langner & Schreiner speak of a proposed model in their paper. However, their model is not described explicitly or graphically. It is suggested that the models presented in this work apply specifically to the findings of Langner & Schreiner.

Repeating the experiments of Langner & Schreiner using more sophisticated signal generating equipment, and plotting the histograms using the proposed vertical scale, would provide a clearer description of the signals found at the IC. The primary change would be to insure the tone burst began at the point where the modulated signal was at zero signal level with respect to both the modulation frequency and the carrier frequency. In general, this would require two conditions. First, the modulation must be at a frequency that is a sub-harmonic of the carrier. Second, the start of the tone burst must be synchronized with the minimum in the envelope of the modulated

tone.

Moller has provided some references to other work related to the IC (1983, page 170).

9.4.8.2 Stage 4 (ancillary signal processing in the PAS)

The precision acoustic servomechanism, and its associated **alarm mode** signaling channels, are the only elements of the auditory system that are operational 24 hours per day. They are focused on threat determination using angle location techniques.

9.4.8.2.1 Overview of source location capabilities in hearing

[xxx see chapter 12]

9.2.5.2 xxx

9.4.9 Overview of cumulative performance ending within Stage 5 EMPTY

A review of the majority of the very little information in the literature concerning the cognitive functions associated with Stage 5 suggest this part of the neurological system is totally abstract, with no unique tonotopic, cochleotopic or retinotopic organization. It also suggests that the information within this stage is processed and stored in a holonomic form. This means that considerable progress can be made in localizing specific processing engines associated with a task using fMRI, and similar, techniques but that the detailed understanding of the functional processing at the circuit level will require major breakthroughs in data collection and data analysis before significant progress can be expected at that level.

Whitfield & Evans performed electrophysiological experiments on awake and active cats seeking to understand the role of the cortex as a frequency analyzer[103]. While their experiments were done with great care, their results were unable to confirm the operation of the cortex as a frequency analyzer. Their summary is useful. The complexity of the signal environment within the cortex is daunting. They found most of the neurons they investigated had properties that could not be interpreted based on steady state stimulation. They also found many other complexities they were not in a position to explain. The use of the Bode plot in conjunction with an AM modulated stimulus as an analytical tool (**Section 9.4.8 xxx**) could help considerably in the test environment Whitfield and Evans entered. However, it is likely that the signals are generally associated with the holonomic framework and individual probes will not provide understandable results. They may not even provide results traceable to a recognizable receptive field.

9.4.9.1 Stage 5 (interp and percept cognition)

As noted in Section xxx, the majority of the percepts delivered to the saliency map for accessing by the cognitive stage are probably in the form of analog information delivered over multiple parallel fibers from the appropriate Stage 3 decoding circuits. In general, none of the ensembles of channels delivering these percepts to the saliency map, or to the frontal lobe, have been isolated to date.

---

[103]Whitfield, I. & Evans, E. (xxx) Responses of auditory cortical neurons to stimuli of changing frequency xxx pp 655-672

9.5 The end-to-end performance of hearing (psychophysical data)

The psychophysical performance of hearing has often been subdivided into two categories, that associated with the information content of the stimuli (the "what") and the location of the source of the stimuli (the "where"). Rauschecker & Tian have discussed the "what" and "where" channels of hearing conceptually[104] and Smith, et. al. have recently proposed an interesting way to separate these channels psychophysically for purposes of evaluation[105]. This section will focus on the "what" aspect of hearing. **Section 9.9** will focus on the "where" aspect.

[xxx suggest this is the performance related to and measured by psychology]
Care is necessary in interpreting psychophysical data involving distinctly different sensory systems and/or the cognitive and motor response systems. Significant time delays can be introduced into the data due to the participation of these other stages and systems in forming the response to the test stimulus.

9.5.1 Describing the intensity performance of human hearing

This work has described four regimes of hearing based on stimulus intensity level, kaumotopic, mesotopic, phonotopic and hypertopic. The scotopic regime was described as largely noise dominated and involving significant cognitive activity (attention) in order to follow a conversation or other signal. The mesotopic regime was described as one that was still noise limited but performance was based on the stimulus level. The phonotopic regime was defined as the normal operating regime where adaptation maintained the performance of the hearing system as nominally independent of stimulus level. Finally, the hypertopic regime was defined as one above the stimulus level where the adaptation mechanism could control the perceived intensity level, and one leading signal level overload within the output circuits of Stage 1 (and subsequent stages) and the onset of pain. The demarcation between these regimes has been postulated as 20, 45 and 90 dB SPL.

Nelson & Schroder have recently provided data describing the input-output performance of the human hearing system[106]. The data appears in good agreement with the above definitions and may provide a protocol leading to more accurate definition of the above theoretically derived levels. **Figure 9.5.1-1** reproduces one frame of their paper. Additional construction has been added to indicate the theoretical performance based on this work. The dashed lines represent regions of linear transfer function. The solid lines represent the mesotopic regime (sloping solid line on the left) and the hypertopic regime (sloping solid line on the right). The horizontal line represents the phonotopic regime under the assumption of perfect adaptation. A slope of the phonotopic regime response represents less than perfect feedback compensation in the adaptation amplifiers of the OHC at 1000 Hz. Other frames of the Nelson & Schroder paper provide additional data on the efficiency of the compensation.

---

[104]Rauschecker, J. & Tian, B. Mechanisms and streams for processing of "what" and "where" in auditory cortex *Proc Natl Acad Sci USA* vol. 97, pp 11800-11806
[105]Smith, Z. Delgutte, B. & Oxenham, A. (2002) Chimaeric sounds reveal dichotomies in auditory perception *Nature*, vol. 416 (7 March), pp 87-90
[106]Nelson, D. & Schroder, A. (2004) Peripheral compression as a function of stimulus level and frequency region in normal-hearing listeners *J Acoust Soc Am* vol. 115(5), pp 2221-2233

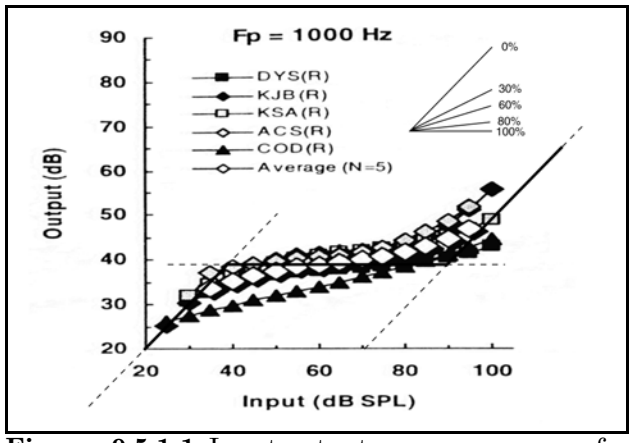


Figure 9.5.1-1 Input-output response curves for human hearing at 1000 Hz. Data was averaged along individual vertical indices. Construction lines describe the proposed theoretical performance. Solid line is the proposed input-output characteristic for a system with perfect adaptation. A tilt in the central region is indicative of the actual feedback gain achieved in the system. The dashed lines indicate a linear transfer function. The plimsoll mark at upper right indicates the slope of the transfer function for a given degree of negative feedback. Data from Nelson & Schroder, 2004.

The plimsoll mark at the upper right shows the theoretical slope of the adaptation amplifier response as a function of the percent feedback. The typical percent feedback in the adaptation amplifier is between 60% and 80% regardless of the frequency within the range explored. This degree of feedback is directly associated with the magnitude of the common emitter to ground impedance ((2) in the circuit diagram of the sensory neurons (**Section 5.2.6 xxx**).

Averaging the data points along vertical indices may not be optimal in this situation due to subject-to-subject variations. In addition, the OHC supporting operations at 1000 Hz for subject COD(R) exhibit less than nominal adaptation performance. COD(R) should be excluded from the averaging procedure. Re-averaging the data for the other 4 subjects appears to give a better fit to the theoretical response.

Nelson & Schroder have provided additional data for frequencies from 500 Hz to 6000 Hz. Some of these frames indicate the OHC of other subjects perform poorly at specific frequencies. **Table 9.5.1-1** tabulates the outliers in the responses of five subjects found in their paper. By eliminating these outliers, a re-computation of the nominal performance at each frequency will produce an improved version of their figure 11.

TABLE 9.5.1-1 xxx
Outliers in input/output performance in selected human subjects

Frequency	500	1000	2000	3000	4000	6000
Subject						
DYS(R)					xxx	
KJB(R)			xxx	xxx		
KSA(R)						
ACS(R)						
COD(R)		xxx				

9.5.2 Describing the intensity performance of hearing in other species

The intensity performance of hearing in the turtle, *Pseudemys scripta elegans*, has been studied in

detail by Crawford, Evans & Fettiplace[107]. They used the axonal patch clamp voltage technique under a variety of conditions, including varying the temperature of the specimen. The data is well represented by the P/D Equation and the expected variation in adaptation level with prior (and/or extended)stimulus. The overshoot expected in the response due to phonon-excitation/de-excitation was not observed in their figure 5 due to the slow risetime of their test stimulus. It was observed in their many other figures where stimuli with more rapidly rising leading edges were used. The data in figure 8 is particularly valuable because it contains considerable data related to the adaptation process obtained under well documented conditions.

Their equation (5) is a simplification of the P/D Equation frequently proposed from empirical evidence by investigators. However, it does not include the temperature parameter, does not account for the intrinsic latency of the circuit and does not reflect the variation in the apparent risetime of the waveforms as a function of the time constant normally associated with the rising portion of the waveform. This time constant is a function of the intensity of the stimulus, and any prior adaptation. Crawford, et. al. noted their insecurity with respect to this time constant in their text (pages 416-417).
They noted the characteristic feature predicted by the P/D Equation (besides the variation in the risetime of the attack portion of the waveforms. Both the onset and decay time constants were a strong function of temperature. They varied by a factor of fourfold for a 20 C drop in temperature (20-26 C down to 3 C) precisely as predicted by the P/D Equation.

Their discussions of chemical kinetics can be ignored if the Electrolytic Theory of the Neuron, and the P/D Equation resulting from it, are adopted as the baseline (as done here). The diode characteristic displayed in their figure 11 is also indicative of the applicability of the Electrolytic Theory. The variation in the decay time constant, measured by eye in figure 15 is the normal result predicted by the P/D Equation. The initial transient has not completely died out before the beginning of the falling portion of the shorter duration waveforms. The actual decay time constant has not changed although the character of the individual cumulative traces have.

Considerable data applying to both the visual and hearing systems of this turtle species are available and can be readily compared. {xxx give reference from other works] There is also similar data available for the saccular hairs of bullfrog[108,109].

---

[107]Crawford, A. Evans, M. & Fettiplace, R. (1989) Activation and adpatation of transducer currents in turtle hair cells *J Physiol* vol. 419, pp 405-434
[108]Assad, J. & Corey, D. (1988) Voltage sensitivity of the adaptation rate in isolated vertebrate hair cells *Biophys J* vol. 53, pp 429a
[109]Assad, J. Hacohen, N. & Corey, D. (1989) Voltage dependence of adaptation and active bundle movements in bullfrog saccular hair cells *Proc Nat Acad Sci USA* vol. 86, pp 2918-2922

9.5.3 Describing the chromatic performance of hearing in humans

Describing the chromatic performance of hearing is very involved. At least seven different mechanisms must be considered. These mechanisms do not all operate over all frequency regions. The overall system performance is not obtained by simply concatenating the response associated with each mechanism.

Furthermore, it has become quite apparent that the CNS of hearing employs two distinctly different fusion processes. The fusion process of the cerebral cortex, and possibly the diencephalon, has a time constant similar to that of vision, about 50 ms (or a fusion frequency of 20 Hz). Selected circuits in the feature extraction engines of the lower brain supporting source location have much shorter fusion time constants, on the order of 1 ms (for a fusion frequency close to 1 kHz).

In addition, it has become quite clear that the individual sensory neurons as a group act as integrators with an integration corner frequency of 600 Hz (a nominal time constant of 1.6 ms).

[xxx review and rewrite based on the material in chap 6 re: architecture & masking etc.]
As noted in Chapter 6, a large dichotomy currently exists between the data describing the critical band width of human hearing and the ability of a human to discriminate adjacent frequencies. Explaining this dichotomy, other than to say it arises because of differences in test protocols, requires careful analysis. The critical band width determination based on masking measurements has been championed by Zwicker following the introduction of the term by Fletcher in the 1930's.. The much smaller just perceivable frequency difference measurements have been gathered by a variety of investigators. It is obvious that the masking experiments do not provide the complete answer to frequency discrimination in hearing.

In the early days, a concept was developed to estimate the width of a given auditory channel observed during very crude masking experiments. It was called the critical ratio. It expressed, in dB, the level the threshold of a specific tone was raised by the presence of a given masking signal. With time, it was found that the critical ratio was a function of the bandwidth of the masking stimulus. It was natural to expand the critical ratio to reflect this dependence. The available data, although inconsistent, suggested the masking involved the root mean square of the amplitude of all of the frequency components within the masking stimulus. If such an RMS energy summation is applied to a signal of many independent frequency components (it is defined as a noise), the total power (or energy in a tone burst) can be shown to be proportional to the square root of the bandwidth involved.

Scharf provided a simple description of the critical ratio along with an elementary formulation[110]. However, his description fails to highlight the square root relationship noted above, although it does highlight the necessity of specifying the signal to noise ratio used as a reference in establishing the relative power levels. A better understanding of the critical ratio can be obtained by reviewing Fletcher[111]. The inconsistencies in the literature are due to the lack of a model. If the bandwidth of the masking stimulus does not fall completely

[110]Scharf, B. (1970) Critical bands. *In* Tobias, J. *ed*. Foundations of Modern Auditory Theory. NY: Academic Press pg 174-175
[111]Fletcher, H. (1940) Auditory patterns *Rev Mod Phys* vol. 12, pp 47-65

within one critical band of the subjects auditory system, this fact must be accounted for in the associated mathematical analysis (**Section 9.7.4 xxx**)
.

Zwicker introduced his "critical band" as a translation from the German of "frequenzgruppe" in 1957. The terms critical band and critical bandwidth have been used largely interchangeably in the literature subsequently. This is unfortunate. The term frequenzgruppe does not imply anything critical about a channel, and the term critical band (or bandwidth) doesn't either. Neither does the term critical band relate directly to the normal full-bandwidth-half-amplitude (FWHA) or -third-amplitude (FWTA) bandwidth criteria. The term critical is used only as a term of convenience. Zwicker did not offer a model of the auditory system, or any other theory, to explain the significance of his measurements.

The mechanisms related to the chromatic response of hearing to be considered here include:

1. the conductive performance of the outer and middle ears,
2. the dispersion window of the two-dimensional surface acoustic wave filter of the cochlea partition,
3. the frequency parameter of the integration circuit associated with the adaptation amplifier,
4. any frequency limiting circuits of Stage 2, including any lead-lag network,
5. the asymmetrical frequency response of the Stage 3 encoding mechanism,
6. any frequency limiting mechanism associated with the two-dimensional correlator of the TRN,
7. any interpolation (or cross correlation) performed as part of the interp formation process,
8. the fusion time constant(s) of the CNS.

As shown in Section xxx, the capture bandwidth associated with the capture area of an individual sensory neuron with respect to the SAW filter is about xxx [xxx express in parts per thousand of frequency or parts per thousand along the tectorial membrane.]

As seen in chapter 4, the capture bandwidth of an individual sensory neuron in cat appears to follow a logarithmic scale with distance along the tectorial membrane. The physical ratio corresponding to the center to center spacing of OHC cilia to the total array length would suggest a very fine capture bandwidth of about 0.5% of the center frequency of interest. This value (five Hertz at 1000 Hertz) is far narrower than any perceived critical band width from psychophysical measurements.

The energy collected by each individual stage 1 sensory neuron is treated differently if it is above or below the low-pass characteristic of the cell. Below this frequency, the sensory neuron acts as an amplifier and provides an output that tracks the (rising (compression) xxx) phase of the stimulus. At frequencies above that critical frequency, the output is the sum of the energy associated with the (rising (compression) xxx) phase sensed by each sensory neuron acting as an integrator.

Based on Chapter 6, the stage 2 signal processing sums the signals from xxx adjacent sensory neurons in order to achieve an output signal that can be used to encode the action potential stream created by the stage 3 projection neurons. This summation results in a processed bandwidth. [xxx put in glossary?] This bandwidth appears to me a constant multiple of the capture bandwidth (xxx over what frequency range).

The stage 2 signal processing may incorporate a lead-lag network in the circuit leading to the stage 3 ganglion neurons. If it does, the frequency response of this network could affect the overall

performance of the chromatic channels. [xxx omit this reference if no effect is recognized]

The stage 3 projection neurons encode the signal using phase modulation (unrelated to the phase tracking described above). At the CNS, these signals are decoded faithfully. [xxx is this true?]

The stage 4 signal manipulation is responsible for evaluating the information contained within the signals received from the PNS. As shown earlier, it appears to do this by employing a correlation process for at least some of the information [xxx explain results of stage 4 more fully or provide section reference]. The result is a signal obtained by subtracting the signal received from spectrally adjacent chromatic channels. The process is limited by a noise component present in the stage 3 decoding circuit or within the stage 4 circuits. The result is a noise limited abstract representation of the frequency of the stimulus.

It should be noted that the fusion time constant of about 50 ms is associated with the "conscious brain." The human can be made aware of this limitation in beat frequency experiments. The shorter fusion time constant associated with the lower brain is probably only of importance to the "subconscious brain." While experiments can be performed where the human reports the results of source location calculations involving time constants on the order of one millisecond, these reports involve signal manipulation, cognition and response circuits that operate at a significantly lower bandwidth and exhibit significantly longer latencies.

[xxx is the following correct?]
This section will discuss the frequency spectrum associated with each of the above stages before presenting a graph describing the overall system performance.

9.5.3.1 Describing the "characteristic bandwidths" of the fine tonal channels EMPTY

Just noticeable frequency difference experiments, along with a large group of experiments based on the ability of the human ear to recognize fine tonal differences strongly suggest that the hearing system incorporates a discrimination mechanism limited by the frequency response exhibited by individual (or at most a few) OHC neurons. This mechanism will be associated with the fine tonal channel in this work.

[xxx add material]

9.5.3.1 Describing the just noticeable differences for the tonal channels EMPTY

Determining the just noticeable difference (jnd) of human hearing is difficult to measure but provides meaningful insights into the hearing modality. The first difficulty concerns the variety of possible jnd's. Just noticeable differences can be measured as a function of changes in stimulus loudness, changes in stimulus frequency change, changes in stimulus loudness in the presence of a masking tone, and changes in stimulus characteristic in the present of noise of various characteristics among others. The second difficulty involves the fact that these measurements are normally performed psychophysically and involve a perceived difference that may differ significantly from what the investigator expects (in the absence of a realistic model. The third difficulty involves the failure to distinguish clearly between the perceived difference which involves the psychophysical concept of loudness and the applied difference which involves stimulus intensity. The fourth third difficulty may arise because of the compression normally introduced by

the feedback mechanism within the sensory neurons (**Section 5.xxx**).

Figure 9.5.3-1 describes some of the different regimens that can be followed using different protocols. [XXX WORDS] Examples of these different experiments are listed on the right. This is the first known tabulation that categorizes these apparently disparate experiments.

MEASUREMENT	PROTOCOL	EXAMPLE
Δ Loudness (based on threshold)	1A internal threshold	
Δ Loudness (based on random noise)	2A internal noise background	Zwislocki & Jordan, 1986
	2B external in–correl band noise	
	2C external out–correl band noise	Small & Campbell, 1961
	2D external harmonic–correl band noise	
Δ Loudness (based on structured external tone(s))	3A external in–correl band tone	
	3B external out–correl band tone	
	3C external harmonic–correl band tone	
	3D external in/out–correl band tones	
Δ Loudness (based on structured external tone differences)	4A external in–correl band tone differences	
	4B external inout–correl band tone differences	

Figure 9.5.3-1 Regimens producing potentially different psychophysical results ADD .

Interpreting the jnd data requires the analyst to recognize the majority of the data was acquired under small signal (threshold) conditions. While the stimulus may appear to be at a high average sensation level, the actual measurement is of a small signal related to the stimulus in the presence of a second small signal at the point of "detection" within the hearing modality. The latter signal may be of internal origin or a masking external signal. The analyst must also be aware of the differences in sensation duration used in the various experiments. The adaptation time constant of hearing as well as the time constant of the perception apparatus can impact the perception of short duration stimuli.

The paper by Zwislocki & Jordan is a good example of these experiments. They applied an external stimulus to node B and measured thresholds created at node using responses reported at node xxx. The threshold represented the limiting condition, an unknown signal to internal noise ratio. In this case, the signal was a perceived loudness. The nature of the internal noise was not characterized but it can be assumed to be a random band-limited white noise.

9.5.3.1.1 Just noticeable differences in loudness at threshold EMPTY

Zwislocki & Jordan have provided excellent material and useful argument in this area[112]. They also attempt to sort out the difficulties inherent in rationalizing Fechner's law, Weber's law and the

[112]Zwislocki, J. & Jordan, H. (1986) On the relations of intensity jnd's to loudness and neural noise *J Acoust Soc Am* vol 79(3), pp 772-780

more recent Steven's power law. A major problem in this area is the incompleteness of these laws. They generally apply to only the sound regime the named individual concentrated on, phonotopic, mesotopic or scotopic. the earlier the work, the more likely it is to apply only to the phonotopic regime. Efforts to coordinate these laws and the related databases have also taken place in the visual regime. **Figure 9.5.3-2**, taken from a companion work, shows the "state of the art" of this research in vision.

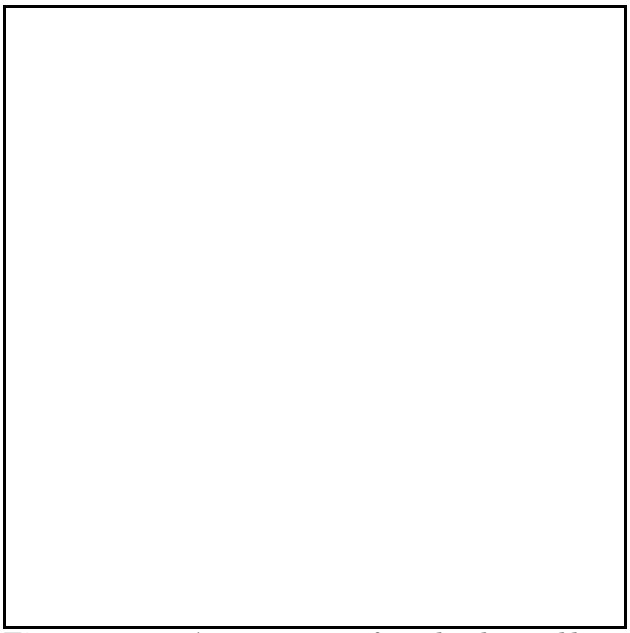


Figure 9.5.3-2 A comparison of psychophysical laws in the visual domain EMPTY.

6.2.2.4 Masking experiments as a corroboration of the critical band model EMPTY

Chapter 10 of Gelfand is particularly clear in its discussion of masking[113]. While only concerned with phenomena and not mechanisms, it may hold critical insights.

The challenge is to isolate the actual critical bands from a wide variety of available data that primarily suggests the existence of one or more critical bands based on local measurements, not across the complete audio spectrum. The text by Zwicker & Fastl is a fundamental source and bibliography in this area[114]. The Zwicker & Fastl material also describes the artificiality of their definition of critical bands at low frequency (page 158).

Small has provided excellent material on a group of noise masking experiments[115]. However, he did not associate them with the particular parameters associated with the critical bands. His goal was primarily related to the study of periodicity pitch. He had earlier found that low frequency noise that saturated the presumed low frequency critical band had no effect on the presumed periodicity pitch of a signal[116]. The curves in the 1970 material refer to the corrected curves in the 1961 paper.

Small chose tones at 2200 Hz and at 150 Hz for unspecified reasons.

Figure 9.5.3-3 is created by combining figures 1 and 3 from Small & Campbell. They showed that several types of tones at 2200 Hz could be masked by a band-limited noise mask approaching 2200 Hz from either side. In the absence of some sort of filter characteristic in the vicinity of the 2200 Hz test signal, the test signal could not be isolated from a noise mask operating at any frequency

---

[113]Gelfand, S. (1981) Hearing. NY: Marcel Dekker
[114]Zwicker, E. & Fastl, H. (1999) Psychoacoustics, Facts and Models, 2nd Ed. NY: Springer Chapter 6
[115]Small, A. (1970) Periodicity pitch *In* Tobias, J. *ed.* Foundations of Modern Auditory Theory, Vol. I. NY: Academic Press Chapter 1
[116]Small, A. & Campbell, R. (1961) Masking of pulsed tones by bands of noise *J Acoust Soc Am* vol. 33(11), pp 1570-1576

band in the audio spectrum. By overlaying their responses, the composite frequency response shown by the heavy line is formed. This response clearly shows the presence of a "critical band" delimiting and controlling the masking of the test signal by the noise signal. Unfortunately, as noted by Small & Campbell, the skirts of the filter used to create their band limited noise approximated 24 dB/octave. Thus the skirts of the "composite" filter were determined by the test set and not the critical band itself. The nominal critical band of Zwicker is shown by the narrow shaded box in the figure. It extends from 2000 to 2320 Hz with a center frequency at 2150. Zwicker did not define the slope of the skirts of his critical bands. It is suggested that the critical band of Zwicker is too narrow to include the sidebands of the test signal (2200 ± 150 Hz). The alternate critical band proposed in **Section 9.5.3** xxx is wider. It extends from 1863 to 2428 Hz with a center frequency at 2127 Hz. [xxx check numbers for my band]

> This figure highlights the dilemma that remains regarding the critical bands. While the width of a "critical band" can be determined in an experiment, its center frequency generally cannot. By moving the test signal incrementally in center frequency, a new critical band will be determined by experiment. This new band appears to overlap the previously determined band and exhibit a new center frequency. The underlying physiological methodology providing this characteristic remains unknown.

The slopes of this wider critical band are also defined as the slopes of a single OHC forming the upper and a single OHC forming the lower skirts of the band. [xxx give a number here]

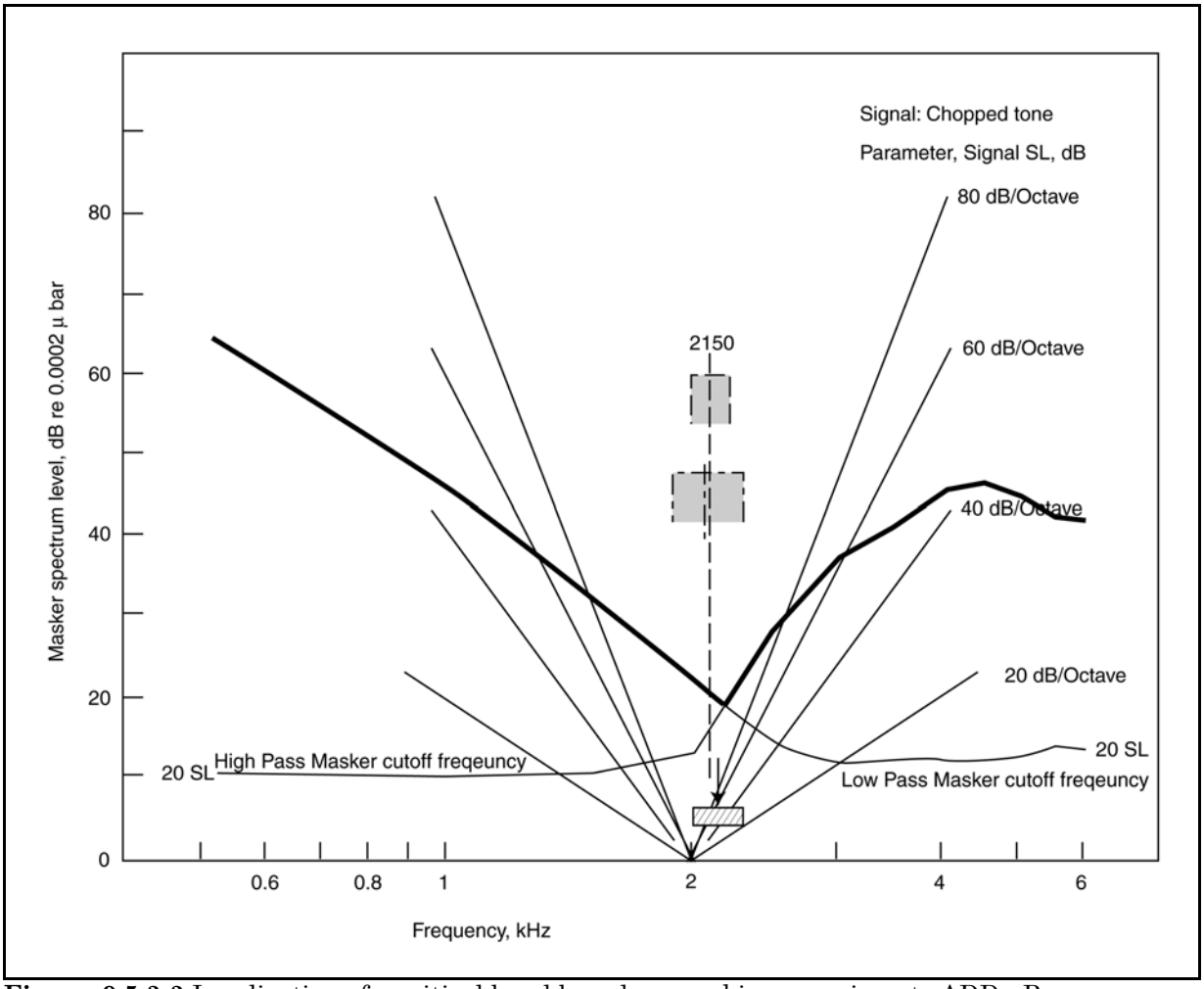


Figure 9.5.3-3 Localization of a critical band based on masking experiments ADD. Remove excess nominal filter slopes. Quantify which of my critical band tables is used. The test signal (shown by arrow and hatched box) consisted of a 2200 Hz tone AM modulated by a 150 Hz square wave. It was also FM modulated at 1 Hz to aid detectability. Assembled using Small & Campbell, 1961.

The Small & Campbell experiments need to be repeated using a protocol more optimized for critical band isolation and new instrumentation, particularly regarding the steepness of the skirts of their band-limited noise signals. Skirts of greater than 60 dB/octave xxx for the noise signal would provide definitive information concerning the structure of the critical bands.

Besides the laboratory investigations into critical bands, an interesting division of the audio spectrum has occurred in the practical world. One third octave audio filters have become largely standard in the audio recording and reproduction world. This has led to an audio spectrum protocol as shown in Zwicker & Fastl (page 233-238) that has even been incorporated into DIN 45 631 and more recently ISO 532 B. Both analog and digital loudness meters have been fabricated and standardized based on this one-third octave scheme.

9.5.3.2 Describing the “correlation range ” of the tonal channels

REWRITE

[xxx use my computed values based on the linearizing of the 1961 proposal by Zwicker.]
The hearing community has had great difficulty in defining the critical bands of human hearing. Although appearing in a wide variety of experiments, the critical band structure has eluded association with the physiology of hearing. The wide variation in proposed widths for the bands is partly understandable due to a lack of a physiological foundation and the limited performance of the test equipment used before the 1980's. However, the basic cause has been a misinterpretation of the data. As developed by analogy in **section 4.6.8.2.2** and in detail in **section 8.5.2**, the "critical band" of the critical band concept can be replaced by a functional parameter called the "correlation range" or equivalent frequency correlation range. This correlation range exhibits a geometric center because of the logarithmic frequency to place conversion in hearing. The correlation range is always centered on this geometric center.

Greenwood has discussed the correlation ranges of human hearing using the term critical band widths based on his psychophysical masking tests and the data of Zwicker, Flottorp & Stevens[117]. His table is reproduced here.

TABLE 9.5.3-1XXX Measurement of correlation range widths

Lower frequency limits of critical bands under consideration, Hz.	Estimates of critical band width derived From Zwicker, et. al., Hz.	Estimates of critical band width based on Greenwood, Hz.	$Q_{critical}$
395	108	Less than 104	~4
500 (center freq.)	110	84<x< 104	~5
1030	188	175	~6
2130	350	288<x	~6
3030	535	425–470	~6
5700	1150	800<x<900	

Greenwood provided an analysis of the data. However, a different interpretation is appropriate if the nominal bandwidth limit of the neural system is assumed, resulting in phase tracking at low frequencies but energy integration at higher frequencies. Under this interpretation, two different mechanisms contribute to the graphs of critical frequency as a function of frequency or basilar membrane position.

[117]Greenwood, D. (1961) Critical bandwidth and the frequency coordinates of the basilar membrane *J Acoust Soc Am* vol. 33(10), pp 1344-1356

Figure 9.5.3-4 reproduces figure 1 from Zwicker, 1961. It surfaces the problem of "critical band" width. The band widths associated with this Zwicker paper do not agree with the earlier values presented in the Greenwood paper based on Zwicker, Flottorp & Stevens. The OHC sensory neurons operating as integrators above 600 Hz, will report a stimulus has been applied within the energy band to which they are sensitive. Alternately, if several OHC are grouped in stage 2 to form a single signaling channel, the report will indicate a stimulus within the bandwidth associated with that group. Zwicker has attempted to define these channels consecutively based on a specific set of psychological tests. The problem is the widths of the individual defined bands are very wide. Each band is far wider than the characteristic bandwidths of individual signaling channels reported in the electrophysiological literature and surmised from difference limen experiments. [xxx or there must be a differential mechanism to specify the actual frequency.]

Zwicker noted the experimental difficulties encountered in measuring the critical bands at very low center frequencies. He also noted the conceptual problem of specifically defining a critical band in this area. He stood by a critical band width at 50 Hz that was extremely wide compared to the other bands. The figure shows an alternate set of critical bands from a figure from Zwicker 1988. It also shows an alternate critical band function drawn through the alternate critical bands at low frequencies and the rest of the critical bands. This alternate critical band function makes the center frequency of each band a simple exponential of the band number(plus a constant). It also makes the the ratio of the critical band width to the center frequency of the band nearly constant. This strongly suggests the critical band structure is closely aligned with the place-frequency relationship found in the cochlear partition.

Zwicker also caveats that his bands may vary from person to person. However, this statement leaves open a major question. If these are in fact discreet bands, there must be an additional mechanism that allows interpolation of their channel value in order to perceive finer frequency intervals.

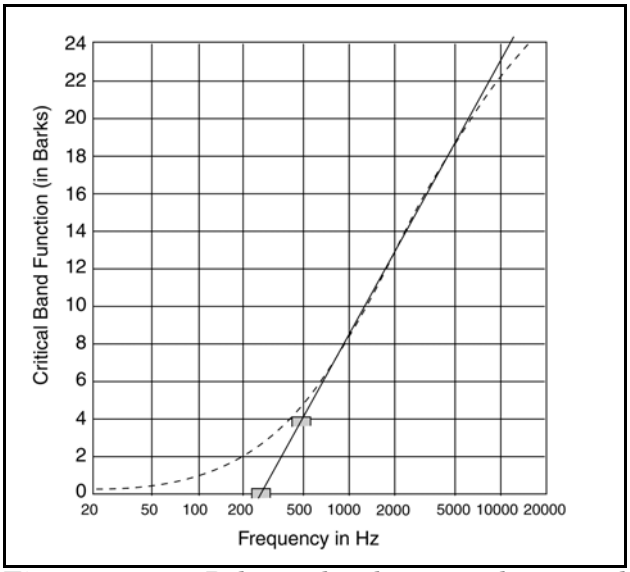


Figure 9.5.3-4 Relationship between the critical band function, critical band number and the frequency consistent with the theory of this work. Shaded hats represent later data points based on Zwicker 1988. The straight diagonal line represents an alternate critical band function drawn through the expanded data set. Modified from Zwicker, 1961.

[xxx need to look at the following is interpretation more closely, the perceived CF becomes a response as a function of center frequency]
Looking at figure 8 (, reproduced from Zwicker, Flottrop & Stevens), 9 and 10 of Greenwood, it would appear that the critical frequency function for humans is based on at least two distinct mechanisms.
The overall response resembles a first order low-pass filter type. In figure 8 , the first pole of the filter is at 600 Hz. Looking more closely, there appears to be a second pole near 5000 Hz. A first pole at 600 Hz is compatible with the bandwidth of the individual sensory neurons of the auditory system. The source of the second pole is unclear. [xxx]

In 1988, Zwicker provided **Figure 9.5.3-5** showing the raw data supporting the critical band concept. The meager database appears to deviate significantly from his 1961 proposal at low frequencies. In the caption to the figue, he says, "The solid line indicates the proposal agreed on by an international commission." However, his 1961 paper is clearer on this point. The proposal was not agreed upon by the I.S.O. but they agreed the proposal should be circulated for comment.

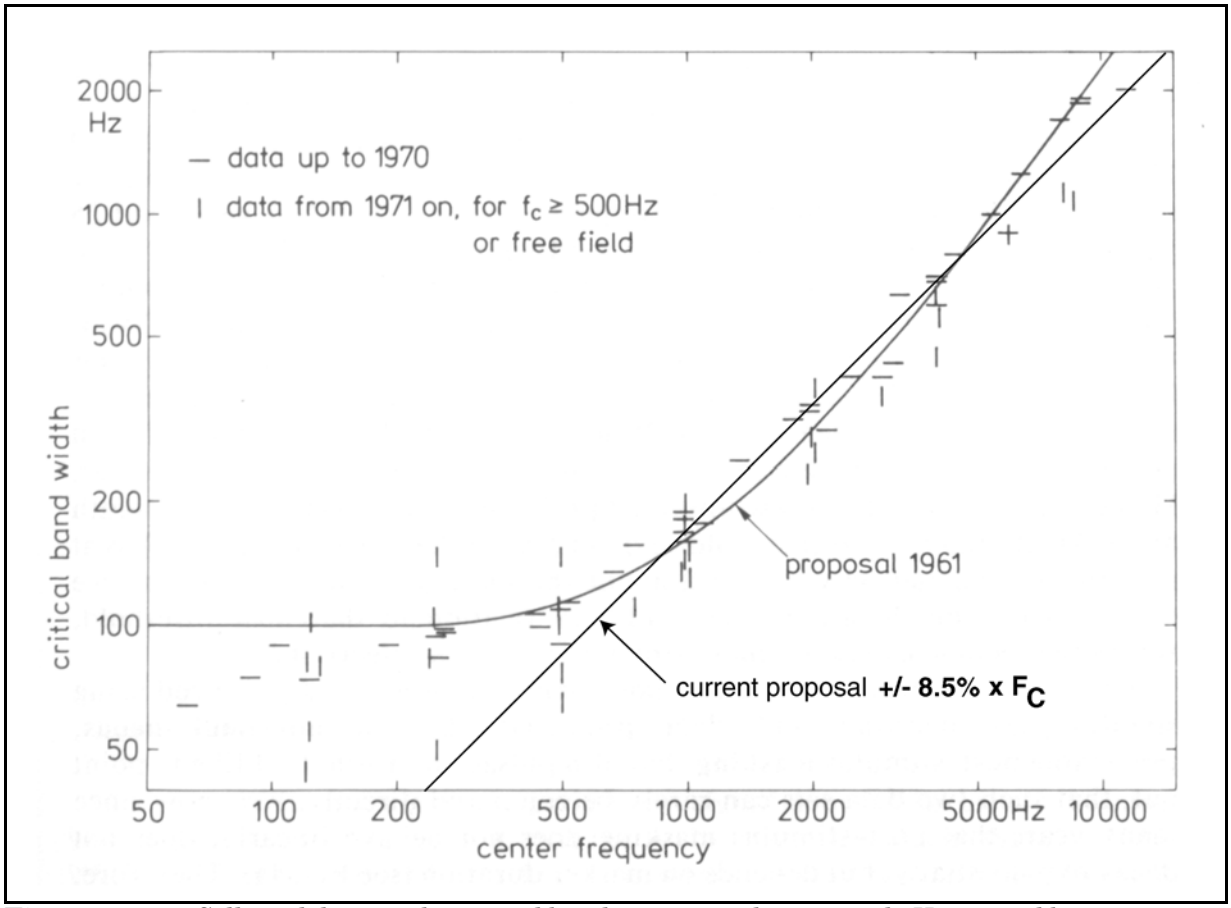


Figure 9.5.3-5 Collected data on the critical band concept and a proposal. Horizontal bars represent data acquired before 1970. Vertical bars represent data published after 1970. Below a center frequency of 500 Hz, only data produced in free-field or equivalent free-field conditions using equalize earphones are plotted. The current proposal is a first order relationship between the critical band function and the center frequency. Data and original proposal from Zwicker, 1988.

The data strongly suggests the baseline should be changed from Zwicker's 1961 proposal to reflect narrower critical bands for center frequencies below 400 Hz. Using a new baseline following the lowest set of vertical bars essentially linearizes the critical band function of the earlier Zwicker graph with respect to log frequency. As shown in **Figure 9.5.3-6**, it also linearizes the critical band function with respect to the position of the source energy along the cochlear partition. It also makes the ratio of the critical band width to the center frequency of that band a constant for all bands. The resulting proposal associates each approximately 1.5 mm section of the cochlear partition with a separate critical band, associates one second stage multiplexer with each section. This proposal also provides a logical reason for the spiral neural paths shown in the graphics of the PNS by de No and by Spoendlin (**Section xxx**). The neurons in these groups are being funneled to the summation nearest multiplexer, whether first or second order.

No direct evidence has been found defining the frequency of the lowest edge of the lowest critical band or the highest edge of the highest critical band. The proposed theoretical bands have been developed based on several nominal values. If these are found to be inappropriate, other values can be adopted and the associated parameters calculated. The upper limit of band number 24 has been kept at 15,500 Hz in accordance with Zwicker. The total number of OHC along a row has been kept at 3456 as the nearest value to the number 3500 of Spoendlin that can be factored by 24. The residue of that factoring is 144. This number can be factored by six, the nominal number of OHC in a row that are combined in one initial channel according to Spoendlin's neuron tabulations. This leaves a residue of 24 which can be related to the first multiplexer in the coarse tonal signaling channel (**Section 9.2 xxx**). By combining these factors, it is useful to assign the lower band edge of the first critical band in humans a frequency of 27 Hz.

Based on this proposal, **Table 9.5.3-2** describes the theoretical values for a set of functions related to the critical bands found in human hearing. As a result of these calculations, the theoretical bandwidth of each critical band remains 30% of its center frequency (rounding changes some of the values in the table slightly). The center frequencies are calculated as the geometrical mean of the channel following the normal convention. However, as shown in **Section 9.2**, these bands are not formed in the normal manner. They are assemblies of subchannels as conventionally used in man-made communications equipment. It is not appropriate to speak of the Q of such assembled channels. However, the slope of the edge of the assembled channel remains quite steep. It is the slope of the individual channel defined by the lowest or highest OHC neuron in the group (**Section xxx**).

TABLE 9.5.3-2
Theoretical Parameters of the Critical Bands of Human Hearing

Band #	Lower Freq.	Upper Freq.	Center Freq.	Bandwidth
1	27	35	31	8
2	35	46	40	11
3	46	60	52	14
4	60	78	68	18
5	78	101	89	24
6	101	132	115	31
7	132	172	150	40
8	172	224	196	52
9	224	292	256	68
10	292	380	333	88
11	380	496	434	115
12	496	646	566	150
13	646	842	737	196
14	842	1097	961	255
15	1097	1429	1252	333
16	1429	1863	1632	433
17	1863	2428	2127	565
18	2428	3164	2772	736
19	3164	4123	3612	959
20	4123	5374	4707	1250
21	5374	7003	6135	1629
22	7003	9127	7995	2123
23	9127	11894	10419	2767
24	11894	15500	13578	3606

It would be useful to compare these theoretical parameters with those for the numbers for the higher critical bands of Zwicker before he arbitrarily adjusted the widths of the lower critical bands[118] and introduced "generous rounding" (page 248 of 1961 paper) into the data set.

9.5.3.2.1 Loudness related to critical bands EMPTY

[xxx Moore, 1977, pages 96 and 101 and 118]

9.5.3.2.2 Noise masking as a method of critical band determination

Only a very few remarks appear in the literature concerning a critically important but totally unexplained phenomenon of hearing. When a noise mask of ever increasing bandwidth is used to mask another stimulus, the effectiveness of the masking ***increases*** when a certain bandwidth is exceeded. Roederer noted the findings of Zwicker & Scharf in 1965. They noted this phenomenon and highlighted the fact the effect of masking increases instead of decreases (or stays the same).

[118]Zwicker, E. & Fastl, H. (1999) Psychoacoustics, Facts and Models, 2nd Ed. NY: Springer pg 158

Roederer attempted to explain this effect in mathematics without clearly associating that explanation with the physiology involved. He attempted to use an empirical relationship from Stevens. This relationship actually obscured the situation. The correct relationship is developed below. The problem is that Stevens developed an empirical formula that only applies within the phonotopic regime of hearing. The theory is better developed within the mesotopic regime where the system is operating linearly with respect to intensity. The same situation can then be applied to the phonotopic regime by introducing the effect of the internal feedback present at this level of stimulation.

When testing with a noise source(or a complex source of multiple unrelated sinusoids), the energy within the band is computed based on the root-mean-square relationship. The effective amplitude of the energy (relative to an equivalent DC signal) within the band is given by the formula;

$$L = C(A_1^{\,2} + A_2^{\,2} + A_3^{\,2} + \cdots)^{0.5}$$

This equation is different from that of Stevens. The equation takes the square root of the sum of the peak amplitudes, not the cube root. The coefficient, C, is independent of the frequency band as long as all of the components are within the specified band. Within the hearing system, this equation applies to the effective amplitude of any complex signal prior to its being passed through an integrator acting as a rectifier. At that point, the value of L is calculated. Subsequent circuits use the value of L obtained from different subchannels rather than values of the individual A's. In this case, the signal at the output of a summing circuit is given by;

$$L_T = L_1 + L_2 + L_3 + \cdots$$

If the component A_1 is outside of the band integrated by the integrator processing A_2 and A_3, the summation becomes;

$$L_T = L_1 + L_2 = C(A_1^{\,2})^{0.5} + C(A_2^{\,2} + A_3^{\,2} + \cdots)^{0.5}$$

Using the Theorem of Pythagorus, the length of the two sides of a triangle is always larger than the length of the diagonal. This is the case here. The amplitude L_T in the second equation is always larger than that calculated in the first. Thus the rise in the signal level following a summation circuit (first multiplexer) in hearing is indicative of the fact that one or more components of the stimulus was processed by a separate integrator. In the case of a noise signal, the summation will show a change in slope of the measurement of L_T as a function of the width of the noise signal in relation to the edge of the critical band involved. In the case of a series of tones, the response may show a level shift rather than a change in slope. This relationship is shown in **Figure 9.5.3-6** from xxx.

This figure shows two distinctly different situations. Xxx [xxx find a better figure than in Roederer, pg 34.]

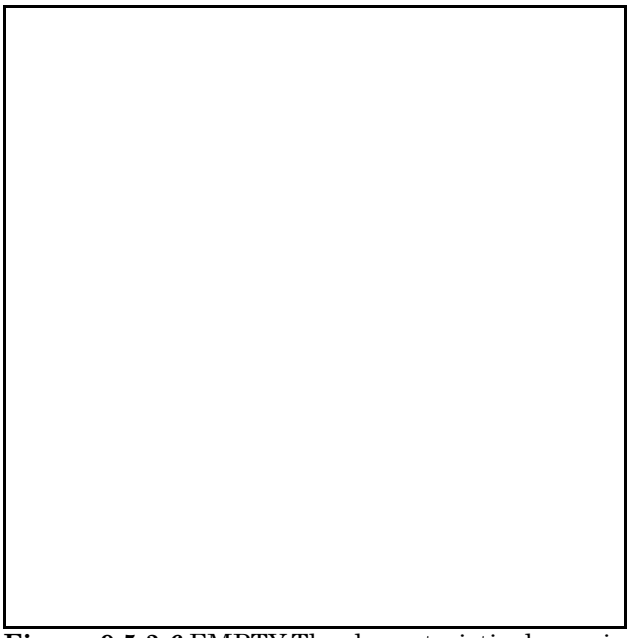


Figure 9.5.3-6 EMPTY The characteristic change in the signal response in critical band measurements due to part of the test signal being outside the critical band under evaluation.

9.5.3.3 Critical band width as a function of center frequency

The frequency performance of the human auditory system can be predicted based on the model presented in **Section xxx**. Three major performance parameters have been addressed in the literature, the center frequency of the audible spectrum as a function of cochlear position, the bandwidth of the signaling channel as a function of the center frequency, and the rate of change of the signaling channel bandwidth as a function of center frequency.

The signaling channel bandwidth is usually described in the hearing literature using the term, "critical bandwidth." As noted earlier, the critical bandwidth is typically defined at the 10% points (FWTA) of the signaling channel frequency spectrum.

The model presented here has provided an explanation for the variation in spectral bandwidth versus center frequency in human hearing. It agrees well with the published values[119],[120].

Zwicker & Terhardt have provided several mathematical expressions approximating the variation in bandwidth and the rate of change of bandwidth as a function of center frequency[121]. The goal of that paper appears to be to provide a convenient equation that can be used in further exploratory research. They sought to develop an equation that closely matched the median values measured in the earlier Zwicker (1961) paper. However, that paper did not include any original test data. It offered a rationalized function "based upon the natural division of the audible frequency range by the ear" and using a set of preferred frequencies. The values associated with these preferred frequencies were "generously rounded." No standard deviations associated with the median values as a function of frequency were given. This makes it difficult to justify the new proposed equation. Its calculated values were compared with the nominal values from the above rounded values. A more meaningful comparison would be to plot the selected curves as overlays on original data, including its associated error bars.

The primary equation developed by Zwicker & Terhardt is based on a transcendental expression (an arctangent) over a limited, but not clearly defined, range. Based on physiological considerations, it can be said that auditory mechanisms, and neural mechanisms in general, do not

---

[119]Zwicker, E. Flottorp, G. & Stevens, S. (1957) *J Acoust soc Am.* vol. 29, pp 548

[120]Zwicker, E. (1961) Subdivision of the audible frequency range into critical bands (Frequenzgruppen) *J Acoust Soc Am* vol 33, pp 248

[121]Zwicker, E. & Terhardt, E. (1980) Analytical expressions for critical-band rate and critical bandwidth as a function of frequency *J Acoust Soc Am* vol. 68(5) pp 1523-1525

result in equations of the transcendental class. This is particularly true with respect to the logarithmic helix formed by the cochlea in humans. It is more appropriate, when moving into the area of applied research related to hearing to use equations compatible with the underlying mechanisms and the resulting empirical data, rather than transcendental equations. Adopting the Zwicker & Terhardt transcendental equation in the applied research area, as done in a recent paper[122], appears unwarranted.

[xxx do not provide a new equation unless it is really solid based on theory and preferably based on good data to support it.]

9.5.3.4 Beat creation and cross-modulation distortion within a critical band EMPTY

The problem can be illustrated using **Figure 9.5.3-7**. It expands the previous figure from Zwicker to show how multiple tones are reported by the human auditory system. It can be compared to the presentation in Roederer[123] reproduced in Deutsch[124].

[122]Deutsch, W. & Fodermayr, F. (undated but after 1993) Visualization of multi-part music (acoustics and perception) www.kfs.oeaw,ac.at/fsf/mus/Poly1.htm
[123]Roederer, J. (1975) Introduction to the physics and psychophysics of music, 2nd Ed. NY: Springer.
[124]Deutsch, D. *ed*. (1982) The Psychology of Music. NY: Academic Press pg 15

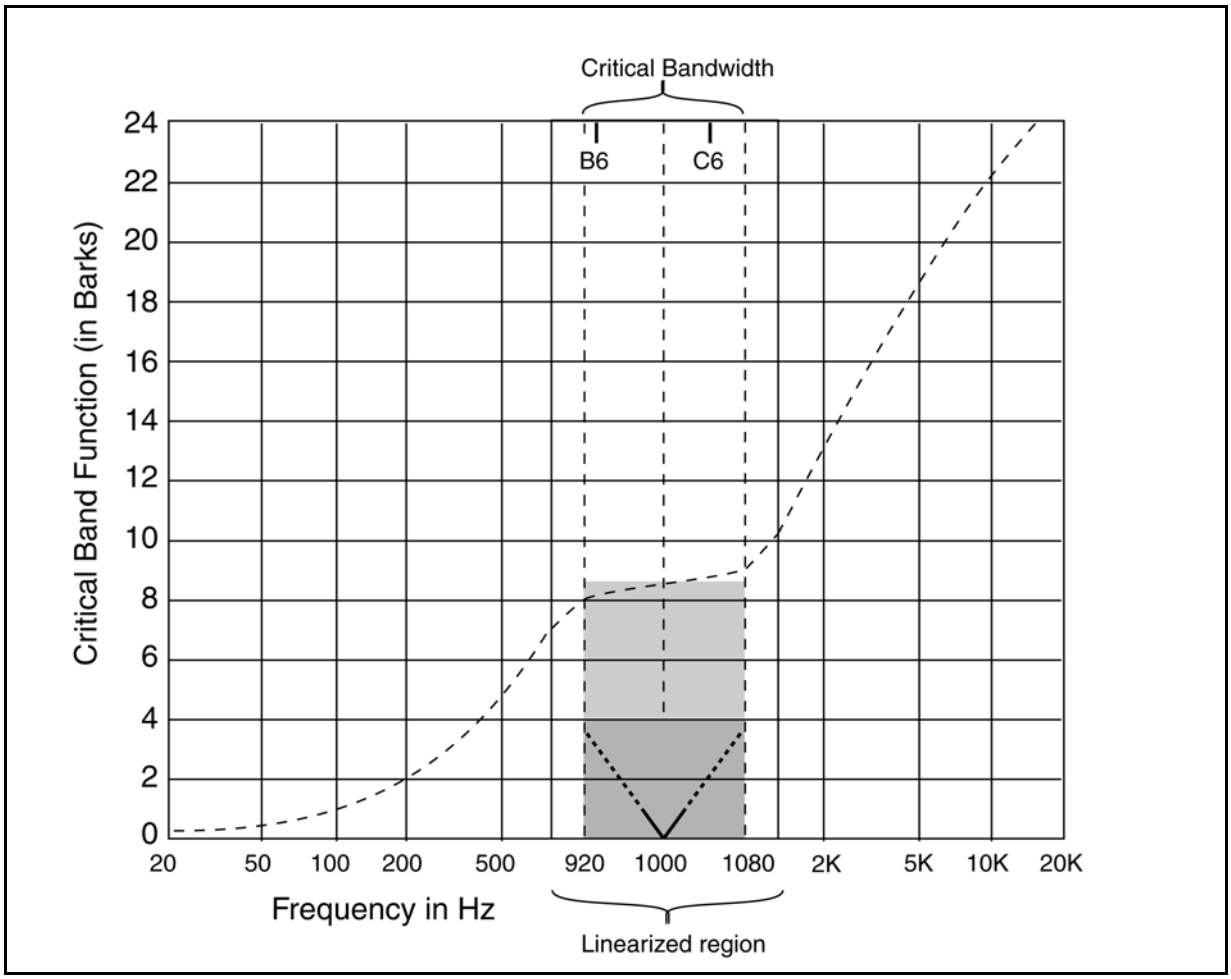


Figure 9.5.3-7 Expanded diagram of critical band widths to show processing of multiple tones. Critical band #9 has been expanded. See text.

Roederer used the simple trigonometric identity:

$\sin \omega_a \cdot t + \sin \omega_b \cdot t = 2\cos \frac{1}{2}(\omega_b - \omega_a)t \times \sin\frac{1}{2}(\omega_b + \omega_a)t$

9.5.3.5 Chromatic performance based on 2IFC experiments EMPTY

Feth & O'Malley have provided significant data concerning chromatic performance and compared it with earlier results using other methods[125].

9.5.3.6 Describing the chromatic performance of hearing in other species EMPTY

9.5.4 Assembled data from stage 3 and 4 recordings–no cognition

[125]Feth, L. & O'Malley, H. (1977) Two-tone auditory spectral resolution J Acoust Soc Am vol 62, pp 940-947

Figure 9.5.4-1, reproduced from figure 1.4 in Jesteadt, provides valuable performance data for the xxx. The frequency response of the system, from the point of excitation (xxx where was it] to the pulse trains generated by stage 3 ganglion cells reproducing data from IHC, is nearly flat.

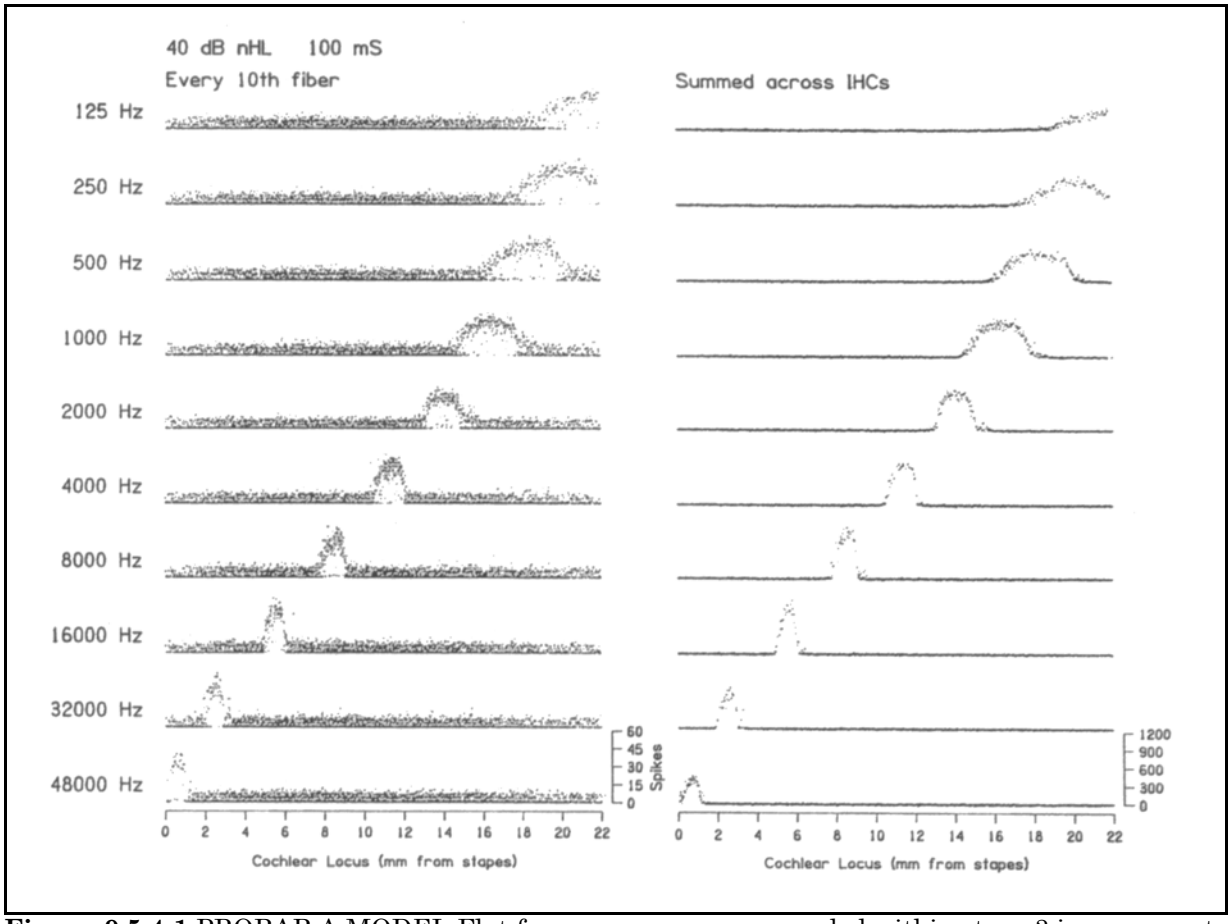


Figure 9.5.4-1 PROBAB A MODEL Flat frequency response recorded within stage 3 in response to excitation of the system at xxx. [xxx get better copy] From xxx, 1997.

The turn of the century witnessed an increase in interest in recording action potentials from stage 3 neurons. The signals were recorded primarily from the auditory nerve but some were evoked potentials recorded from the surface of the CNS. Of particular interest were pulse trains resulting from a single low frequency stimulation. Some investigators recorded signals from multiple neurons simultaneously in attempting to illustrate a "volley theory" approach (Section xxx) to neural signaling.

10.7.2.1.1 Stage 4-5 signals recorded from acoustic and electrical stimulation SPECIAL

Before exploring the data available from investigators recording stage 4 signals, it is important the reader studies the 2002 review by Eggermont and Ponton[126]. The review is excellent but needs to be read critically. The paper is focused on the performance of the human auditory system. However, the simple conceptual model does not appear compatible with the human system. As an example, the paper does not discuss the circuits between the cochlea and the entrance to the primary auditory cortex (AI). Their figure 4 shows a variety of signals originating in the inferior colliculus and terminating in multiple regions of the AI. However, the literature is unclear concerning the inferior colliculus in humans. Some authors assert it is morphologically absent or at best rudimentary in humans. The inferior colliculus is primarily associated with source signal location via binaural timing calculations. It is a massive morphological structure in animals dependent on such source location information for survival in their ecological niche, i.e., the dolphins and the bats. The basic assumption of Eggermont and Ponton, that the IC is critically important in the signal processing associated with speech perception of the human auditory system, is not supported by this work. A human model must be used that recognizes the multiple signal paths created and formatted by the cochlear nucleus, and largely ignoring the IC.

[**Figure 2.4.2-1**] is offered as a more insightful description of the major signaling paths within the auditory modality. It shows the three primary paths, the LOC path dedicated primarily to temporal (time and intensity) data, the MOC path dedicated primarily to tonal data, and the IC related paths (associated with source location and alarm mode activities).

Figure 4 in Eggermont and Ponton displays a maze of interconnections between the IC and various elements of the diencephalon and the cerebral cortex. Other paths are not shown explicitly (page 80). The paths appear to be the result of "traffic analysis." It is not clear how much of this traffic analysis is based on topological dissection and how much is due to neural signal tracking. No description is given of the character of the signals passed along these various paths (with the exception of some discussion on pages 90-91). The various elements appear to be labeled based more on the topography of the brain than on their functional significance. All three identified segments of the MGB pass signals to all three of their topographically identified "core" elements.

No source of stimulus to either the supra geniculate or medial pulvinar is shown in their figure 4. The combination of the supra geniculate and medial pulvinar of that figure appears to be the equivalent of the functionally defined auditory PGN and pulvinar of this work. If correct, the stimulus to the combination should originate in the MOC path from the cochlear nucleus. The

---

[126]Eggermont, J. & Ponton, C.(2002) The neurophysiology of auditory perception: from single units to evoked potentials *Audiol Neurootol* pp 71-99

output of the medial pulvinar would consist of tone oriented percepts of a high order, possibly describing a time sequence of individual stimuli using only a single encoded "word."

As noted in **Section xxx**, the auditory PGN is typically located adjacent to the MGN in the direction of the IC. The prominence of the auditory PGN may exceed that of the IC in humans, leading to confusion in the morphological identification of these two features.

A clear distinction needs to be made between the tonotopic projection of signals from the cochlea to higher brain centers and the alternative, the receptive field of a higher brain center neuron projected back to the cochlea. The receptive field projection may provide different results for continuous tones versus brief tones versus various clicks. Such differences are due to signal processing at various intermediate locations within the auditory modality. Additional effort is needed to differentiate between the various signal types carried by the principle signaling paths. For example, signals passed from the correlation circuits within the auditory PGN/pulvinar combination may consist of percepts describing a particular musical note (irrespective of octave content) or series of notes. The percept may exhibit strong correlation with the start time of a stimulus but otherwise be largely independent of the stimulus. In crude experiments, the perceived note may be associated with a "missing fundamental" that is not present in the stimulus (**Section 9.xxx**).

[xxx combine with the next paragraph and move to the Performance or ganglion cell operation] Several experimentalists have made recordings of signals acquired from the primary acoustic cortex (AI) following acoustic or electrical stimulation of the peripheral auditory system. The experiments have been largely exploratory research. They have not described which of the multiple signaling channels between the cochleas and the primary auditory cortex (AI) they were investigating. In any case, their data describes neural channels traversing large portions of the auditory modality. Thousands to millions of neurons are involved in each channel. As a result, no functional description of these channels can be defined and the recordings only show global parameters of the system.

Future investigation of signals recorded at the surface of AI should make every effort to describe the neural channel, **Figure 9.5.4-2**, they are investigating. [xxx Use a modification of Raggio.]

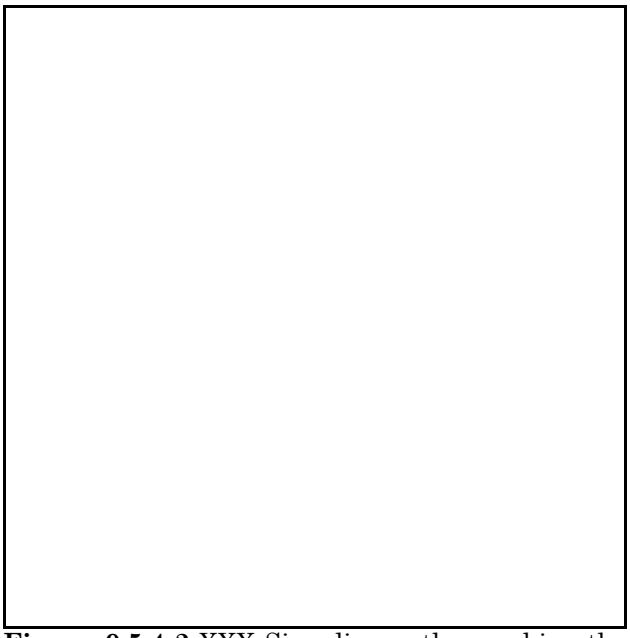


Figure 9.5.4-2 XXX Signaling paths reaching the Primary Auditory Cortex EMPTY.

Before analyzing the reported data, it is important to review the terms associated with quasi-random events. The terms of interest are stochastic, pseudorandom and chaotic.

Stochastic is frequently used in the hearing modality literature to describe a process that is conceived as random. However, the randomness of the process is seldom demonstrated.

As a guide, if a presentation exhibits any recognizable pattern, the underlying mechanisms are necessarily not stochastic (random) from the mathematical perspective.

A random process must have a very specific set of inter-symbol probabilities and the amplitude of the output must exhibit a Gaussian Distribution. The process and output amplitude distribution must conform to the requirements of the Poisson Equation.

In the absence of the precise inter-symbol probabilities required, a quasi-random process must be described as pseudo-random. Most encryption algorithms are pseudo-random codes that do exhibit a unique (eventually) repeating structure.

For the discussion at hand, a more important process is a chaotic process. Such a process may appear random but it actually generates a repeating structure. Such chaotic processes are similar to xxx.

A stage 3 encoding neuron (a ganglion cell) excited by a pulse input may operate in a variety of modes, depending on the amplitude of the excitation.

When driven by long, low repetition rate pulses, the encoding neuron may generate more than one action potential during the pulse interval. The generation of multiple action potential pulses in response to a single drive pulse is often described as entrainment. It occurs when the excitation amplitude is high and the excitation pulse duration is long compared to the duration of an action potential plus its associated refractory period. At a constant excitation pulse amplitude, the interval between the individual action potential pulse pairs remains fixed. However, these intervals need not be a submultiple of the excitation pulse interval. Thus, when overlaid on an oscilloscope, the action potential pulse pattern may appear aperiodic.

When an encoding neuron is excited by a long pulse with a duration less than the duration of an action potential plus its associated refractory period. No entrainment will be observed. Only a single action potential will be generated in response to each excitation pulse.

If the encoding neuron is excited by a pulse short with respect to the duration of an action potential plus its associated refractory period, the output of the neuron may vary considerably based on the interval between excitation pulses. For long interpulse excitation intervals, one action potential will be generated for each excitation pulse and the output can be described as synchronous with the excitation.

If the encoding neuron is excited by pulses that are short with respect to the duration of an action potential plus its associated refractory period, and occur at interpulse intervals shorter than the action potential duration plus its associated refractory period, the output may become very complex. The output will typically consist of an initial action potential, with the generation of a second action potential dependent on the precise state of the encoding neuron within its refractory interval at the time the second excitation pulse is initiated. The second excitation pulse may change the refractory state of the neuron without generating an action potential. If the refractory state is changed, the generation of an additional action potential may depend on the time before the arrival of an additional excitatory pulse. If the pulse arrives before the end of the refractory period, no action potential will be generated. The result is an output that, when overlaid on an oscilloscope that is synchronized to the excitation pulse train, consists of small groups of unevenly spaced pulses. If the output. is overlaid on an oscilloscope, that is synchronized to the output pulse stream, the display will appear chaotic. It will appear to contain a very complex internal structure

that is clearly not random.

The following figure from Hartmann and Kral illustrate these situations using a parametric presentation of frequency and time. The stimulus intensity is not given explicitly in this form..

10.7.2.1.1 Stage 4-5 signals recorded from acoustic and electrical stimulation

[xxx move a variant of this paragraph to the neuron chapter on ganglion cell operation]
Figure 9.5.4-3 reproduces a figure compiled by Hartmann and Kral as part of a discussion on the design of electrophonic prostheses[127]. The figure shows a variety of interesting features. Note the difference in the timing of the initial pulses (curved line between frames). The ordinates of the two frames are different. The upper ordinate is logarithmic but the lower ordinate is linear.

Part of the figure is based on Eggermont[128]. His paper expanded on a 1994 study of signals recorded at a fixed depth below the cortical surface in cats in response to a variety of acoustic stimulant waveforms. This author suggests the depth corresponds to the location of stage 3 decoding neurons. The action potential pulse trains resulted from clicks. Eggermont did not explicitly note the delay between his acoustic stimulus and the time of signal recording in the CNS. This delay is on the order of 15 ms as seen more clearly in frame B of his figure 2. Only 1.6 ms of this is attributable to the acoustic signal traveling through the air. Eggermont employed considerable mathematical manipulation without describing a neural model supporting those manipulations.

Hartman and Kral described these graphics as dominated by stochastic processes with little additional analysis. Interestingly, Eggermont did not use either of the terms stochastic or random in his interpretation of his data. It could not be determined if Fiseifis and Scheich used either of these terms. In fact, frames A & B exhibit a variety of statistical features that are not compatible with stochastic processes.. As noted in the parallel frames C & D, frames A & B are dominated by the pulses arriving at the primary auditory cortex after a constant delay determined by the pulse frequency of the stimulus. These features would be continuous in frames A & B if the excitation frequencies had varied continuously. Frame A is a composite of the responses of four different cats at 65 dB SPL. As Eggermont noted in his figure 2(B), one cat exhibited considerably less pulse to pulse jitter than did the others. His technique was to fully reduce the data for individual cats and then combine the data for the four into what he calls a dot-raster display.

Eggermont describes his method of data capture. He pressed multiple probes (3 to 8) into the primary auditory cortex (AI) of the cerebral cortex to a depth of between 600 and 1200 microns. He then performed a search for neurons that responded to his search stimulus, a complicated sequence of tones, clicks and noise bursts. Thus, he located neurons in stage 4 of the CNS that responded to his acoustic stimulus. He characterized the neural channel between his acoustic stimulus and the selected AI neurons, over a wide frequency range (typically 625 Hz to 20 kHz) to determine a tuning curve and a characteristic frequency.

---

[127]Hartmann, R. & Kral, A. (2004) Central responses to electrical stimulation *In* Zeng, F-G, Popper, A. & Fay, R. *eds*. Cochlear Implants: Auditory Prostheses and Electrical Hearing. NY: Springer pg 237
[128]Eggermont, J. (2002) Temporal Modulation Transfer Functions in Cat Primary Auditory cortex: Separating Stimulus Effects From Neural Mechanisms *J Neurophysiol* vol 87, pp 305–321,

He then presented his gamma tone test stimulus 20 times and averaged the response of each selected electrode in each cat before combining the data as noted above.

Looking at frame B, three distinct regions are noted. The primary feature is the responses to the individual pulses in the test stimulus pulse train at each repetition frequency. Note the very low pulse frequencies (repetition rates) used. This is tied to the fact the neurons employed occurred late in the channels of the auditory modality. In many cases, the recorded signals represented high level percepts extracted from the more primitive signals passed to the CNS. Eggermont described these channels at the recording point as low or bandpass channels. "The responses to click trains (Fig. 6A) show good locking throughout the entire train for repetition rates 13.44 Hz, but only for the first two clicks at a click repetition rate of 16 Hz. For higher repetition rates, 32 Hz, the responses skip the second click and respond again to the third click in the train."

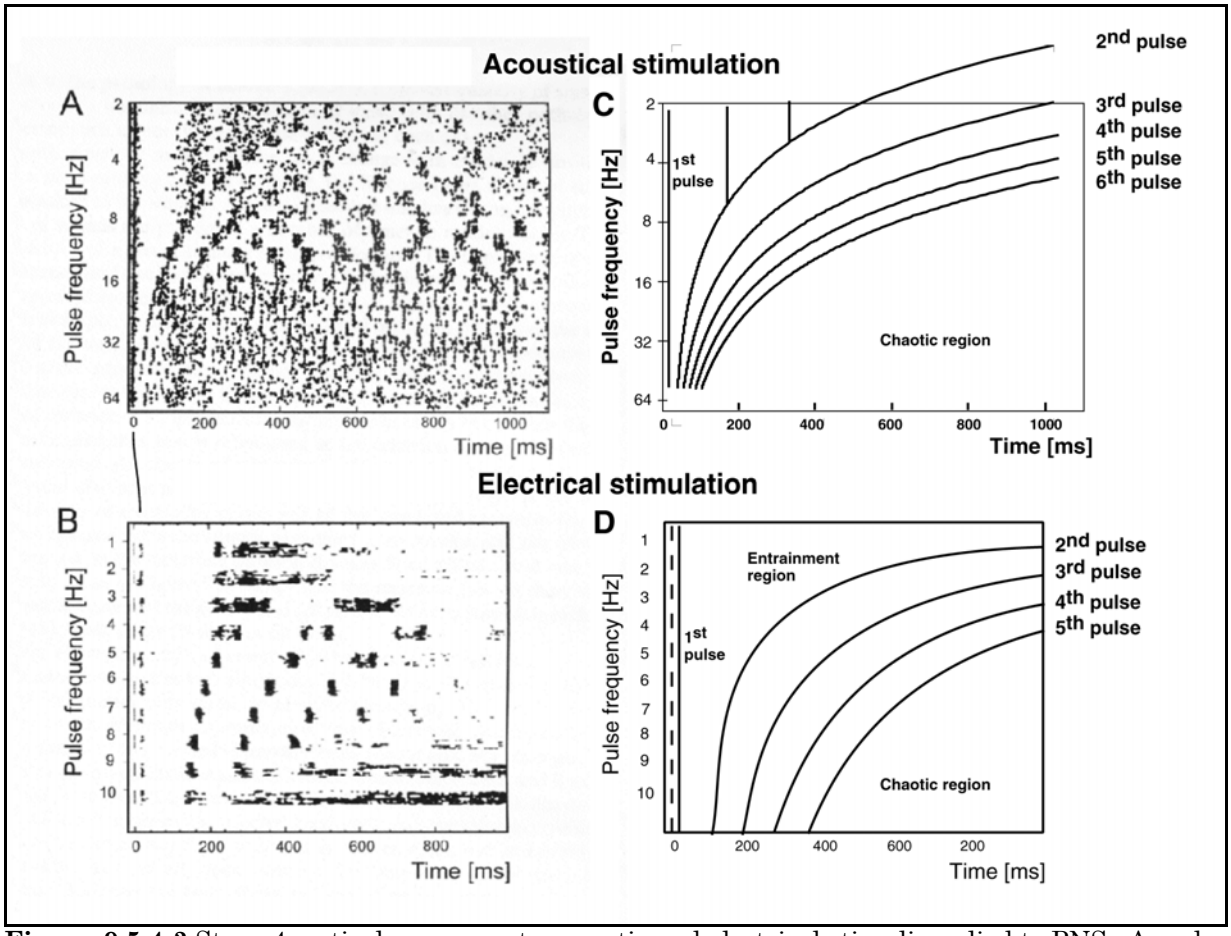


Figure 9.5.4-3 Stage 4 cortical responses to acoustic and electrical stimuli applied to PNS. A; pulse patterns in the primary auditory cortex (A1) of cat following acoustic stimulation (From Eggermont, 2002). B; pulse patterns in A1 of a gerbil due to electrical stimulation of the auditory nerve (From Fiseifis & Scheich, 1995). Note the difference in timing of the initial pulses. C & D; form of the graphs in A & B as a result of plotting the time delay between the first pulse and the n^{th} pulse. See text.

Raggio and Schreiner presented two papers on stage 4 and/or stage 5 signals recorded as a result of acoustic and electrical stimulation in the periphery elements of the hearing modality in cats[129,130]. They did not delve deeply into the signal processing by the channels connecting these two points. They noted (with citations), "Using electrical pulse stimuli, Hartmann and colleagues (1984) found strong phase-locking in auditory nerve fibers if the interpulse interval was >2 ms. Strong phase-locking to sinusoidal and pulsed stimulation in excess of 600 Hz has been reported that was similar or slightly better than following rates seen for acoustic stimulation." They noted that entrainment disappeared at a repetition rate of 38 Hertz and above (page 1294).

Malone et al. have presented similar data for the rhesus monkey[131]. Their introduction begins, "In many animals, the information most important for processing communication sounds, including speech, consists of temporal envelope cues below 20 Hz." This statement indicates a lack of understanding how the auditory modality operates. They did not present any block diagram or schematic of the auditory system they were attempting to interpret. Thus, it is hard to understand what they expected their data to show. They did include one important insight, "Rees and Moller (1987) emphasized that neurons of the inferior colliculus 'do not function as an array of stimulus invariant modulation frequency detectors" of the sort appropriate to a modulation filterbank, but rather "carry a selectively emphasized version of the input signal's amplitude envelope which is modified by the prevailing stimulus conditions.'" The same statement can be made concerning most of the signals emanating from the cochlear nucleus.

Their investigation using two animals was extensive. Most stimuli were presented binaurally even though only one hemisphere of the animals brain was instrumented. They focused on sinusoidally amplitude modulated stimuli. "Long stimulus durations were chosen to minimize the effects of onset responses while maximizing the number of modulation periods. This choice was crucial for the low modulation frequencies emphasized in this study." The carrier frequency used was nominally equal to the best frequency of the neuron in contact with the probe. Lacking a model, their analysis was limited largely to interpreting modulation period histograms. They used carrier tones between 100 Hz and 32 kHz that were always substantially higher than the modulation frequency. Their highest recorded pulse rate at the AI was 50 pps, even for modulating tones of 100 Hz. This suggests their recording sites represented neurons portraying stage 4 extracted information largely independent of the input signals amplitude and frequency characteristics.

They did note (with citations), "In addition to characterizing cortical responses to tones of short duration (typically 100 ms), we also measured responses to an unmodulated tone of the same duration as the modulated stimuli (10 s). The response to a pure tone of the same carrier frequency and level as the SAM stimuli served as a reference for the responses to modulated tones Fig. 1A). A striking aspect of cortical responses in awake animals is the fact that roughly one third (113/361; 31%) had significantly elevated firing rates relative to the spontaneous rate when calculated over the duration (10 s) of the control tone. In an additional 12% (42/361) of neurons, the firing rate was significantly suppressed over the duration of the control tone. Thus the generally accepted notion, derived from studies of anesthetized animals, that cortical neurons do not give sustained responses to pure tone stimuli of long duration does not hold for nearly one half (43%) of our data sample.

---

[129]Raggio, M. & Schreiner, C. (1994) Neuronal responses in cat primary auditory cortex to electrical cochlear stimulation. I. intensity dependence of firing rate and response latency *J Neurophysiol* vol 72(5), pp 2334-2359
[130]Schreiner, C. & Raggio, M. &(1996) Neuronal responses in cat primary auditory cortex to electrical cochlear stimulation. II. Repetition rate coding *J Neurophysiol* vol 72(5), pp 1283-1300
[131]Malone, B. Scott, B. & Semple, M. (2007) Dynamic amplitude coding in the auditory cortex of awake rhesus macaques *J Neurophysiol* vol 98, pp 1451–1474

Their data can be reinterpreted based on the Electrolytic Model of Hearing presented here. That interpretation requires the applied acoustic signals be processed separately by the OHC and IHC channels.

9.5.5 Assembled data from the complete system

Figure 9.5.5-1

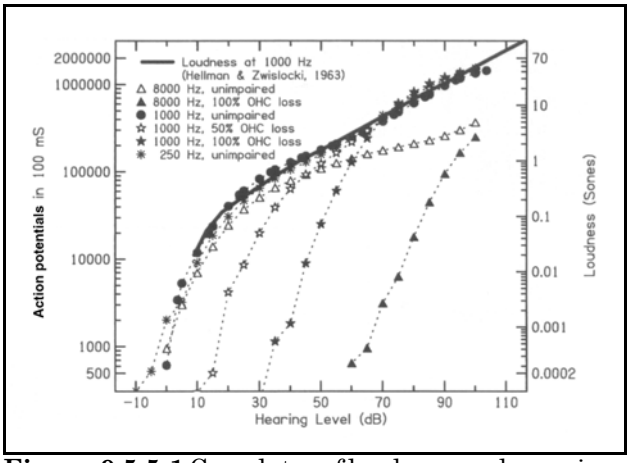


Figure 9.5.5-1 Correlates of loudness under various states of damage ADD. Commentary. From Javel in Jesteadt, 1997.

9.6 The parametric psychophysical performance of the auditory system

Exploring the performance of the auditory system in the laboratory requires very close attention to the features of the system when developing the test protocol. Unfortunately, this level of attention has not usually been provided in the papers presented in the literature. The value of many critical parameters frequently go unreported. Similarly, the protocols frequently introduce signal components that are not passed by the system, or are not passed by the system in the form expected. Finally, the data is frequently not recorded in sufficient detail to demonstrate conclusively the operation of the system. This latter problem has been glaringly frequent in the use of the conventional AP–frequency based histograms rather than the required AP–interval based histograms. Other similar problems will be highlighted in the following sections.

9.6.1 Intensity threshold and just noticeable intensity differences in hearing

9.6.1.1 The mean detectability of a tone burst

Moore has reproduced **Figure 9.6.1-2** from

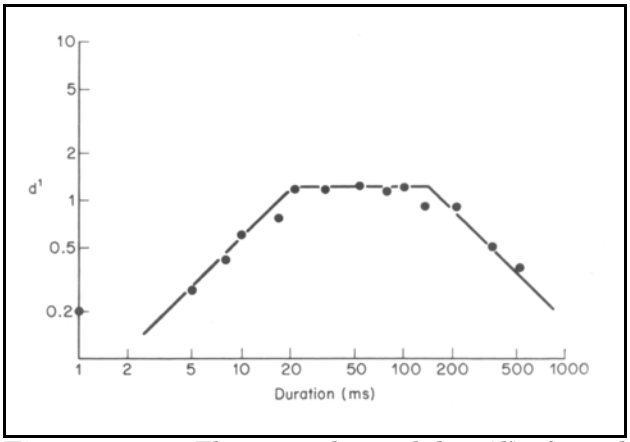


Figure 9.6.1-1 The mean detectability (d') of equal energy (? intensity) tone bursts of frequency 1 kHz as a function of duration. See text. From Stephens, 1973.

Stephens[132]. There is a question mark in the caption presented here. Moore did not differentiate carefully between the concepts of power and energy[133]. Moore was starting a discussion using only psychological evidence and early analyses dating from the 1940's. They were based on the idea that the auditory system was a perfect energy integrator. This concept clearly fails after 20 ms. The system is clearly responsive to the power (or intensity level) of the stimulus, not its energy. His discussion did develop an exponential expression for the rising waveform. The expression plots as a straight line on log-log coordinates. However, the measured response soon departs from that equation. Moore proceeds to discuss the flat top and trailing edge in more general terms. No equations were developed in support of these later discussions.
The figure clearly shows two exponential functions for this response at 1000 Hz. Moore discusses the database in the literature up to 1977.

In this work, the rising edge response is associated with the phonoexcitation/de-excitation (P/D) equation. The precise slope of the response depends on both the frequency of the tone and the risetime of the envelope of the stimulus. In general, the slope rises with intensity (and the latency becomes smaller) at a constant frequency. The falling edge response is associated with the time constant of the electrostenolytic supply to the adaptation amplifier circuit. It varies only slightly with frequency (particularly above 600 Hz). Below 600 Hz, the sinusoidal waveform at the collector of the Activa causes a more complex response than a simple exponential.

9.6.1.2 The just detectable intensity difference in hearing EMPTY

9.6.2 Relationships between frequency difference in hearing

9.6.2.1 The just detectable frequency difference for humans

[Zwislocki & Hordan '85]

Figure 9.6.2-1 provides some performance data of interest. The two upper continuous curves are from Zwicker, Flottorp & Stevens[134]. They interpreted the expression *frequenzgruppe* found in earlier German publications as "critical band." Zwicker, et. al. did use the term "critical band width" (with a space) in the text of their paper. They did not define their criteria for determining their bandwidths. Other papers have provided more specific data related to the xxx bandwidth in (species). It is important to note their critical band is much wider in the area of 500 Hz than the vertical distance between the musical notes A5 and C5.

Their data was apparently obtained within the phonotopic range since they say the critical band width data was invariant with stimulus level over a range of 80dB. They obtained their difference limen data at 80dB SPL. They did not provide data points and noted that their data was amalgamated from a variety of sources.

---

[132]Stephens, S. (1973)Auditory temporal integration as a function of intensity *J Sound Vib* vol. 30, pp 109-126
[133]Moore, B. (1977) Introduction to the Psychology of Hearing. NY: Macmillan Press pp 62-72
[134]Zwicker, E. Flottorp, G. & Stevens, S. (1957) Op. Cit. pg 56

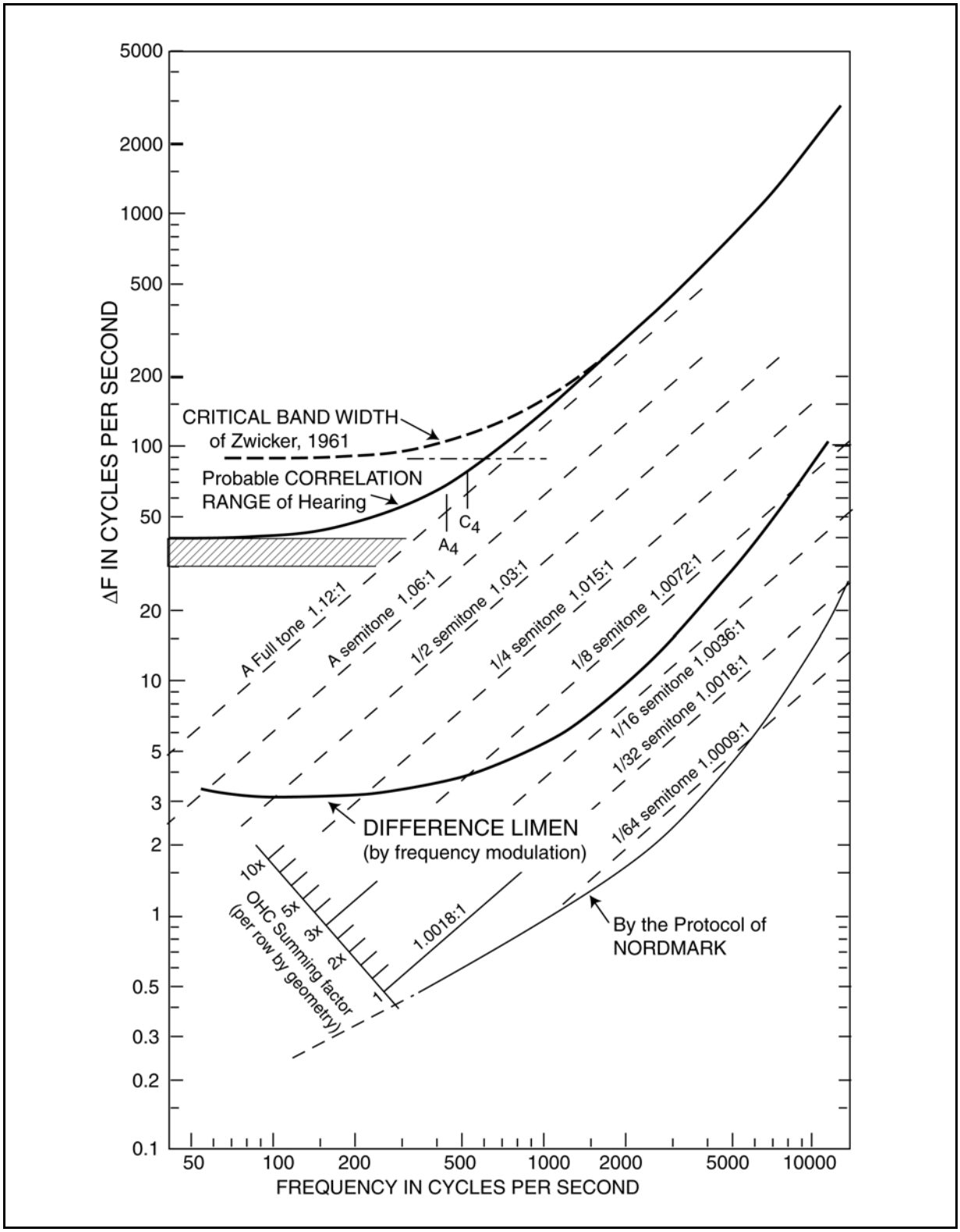


Figure 9.6.2-1 The just noticeable frequency difference and critical band widths for humans. The critical band is derived from multiple experiment protocols. The difference limen is derived from a low deviation frequency modulation experiment. Data from Zwicker, et. al., 1957 and Nordmark, 1968.

The cited paper is exploratory in character and leaves the needed parameters related to the acquisition of the data unclear. An example of this ambiguity will be illustrated below in describing their frequency modulation experiments.

> The reader is cautioned concerning their Table I. The values therein are illustrative only and do not apply specifically to the human ear. The center frequencies of each band are calculated as the arithmatic mean rather than the geometric mean. While this makes the table more readable, it is an unlikely case for a logarithmically organized sensory system.

The data related to the difference limen is perceptual in character. This suggests it may be a function of training. The critical bandwidth is also based on perception but is believed to be representative of the individual neural signal channels at or near the OHCs. [xxx OHCs or at the output of the spiral ganglia] The broader mechanical response measured at the membranes within the cochlear does not involve either the neural system or perception. Rhode says the slopes of the mechanical tuning curves are typically two to ten times less than that of the neural tuning curves[135]. He provides data and references supporting this proposition. "The tip of the neural tuning curve is much more pronounced than for the mechanical curve." In summarizing his data, he noted the generally observed reduction in the Q of the mechanical circuits at low frequencies (near the apex).

When overlaid with dashed curves denoting the center-to-center spacing of musical scales, the performance of the human ear, as reported by Zwicker, et. al., is clarified. While the critical bandwidth of the individual frequency channel in human hearing rises exponentially above 600 Hz, it shows a constant width below that frequency. This suggests the performance of the individual signaling channels change from those performing signal integration above 600 Hz and those providing signal reproduction below 600 Hz. The frequencies of middle A (440 Hz) and middle C (523 Hz) on the tempered musical scale are shown for reference. The psychophysical frequency discrimination sensitivity of the human far exceeds the bandwidth of the individual signaling channels. It is between 1/16 and 1/32 of the critical bandwidth over the frequency range explored in Zwicker et. al. This performance suggests the presence of an additional mechanism as proposed above. The psychophysical discrimination capability of hearing relies upon a differential signal derived from pairs of adjacent frequency signaling channels. This discrimination capability appears to be signal-to-noise dependent and therefore also a function of the stimulus intensity.

This data suggests the signal processing within the auditory system goes well beyond that suggested by Bell based strictly on the geometry of the outer hair cells[136]. The measured JND with frequency is well below his postulated capability of a full semitone, again suggesting an additional mechanism is present.

Their method of measuring the just-noticeable-difference frequency employed the frequency modulation of their reference oscillator. They gave a brief, and elementary description of the difference between frequency modulation (FM) and amplitude modulation (AM), but did not address phase modulation (PM).

> Phase modulation has the unique property that a fixed DC level signal can cause a repetitive

---

[135]Reference Data for Engineers (1985) Indianapolis, In: H. W. Sams

[136]Bell, A. (2003) Are outer hair cells pressure sensors? Basis of a *SAW* model of the cochlear amplifier *In* Gummer, A. Biophysics of the Cochlea. Singapore: World Scientific pp 429-435

change in phase applicable to the spacing (the period) between a series of action potentials. Thus, it also results in a frequency change when the spacing between multiple pulses is considered. However, the phase change can be for only one period. Under this condition, the frequency associated with the two adjacent pulses is not rigorously defined or easily measured. However the phase change is easily measured. The neural system uses the phase change between the first two action potential pulses in a stage 3 intensity-related-channel signal as an indication of danger. The signal is immediately passed to the **Alarm** mode circuits. The entire pulse train is passed to the **Awareness** and/or **Analytical** mode circuits for further interpretation. Nordmark provided a valuable discussion of phase versus frequency modulation and the methods of measuring each using the names phase frequency and group frequency[137].

Zwicker, et. al. did not demonstrate that they were using frequency modulation as opposed to phase modulation. They also introduce the difference between a "small" frequency modulation and a "small" amplitude modulation. The more appropriate terms in their context are a low deviation ratio frequency modulation and an amplitude modulation of any degree. The frequency components in the sidebands of amplitude modulated signals do not change with modulation depth. However, the sidebands of frequency and phase modulation change dramatically if the deviation ratio, the ratio of the change in carrier frequency to the baseband frequency, exceeds one. When this ratio is below 1.00, the modulation is called narrowband frequency modulation (or NBFM in that technology). For narrowband FM, the relationship between the sidebands in FM and AM are approximately as described in their paper. The actual temporal phase between the sidebands and the carrier are more complex than they indicated. For higher deviation FM, the sidebands of FM modulation are quite different.

A difficulty arises in interpreting their data when they describe "the rate of modulation is about 4 per second." If they mean the deviation ratio were four, the resultant sidebands would not even approximate the condition they described. Unfortunately, if they mean the oscillator frequency was made to swing by a given frequency, say 100 Hz, at a rate of four cycles per second, the resulting sidebands would also differ from what was described. They would consist of a group of sidebands closely spaced about the desired frequency change, 100 Hz, with individual sidebands within the group separated by four Hz. The total energy within each group, centered at 100 Hz each side of the carrier, would approximate what those authors had in mind. If similar experiments are repeated, future experimenters would do well to consult a text on modulation theory when developing their protocols. This will insure all of the energy in the sidebands occurs at the modulating frequency on each side of the carrier. However, this technique is fraught with another challenge. If the center frequency of the initial oscillator is not matched to the center frequency of the signaling channel, one of the sidebands will fall outside of the bandwidth of the signaling channel before the other one does. This will result in considerable distortion in the resulting signal within the neural system (distinctly noticeable in signaling channels below 600 Hz center frequency). A solution to this problem is to use a deviation ratio that causes the carrier frequency to disappear completely.

A major problem with the use of narrowband FM as an investigative technique (assuming a sinusoidal modulation in order to restrict the sidebands to a single pair) is the relatively small amount of time, the applied baseband signal remains at either of the limits of the frequency excursion during the test interval. Much of the time, the signal is at an intermediate level. For frequencies below 600 Hz, this can cause a problem in interpreting the data since the auditory

---

[137]Nordmark, J. (1968) Mechanisms of Frequency Discrimination *J Acoust Soc Am* vol 44(6), pp 1533-1540

system can track this variation in frequency during the test interval.

Stebbins provided data from Nordmark indicating a much greater frequency discrimination sensitivity for the human at 45 dB SPL[138]. The difference is about 3:1; a ratio compatible with the above discussion of the dwell time associated with the FM technique. This level is near the mesotopic-phonotopic interface and implies "linear" operation of the auditory system. Nordmark used an entirely different protocol. The frequency discrimination sensitivity shown is in much better agreement with the ability of well trained individuals (and is similar to that associated with a good piano tuner). This threshold level corresponds to about 1/64 of a semitone or a frequency difference of only 0.1% for frequencies above 600 Hz.

The above figure can also be evaluated using the discussion and conclusions in Rhode[139]. [xxx words, especially conclusions about width of mechanical versus neural versus perceived]

Roederer has provided a composite figure attempting to describe how a mixture of two tones is perceived under a variety of conditions[140].

9.6.2.1.1 A mathematical representation of frequency sensitivity

By combining the material of the previous section and the related material in **Chapter 8**, a composite presentation illustrates a variety of important relationships. **Figure 9.6.2-2** represents the performance of the multidimensional correlator of the PGN in circular coordinates. Each octave range is represented by a nominally equal distance along the cochlear partition. It has been associated with a loop of a spiral to provide a frequency continuum. The pitch of the spiral is unimportant. Nominal frequencies have been associated with each loop as indicated at the bottom of the spiral. The frequency advances exponentially based on two to the nth power. The lowest frequency shown is arbitrarily taken as 125 Hz. The scarcity of spectral channels below this value and the secondary order curvature of Hensen's stripe in the cochlear partition make analyses below this frequency more complicated. The highest relevant frequency is above 16 kHz but less than 22.6 kHz. The highest spectral channel of the human auditory system is not well documented. The notes of the musical scale are shown surrounding the spiral and show how the same note is associated with multiple frequencies (depending on their octave range).

The orientation of the figure relative to the named musical notes is arbitrary. The precise frequency associated with these names has changed irregularly throughout history and as recently as the 1930's.

Every radial in the figure represents a single perceived pitch. That pitch can be generated by a single tone at the intersection of the radial and a specific loop of the spiral. Alternately, it can be generated by a series of harmonics appearing at the intersections of the radial with one or more of the loops. The series need not be regular or complete. The series need not start with the lowest frequency intersection.

---

[138]Stebbins, W. (1975) Hearing of the anthropoid primates: A behavioral analysis *In* Tower, D. *ed*. The Nervous System, Volume 3. NY: Raven Press pp 113-123
[139]Rhode, W. (1978) Some observations on cochlear mechanics *J Acoust Soc Am* vol. 64(1), pp 158-176
[140]Roederer, J. (1995) Op. Cit. pg 33

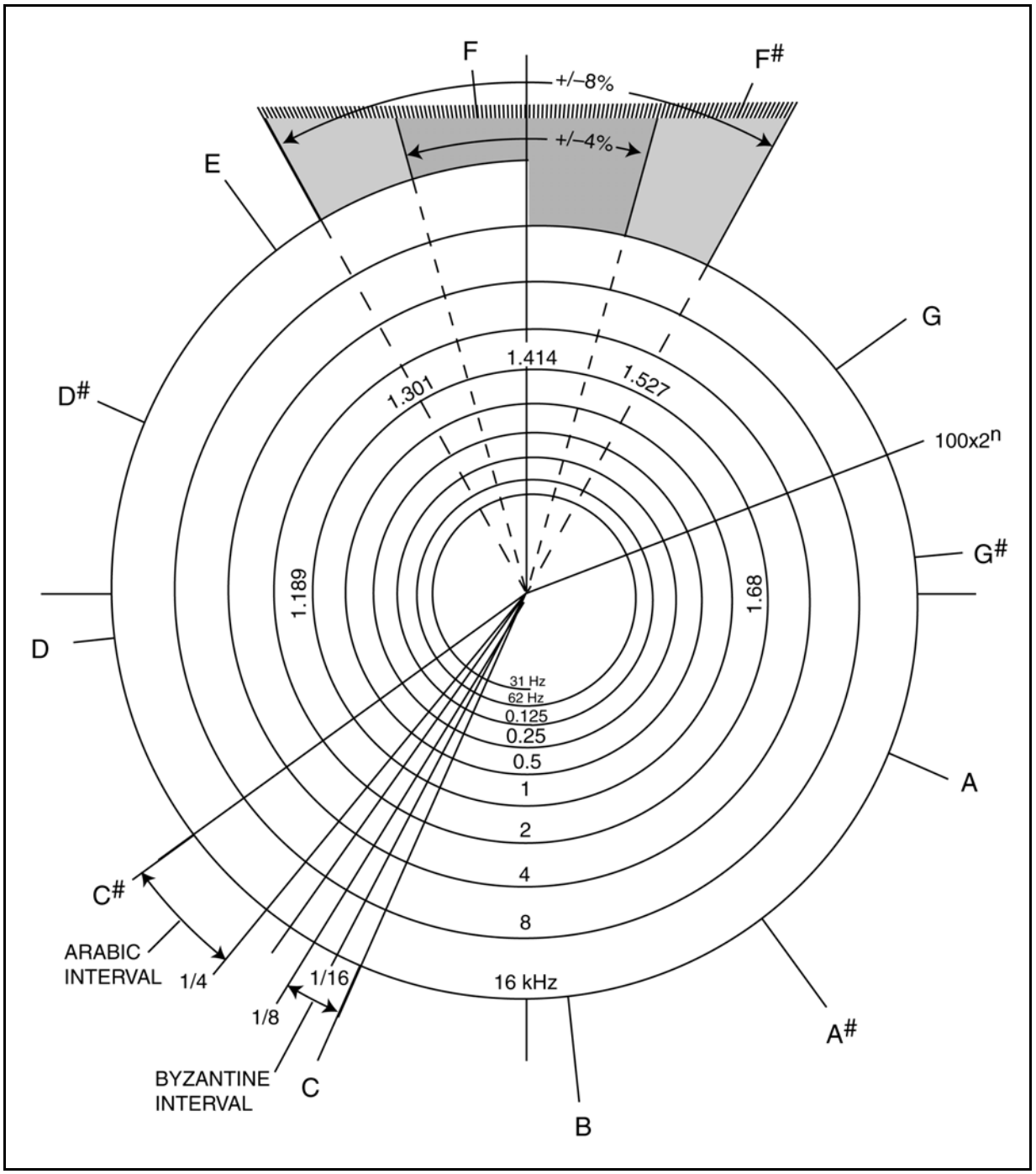


Figure 9.6.2-2 A structural description of frequency sensitivity of human hearing. The figure uses a spiral to represent the entire frequency range of human auditory sensitivity. All frequencies are shown in kilo-Hertz. The octave organization of hearing is illustrated along with its relationship to the musical scale. The upper quadrant includes two estimates of the "attention span" of human hearing along with a vernier representing 1/64th semitone steps. The two estimates and the vernier can be rotated about the center of the spiral. If the two estimates do in fact represent attention, the two estimates and the vernier could be rotated at the volition of the cognitive system of the brain. See text.

Presenting the human frequency space in a spiral format is not new. Shepard reviews the subject dating from 1855[141]. However, the earlier work was all empirical. This is the first discussion marrying physiology and psychophysics in the context of music, and the critical band.

Section xxx introduced the idea that the so-called critical band might be an expression of the attention span of the multi-dimensional correlator of the PGN. The upper quadrant of the spiral has been overlaid with two estimates of this "attention span."

> In the auditory environment, the place parameter is a linear function and a normal distribution is symmetrical about the mean. However, the frequency parameter is exponential with respect to place. As a result, a normal distribution about the mean of place is an asymmetrical distribution (a log-normal distribution) in the frequency domain. The use of the expression ±x% to describe a range about the mean of an exponential environment is inappropriate (unless the value of x is a very small number). The correct expression is k·mean on the "high" side and (1/k)·mean on the "low" side where k can have any value. The error at 8% is small, the correct values being –8% and +8.6%. However, the error grows rapidly with -30% pairing with +43%.

In both examples, the attention span represents the apparent noise bandwidth of the attention mechanism associated with the Pulvinar. This putative attention mechanism can be directed to any perceived pitch by the TRN or under the volition of the cognitive mechanism. Thus, the shaded sectors can be rotated around the spiral to center on any particular perceived pitch. The resulting noise mechanism is symmetrical about that pitch in place coordinates.

> It is important to note the noise bandwidth represented in this way is not defined by any filter with discrete edges. The equivalent filter is always represented by the edges of a distribution about the mean (the perceived pitch). The physiological system does not support a description of a series of critical bands with finite and stationary borders such as proposed by Zwicker and Zwicker et al.

Also shown in the figure is a vernier spaced at 1/64th of a semitone. This is the potential limit of frequency discrimination (segregation) by a well trained human ear using the Nordmark protocol. A semitone is 1/12th of an octave (the range between A and A#). Alternately, the limit may be four times coarser, 1/16th of a semitone using the FM protocol. In either case, it must be recognized that the sensitivity of the spectral channels of hearing cannot be represented as a constant level continuum. At the detailed level, the overall spectral sensitivity of hearing is serrated, in accordance with the spacing of the OHC along the cochlear partition and a yet unknown interpolation function providing information about stimuli presented at frequencies between the responses of two adjacent OHC (or groups of OHC). The potential mechanism providing such fine frequency discrimination in humans is yet to be described.

- - - -

The mathematical description of the above figure is unconventional in the context of much auditory research with tones, but not in the context of music. Psychophysical researchers have focused on a series of harmonic frequencies (a linear mathematical sequence) to investigate hearing. However, the musical theorist has focused on a sequence of frequencies based on an octave interval (an

---

[141]Shepard, R. (1982) Structural representations of musical pitch *In* Deutsch, D. *ed*. The Psychology of Music. NY: Academic Press Chap 11

exponential mathematical sequence). As in vision, the hearing researcher tends to focus on stimuli which are easy to create in the external environment while the theorist tens to focus on the perceived features of the stimulus. The differences between the researchers linear sequence, f, 2f, 3f, · · · and the theorists exponential sequence, f, f·2, f·2^2, f·2^3, · · · are profound. The difference is illustrated in **Figure 9.6.2-3**.

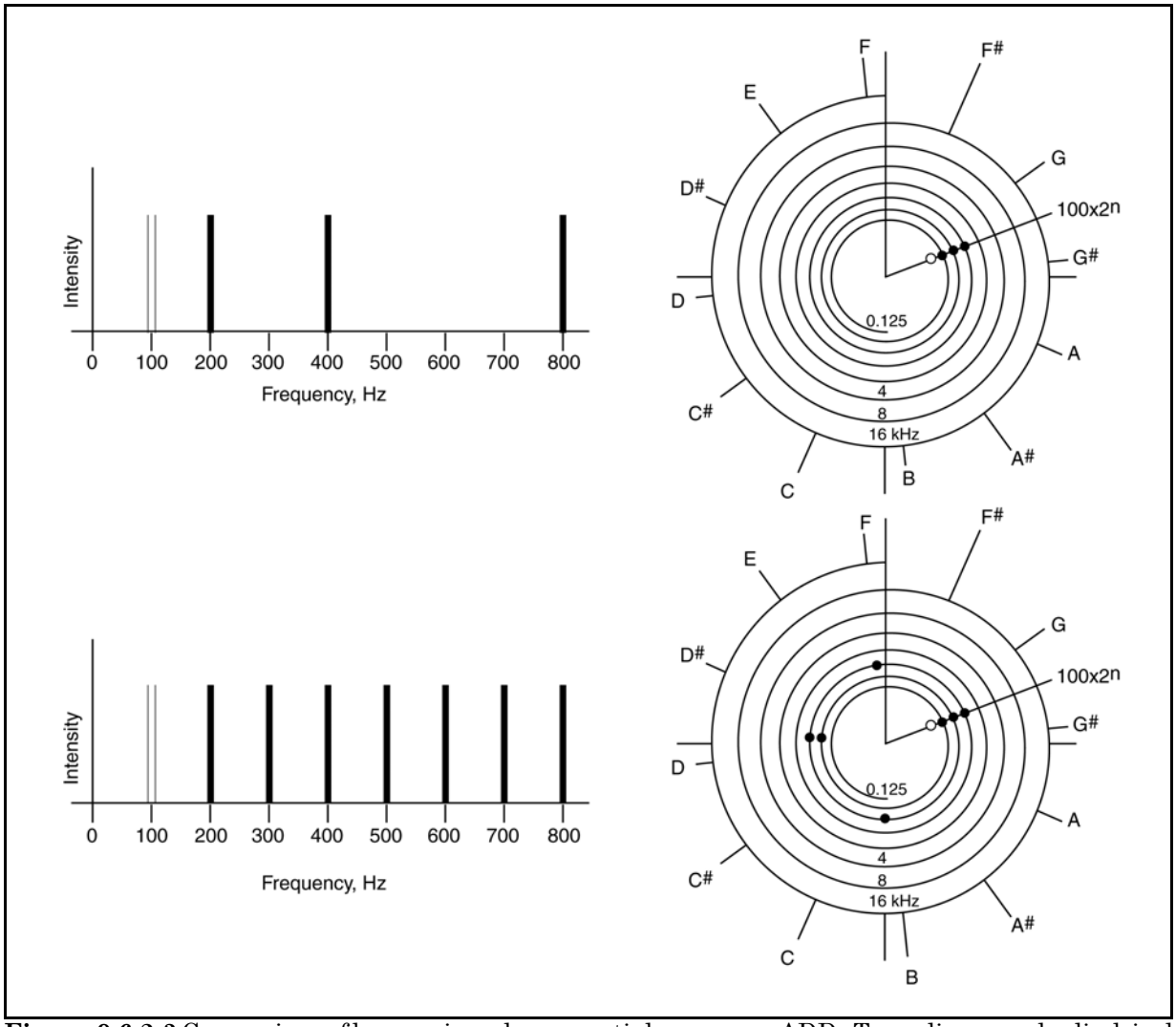


Figure 9.6.2-3 Comparison of harmonic and exponential sequences ADD. Top; a linear and cylindrical representation of frequency space with an exponential sequence overlaid. The human perceives a single note between G and G#. The fact that the 100 Hz component is missing from the sequence has no impact on the human's perception. Bottom; the same representation with a harmonic sequence overlay. The human perceives a chord consisting of the note between G and G# as well as a note near D, note near B and a note near F. The missing fundamental plays no role in this perception.

In the upper figure, an exponential sequence based on (100 Hz)·2^n is shown for n=1,2,3. In the cylindrical representation, these tones all represent, and are perceived as, a *single* musical note located between G and G#. Any discussion concerning the fundamental frequency at 100 Hz is

superfluous. The human perceives the single note. He/She may be taught that this note corresponds to a fundamental of 100 Hz but this is largely a meaningless relationship. The lower part of the figure shows a similar presentation for a harmonic sequence based on (100 Hz)·n for n=2, 3...8. The representation is cylindrical coordinates is quite interesting. The human perceives a note between G and G# as before. However, he also perceives a note near B, a note near D and a note near F. He perceives a *chord* not a *single* note.

In both these examples, the presence or absence of the fundamental is insignificant. Any question concerning a "missing fundamental" is less than academic, it is misleading relative to the phenomena involved. The neurological system is focused on a group of signatures that exhibit a unique relationship. They all reinforce each other in the exponential example, whereas they generate multiple perceptions in the harmonic case.

The use of an algebraic (harmonic) sequence instead of an exponential sequence in the research environment appears to be significant and inappropriate.

9.6.2.1.2 Rationalizing the circular and rectilinear acoustic ranges

[xxx this paragraph does not surface anything startling.]
A clear understanding is needed of how the spiral presentation above can represent the actual neural network used in the PGN. The two presentations in the above figure are equivalent (when the linear graphic is expanded) and linked by the exponential conformal transformation, w–>e^Z.

The features of this transformation are illustrated in **Figure 9.6.2-4**. The transform represents an exponential spiral. It has not been modified to incorporate the effects of the apical hook in hearing.

Both frames are projections of a sphere onto a planar surface. The circular projection is known as the stereographic projection. The rectilinear projection is known as the Mercator projection. A spiral at a constant angle with reference to the longitude lines on a sphere is known as a loxodrome. The projection on a stereographic projection appears as a similar spiral. The equivalent on the Mercator projection is the (straight) Rhumb line. In either case, the locus intersects the radials (or horizontals) at a constant angle. [xxx vert or horiz in final figure ?]

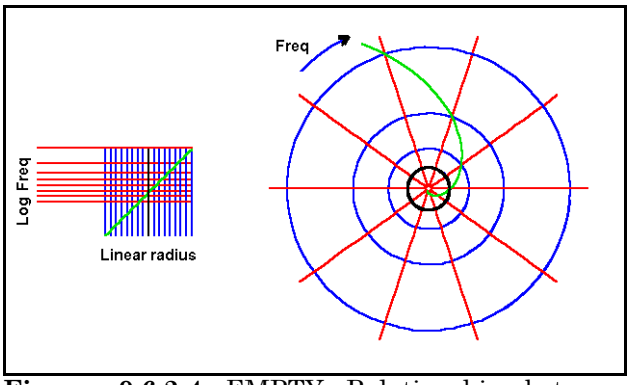


Figure 9.6.2-4 EMPTY Relationship between circular and rectilinear audio perceptions.

The mathematics of this transformation are developed in the literature[142,143,144]. Kober identifies the relationships between the polar and the rectilinear representations of the exponential transformation, w–>e^Z.

---

[142]Kober, H. (1957) Dictionary of Conformal Representations, 2nd Ed. Dover Publications Page 85
[143]Director, B. (undated) Riemann for Anti-dummies, Part 48: Riemann's Roots
http://www.wlym.com/antidummies/part48.html
[144]Jones, F. (2004) Honors Calculus III/IV. Rice University. http://www.owlnet.rice.edu/~fjones/chap6.pdf

The multiple revolutions of the spiral can be related to the spiral ganglia of the neural system. The overall locus suggests the neurons emanating from the cochlea are grouped into sets drawn from fixed length segments of the cochlear partition before being forwarded to the CNS. Each set appears to serve a (one-dimensional) length equal to one octave of frequency, nominally 2.3 mm in humans (**Section 6.2.3**). The output signals of each set are forwarded to the CNS where they are reassembled into a nominally rectilinear two-dimensional array. This processing is an example of "anatomical computation" in hearing. A remaining open question is what is the lowest frequency signal propagated by each set of signals.

The creation of these sets by the highly divergent spiral ganglia suggests the MOC pathway originates at the spiral ganglia, instead of the cochlear nucleus, and the neurons of the MOC pathway travel through the cochlear nucleus largely undisturbed.

The curvature of the cochlea, the length of the cochlear partition accessed by each set of spiral ganglia, and the precise alignment of the neurons forming the two-dimensional array in the PGM must be matched in an individual with significant musical talent.

9.6.3 Resolving the pitch versus harmonic frequency content dichotomy

Presenting the human spectral performance using cylindrical coordinates as above puts the discussion of pitch in a more understandable context that resolves the question of the "missing fundamental." It also resolves a variety of other terms introduced during the development of the mathematical theory of music. Several points are of major importance. The primary fact is that the CNS only receives signatures from the periphery that can be traced to locations along the cochlear partition. It does not receive any sinusoidal waveforms (no matter what the frequency or complexity of the stimulus). Second, the CNS does not make any attempt to calculate the least common multiple of the frequencies represented by the signatures it receives. Finally, humans perceive a pitch when ever one or more harmonically related tones are used as a stimulus. Some humans are taught to relate that perception to a specific frequency tone at an early point in their musical education. The specific tone is usually associated with the lowest common multiple of the tone(s) in the stimulus.

de Boer presented a paper with an ambitious title in 1977. He attempted to resolve the perceived pitch versus harmonic content dilemma[145]. It was not well received (comment on page 335 of his paper). The paper followed the wisdom of the time that the sinusoidal components of the stimulus were passed to the "central pitch processor" of the CNS for cross-correlation. The paper was highly mathematical and light on physiology. It followed an exhaustive analysis in 1976 that is of considerable historic and pedagogical interest[146]. However, it also failed to resolve the subject of pitch perception. The problem in these analyses appears to be the lack of a sufficiently detailed physiological model of hearing.

Yost presented **Figure 9.6.3-1** in an introductory text in 2000 that attempted to explain the pitch situation based on the common wisdom. The caption to his original figure has been changed

---

[145]De Boer, E. (1977) Pitch theories unified *In* Evans, E. & Wilson J. *eds.* Psychophysics and Physiology of Hearing. NY: Academic Press pp 323-335

[146]De Boer, E. (1976) On the "Residue" and auditory pitch perception *In* Keidel, W. & Neff, W. *eds*. Auditory System: Clinical and Special Topics. Vol. V/3 of Handbook of Sensory Physiology. NY: Springer-Verlag Chapter 13

intentionally.

[xxx See material in Yost 2000, pages 193 on First introductory text that addresses this vexing problem.]

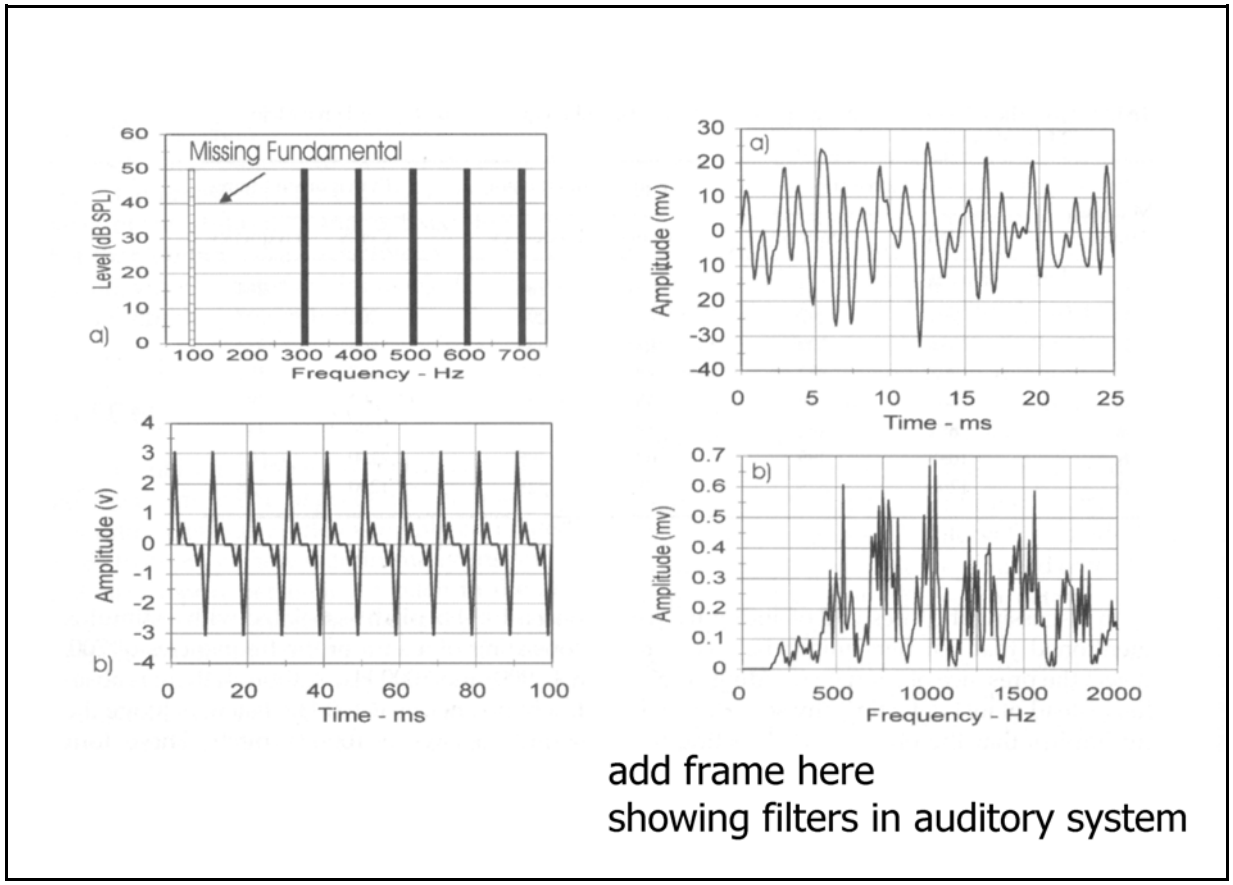


Figure 9.6.3-1 Examples of perceived pitch ADD. Upper left; the tones present are harmonics of 100 Hz. The perceived pitch can be described as half way between the musical notes G and G#. From Yost, 2000.

[xxx develop pitch in terms of periodicity pitch theory of Schouten (1938) in (page 8 of Deutsch.) In the more focused text of Deutsch, the subtle difference between the frequency of a single fundamental tone, and the periodicity of a complex waveform is developed. Both of these concepts are frequently labeled pitch. To do so clouds the discussion beyond any ability to clearly understand the concepts involved.

The definition of pitch found in dictionaries is so muddled as to defy description. Even for its use in music, a significant series of alternative meanings is given. Rasch & Plomp say, "*Pitch* is the most characteristic property of tones, both simple (sinusoidal) and complex." and "Pitch is related to the frequency of a simple tone and to the fundamental freuqncy of a complex tone." The problem here

is pitch is used in two different senses in each of these sentences. The pitch associated with a single tone is frequently not the pitch derived from a complex waveform consisting of multiple tones. This is particularly relevant if the complex waveform does not contain the fundamental used to generate the other tones in the waveform.

Efforts have also been made to use pitch to describe the perceived response to a physical stimulus, using what is called the *mel* scale[147]. Fortunately, this nomenclature has disappeared into the ashcan of history.

It is to the benefit of all to restrict the term pitch to the description of complex waveforms. Otherwise, investigators should be careful to provide adjectives to further define their use of the noun pitch. Here the periodicity of a complex waveform will be defined as its pitch (worst case, periodicity pitch). The quality of a single tone will be described by its frequency or musical designation, not by using the expression, pitch (worst case, fundamental pitch). Unfortunately, the use of the term pitch to describe a single tone has a long history. This usage cannot be avoided when reviewing the literature of others. As an example, the subject of "perfect pitch" in the vernacular generally relates to the ability of a person to identify a pure tone, not a complex waveform.

In this context, pitch in the musical sense has a range of about 20 to 5000 Hz, roughly the range of fundamental frequencies of piano strings and pipe organs. At higher frequencies, the human ear cannot perceive sufficient harmonics, the residue, to derive a pitch. At low frequencies, in the range of 10-50 Hz, the perceived pitch suffers from a "rattling character."

In the following summary, the quotations are from Rasch & Plomp.

"In this theory pitch is derived from the waveform periodicity of the unresolved higher harmonics of the stimulus, the *residue*. This periodicity does not change if a component (e.g., the fundamental one) is removed."

"In musical practice complex tones with weak or absent fundamentals are very common." This does not prevent the perception or determination of their pitch.

Figure 9.6.3-2

[147]Stevens, S. Volkmann, J. & Newman, E. (1937) A scale for the measurement of the psychological magnitude of pitch *J Acous Soc Am* vol. 8, pp 185-190

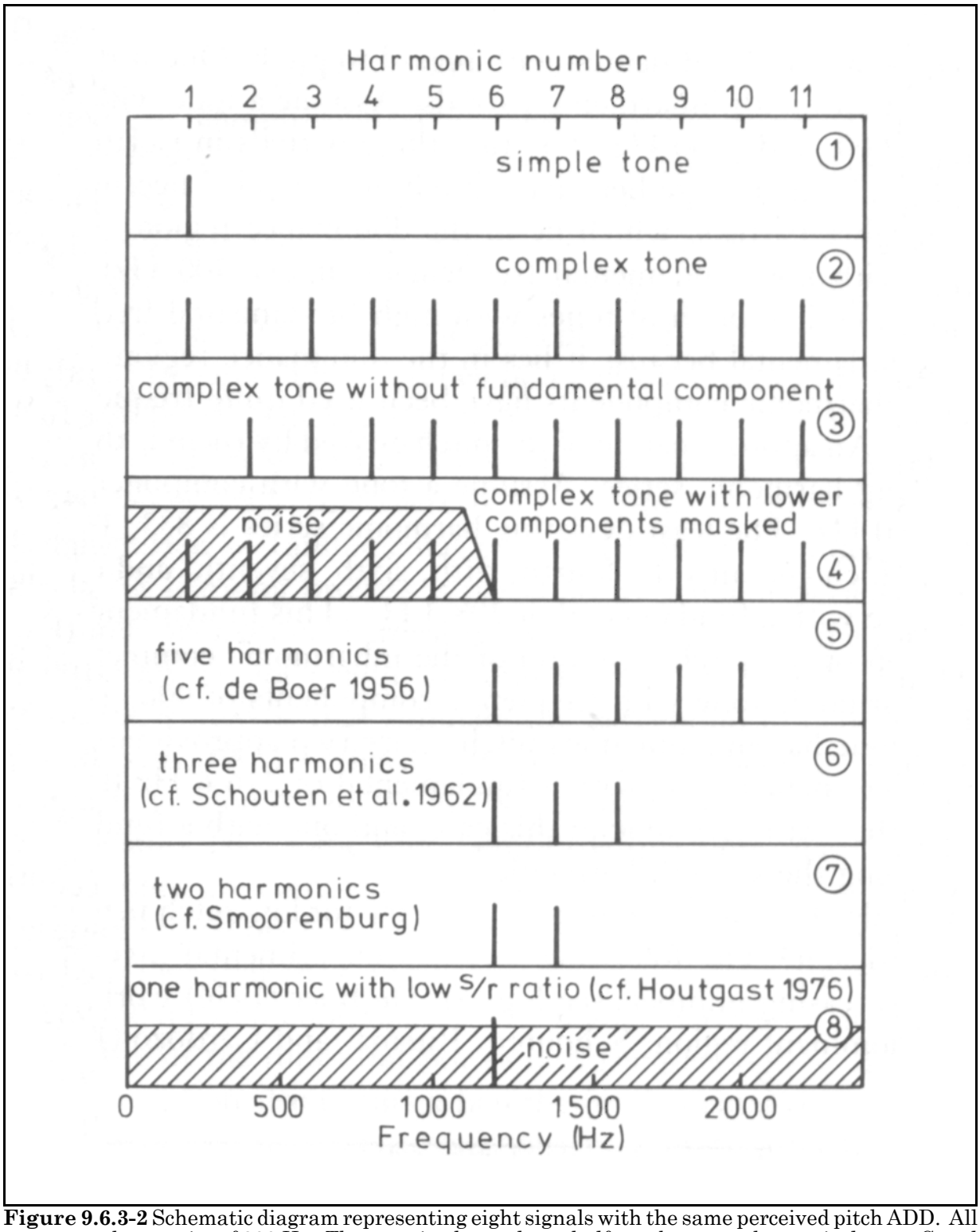


Figure 9.6.3-2 Schematic diagram representing eight signals with the same perceived pitch ADD. All tones were harmonics of 200 Hz. The perceived sound was half way between the musical notes G and G#. From Rasch & Plomp in Deutsch, 1982.

9.6.3.1 The range of pitch determination

Moore provided a relevant paragraph in 1977 (page 166). "The neurophysiological evidence in animals indicates that synchrony of nerve impulses to a particular phase of the stimulating waveform disappears above 4-5 kHz." The inflection point is actually near 600 Hz. Above this frequency, the amplitude of the relevant phase information is reduced considerably. He goes on. "Above this frequency our ability to discriminate changes in the frequency of pure tones diminishes, and our sense of musical pitch disappears. It is likely that this reflects our use of temporal information in the frequency range below 4-5 kHz." In a later sentence, he qualifies the above material. "For complex tones phenomena involving periodicity [sic] mechanism (e.g. the missing fundamental) are only observed for fundamental frequencies below about 1400 Hz, whereas for pure tones, the upper limit for the operation of temporal mechanism probably lies around 4-5 kHz."

9.6.3.2 On the subject of "perfect pitch"

Moore has provided a short discussion of highlights concerning the terms absolute pitch and tone deafness[148]. Ward & Burns develop the subject of absolute pitch in some detail[149]. Musicians speak of absolute pitch, perfect pitch or positive pitch when they purport to describe a person with the ability to identify a specific frequency of tone by its musical designation. This capability goes beyond the ability to identify the musical designation alone. It requires the person to also identify the octave the tone is associated with (440 Hz = A5 or A above middle C). As Ward & Burns stress, the ability of a person to identify a specific tone in isolation is intrinsically limited by the width of the critical bandwidth of his auditory system at that frequency. Thus, a person with perfect pitch is in reality accurate to about ± 6 %, a semitone when presented with a single tone at a frequency above 600 Hz. His performance is significantly poorer at frequencies lower than 600 Hz (**Section xxx** [xxx section on JND above]. Ward & Burns discuss a wide variety of attempts to teach "perfect pitch," etc. The conclusion appears to be that it is more an innate skill than a teachable skill.

Ward and Burns also note the statistical difficulty of defining the ability of a person to demonstrate their pitch ability. They frequently adjust an oscillator to frequencies such as 444, 432, 449, 438 and 882 Hz when asked to demonstrate the note A4. The major error associated with 882 Hz suggests they are able to define a tone within an octave range but not a specific tone. Houtsman has provided a summary discussion and graph based on the early work of Bachem[150]. The graph suggests that one with "perfect pitch" displays an error of about 3% after a period longer than one minute from hearing a recognized reference tone. He notes that those claiming absolute pitch typically make quick absolute identifications, accurate within a semitone, with octave confusions being the principal source of errors. Such findings would support the Stage 4 correlation procedure defined in this work for the fine tonal channel (**Section 8.1.6 xxx**).

Deutsch discusses the frequent proposals to consider pitch as multidimensional. She describes subdividing the subjective classification of musical pitch into a tone chroma within an octave and a tone height describing which octave (page 272). Shepard follows this concept later in the same work (Chapter 11, pg 384).

---

[148]Moore, B. (1977) Op. Cit. pp 157-158
[149]Ward, W. & Burns, E. (1982) Absolute pitch *In* Deutsch, D. *ed*. The Psychology of Music. NY: Academic Press. Chapter 14
[150]Houtsma, A. (1995) Pitch perception *In* Moore, B. *ed*. Hearing. NY: Academic Press Chapter 8 pp 288-289

Keidel has made some interesting observations concerning the change in the accuracy of a sense of "perfect pitch" in a trained individual as a function of body temperature[151]. Although the reference suffers from translation, it suggests there is a temperature sensitive master clock within the brain. This work takes a different view. It suggests the performance of the PNS is temperature sensitive, as demonstrated in a variety of cases; the propagation velocity of the SAW filter, the time constants of the adaptation amplifier networks, potential bias points of Stage 3 circuits, etc.

9.6.4 The perception of periodicity pitch EMPTY

While the subject of the missing fundamental has a long history of indecisive discussion in the hearing community, it is a fundamental feature of music theory.
[xxx look closely and use the words of Roederer on page 45 particularly & 43-50 in general]
[xxx Review Langner & Schreiner re signal periodicity and periodicity coding[152].]

Figure 9.6.4-1(A) provides the quintessential description of the periodicity pitch based on a common musical form the "perfect fifth." It is the description found within conventional hearing theory based on Fourier Analysis. As shown, the periodicity pitch is given by the time interval, τ_0, before the summation waveform repeats itself. The shape of the actual waveform, but not its period, is highly dependent on the phase relationship between tones f_1 and f_2. However, ***the CNS of the auditory system does not rely upon Fourier analysis, it correlates amplitudes based on addresses***.

Figure 9.6.4-1(B) shows the actual situation applicable to hearing for both the perfect fifth as well as a variety of other consonant musical intervals. [xxx B is missing]

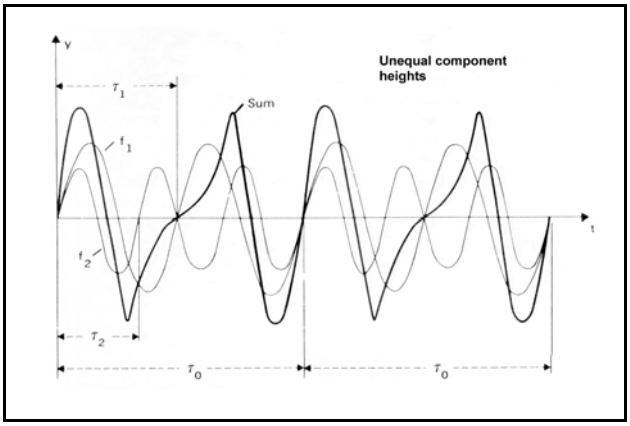


Figure 9.6.4-1 Superposition of two pure tones a musical fifth apart. For a given phase relationship between the two tones, f_1 and f_2, the left and right halves of the figure are identical. A periodicity pitch can be defined as the reciprocal of the period, τ_0. However, note the lack of any sinusoidal component at this frequency. Modified from Roederer, 1995.

Figure 9.6.4-2 is modified from Zwicker & Fastl. It clarifies the fact that a pitch can be perceived that can then be matched to a separately presented fundamental tone. However, the perceived periodicity pitch is not that tone. Based on their findings, the generation of a perceived periodicity pitch is limited to the range shown. If correct, it constrains the writings of others in a fundamental way. Many authors speak of the

[151]Keidel, W. (1974) Information processing in the higher parts of the auditory pathway *In* Zwicker, E. & Terhardt, E. *eds*. Facts and Models in Hearing. NY: Springer-Verlag pg 216+
[152]Langner, G. & Schreiner, C. (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms *J Neurophysiol* vol. 60(6) pp 1799-1822

concept of periodicity pitch without specifying the upper frequency limit of the mechanism involved. Zwicker & Fastl appear to limit the range of the mechanism to the frequency band defined loosely by the low pass characteristic of the adaptation amplifier. This suggests that an actual sinusoidal component of the stimulus must be present at the output of the sensory neurons if a periodicity pitch is to be perceived. It also suggests discussions of periodicity pitch should be restricted to frequencies where the lowest frequency present in the stimulus is below five kilohertz.

This limitation of periodicity pitch determination to frequencies in the range below five kilohertz is supported by Rasch & Plomp[153].

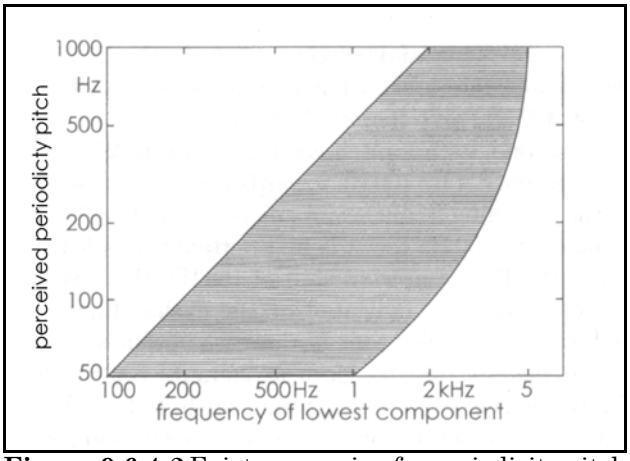


Figure 9.6.4-2 Existence region for periodicity pitch based on Zwicker & Fastl. Shaded area: Spectral region in which at least one spectral component must be found for a periodicity pitch to be perceived. The ordinate label has been changed to agree with the context of this work. Modified from Zwicker & Fastl, 1999.

9.6.4.1 Comparing the critical bands of man and cat

[xxx rewrite, integrator makes little sense in this context]
Figure 9.6.4-3, provided by Evans, compares the so-called critical bands as a function of CF for the human and the cat. The data suggests the Q–values for humans and cats are not substantially different. It also suggests the integrator breakpoint is similar between the two species.

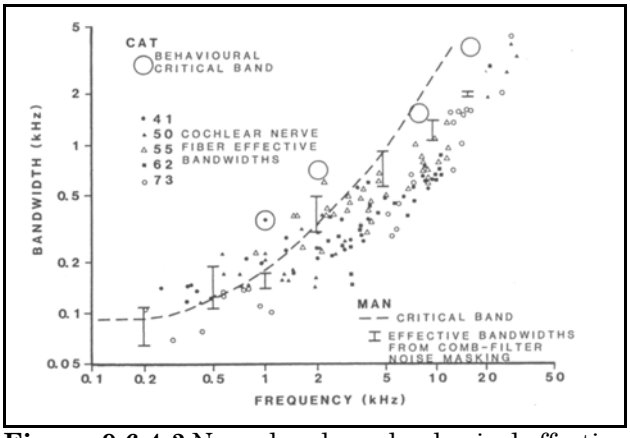


Figure 9.6.4-3 Neural and psychophysical effective bandwidth measure in cat and man. Adapted by Harrison, 1988, from Evans.

[153]Rasch, R. & Plomp, R. (1982) The perception of musical tones *In* Deutsch, D. *ed*. The Psychology of Music. NY: Academic Press pg 7

9.6.4.2 The generation of perceivable beats due to frequency differences

[xxx pgs 14-17 in Deutsch]
[xxx consolidate in 9.7 ?]
The experimental literature contains a variety of papers attempting to describe empirically the results of the interference between two tones and/or their harmonics. The role of the critical bandwidth of the sensory system is not always identified in these analyses. The key fact is that any two tones, whether each tone is at a fundamental frequency or at a harmonic of a fundamental frequency, will interfere at the signal generation level (Stage 1) only if the two components fall within the critical bandwidth of a single signaling channel. Interference at later stages in hearing is possible but this interference involves different mechanisms than Stage 1 interference.

Stage 1 interference can occur under two significantly different conditions. The first case involves mixing, the process of adding two or more tonal components within a linear environment. No harmonics are generated in this process, only the sum and difference frequencies derived from the stimuli. This is the case described on pages 14-16 of Rasch & Plomp writing in Deutsch. They use the term linear transmission to describe the character of the channel.

The second case involves modulation, frequently called cross modulation. In this case, the two (or more) stimuli are multiplied together. The result is a set of new frequency components in the output near the frequency of each input, and near the sum and difference of each component pair, and . . . . The complete set can be very complicated and theoretically extends to infinity, although the amplitude of the components falls rapidly with order. Rasch & Plomp describe the output of this process as combination tones (pages -19) relating to a nonlinear channel. They suggest (in a simplified form) the major components of interest are the positive values given by the difference tone, ±(f – g), the second order difference tones, ± (2f– g) and ±(f–2g) and the third difference tones, ±(3f–2g) and ±(2f–3g).

The theoretical description of these tones within the hearing system can be very complex. However, most of the tones above 600 Hz will be integrated in the sensory neurons by the integrator associated with the collector of the adaptation amplifier. An important consideration is whether the nonlinear element is found before or after the piezoelectric transducer. In general, the piezoelectric transducer associated with a particular sensory neuron will operate linearly up to the point of physical destruction of the neuron and its stereocilia. If the nonlinearity were introduced before the SAW filter of the cochlea, only those components within the critical bandwidth of a particular sensory neuron would be processed by that neuron. In a broader context, only those components within the critical bandwidth of a sensory channel (resulting from Stage 2 processing) would be processed by the neurons of that channel. Finally, a perception of a difference frequency can be generated within the CNS that relates to the signals projected through different channels due to the original stimulus.

These different restrictions suggest why Rasch & Plomp note the following. “However, the correspondence between the relative amplitude predicted and the subjective loudness measured is far from perfect. Clearly, the phenomenon of combination tones is more complicated than can be described in a simple formula.” Very careful experiment design, using a comprehensive model, is required to demonstrate the effects of cross modulation in hearing.

9.6.5 Ascertaining patterns in hearing

Bregman has provided a very interesting discussion of the recognition of patterns in hearing based on psychophysical experiments[154]. He defines his work using the title "Auditory Scene Analysis" and proceeds to establish a framework within which to accomplish this broad task. He suggests the system can be divided into two major activities. First, the use of primitive processes of auditory grouping and second, by governing the listening process by schemas that incorporate our knowledge of familiar sounds. His book focused on the first activity.

Bregman suggests the primitive processes are quite varied and the system initially breaks the incoming energy down into a large number of separate analyses. These correspond to the stage 2, signal processing, and stage 4, signal manipulation, tasks of this work. Further, he suggests these are broken down into temporal domain tasks and frequency domain tasks. These correspond to the LOC channel tasks and the MOC channel tasks of this work.

Bregman then assigns the label, "sequential integration" to the task of extracting information (primarily patterns) from the MOC channel signals and the label "simultaneous integration" to the task of extracting information (again, primarily patterns) from the LOC channel signals.

After establishing the above framework, he proceeds to provide the results of a large variety of experiments and his interpretation of their results, only in text form unfortunately. He then closes with a discussion of attention and its limitations when interpreting complex mixtures of sequential and simultaneous patterns. He seems to indicate the attention span of human hearing is largely limited to the results of either the sequential or simultaneous tasks at a given instant. He notes, "According to this view, only one stream of sound exists at a time, the one you are paying attention to. There is no such thing as a second grouping of perceptual evidence that is structured even if you are not paying attention to it. This is one important way in which this class of theory differs from the theory of primitive scene analysis."

This work takes a different perspective. It suggests both the results of sequential and simultaneous integration (the results of his primitives analysis) are provided to the saliency map by the stage 4 signal manipulation (information extraction) circuits, but the stage 5 cognition capability can only apply its attention to one of these results at a given time. It is suggested by this work, that complex patterns of music involving both temporal and spectral sequences can be appreciated simultaneously. In this author's experience, it is suggested the tone poems of Dvorak are appreciated in this way.

His set of relatively simple examples will not be repeated here. They are largely in agreement with the material presented above. His explanation of the "missing fundamental problem" is the historical one.

> Bregman is struggling with a lurking variable in his discussion of the missing fundamental problem. He is basing his analyses on non-tonal language speakers. They are educated to focus on discrete musical notes, regardless of octave, and not musical frequencies. As developed above, if he examined tonal language speakers, or musicians brought into music at a very early age, his results would be quite different. This fact is demonstrated by the

[154]Bregman, A. (1990) Auditory Scene Analysis. Cambridge, MA: MIT Press

preponderance of orientals exhibiting “perfect pitch.”

9.6.5.1 An example of sequential-simultaneous integration

Bregmen did address one experiment of particular interest. Three tones; a single tone, A, and a pair of tones, B & C where the pair was presented concurrently but alternated with A. It was found that if B and C started and ended at the same time, they tended to be treated as two components of a single complex tone, BC, that was perceived as righ in quality. On the other hand, there was a tendency to tret B as a repetition of A whenever A was close in frequency to B. B seemed to be the object of a rivalry. When it was captured into a sequential stream with A, it was less likely to be heard as part of the complex tone, BC. Conversely, when it was captured by C and fused with it, it was less likely to be heard as a repetition of A. It seemed that sequential grouping and spectral grouping were in a competition that served to resolve competing evidence concerning the appropriate grouping of sensory material.”

9.6.6 Performance in the face of masking

[xxx reference framework in figure 5.2.3-2 The fundamental test sequence showing pre peri and post time intervals in both masking and adaptation]
Without a firm model of the hearing modality, it is difficult to describe the phenomenon of masking precisely. **Section 6.2.1** developed the concept of masking as a tool in hearing research. It also restated the conventional wisdom concerning masking based on the model and theory of this work [xxx did it?].
- - - -

9.6.6.1 Background related to Masking

The hearing modality relies upon two separate and distinct signaling channels, the MOC (tonal information) and LOC (pulse & transient information) paths to deliver signals to the Thalamic level of the brain (where the predominant perception of both music and speech occurs). There is additional extraction of information by areas 41 & 42 of the cerebral cortex. There is also some signal processing within the lower brain (in the mini-brain of Lorento).

So, the question is; which one of these paths and/or processing centers is being masked by a particular type of masking signal?

Here again, a null hypothesis is needed in order to develop a sophisticated laboratory protocol. Otherwise, all laboratory work must be considered exploratory rather than basic research.

- - - -

The following discussion relies upon the mathematical concept of orthogonal functions[155]. An arbitrary waveform can always be described mathematically as the summation of a set of orthogonal functions with appropriate coefficients. Mathematically, the appropriate expression is;

[155]Storer, J. (1957) Passive Network Synthesis. NY: McGraw-Hill pp 262-276

$$F(x) \simeq F_N(x) = \sum_{0}^{N} A_n f_n(x)$$

where the function $F_N(x)$ approximates the desired function, $F(x)$, and is given by the summation.

Any set of orthogonal functions, $f_n(x)$ can be used in this representation. The error between the actual function and its series representation is easily found. The error decreases monotonically with the increase in N.

The most common set of orthogonal functions is the family of sinusoidal waveforms described as individual tones. This is the only family of periodic solutions that satisfy the orthogonal criteria. However, there are many other families. Those families useful in approximating a band limited signal, F(x), are known as the Gegenbauer polynomials. A well known series within this family are the Hermite and Modified Hermite polynomials. The Hermite polynomials, $He_n(x)$, are a classical orthogonal polynomial sequence that arise in probability theory, in physics, in numerical analysis and in systems theory in general. They are related to the Poisson Distribution. The properties of the modified Hermite polynomials have been developed by Chang & Wang using different notation[156]. Note the form of the derivatives in the family. A caution; there are a variety of functions named modified Hermite polynomials in the literature.

[156]Chang, R. & Wang, M. (1986) The properties of modified Hermite polynomials and their applications to functional differential equations, *J. Chinese Inst. Engrs.* Vol 9 , pp 75-81.

The modified Hermite polynomials, introduced in [3], are defined as

$$\tilde{H}_k(x) = \alpha_k H_k(\beta x), \tag{9.1}$$

$$\alpha_k = \frac{1}{(2\beta)^k k!}, \qquad \beta \in R. \tag{9.2}$$

They satisfy Bonnet's recurrence relation

$$x\tilde{H}_k(x) = (k+1)\tilde{H}_{k+1}(x) + \frac{1}{2\beta^2}\tilde{H}_{k-1}(x), \tag{10.1}$$

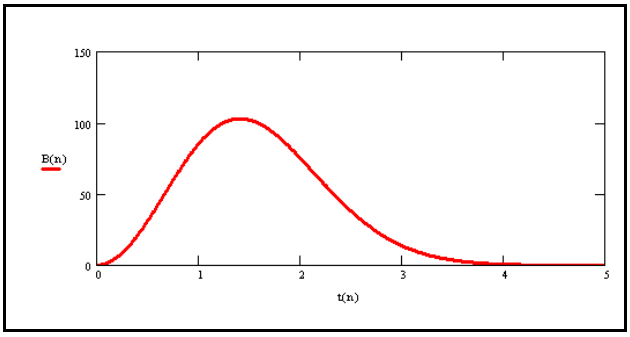
and for the derivatives we have

$$\tilde{H}'_k(x) = \tilde{H}_{k-1}(x), \qquad \tilde{H}''_k(x) = \tilde{H}_{k-2}(x). \tag{10.2}$$

From Chang & Wang, 1986

Figure 9.6.6-2 The modified Hermite polynomials.

The modified Hermite polynomials are finding current use in impulse-based radio communications systems. The first member of the family has the form;



- - - -

The two distinct signal processing channels (LOC and MOC) strongly suggest separate and distinct information extraction mechanisms. Conceptually, it is easy for most investigators to envision information extraction related to sinusoidal waveforms or tones. These tones are mathematically orthogonal and therefore uniquely separate from each other. They appear to underpin the operation of the MOC paths. However, there are many other sets of waveforms that are orthogonal. Those of the band-limited type of interest in hearing are known as Gegenbauer polynomials. As an example, the set of modified Hermite polynomials are orthogonal. It appears a set of this type of functions underpin the operation of the LOC signal paths. Each of the individual band-limited channels formed along Hensen's stripe, as a result of its curvature and the Marcatili Effect, appear to be optimized for detecting a different member of an orthogonal

family of waveforms. The modified Hermite polynomials provide an easy visualization of such a family, but there are many other potential families that could be employed.

To determine the actual family of waveforms used within the LOC channels, it will be necessary to determine the method of signal processing employed in the stage 2 signal processing of hearing. Implementation of this method appears to be a major responsibility of the cochlear nucleus of the hearing modality. To investigate this signal processing, it will be necessary to record stage 2 analog waveforms from the cochlear nucleus, not just the stage 3 action potentials typically recorded.

It is obvious that the broadest band IHC channel (near the base of the cochlea) is most effective at sensing the shortest duration impulses, such as the click heard at the instant of a nearby lightning strike. The narrowest IHC channel (near the apex) is most effective at sensing a long pulse , such as the rumble of thunder associated with a far away lightning strike.

The above analysis and the more commonly recognized tonal performance of the hearing modality suggests, the OHC are uniquely capable of analyzing the frequency content and the IHC are uniquely capable of analyzing the impulse content of acoustic stimuli. From these facts, it becomes clearer that the OHC are very capable of analyzing the vowel sounds of speech while the IHC are particularly good at analyzing the consonant (fricative) sounds of speech. As a result of employing these two channels, the CNS, and particularly the auditory perigeniculate nucleus (PGN) is presented with information quite similar to the typical auditory spectrogram recorded in the sound laboratory. Since the neural signals are quantized by the dispersion process within the cochlea, the neural spectrogram formed within the PGN is also quantized. It is more like the more recent ones created using digital oscillographs than that of the early audio researchers using analog (mechanical and electronic) oscillographs.

There is one more degree of freedom introduced into the PGN spectrogram. As noted in **Section 8.4.2**, a Riemann Transform is used to divide the frequency spectrum at octave intervals. These intervals appear to be presented neurally in a separate dimension. As noted in **Sections 9.6.2 through 9.6.5**, the resulting "information space" is interpreted differently by tonal language speakers and atonal language speakers.

> As noted in Chapter 8, a separate Riemann Transform may be used in the inferior colliculus to create a similar array of 1/4–1/3 octave intervals in the IC.

- - - -

9.6.6.1 Designing Masking Experiments

[xxx masking versus eliminating the first fricative in speech communications]
[xxx separate experiments into tonal and impulse masking]
[xxx develop figure incorporating transversal filter and multiple individual outputs. Figures beginning with 4.3.2-2, figure 6.1.1-1 & 6.2.2-1, 8.2.3-3, chapter 8. Probably 9.2.1-1 & 9.2.2-1 are best place to start.]

While masking experiments are very easy to conceptualize, it is difficult to design meaningful experimental protocols. The dual character of the hearing modality requires very close attention to the protocol. Some signals are generated by the narrowband OHC and processed via the MOC

tonal path. Others are generated by the wideband IHC and processed via the LOC (time & intensity) path(s). Only in the CNS are these signals merged into a complete representation of the external environment. Thus, any protocol needs to be based on a null hypothesis as to what neural paths and signals are to be, or are being, masked and what the expected result will be. If the masking is designed to interfere with the stage 4 correlation within the PGN, that must be stated. If it is designed to interfere with later stage 4 information extraction (the meaning of a sound pattern rather than just the pattern) associated with xxx's area of the temporal lobe, that needs to be developed.

Section 9.2 described the actual circuitry involved in masking. This discussion surfaced a feature not previously discussed in the literature. Masking can occur within individual OHC and at any level of multiplexing. To interpret the empirical literature, it is first necessary to determine where in the system the masking is occurring or is designed to occur. By analyzing the circuit structure involved, it is possible to explain why some masking signals are more effective if they contain significant energy ***outside*** of the identified channel.

Masking by both pure tones and by noise is of interest here. Masking by a pure tone is largely a man-made phenomenon, found mostly in the laboratory, except for various alarm oriented stimuli found in the community. Natural masking usually involves a broadband background noise. Both of these types of stimulus interference were thoroughly documented by Fletcher[157] and others long ago.

Before proceeding, recall that *the auditory signals passed to the brain by the neural system are fundamentally signatures, **not tones***. These signatures are fundamentally different for pure tones and for noise. At the applied research level, there are also differences in the signatures of "white noise" (noise that has a power density that is independent of frequency), compared to "colored noise" (noise with a power density that is not constant with respect to frequency).

[xxx do the following chapters belong in 9.6.6.2 or will it address IHC as well?]
When discussing masking, it is again necessary to treat the OHC as integrators if the correct conclusions and relationships are to be derived from work in the laboratory. The combination of the frequency selectivity of the SAW filter at the input of the OHC and the integrator element at the output of the OHC further complicates the analyses of noise masking conditions. It virtually assures that the noise performance will not be determined by the features of the white noise usually used as a source in the laboratory. The noise signal generated within the neural system is invariably pink. It must be so treated mathematically.

Based on the model of this work, the stage 4 correlation processes occur in the spatial, not the temporal, domain. Lewis has described the problem of interference in a spatial correlator with reference to the recent animation efforts in Hollywood[158]. He notes the conventional cross-correlation function for template matching is very susceptible to "hot spots" in the table of values. Such hot spots can cause false correlation values. To minimize the effect of such hot spots, he proposes introducing a normalized cross-correlation function. While producing a less sharp correlation function, it tends to perform better a higher percentage of the time. It is quite possible the neural system also uses such simple alternatives to the simple cross-correlation function. Some of the data from experiments described in Section 9.5.3 may suggest whether the neural sytem follows the guidelines suggested by Lewis.

---

[157]Fletcher, H. (1953) Op. Cit. Chapter 10, pp 153-175
[158]Lewis, J. (1995) Fast normalized cross-correlation *Vision Interface* Ref 101

One feature described in detail by Fletcher was the relative ability of a low frequency tone to mask a high frequency tone versus the ability of the high frequency tone to mask the low frequency tone. This ability varied greatly with the frequency difference between the tones. Such differences were used to define both the characteristic frequency of the accumulation channels and the critical frequency (frequency group) of the summation channels of tonal hearing (**Section 6.2.2 xxx**).

The ability of band-limited white (thermal) noise to suppress normal hearing (based on threshold measurements) is well portrayed by Fletcher (pages 206-207).

Each level of multiplexing within the coarse tonal channels of hearing is saturable. The auditory system is designed with a variable gain first stage, the adaptation amplifier in each sensory neuron, followed by a system designed to operate at constant average signal level. Variations in signal level about this average are limited to a dynamic range of about 200:1 in voltage (46 dB). A further complication, the adaptation mechanism is time sensitive.

[xxx Moore, 1977 has provided material worthy of reinterpretation based on the theory of this work sections 9.5 & 9.6 . The work of this team has recently involved very complex masking configurations[159]. The results have not shown sufficient differences or been reported for enough individuals to achieve statistical relevance.

Plack & White have also carried out very complex masking experiments related to adaptation[160]. However, they have not employed any physiological model. Their work can be reinterpreted productively based on this work.

9.6.6.1 Masking by broadband noise EMPTY

[xxx don't implement these two subjects in the printed edition.]

9.6.6.2 Masking by other tones EMPTY

9.6.6.2.1 Masking by tones of similar frequency

[xxx data showing downturns Fahey & Allen] **Figure 9.6.6-4** displays the masking of a tone at xxx by a tone of frequency specified by the abscissa scale. xxx pg 94 in Roederer.]

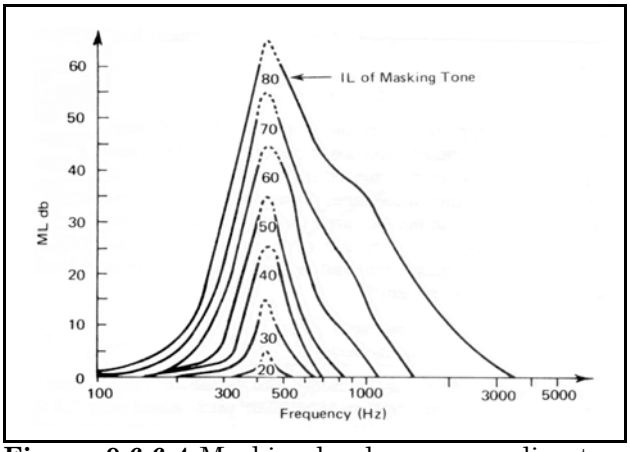


Figure 9.6.6-4 Masking levels corresponding to a pure tone of 415 Hz, for various sound level values of the masking tone. From Egan & Hake, 1950.

9.6.6.2.2 Masking by tones of dissimilar frequencies

[159]Fantini, D. Moore, B. & Schooneveldt, G. (1993) Comodulation masking release as a function of type of singal, - - - *J Acoust Soc Am* vol. 93(4) pt. 1, pp 2106-2115

[160]Plack, C. & White, L. (2000) Perceived continuity and pitch perception *J Acoust Soc Am* vol. 108(3), pp 1162-1169

The literature represents that low frequency tones mask high frequency tones effectively, whereas the efficiency is much less in the opposite case. [xxx add words]

9.6.6.3 Masking, suppression and interference as psychophysical phenomena

[xxx Fletcher chap 10]
Combining the lack of a viable model of the hearing system with the low cost of performing masking and suppression experiment, has led to the generation of a large empirical literature. Much of it appears conflicting, frequently due to problems with the protocols used.

The term masking is frequently used to describe experiments where a test signal occurs within the time interval of a second signal. Alternately, it is frequently used to describe experiments where a test tone occurs following the cessation of a second signal. This latter case will be labeled suppression instead of masking. Suppression of the test signal is a factor of both the intensity and the duration of the preceding signal. It is also a function of the time interval between the cessation of the preceding signal and both the occurrence and duration of the test signal.

9.6.6.4 Masking experiments in hearing

A large amount of laboratory results, and many references, related to masking appeared in Carterette & Friedman[161].

The previous work in masking is a prime example of exploratory research. It has created its own vocabulary based largely on procedural names for types of masking rather than on functional names. As Zwislocki has noted, the definition of maskin is strictly operational. These names reflect the empirical protocols used. The following three term were defined by Jeffress in 1970.

[xxx compare with glossary]
Remote masking– Masking by an interfering signal outside of (even above) the critical band containing the test tone.

Backward masking– Masking of a signal by a noise that occurs later in time than the test signal.

Forward masking– Masking of a signal by a noise that terminates before the start of the test signal.

That paper, and the paper by Scharf in the same compendium, illustrate the exploratory nature of the work of that time and earlier.

Scharf contributed additional terms.

Two-tone masking–Masking of a narrowband noise by two tones arranged symmetrically about the noise. This term describes a protocol used by Zwicker in 1954.

The expressions “narrowband masking” and “wideband masking” are also encountered frequently.

---

[161]Carterette, E. & Friedman, M.eds. (1978) Handbook of Perception, Volume IV: Hearing. NY: Academic Press Chapters 8 & 9

The expressions suggested by Zwicker & Fastl for these methodologies are not recommended[162]. Their choice of English prefixes can be misinterpreted based on the usage of others. It is suggested that their use of "pre-" be replaced by "prior to" and "post-" be replaced by "after" when referrring to the masker. This allows the terms pre- and post- to be associated with the maskee rather than the masker, as is the general practice. It also allows the maskee and masker to exist during different time intervals without introducing confusion.

9.6.5.5 Sensitivity suppression in hearing

[xxx 9.7.4.4 and 9.7.4.6 share the same figure 9.7.4-1 the recovery of hearing sensitivity -] Harrison has provided a discussion of the two-tone sensitivity suppression data available up to 1988[163].

Nelson & Schroder have provided recent psychophysical data on the suppression of a signal by a preceding signal in humans[164]. They frame their experiments in terms of temporal masking curves (TMC) under fixed-probe-level conditions. Their protocol is similar to that shown in **Section 5.2.6 xxx** except the test tones occur individually in each set of tests, thereby avoiding the internal interference caused by using a continuous series of test tones during one experiment. By using separate test tones, the procedure is identical to that used to determine the dark adaptation characteristic in vision. In fact, if the curve is inverted, the kaum adaptation characteristic of hearing is obtained with precision. They used a three-interval forced-choice (#IFC) adaptive test procedure where the test tone level was maintained at 10 dB SL and the exposure (or suppression) tone was varied.

Nelson & Schroder use the term compression without providing a precise definition. This work will continue to restrict the term compression to time invariant modifications to acoustic signals, typified by the logarithmic conversion of signal current to signal voltage at the pedicle of the sensory neurons. Any time variant features of the same signals will be characterized as an adaptation.

Figure 9.6.6-5 presents one frame of their Figure 1. It has been modified to include zero time and the necessary changes in the intensity scale. It also show two theoretical lines. The first is a simple exponential passing through zero time. The second is an approximation accounting for the non-rectangular shape of the suppression signal and the test probe signal. The test probe tone was F_p = 4000 Hz. The figure shows the recorded signal following the application of an on-frequency (F_m = 4000 Hz) and an out-of-band suppression tone (F_m = 2400 Hz = 0.6 x F_p) for 200 ms. The dashed line representing the median line for an ensemble of out-of-band tests has been extended to cross the vertical axis.

---

[162]Zwicker, E. & Fastl, H. (1999) Psychoacoustics: Facts and Models, 2nd Ed. NY: Springer pg 78
[163]Harrison, R. (1988) The Biology of Hearing and Deafness. Springfield, Il: Charles C. pg 79-81
[164]Nelson, D. & Schroder, A. (2004) Peripheral compression as a function of stimulus level and frequency region in normal-hearing listeners *J Acoust Soc Am* vol. 115(5), pp 2221-2233

Nelson & Schroder describe the on-frequency suppression as trisegmental. However, it is suggested that the first order mechanism, involving the first four stages of neural hearing, follows a simple exponential. Any additional variation is presumed to be due to the protocol of the experiment and possible perceptual and cognitive activity within Stage 5 of the hearing system. Their use of "10 ms raised-cosine rise and decay times" clearly impacts the perceived response at times near zero. No estimate was given of the time involved in any cognitive activity related to their protocol. However, the wide statistical variation among their subjects was indicated in their Figures 2 & 3. Many of the curves do not show the degree of curvature near 40 ms shown in the above figure.

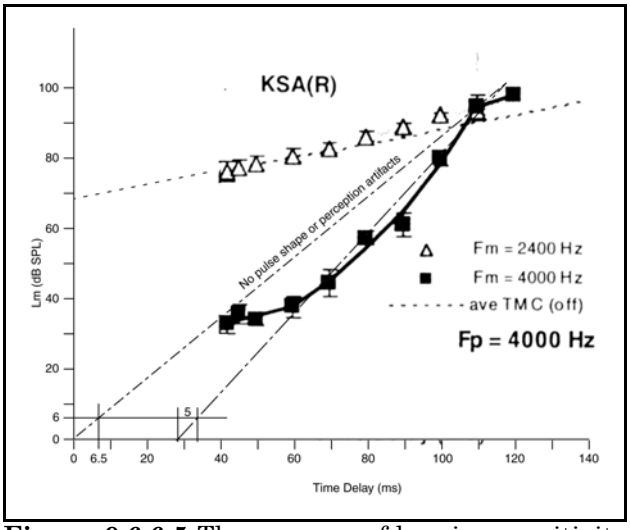


Figure 9.6.6-5 The recovery of hearing sensitivity following suppression by a pure tone. Lm; level of 200 ms suppression tone. Solid squares; response for a 20 ms, 4000 Hz, probe following cessation of a 200 ms, 4000 Hz adapting tone. Open triangles; response for the same probe following cessation of a 2400 Hz (out-of-band) adapting tone. Three theoretical constructs are also shown. See text. Data from Nelson & Schroder, 2004

Locating the original justification for using trapezoidal, instead of rectangular, shaped acoustic signals in hearing research is difficult. It may have become a custom following the limited capability of early signal generation equipment. The use of a trapezoidal stimulus envelope introduces a significant additional complexity into the data analysis function. Impulse, step and rectangular shaped stimuli enjoy unique mathematical characteristics not shared with trapezoidal waveforms.

Future experimental protocols should avoid the use of trapezoidal, or other extended edge stimuli envelopes wherever possible.

The level of adaptation, about 40 dB in their experiments, remained relatively constant regardless of test frequency between 500 Hz (the beginning of the integrating range of acoustic signals) and 6000 Hz. The full data set confirms that adaptation is performed independently within each frequency band of the sensory system (within the OHC and IHC of Stage 1 according to this work).

Nelson & Schroder offered no analysis of the effect of the broadband signaling channels associated with the IHC neurons, versus the OHC tonal channels, since the experiments were psychophysical.

9.6.6.6 Interference in the high level channels of hearing

[xxx reoutline]
[xxx divide into two sub-sections or move binaural part out]
Interference is defined by many authors as a phenomenon involving binaural hearing. Others describe interference as it relates to cross-modulation. Separating the subjects of binaural and monaural interference, and interference from adaptation resolves many of the conflicts found in the literature of binaural hearing. A particular point of contention is whether interference varies with time. This has led Stokinger, et. al. to assert that, based on their experiments, they have concluded that monoaural perstimulatory loudness adaptation does not exist in most subjects with normal

hearing. This work takes a different view (see **Section xxx**).

Bray, Dirks & Morgan have provided data on interference (using the term adaptation) in human hearing under dichotic (binaural) conditions and a brief list of citations[165]. Their experiments involved periods longer than the adaptation time constants of hearing and stimulation at the 70 dB SPL level (in the phonotopic regime- - adaptation as a mechanism is active). Thus, the data relates to essentially steady state conditions in spite of the fact that pulse techniques were used as a test probe. Under these conditions, their findings appear correct. "Under these conditions we have confirmed that loudness adaptation (here interference) as a peripheral phenomenon free from central effects does not occur."

Any effect of interference appears to be due more to a question of redirection of attention in Stage 5 rather than any physiological interference phenomenon as a result of signal manipulation prior to Stage 5.

9.6.6.7 The time constants associated with kaum adaptation

[xxx 9.6.3.4 and 9.6.3.6 share the same figure 9.X.X-1 the recovery of hearing sensitivity -]
The OHC and IHC are bathed in a liquid nutrient bath that supplies their electrostenolytic mechanisms. The sensory neurons of vision are embedded in a tissue, that is expected to exhibit much lower diffusion rates. It would be expected that the time constants related to the power supplies in hearing would be considerably shorter than in vision.

The dash-dot construction lines of the above figure provide estimates of the time constant of the kaum adaptation mechanism in human hearing based on psychophysical experiments. The "artifact free" construction line suggests this time constant is 6.5 ms, while the line to the right suggests this time constant is about 5 ms.

Young & Sachs found a significantly longer time constant related to kaum adaptation in the cat based on electrophysical experiments (measurements at node F). Their data, using a different protocol is discussed in **Section 5.2.6** xxx. It gave a time constant of about six seconds.

The three order of magnitude difference between these numbers suggests more research is necessary. The 6 ms number probably applies to a transient due to stray capacitance within the sensory or propagation sections of the neural system. The six second value is more likely representative of the true kaum adaptation time constant of the adaptation amplifier within the sensory neurons.

9.6.7 The automatic adjustment of sensitivity in the auditory system

As recently as 2008, Allen et al. have described what they call "a basic paradox[166]." "How can the basic cochlear detectors (the IHCs) have a dynamic range of less than 50 dB (a factor of 0.3×10^2), and yet the auditory system has a dynamic range of up to 120 dB (a factor of 10^6)?" A

---

[165]Bray, D. Dirks, D. & Morgan, D. (1973) Perstimulatory loudness adaptation *J Acoust Soc Am* vol. 53(6), pp 1544-1548

[166]Allen, J. Regnier, M. Phatak, S. & Li, F.(2009) Nonlinear cochlear signal processing and phoneme perception *In* Cooper, N. & Kemp, D. *eds*. Concepts and Challenges in the Biophysics of Hearing. Singapore: World Scientific

comprehensive review of the overall hearing modality shows the operation of the piezoelectric transducers provide an answer to Allen's paradox.

[xxx rewrite the following, replace with piezoelectric replacement equation.]
The data supports two major mechanisms of sensitivity adjustment within the auditory system. The most prominent and well documented is that associated with the phenomenon of adaptation. The sensitivity is reduce automatically through a unique mechanism, avalanche gain as a function of the collector potential, in the adaptation amplifier of the sensory neurons. The mechanism is also found playing a major role in vision. The mechanism is a specialized form of negative internal feedback. The second mechanism is more easily recognized. It involves external feedback through the superior olivary complex and probably the precision acoustic servomechanism under the supervision of the thalamic reticular nucleus. As mentioned in Art & Fettiplace, it is quite possible the exterior feedback loop serves to desensitize the hearing system to the sounds associated with chewing. It has also been suggested that the loop may include a reflex element. This would involve a shorter loop using the afferent signal paths only as far as the inferior colliculus. This loop would be operative even in the presence of anaesthesia.

9.6.7.1 The adaptation performance of the auditory system

While the electrophysiological data available describing the adaptation performance of the human system is sparse, the data available for the cat is quite good. The data of Smith is good and accompanied by a thorough discussion, but his experiments covered a limited range and are difficult to relate to an absolute calibration scale[167]. A later paper by Smith & Brachman, using gerbils, is of less value[168]. The paper of Young & Sachs, based on cats, is much more comprehensive[169]. It will be shown that the data of Young & Sachs is very similar to that for the visual system, again confirming the similarity in architecture and circuitry between the two systems. Young & Sachs make an interesting statement appropriate to the 1970's. "It is not possible to specify the nature or the locus of the physiological changes underlying the post exposure decrease in responsiveness which has been described."

[xxx review data from both authors] [put in graph that looks like visual graph and discuss whether it is first, second or third order]

Elliott & Fraser provided a good discussion of human adaptation up through 1970 based on psychophysical data[170]. They stressed the many methods and techniques of acquiring psychophysical adaptation data. Their introductory remarks attempted to highlight both the many types of adaptation encountered and the many descriptive labels used by different investigators. Their figure 11 is pedagogical in character. It does not show the transient characteristics of either the test tones or control tones. Harrison provided some data on human adaptation in 1988 but it does not appear to be fully developed[171]. Neither author was able to explain the underlying mechanisms controlling the various facets of adaptation. Using the description in **Section 5.4 xxx**,

---

[167]Smith, R. (1977) Short-term adaptation in single auditory nerve fibers: some poststimulatory effects *J Neurophysiol* vol 40, no 5, pp 1098-1112
[168]Smith, R. & Brachman, M. (1980) Operating range and maximum response of single auditory nerve fibers *Brain Res* vol 184, pp 499-505
[169]Young, E. & Sachs, M. (1973) Recovery from sound exposure in auditory-nerve fibers *J Acous Soc Am* vol 54, no 6, pp 1535-1543
[170]Elliott, D. & Fraser, W. (1970) Fatigue and adaptation *In* Tobias, J. *ed*. Foundations of Modern Auditory Theory. NY: Academic Press Chapter 4
[171]Harrison, R. (1988) The Biology of Hearing and Deafness. Springfield, Il: Charles C. Thomas pp 164-165

the nature of the changes can now be described in considerable detail. Adaptation can be described as prestimulus, peristimulus and post stimulus. Within these broad titles, major sub-classifications exist. During the peristimulus interval, both an attack time constant, a first order decay time constant and a second order decay time constant must be recognized. Since the adaptation amplifier is conducting during this interval, the first order time constant is variable function of the stimulus level. During the post stimulus or kaum-adaptation period, both a first and second order time constant must be recognized.

The concept of fatigue is frequently discussed along with adaptation. Elliott & Fraser were unable to offer any clear distinction between fatigue and adaptation except to suggest fatigue is usually encountered at moderate to high stimulus levels and adaptation is usually encountered at moderate to low stimulus levels. With respect to auditory signaling (as opposed to cochlear damage), the two phenomena, adaptation and fatigue, are the same.

A significant complication in hearing is the tendency to make binaural adaptation measurements. The investigators do not always recognize or discuss the high level of control-ear stimulation present due to bone conduction of a stimulus applied to the test ear (Elliott & Fraser, page 141). Furthermore, the same investigators do not always consider the attack and decay time constants of the signal applied to the control-ear when developing these same parameters in the test-ear.

9.6.7.1.1 The human peristimulus interval attack time constant

The peristimulus adaptation characteristic of humans plays a major role in musical composition and performance. Within the range of normal communications, the effects of adaptation are well known. If a note is presented for only a short interval, it will not appear as loud as for a longer interval. On the other hand, a note held for an excessive interval will appear to decrease in loudness due to adaptation. The first effect is due primarily to the integration characteristic of the sensory neurons. The height of the output pedestal (the primary signal passed to the CNS in musical appreciation) of the sensory neurons is strongly affected by the frequency of the stimulus frequency. The second is due primarily to the adaptation introduced by the limited performance of the cardiovascular system supplying those same neurons. See **Section xxx**.

The interplay of the slow rate of buildup of a specific tone in some musical instruments and the limited response of the human hearing during such slow buildup can have a significant impact on how a musical selection is played.

9.6.7.1.1X The nominal delay in recognizing a tone

Determining the rise time of the auditory system by psychophysical experiment is difficult for two reasons. The delay associated with cognition and effector deployment by the motor system are included in the overall response time. The initial response of the sensory system is frequently limited by the rise time of the acoustic stimulus. However nominal values may be useful. **Figure 9.6.7-1** shows the relative response of pure tones of short duration, a condition involving signal integration during the peristimulus interval, from Roederer[172]. It does not show the subsequent adaptation. Roederer noted that the response was related to the total energy delivered by the tone rather than the power level of the stimulus. [xxx his words on page 95 are messed up. Responses do not go through zero time. Xxx figure may not belong here.]

---

[172]Roederer, J. (1995) The Physics and Psychophysics of Music. NY: Springer-Verlag pp 95-96

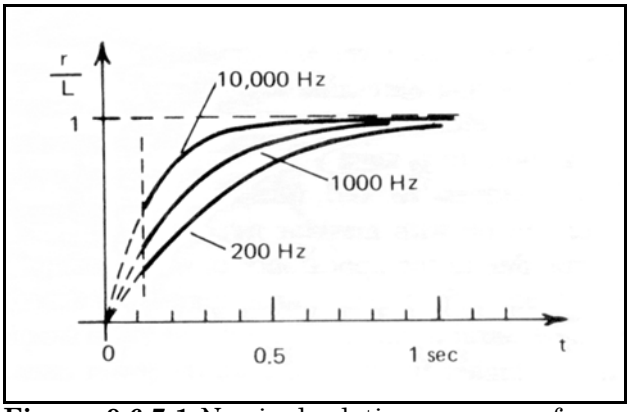


Figure 9.6.7-1 Nominal relative response of pure tones of short duration. r/L; ratio of actual loudness (r) to loudness (L) of a steady tone of the same frequency and amplitude. Based on the integration of the envelope of the stimulus by the auditory system. This figure appears to be dominated by internal signal manipulation and motor response delays. The underlying effect is stimulus level dependent. From Roederer, 1995.

9.6.7.1.2 Human peristimulus adaptation

While the attack time constant of hearing is in the fraction of a second range, the peristimulus adaptation mechanism has a much longer time constant. Small & Minifie have reviewed the wide variation in values presented for peristimulus adaptation by different investigators. They also reviewed the different techniques used. Comments concerning the impact of the dynamics of the test tone itself were included. A series of interesting conjectures are presented in the absence of a model of the mechanisms under discussion. These generally agree with the more specific statements of Chapter 5 based on the model of this work. They provided a graph similar to **Figure 9.6.7-2** xxx that shows both the first order and second order time constants[173]. Notice the lower curve in each frame exhibits a sinusoidal component. A similar situation is encountered in vision when measuring adaptation parameters. The function the data is following is called an exposine, the product of the first order exponential and the second order sinusoidal component[174].

The first order time constant of the adaptation amplifier circuit is read from the individual exponential responses. Since this time constant is dominated by the product of the adaptation collector impedance times the capacitance of the collector circuit, it varies with the average signal current through the adaptation amplifier. Unfortunately, the graphs have been drawn to show adaptation relative to an arbitrary zero. This is commonly done to avoid showing the rollover of the curves and the attack portion of the response to the left of the peak. It appears that this zero is about 5 dB below the true peak value of these curves. Making this assumption, the zero dB level shown approximates the 1/e level of each exponential and thereby defines the time constant of that exponential. They vary from 9 to 21 seconds in the upper frame and 6 to 16 seconds in the lower frame. These time constants may be impacted slightly by the impedance of the electrostenolytic supply impedance of the adaptation amplifier. The impact of this impedance becomes significant in the lower curve of each data set. Its effect is to convert the entire circuit to a second order system consisting of the collector impedance of the adaptation amplifier the capacitance of the collector circuit with respect to ground and the supply impedance operating in a T configuration. This configuration represents a second order system. The system exhibits the time constant measured above as well as a second time constant. The time constant of the second order mechanism is approximately two minutes. This is the time constant of the cardiovascular system supporting the cochlea.

The variation in the time constant of the peristimulus response has not been previously identified and explained in the hearing literature. Recognizing the mechanisms generating this parameter explains why such a wide variation in values are shown in the literature for this one parameter.

---

[173]Small, A. & Minifie, F. (1961) Effect of matching time on perstimulatory adaptation *J Acoust Soc Am* vol. 33, pp 1028-1033

[174]Fulton, J. (2004) xxx

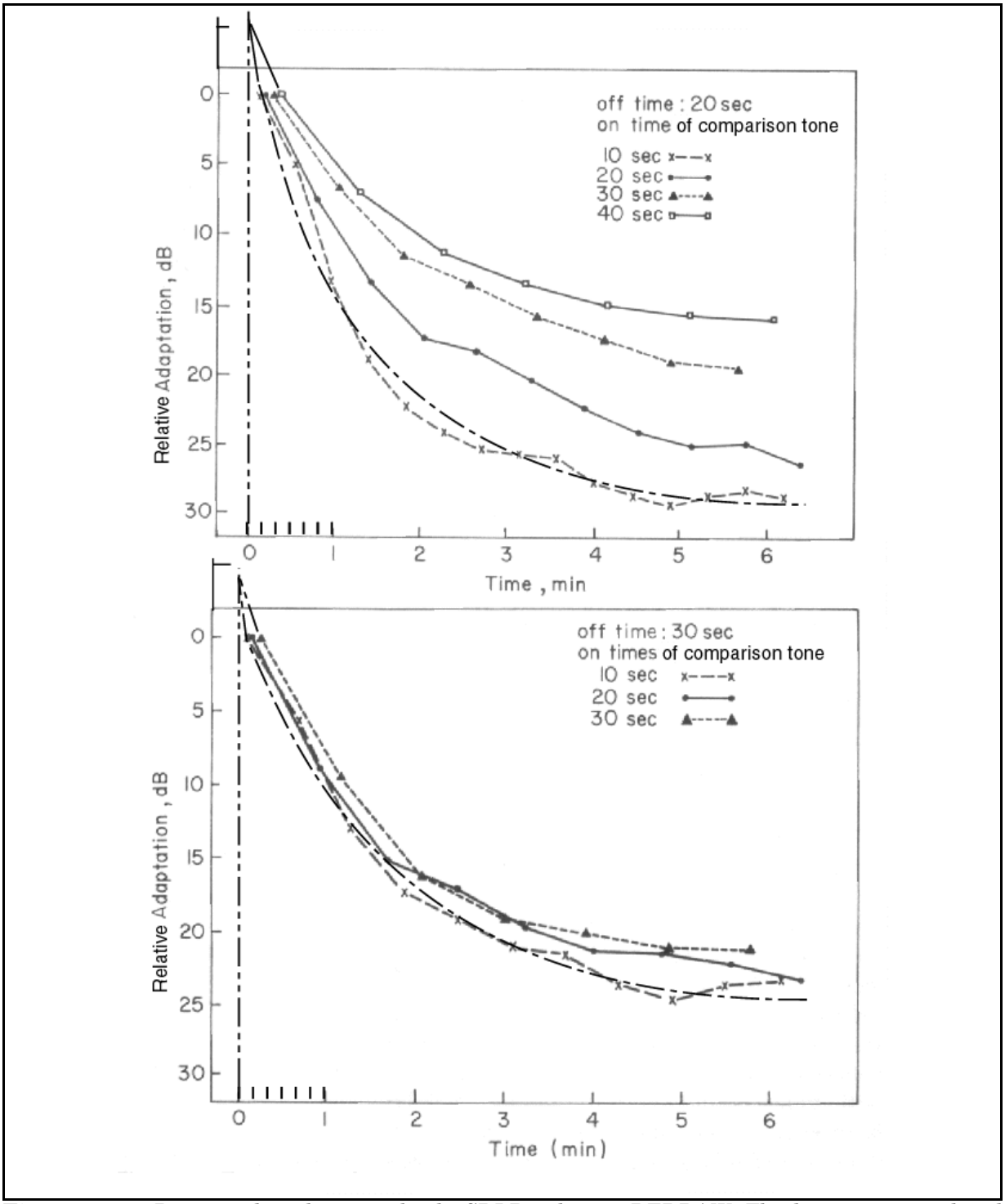


Figure 9.6.7-2 Peristimulus adaptation by the SDLB technique REDRAW. The data points are placed at the center of the respective on-interval and at the mean of the measured values. No values are obtained at zero time using this technique. However, the curves have been extended here to an estimated initial value in order to obtain an estimate of the time constants involved. Note the dash-dot line through the lower graph of each frame. This line shows the adaptation path in the absence of second order effects. The actual curve takes the form of an exposine. The stimulus was at 4000 Hz and 75 dB SL. See text. Modified from Small & Minifie, 1961.

9.6.7.1.2X Peristimulus adaptation in the presence of noise

Carterette has discussed and provided data concerning peristimulus adaption in the presence of noise[175]. The performance is not significantly different than similar adaptation in the presence of a tone.

9.6.7.1.3 Post stimulus adaptation in the presence of masking

Zwicker has provided data on post stimulus adaptation in the presence of masking[176]. [xxx add words] [xxx **Section 5.4.3.3**]

9.6.7.1.4 Post stimulus adaptation under overload conditions

Several investigators have reported on an unusual feature of post stimulus adaptation under overload conditions— conditions of hypertopic stimulation in the region of 120 dB SPL[177]. This level is sufficient to insure the sensory neurons are operating outside of their normal range. It is also possible that subsequent neurons as well as the electrostenolytic support mechanisms are operating outside of their normal range. Under these conditions, nearly any output could be expected at points well removed from the nodes of the PNS. A frequent report is of a "two-minute bounce." an actual additional loss in pure tone sensitivity following cessation of the extended period of excess stimulation.

Figure 9.6.7-3 shows this effect modified from Hirsh & Ward. The threshold sensitivity is seen to make a "bounce" upward around the two minute point and again in the six to eight minute region. The explanation of this response is the same as for the previous figure, except the situation is more severe. Note the response is altered significantly at 4000 Hz even though the fatiguing signal was at 500 Hz.

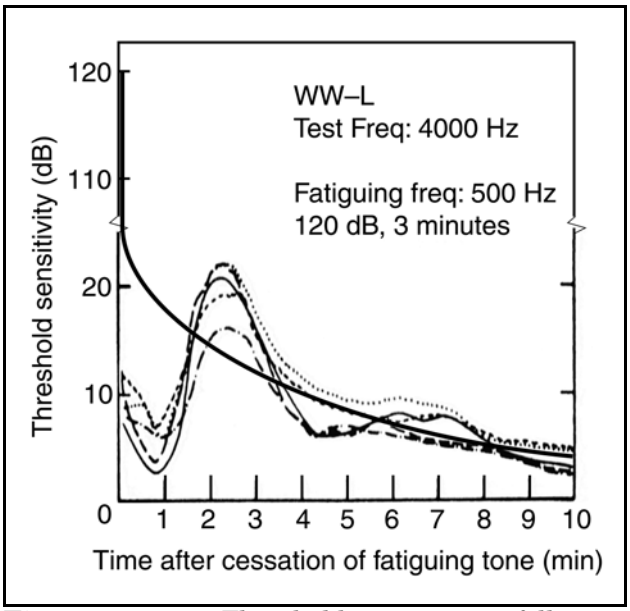


Figure 9.6.7-3 Threshold sensitivity following fatiguing exposure. Note the significant break in the vertical scale. The heavy black line shows the first order exponential curve underlying the overall exposine function. The light black and broken lines represent individual tests under identical conditions. Modified from Hirsh & Ward, 1952.

Elliott & Fraser use the term poststimulatory auditory fatigue for this effect. However, it is quite clearly a fatigue that occurs during the

[175]Elliott, D. & Fraser, W. (1972) Fatigue and adaptation *In* Tobias, J. *ed*. Foundations of Modern Auditory Theory, Volume II. NY: Academic Press Chapter 11 pp 148-151
[176]Zwicker, E. (1988) xxx *In* Duifhuis, H. Horst, J. & Wit, H. *eds*. Basic Issues in Hearing. NY: Academic Press pp 135-142
[177]Elliott, D & Fraser, W. (1972) Op. Cit. pp 121-139

peristimulatory interval. The recovery from the fatigue condition occurs during the post stimulatory period. While Elliot & Fraser have explored this fatigue condition in discussion, and showed the laboratory results of others, they did not present or refer to a model of the relevant mechanisms. They do not provide any indication of the height of the peristimulus stimulation in the original graphs although it can be inferred from the caption. This height is indicated explicitly here. Note the major scale change in the figure.

The two-minute bounce is very suggestive of the time constant of the cardio-vascular system as described in **Section 9.6.4.2**. The same mathematical solution applies to this situation[178]. The response follows an exposine function. The exponential portion of the response is unusual in two respects. It has a time constant of only 6-21 ms. However, it is plotted on a scale that is itself logarithmic (with a base of ten). Notice that the two-minute bounce is accompanied by another "bounce" from two and one-half to five minutes later depending on the unit under test. The important feature here is that the period of the sinusoidal component is approximately 3.5–4.0 minutes. The period in the above figure was about 4.0 minutes as well. The fact that the recovery from a 500 Hz stimulation is different at 1000 Hz and 4000 Hz is indicative of the fact the fatigue was related to the cardiovascular supply supporting the sensory neurons within the cochlea. The same conclusion can be drawn when discussing the graphic of Davis, et. al. appearing on page 139 of Elliott & Fraser.

It appears that the period of this second order sinewave is indicative of the health of the cardiovascular system serving the ear of the subject (**Section 10.XXX**). [xxx drop unless something is in 10] While Elliott & Fraser review earlier speculation that this condition is indicative of the oxygen supply to the ear, it is suggested here that it is the glutamate supply or GABA clearance that is the underlying controlling mechanism.

9.6.7.2 Gain reduction through external feedback in the auditory system EMPTY

9.6.8 The reaction time of the auditory system

[xxx address cognition and motor delay times here.
Reaction time in the auditory system involves the performance of the entire neural system. The reaction time is highly dependent on the character of the stimulus, the response of the sensory neurons, whether cognition is required and/or whether the response has been converted to a reflex through training, and the time required by Stage 6 of the system to report a response.

[xxx combine with above]
When speaking of the reaction time of hearing, it is important to differentiate between the operation of the system through Stage 4, through Stage 5 and through Stage 6. The motor system of Stage 6 frequently introduces a large delay in the overall reaction time of the system. This delay varies with task and the distance of the muscle affecters from the CNS. Stage 5 contributes a variable amount of delay to the overall reaction time depending on the complexity of the choice(s) that must be made. The delay of the system up to the output of Stage 4 (at the saliency map) appears to be quite constant.

Reaction time is a clear function of both the amplitude and the nature of the stimulus. Only the reaction to a sufficiently loud stimulus of duration longer than a few seconds will be discussed here.

---

[178]Fulton, J. (2004) xxx

This is meant to characterize the circuits and response associated with the alarm mode of hearing. More complex stimuli frequently involve interpretation and perception that introduce additional delays compared to those associated with the **alarm mode** alone.

[xxx look at Chocholle, fg 2 & page 490 in Rosenblith if nothing else surfaces. Compare with the decay time constant in the P/D equation.]

Reaction time is a clear function of both the amplitude and the nature of the stimulus. Only the reaction to a sufficiently loud stimulus of duration longer than a few seconds will be discussed here. This is meant to characterize the circuits and response associated with the alarm mode of hearing. More complex stimuli frequently involve interpretation and perception that introduce additional delays compared to those associated with the **alarm mode** alone.

[xxx look at Chocholle, fg 2 & page 490 in Rosenblith if nothing else surfaces. Compare with the decay time constant in the P/D equation.]

Figure 9.6.9-1from Chocholle (1954) shows the gross reaction time of the complete neural system to a pulsed tone. The report consisted of pressing a telegraph key upon perception of the tone. Scharf analyzed how this figure was obtained in some detail[179]. Other subjects exhibited reaction times up to 100 ms longer or shorter in similar experiments. He notes the tension in the subjects finger plays a significant part in the total reaction time.

[179]Scharf, B. (1978) Loudness *In* Carterette, E. & Friedman, M. *eds*. (1978) Handbook of Perception, Volume IV: Hearing. NY: Academic Press Chapter 6 pp 230-234

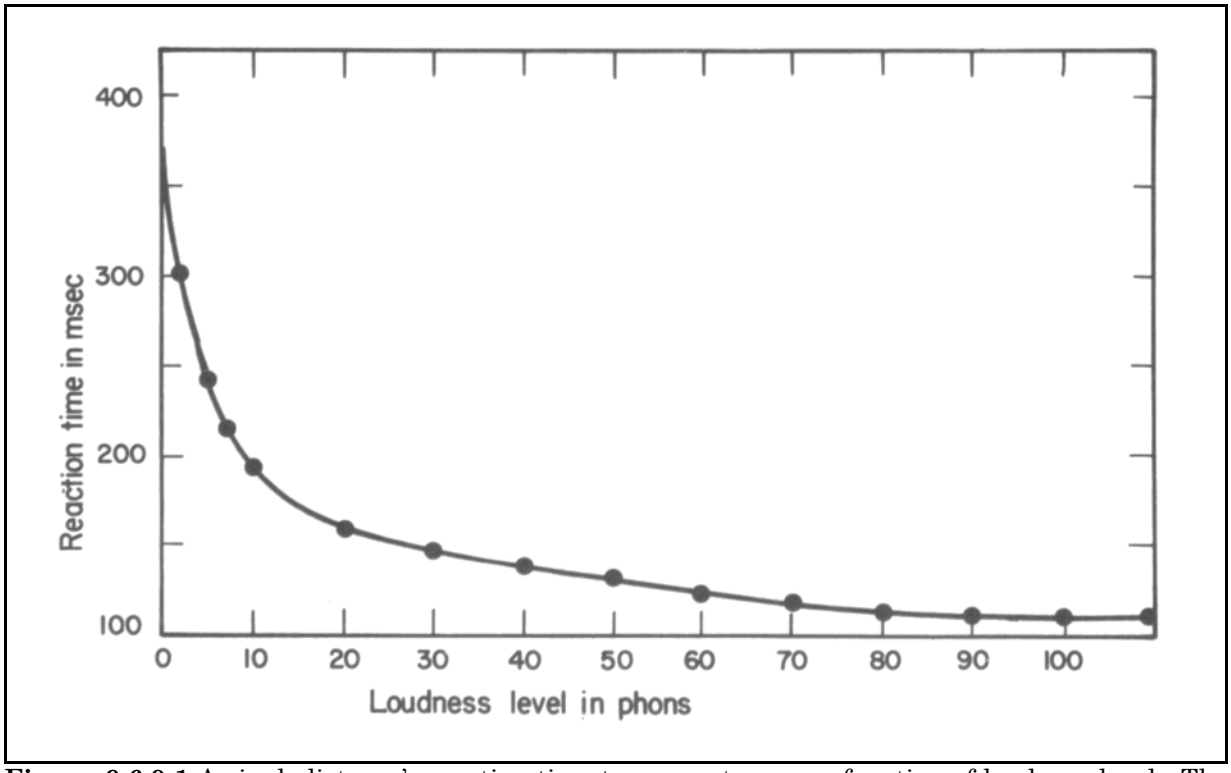


Figure 9.6.9-1 A single listener's reaction time to a pure tone as a function of loudness level. The same curve represents tones varying in frequency from 50 to 10,000 Hz. Adapted from Chocholle, 1954

9.6.9 The behavioral response versus the action potential response

It is a premise of this work that the behavioral response of the organism at threshold levels of stimulation depend on the transmission of an action potential from at least one ganglion neuron of the auditory nerve in response to a voltage change in at least one sensory neuron. Dallos et. al. explored this relationship in 1978 and appear to confirm this relationship[180]. The use of the term "compound" in connection with the action potential appears to be superfluous within the body of the paper.

9.7 EMPTY

9.8 The functional psychophysical performance related to communications RENUMB

The psychophysical performance of hearing related to communications can best be understood by considering both the characteristics and operation of the source of the sound and the operation of the auditory system. This area is best addressed by addressing communications related to speech

[180]Dallos, P. Harris, D. Ozdamar, O. & Ryan, A. (1978) Behavioral, compound action potential, and single unit thresholds: relationship in normal and abnormal ears *J Acoust Soc Am* vol. 64(1), pp 151-157

and to music separately.

When discussing either speech or musical communications, it is important to recognize that most likely the analytical mode of hearing is being discussed. While music can certainly be used as background during other activities (and some people like to use voice communications for this function), it is a matter of attention. When listening to speech or music as a primary task, attention is clearly involved. Attention is a feature of the analytical mode. However, our understanding of attention as a mechanism remains limited.

A common challenge when discussing both speech and musical communications is to delineate those performance features based on the neural circuitry of Stage 4 (pre-wired neural connections) and Stage 5 (neural connections implemented based on learning). The special case of the preprogrammed circuits associated with the neonate may be useful in this area as well. In this case, the analogy is to a boot program in computers. It may consist of a combination of pre-wired connections and the formation of connections in the memory area associated with learning based on the expansion (through computation) of a minimal amount of pre-wired information.

9.8.1 The communications based performance requirements of hearing

The global requirement placed on the hearing system have been plotted by a variety of authors. However, little concerning the specific requirements have appeared in a concise form in the research literature. If the requirements of communications are to be used to specify the performance requirements of hearing, it is the most sophisticated performance that is observed in practice that best describes the capability of the actual system (even if this capability is not used routinely by an individual). If someone is described as having perfect pitch, or being able to recognize tones of only 0.3% difference in frequency, then all other members of that species must have a similar inherent capability. The fact that low-grade communications can be carried on using a lesser grade of performance is largely irrelevant.

9.8.1.1 Performance requirements based on speech

Speech has developed into a very complex form of communications. Its generation has been studied intensely from a myriad of perspectives. **Figure 9.8.2-1** reproduces a figure by Liberman reproduced in Sanders that appears particularly clear[181,182]. There is a tendency to believe that the perception of speech involves a reversal of this sequence of conceptual phases. However, there is significant evidence that the de-convolution of speech is distinctly different from the formation of speech.

The figure highlights the difficulty of encompassing the entire problem of speech generation or speech de-convolution in a single graphic (or within a single scientific specialty). The figure only addresses the conceptual steps of interest to a speech psychologist or linguist. It does not address the neurology involved in physical speech generation.

The complexity of speech generation and deconvolution is highlighted by the jargon used within different specialties. The linguist uses the concept of speech segments, usually referred to as consonants and vowels. Globally, the 40 consonants and vowels are referred to as phonemes.

---

[181]Liberman, xxx (1970) xxx
[182]Sanders, D. (1977) Auditory Perception of Speech. Englewood Cliffs, NJ: Prentice-Hall pg 148

When their more subtle variants are added to the mix, the result is 50 to 60 allophones. When a sequence of these phonemes (or allophones) are assembled, they become a morpheme. A morpheme is a minimal sequence of phonemes that has its own meaning or grammatical function. A word is made up of one or more morphenes. The reading community uses a different jargon. A word is made up of one or more syllables. A syllable is made up of one or more tonations that can be described by their combined temporal and spectral characteristics. A series of words forms a sentence expressing a concise idea. Sentences can be combined into paragraphs expressing a more complex idea. Other psychologists describe their coarser segment of speech as involving interps and percepts. An interp, derived from basic interpretation, can be equated to a morpheme. A percept, derived from a higher level of perception, can be equated to a word or a sentence, occasionally a paragraph. The parsing of a paragraph to the level of phonemes illustrates the difficulty of describing the de-convolution process. In this work, it has become useful to describe a "basic interp," a "complete interp" and a percept. It is the percept that is provided to the saliency map for further cognitive action. This action may well assemble percepts, frequently from multiple sensory modalities, into a "global percept."

Figure 9.8.2-1 describes the specific requirements on the hearing system as it relates to speech. While this field has undergone an explosion in activity in recent years, the fundamental requirements have been visualized for a very long time. This figure, modified by the addition of the critical band borders proposed by Zwicker, has been available since 1959. Note the very short time intervals associated with the fricatives, t, p & c. Their durations are on the order of five milliseconds or less. Note also that the times shown are in free space. The features occur in a slightly different positions within the peripheral neural system. The low frequency components are moved to the right by about five milliseconds. This is due primarily to the finite travel time of acoustic energy along the SAW filter of the tectorial membrane and the physical location of the sensory neurons along that filter.

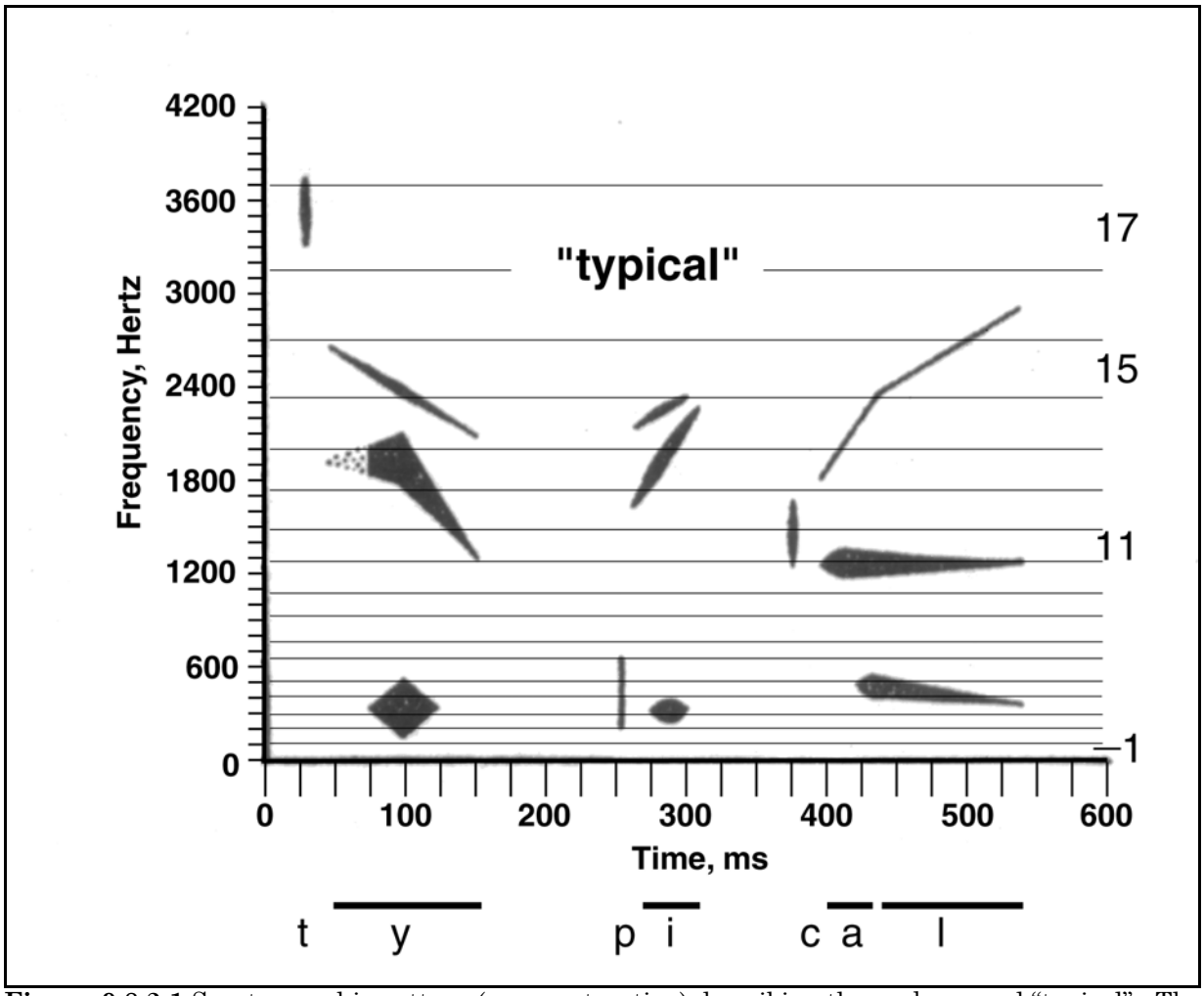


Figure 9.8.2-1 Spectrographic pattern (a reconstruction) describing the spoken word "typical". The bars above the vowels, etc. suggest their length. The numbers at left define the "critical bands" of Zwicker. Pattern from Lieberman, et. al., 1959.

The above figure was prepared to show the gross requirements on the hearing system if it is to perceive the word "typical" in a nominal environment. Liberman, et. al. describe the intelligibility

of this pattern as "of a fairly high order." The frequency band shown is equivalent to that provided by a nominal long distance telephone circuit (typically 6 dB down near 300 and 3,000 Hz).

Figure 9.8.2-2 provides more detail concerning the critical change in frequency associated with the transition between an initial fricative and the following vowel. The rates of frequency change between the first and second formant have attracted some notice in the literature. These rates are on the order of 20-40 octaves/sec and typically last for only a fraction of 50 ms. The duration, starting frequency (locus), and steady-state frequency of the vowel portion of the sound is extremely important in speech recognition. Two loci are shown by dashed lines in the figure. It is absolutely clear that the difference in frequency between the locus of a vowel and its steady state frequency must be discerned in high quality speech recognition.

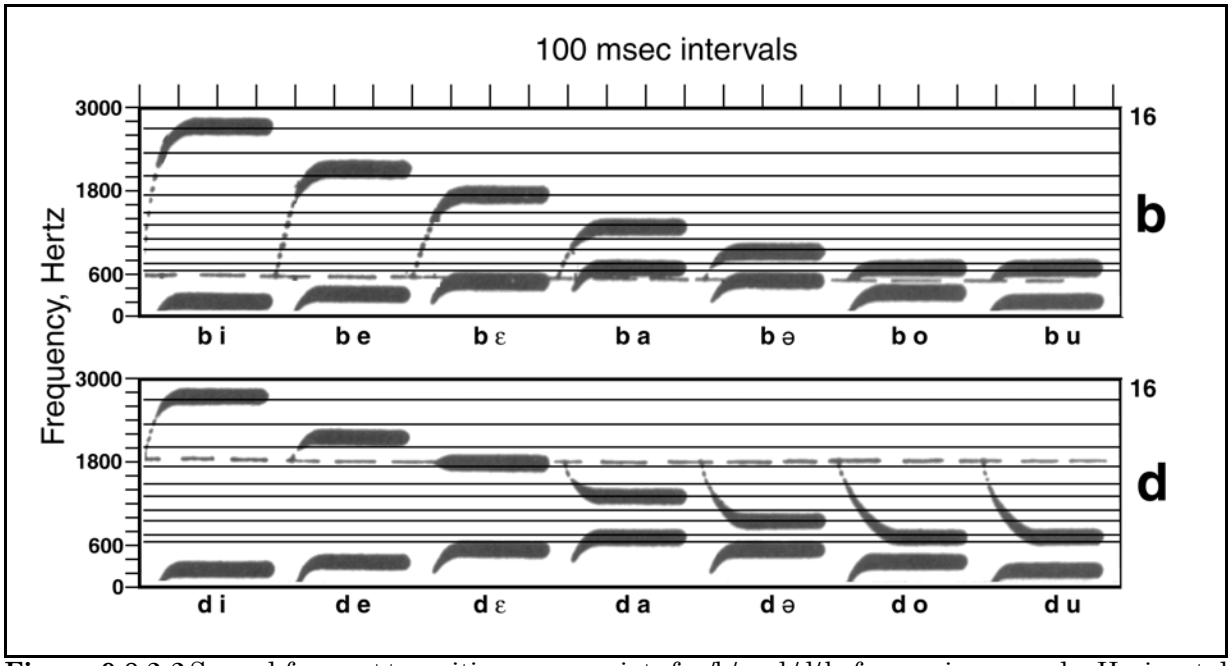


Figure 9.8.2-2 Second-formant transitions appropriate for /b/ and /d/ before various vowels. Horizontal bands are the critical bands of Zwicker ending with number 16. Only at frequencies well below 1000 Hz does the subject have a reasonable chance of perceiving differences between the syllables shown based on Zwicker's critical bands. Notation is that of the linguist. Patterns from Liberman, et. al., 1959.

From these figures, it is clear that the hearing system must discriminate in time with a resolution on the order of 10 ms, and discriminate in frequency with a resolution probably at least two-to four times better than is provided by the critical bands of Zwicker, if good speech recognition is to be obtained.

Sanders provided his summary of both active and passive theories of speech perception in 1977[183]. Miller provided a summary of theories of speech perception from his perspective in 1988[184]. Reviewing the theories of these two summaries is a recreational experience. The perspectives vary

[183]Sanders, D. (1977) Auditory Perception of Speech. Englewood Cliffs, NJ: Prentice-Hall
[184]Miller, J. (1988) Theories of speech perception as guides to neural mechanisms *In* Miller, J. & Spelman, F. *eds*. Cochlear Implants: Models of Electrically Stimulated Ear. NY: Springer-Verlag Chapter 18

widely (from demons to analysis-by-synthesis). Miller did define a three-dimensional auditory-perceptual space (APS) that may be equivalent to a part of the database type structure defined in this work as the saliency map.

The conceptual level of this earlier work is characterized by a brief description of the "motor theory" developed by the psychologist, Liberman, in 1972[185]:

- The speech code involves a special function
- It exists in a distinctive form
- It is unlocked by a special key
- It is perceived in a special mode.

While progress has been made since 1972, the field is still characterized by high-level concepts. Lack of an adequate neurological model prevents the development of a realistic model of speech perception.

9.8.1.2 Performance requirements based on music appreciation

Previous discussion (Section xxx) have shown that the performance achievable in human hearing is from 20 to 40 times better than the bandwidth of Zwicker's critical bands. While this does not make signal discrimination based on band-to-band signal differencing, like that used in vision (PBV, xxx or BV, xxx), it does make it very difficult. Such processing would probably encounter significant signal-to-noise limitations. Similar earlier discussions (Section xxx) have shown that the perceptual frequency difference achieved among trained individuals is considerably better than the bandwidth of the signaling channels measured and reported from the cochlear nerve. While signal perception based on band-to-band signal differencing at the reported bandwidth of the channels in the cochlear nerve would be much more practical, such signal differencing has not been reported in the literature.

Only signal differencing between adjacent small groups of OHC (typically 3 to 6 OHC in a single row along the cochlear partition) appears to have a reasonable chance of providing the just noticeable frequency difference reported by trained musicians at phonotopic intensity levels.

9.8.2 The psychophysical performance of hearing related to speech

Any discussion of the relationship between speech and hearing is complicated by the very large number of variables that have been designed into the process. These variables allow a high degree of communications fidelity under very wide and frequently difficult background acoustic environments. Miller has provided good graphics and a discussion bounding this situation[186].

It is very important that any researcher in this area be familiar with the important writings of Mattingly & Liberman[187]. While not based on detailed physiology, it gives a very useful framework

---

[185]Liberman, xxx (1972)
[186]Miller, J. (1978) Effects of noise on people *In* Carterette, E. & Friedman, M. *eds*. Handbook of Perception, Volume IV: Hearing. NY: Academic Press Chapter 14
[187]Mattingly, I. & Liberman, A. (1988) Specialized perceiving systems for speech and other biologically significant sounds *In* Edelman, G. Gall, W. & Cowan, W. *eds*. Auditory function: Neurobiological Bases of Hearing. NY: John Wiley & Sons Chapter 26

for understanding the perception of speech.

Great progress is currently being made on relating various speech components to their site of processing within the cerebral cortex using both non-invasive techniques like fMRI and invasive techniques based on electrophysiology. This work will not be addressed here. Rauschecker & Tian have provided good material on the gross physiology of the rhesus monkey when responding to vocalization[188].

When discussing speech communications, the importance of both the tonal and fricative portions of both the voice and hearing mechanisms are brought into clear focus. Without either of these major portions, speech communications becomes difficult.

9.8.2.1 The acoustic interplay between speech and hearing

The correlation process defined in Chapter 8 applicable to hearing clearly favors speech containing the even harmonics of speech. The production of speech is a very complex acoustic process that is better understood from an anatomical/morphological perspective than from an acoustic perspective.

The acoustic description of the speech production apparatus requires consideration of the entire aural path from lungs to lips. The system consists of a series of cavities, that can be considered waveguides, separated by a resonating structure, the vocal folds. The sound produced in the far field depends strongly on the pressure waves generated by the muscles of the chest cavity, the dimensions of the pleural cavity, the dynamic properties of the vocal folds and the detailed dimensions of the aural cavity. The dimensions of the aural cavity are critically important because they define the resonant conditions that are excited by the initial sound generated by the vocal folds. The dynamics of the lips and teeth allow the system to operate as a closed end (nearly shorted) resonant waveguide as well as an open ended waveguide. The harmonics created under these conditions varies greatly[189]. By varying the structure of the aural cavity, the system can favor the generation of even or odd harmonics at a given frequency.

9.8.2.1.1 The primary and secondary sources of sound in speech EMPTY

[xxx stuff from Pierce & David pg 56-66

Considerable overlap exists in the terminology used to describe the human voice. Pierce gives the following values for the fundamental frequency of the following voices[190]:

Bass	60–384 Hz
Baritone	76–512
Tenor	96–640
Contralto	105–1280
Mezzo	128–1280
Soprano	128–1450

Sung notes as low as 40 and as high as 2048 Hz have been achieved by accomplished vocal artists.

---

[188]Rauschecker, J. & Tian, B. (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex *Proc Natl Acad Sci USA* vol. 97, pp 11800-11806
[189]Reintjes, J. & Coate, G. (1952) Principles of Radar. NY: McGraw-Hill pg 619
[190]Pierce, J. & David, E. (1958) Man's World of Sound. NY: Doubleday. pg 64

Pierce & David give considerable information on the more subtle qualities that can be achieved in the singing voice.

9.8.2.1.2 The role of the lips, teeth and tongue during phonation EMPTY

The formation of the sounds of speech is a very complex process employing a large number of muscles. **Figure 9.8.3-1** from Fletcher describes the major elements of the vocal system that are controlled by these muscles[191].

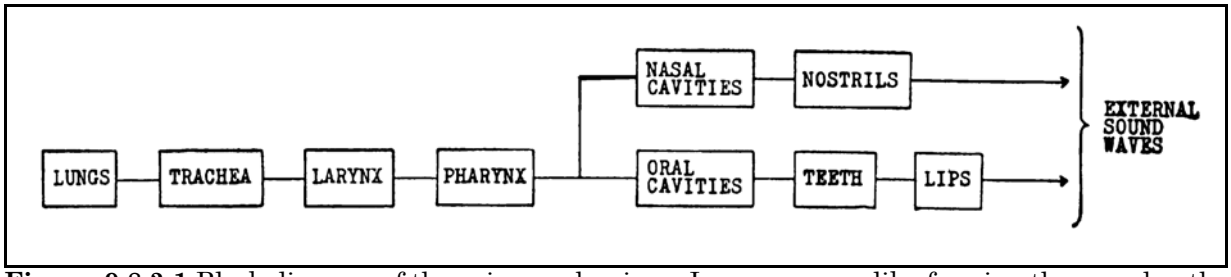


Figure 9.8.3-1 Block diagram of the voice mechanism. In some cases, like forming the sound p, the air stream need not originate in the lungs. It may use the reservoir of air in the mouth driven outward by the contraction of the cheeks. Modern usage frequently defines the pharynx as including the nasal and oral cavities. From Fletcher, 1953.

Moore has provided a detailed figure describing how all of the above elements are configured in the process of generating one sound in speech[192]. He also makes a case for the existence of a "speech mode" in hearing. This speech mode supports the extraction of, and correlation of, both vowel and fricative sounds in a temporally extended context to extract complex interps and/or percepts from speech.

9.8.2.1.3 Spectrograms of human speech EMPTY

[xxx currently a few words in Section 8.6.1 under requirements. That may be the right home.]

A significant feature of human speech is the delay between the initiation of an unvoiced consonant and the initiation of the following vowel in a consonant-vowel phoneme. In English, this voice-onset-time (VOT) is generally 20–50 ms depending on the individual. Eggermont and Ponton have summarized the available research in this area[193]. In high signal to noise ratio situations, the unvoiced consonant is readily identified and the recipient is guided to expect a vowel to follow after the above interval. Under poor signal-to-noise conditions, the unvoiced consonant is frequently masked and the appearance of the vowel may be unexpected.

9.8.2.1.4 Gross histograms of human speech EMPTY

---

[191]Fletcher, H. (1953) Speech and Hearing in Communication. NY: Nostrand Co. pg 8
[192]Moore, B. (1977) Op. Cit. pp 210-234
[193]Eggermont, J. & Ponton, C. (2002) The Neurophysiology of Auditory Perception: From Single Units to Evoked Potentials *Audiol Neurootol* vol 7, pp 92–95

Zwicker & Fastl reported on a statistical study of the rate of information transfer in German, English, French and Japanese[194]. **Figure 9.8.3-2** shows the results of more than 10,000 data points with a bin width of 20 ms. "All of the intervals are concentrated between 100 and 1000 ms. Almost 90% of the data points are found between 200 and 600 ms, and the most frequent temporal interval hasa length of about 300 ms. This means that in running speech, about 2 to 5 events are perceived per second."

When a rate of 2-5 perceived events per second is multiplied by the logarithm, to the base 2, of the number of possible signals based on the number of critical bands, an information transfer rate of about 20 bits per second is obtained, 4 x $\log_2$ 24 = 18.4 bits per second. This is the generally accepted information transfer rate for the human brain based primarily on aural communications experiments.

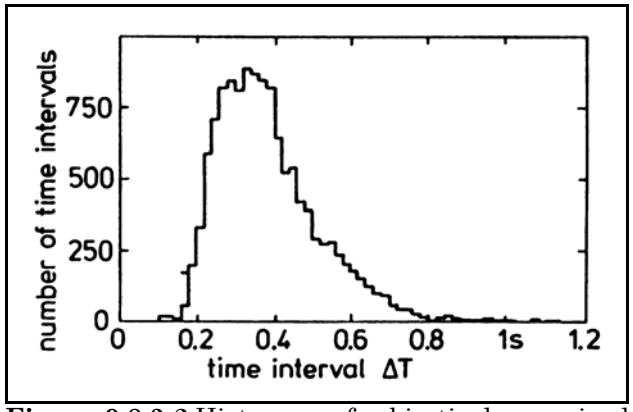


Figure 9.8.3-2 Histogram of subjectively perceived intervals in different languages. The broad maximum near 300 ms indicates optimization of speech to receive information packets at a rate near 4 Hz. From Zwicker & Fastl, 1999.

9.8.2.2 Models of speech recovery in hearing

This work will not review the voluminous psychophysical literature of hearing related to speech recovery. Sanders provided an introductory review of the field in 1977[195]. It can provide a starting point for a broader study. The material presented here has been largely constrained to the physiological sphere.

9.8.3 The psychophysical performance of hearing related to music

If the reader has a limited background in music, the texts by Deutsch[196] and by Roederer[197] are highly recommended. The first chapter by Rasch & plomp provided many definitions. Moore has also addressed music more generally in his 1977 book[198]. Music is defined in terms of its three primary attributes; pitch, loudness and timbre. The multi-dimensional character of the term timbre is illustrated (page 13). Although the concept of timbre is highly subjective, the American Standards Association provided an archaic definition in 1960, "Timbre is that attribute of auditory sensation in terms of which a listener can judge that two steady-state complex tones having the same loudness are dissimilar." In 1991, Bregman cited the then current and very similar ASA definition and noted, "This is, of course, no definition at all (page 92)." Quoting Rasch & Plomp, "Recent research has shown that temporal characteristics of the tones may have a profound influence on timbre as well, which has led to a broadening of the concept of timbre (Schouten, 1968). Both onset effects (rise time, presence of noise or inharmonic partials during onset, unequal rise of partials, characteristic shape of rise curve etc.) and stead state effects (vibrato, amplitude

---

[194]Zwicker, E. & Fastl, H. (1999) Psychoacoustics: Facts and Models, 2nd Ed. NY: Springer pg 353
[195]Sanders, D. (1977) Auditory Perception of Speech. Englewood Cliffs, NJ: Prentice-Hall
[196]Deutsch, D. *ed.* (1982) The Psychology of Music. NY: Academic Press
[197]Roederer, J. (1995) The Physics and Psychophysics of Music. NY: Springer-Verlag
[198]Moore, B. (1977) Introduction to the psychology of hearing. London: Macmillan pp 101-102 & 154-159

modulation, gradual swelling, pitch instability, etc.) are important factors in the recognition and, therefore, in the timbre of tones."

Music can be defined as an acoustic stimulus consisting of the superposition of complex groups of tones(chords) presented in succession according to an equally complex rhythm, where the individual tones may exhibit a temporal variation within each individual interval of the rhythm. The number of variables that can be introduced within the elements of this definition is very large. It is a mark of the complexity, the computational capacity and the storage capacity of the central nervous system that it can perceive and interpret and memorize complex arrangements of this type. Many other definitions of music can be given. However, once you move beyond the above mechanistic definition, emotion, experience and training play a great deal in how one attempts to develop, and how one expresses, a new definition.

Deutsch asserted, a primary relationship exploited in music is that "a strong similarity exists between tones separated by octaves—that is, whose fundamental frequencies stand in the ratio of 2:1 (or a power of 2:1)." **Appendix C** will develop the musical relationships between tones in some detail.

A major goal of this discussion is to delineate between performance based on pre-wired neural circuits (such as the initial circuits of the diencephalon), performance based on training in the musical arts, and performance based on learned preferences. The latter are illustrated by the differences between western, middle eastern and eastern music based on immersion in the local culture. Performance based on training in the musical arts has been explored by many. The studies by Soderquist & Moore[199] and by Scharf[200], and the general background provided by Plomp are particularly relevant. Moore quoted Soderquist (page 102), "the musicians were markedly superior" in their ability to identify the residues (or the partials) presented to them under test conditions. Moore suggested, "This result could mean that musicians have narrower critical bands, but this is unlikely if the critical band reflects a basic physiological process, as is generally assumed. It seems more plausible that some other mechanism is involved in this task, and that musicians, because of their greater experience, are able to make more efficient use of this mechanism." This work proposes that the superior performance of the musicians involves their training to associate the signals generated by the multidimensional correlator of the TRN with specific residues that they have been taught are equivalent to various fundamental frequencies. As a result of this training, the musicians do not rely upon the coarse signal summation channels described by the critical band concept at all. They rely upon their perceptions developed by the more precise signal accumulation channels (**Section 6.2.xxx**)

9.8.3.1 Terminology specific to music appreciation

9.8.3.1.1 Beats and roughness

[xxx page 14-16 Rasch & Plomp
Figure 9.8.4-1

[199]Soderquist, D. & Moore, M. (1970) Effect of training on frequency discrimination in primary school children *J Auditory Res* vol. 10, pp 185-192
[200]Scharf, B. (1970) Critical Bands *In* Tobias, J. *ed.* Foundations of Modern Auditory Theory. NY: Academic Press Chapter 5

Explain how this figure can be used for a variety of more complex tests where f and g are harmonics of a more fundamental tone.

The term roughness is found frequently in physchphysical experiments involving beats and also AM modulated stimuli. It is not clearly defined due to its perceptual nature. No indication that such a phenomenon had been recorded electrophysiologically was found in the literature. Based on this work, at least two sources of roughness can be suggested. At some signal levels, the sampled data form of processing used within the Stage 3 signal decoding circuits can introduce non-synchronous noise of a very coarse character into the signaling channel. Second, the mechanism introducing the limited fusion frequency of the CNS remains unaccounted for. This mechanism could also introduce a harsh perception into the system.

Fastl performed a series of experiments designed to relate roughness to the modulating frequency of a 100% modulated tone at 2 kHz. His results were negative[201].

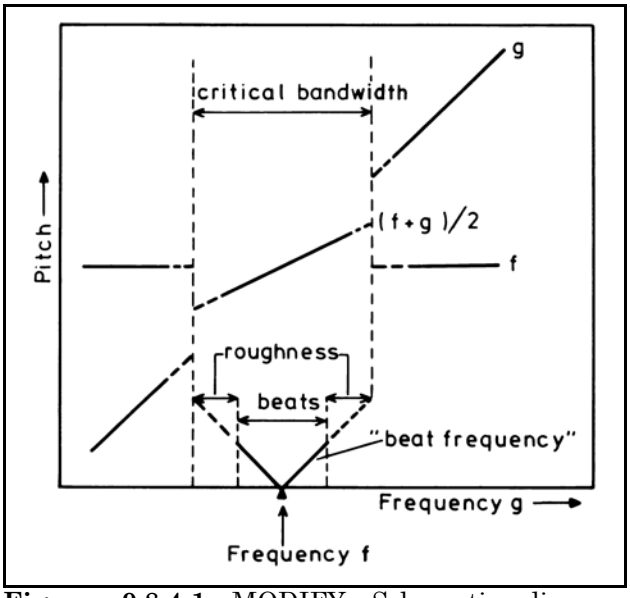


Figure 9.8.4-1 MODIFY Schematic diagram showing perceptual phenomena between two simple tones occurring simultaneously. Change lower scale to show a vernier region and make diagonal continuous but not straight.

9.8.3.1.2 Difference between beats and modulation

[xxx Beats are the result of summation. Modulation is the result of multiplication. The mathematics are more different than apparent in a graphic.]

[xxx Rasch & Plomp pg 14-15

[201]Fastl, H. (1990) The hearing sensation roughness and neuronal responses to AM-tones *Hear Res* vol. 46, pp 293-296

9.8.3.1.3 Consonance and dissonance

When discussing consonance and disonance, it is important to recognize that most texts are limiting their discussion to the eight note "Just Scale." This is the only scale in which the ratios of musical notes are exact ratios of small integers. This scale is only approximated by the white keys on a piano. The Pythagorean and more widely used Equally Tempered scales only approximate these relationships (within the tolerance of the human ear).

In 1977, Terhardt addressed the difficulty of defining consonance and dissonance in a musical context[202]. The discussion appears to be limited to western music.
[xxx address pag 489 by de Boer in Keidel & Neff. in terms of ratios -see also Append on music and confusion between definitions of a fifth, fourth etc.
Figure 9.8.4-2 [xxx Rasch & Plomp pg 19 include comments about fifth etc.]
[xxx talk about different scales in tobias and in Roederer. describe them on a table of frequencies or in terms of my 2-d correlator. Also see their definition based on keyboard location in Appen C.. And the fact they are to a large extent learned.

Roederer has provided a discussion of musical consonance[203].

9.8.3.1.4 Effect of phase relationships in music

[xxx discuss classical versus view based on integration frequency of this work.

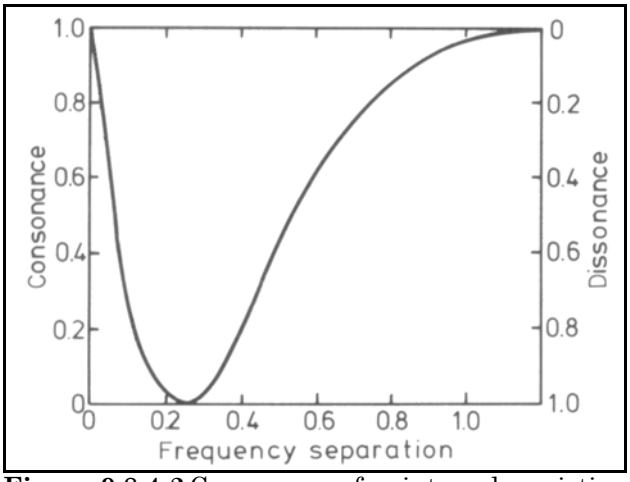


Figure 9.8.4-2 Consonance of an interval consisting of two simple tones as a function of frequency separation, measured relative to the critical bandwidth. Based on Plomp & Levelt, 1965.

Figure 9.8.4-3 [xxx Rasch & Plomp page 27]

[202]Terhardt, E. (1977) The two-compnent theory of musical consonance *In* Evans, E. & Wilson J. *eds.* Psychophysics and Physiology of Hearing. NY: Academic Press pp 381-389
[203]Roederer, J. (1995) The Physics and Psychophysics of Music, 3rd Ed. NY: Springer-Verlag. Chapter 5

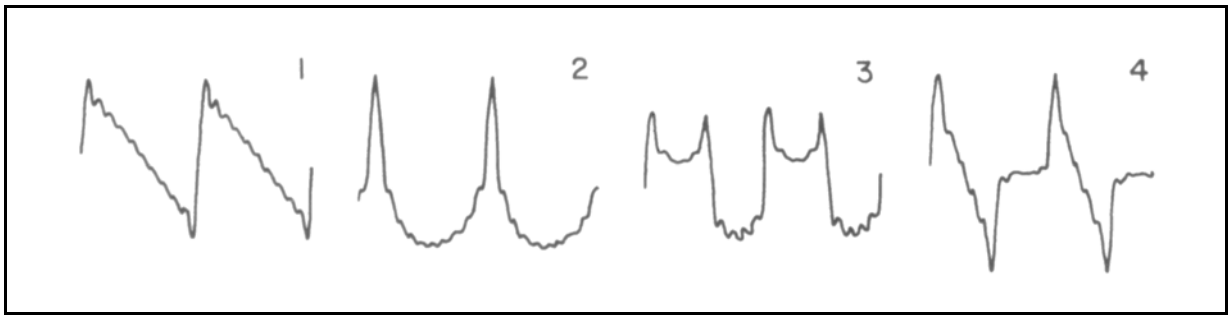


Figure 9.8.4-3 Waveforms sounding very similar for a fundamental frequency near or above 600 Hz. The phase relationships between the fundamentals and the harmonics are grossly different. From Plomp, 1976.

9.8.3.1.5 Rhythm and tempo

[xxx chapter 6 in Deutsch]
Fraisse has provided an in depth discussion of the complex concepts of rhythm, tempo and cadence[204]. The subject rapidly proceeds into discussions of mythology and Gestalt concepts. The initial discussion closes with a definition from Marin (1972). "Inherent in the rhythmic concept is that the perception of early events in a sequence generates expectancies concerning later events in real time." Fraisse defines the simplest rhythm as a cadence, a simple repetition of the same stimulus at a constant frequency. Mursell has provided an in-depth discussion of rhythm in the context of complex music[205]. He shows how these complex rhythms allow one performer differentiate his style from another. This level of detail suggests the level of sophistication of the human hearing system that enables it to perceive such subtleties. [xxx more words]

9.8.3.2 Constructional frameworks associated with musical scales

Investigators have frequently attempted to develop a mathematical or geometrical relationship between music and Euclidian geometry. Roederer walks through the development of three scales of historical as well as practical significance (pages 171-181). He reviews the eight note "just scale" as well as its 12 note cousin, the chromatic just scale. He then proceeds to the Pythagorean (eight notes) and chromatic Pythagorean scales before arriving at the now universally used equally tempered scale, which is also chromatic. The equally tempered scale is a logarithmically based scale that is well matched to the logarithmic organization of the cochlear partition of the ear. As developed in **Section 8.1.6 xxx**, the auditory system appears to use a Riemann Transform to translate the uni-dimensional structure of the cochlear partition into a two-dimensional correlator within the TRN of the thalamus.

Shepard and Ward & Burns, both writing in Deutsch, have developed representations of the musical scales based on a helical notation where the height of individual turns represent increasing octaves with the individual notes within an octave spaced along one revolution. These have been expanded into double helices and helices wrapped around a torus to illustrate different relationships.

---

[204]Fraisse, P. (1982) Rhythm and tempo *In* Deutsch, D. *ed*. The Psychology of Music. NY: Academic Press. Chapter 6
[205]Mursell, J. (xxx) the Psychology of Music. Westport, Ct: Greenwood Press pp 176+

These cylindrical coordinate forms are all conformal transforms relatable to the Riemann transform discussed in **Section 8.1.6 xxx**. All of these transforms are attempts to better understand the computational anatomy (computational techniques based on rearranging the geometric arrangement of neural circuits) used within the human auditory system. The primary question is what transforms can be directly related to the neural system. The Riemann transform is a simple transform that has been found to apply widely in the visual system (ref my work). Very complex transforms may be the result of concatenation of multiple transforms. Some of these may be directly relatable to Stage 4 (and be considered pre-wired) while others may be more properly related to Stage 5 (and be developed based on learning).

The discussion of a variety of Euclidian shapes by Shepard as they relate to music is interesting[206]. However, he addresses the relationship of these shapes to music as would a classical geometrician. There is no particular concern with relating these shapes to the specific problem at hand, understanding the functional structure of human hearing. He develops his geometrical descriptions based on four canonical conditions. However, he notes a number of his conditions can not always be achieved simultaneously in his constructions (page 351).

When reading Shepard, it is important to note he uses the term rectilinear to refer to a one-dimensional line, the logarithmic frequency scale overlaid with musical notation. From there, he folds the line into two-point, three-point, four-point, six-point and 12-point geometrical forms that can be related to different musical features (listed on page 354 and diagramed on page 360).

It appears that many of the more erudite relationships within the theory of music that Shepard discusses are also erudite in the sense of learned by listeners. If so, these relationships are more likely to be learned and therefore associated with Stage 5, Cognition, rather than with Stage 4, Signal Manipulation. The problem is that many of these relationships appear to involve intra-octave tones rather than be pitch based. This fact suggests that the these relationships are evaluated in Stage 4, but in parallel with the more fundamental (octave based) Riemann transform proposed to be found in the TRN of the diencephalon. If true, this proposal would suggest these relationships are evaluated in other signal manipulation engines at a parallel level with the pulvinar/cerebellum. The auditory areas of the temporal lobes immediately come to mind. Their performance in the high level evaluation of highly complex auditory stimuli would be compatible with the results reported in **Section xxx** (concerning commissorotomies). It has been shown that these surgical procedures disconnections or trauma induced failures of the temporal lobes had little effect on the routine processes of hearing associated with daily life.

The suggestion that the temporal lobes contain engines performing complex correlations involving many layers of neural tissue is reminiscent of the "blobs" of the occipital lobe in vision [xxx reference to my material or to Hubel et. al.]

4.1.5.3 Specific terms describing the musical qualities of an acoustic stimulus BELONGS IN 9

Chord– A set of frequencies (of equal and constant amplitude) sounded simultaneously. In music, the individual frequencies are chosen from the musical scales to insure a harmonious sound.

---

[206]Shepard, R. (1982) Structural representations of musical pitch *In* Deutsch, D. *ed*. The Psychology of Music. NY: Academic Press. Chapter 11

intonation–

Just intonation– A scale based purely on harmonics of a given frequency note used as the reference. Found in most ancient instruments. Formal music can only be played in the key it was written for.

Equal tempered– A scale based on logarithmic progression in frequency from a given frequency used as the reference. Used in all modern musical instruments. Allows formal music to be played beginning at any point on the tempered scale without sounding discordant.

Both intonations typically use 440 Hz as the reference frequency.

Interval– An important term in subjective quality assessment. It describes the distance in frequency between adjacent sounds in a chord of two or more frequencies. In music the frequencies are chosen from the musical scales to insure a harmonious sound.

Consonent interval–
Dissonant interval–
Melody– A sequence of stimuli (of equal and constant amplitude) played sequentially.

Quarter-tones–

Scale– A series of individual frequencies (of equal and constant amplitude but defined duration) played sequentially. In music, the individual frequencies can be replaced by chords.

Semitone–

Tempo–

Tempre–

Texture–

9.8.3.3 A summary of conclusions and caveats of others

Sundberg has provided an illuminating discussion of the difference between the speaking and singing voices[207].

9.8.3.4 A summary of conclusions and caveats of others

The magnitude of information developed concerning the psychophysical response of humans to music is immense. The framework to interpret much of this information remains to be developed. As a consequence, this section will present a selection of axioms prepared by different investigators to quantify (or qualify) what is known by these performance features. This author generally agrees with these axioms, and hope additional corollaries will appear to further quantify them.

Burns & Ward, writing in Deutsch (page 264-265), draw some succinct conclusions regarding the perception of musical intervals and scales:

---

[207]Sundberg, J. (1982) Perception of singing *In* Deutsch, D. *ed.* The Psychology of Music. NY: Academic Press Chapter 3

"1. The use of a relatively small number of discrete pitch relationships in music is probably dictated by inherent limitations on the processing of high information-load stimuli by human sensory systems.

2. Natural intervals, in the sense of intervals that show minimal sensory dissonance (roughness) for simultaneous presentation of complex tones, have probably influenced the evolution of the scales of many musical cultures, but the standards of intonation for a given culture are the learned interval categories of the scales of that culture.

> A corollary of this is that the intonation performance of a given musician is primarily determined by his or her ability to reproduce these learned categories and is little influenced, in most situations, by any psychophysical cues (e.g., roughness, beats of mistuned consonances, etc.).

3.The concept of categorical perception, also relate to the limitation on processing of high information-load stimuli, is probably a reasonable description of the way in which intervals are perceived in all but minimal-uncertainty situation, an analogous situation to the perception of phonemes in speech.

4. Quartertone music might be theoretically feasible given sufficient exposure to it, but the present 12-interval Western scale is probably a practical limit. Any division of the octave into intervals smaller than quartertones is perceptually irrelevant for melodic information.

5. Octave generalization is probably a leaned concept with its roots in the unique position of the octave in the spectrum of the sensory consonance of complex-tone intervals."

In the above statements, "categorical perception" appears to relate to the percepts and high level interps of this work. The expression "high information-load stimuli" is suggestive of the most complex melodies (short time sequences of chords) that are documented as specific percepts within the neural system of this work.

They then discuss some caveats based on their studies. The principle caveat is:

"The perception of isolated melodic musical intervals may have little to do with the perception of melody. As several of the chapters [in Deutsch] will indicate, there is considerable evidence that melodies are perceived as Gestalts or patterns, rather than as a succession of individual intervals, and that interval magnitude is only a small factor in the total percept."

9.8.4 More complex performance characteristics related to speech and music

[xxx address the frequency and transient characteristics of both speech and music]

The scientific description of the human ability to evaluate music remains at the exploratory stage, if not the conceptual stage.

[xxx review Weinberger paper from Scientific American, a paper written for a general audience and

providing references to other general works[208].

Eggermont and Ponton have provided good information concerning several complex features of speech recognition[209]. Their comments concerning the characteristic "voice onset time" or VOT are particularly useful. "VOT is the time interval between consonantal release (burst onset) and the onset of voicing. Typical values for VOT in the English language are in the range of 20–50 ms."

9.8.4.1 Perceptual performance related to multiple tones and musical chords

[xxx Fletcher pg 216

[xxx Fletcher pg 217 differentiate between closed auditory duct (binaural systems) and open auditory systems (stereophonic systems).]

9.8.4.1.1 Sensing beats between closely spaced tones EMPTY

Harrison[210] has provided a discussion related to both the perception of difference tones, $f_2 - f_1$, and the tones due to so-called cubic distortion, $2f_1 - f_2$ ($f_2 > f_1$).

9.8.4.2 Distinguishing between like sonorant sounds e.g., /awa/ – /afa/

[xxx section needs more work. It does not lead to any conclusions Look at details of Allen's AI-gram Review Diehl & Lindblom more completely]
The speech interpretation community has been stymied in their understanding of how the brain interprets small semantic differences related to the leading edge of individual sounds, even within a syllable.

The problem is based on the communities failure to recognize the dual character of the auditory signaling and interpretation functions. The IHC relate directly to the temporal characteristics of speech and the OHC relate directly to the tonal characteristics of speech. The OHC signaling channels exhibit a limited rise time to a stable value, just like the P– and Q– chromatic channels in vision. The IHC closest to the vestibule exhibit a much more rapid rise time to a stable value, much like the R– luminance channel of vision.

Until the auditory community recognizes the capabilities of these dual channels, they cannot make significant progress in understanding how the brain perceives the differences between similar syllables.

Diehl & Lindlom have provided an excellent summary of the problem with many graphic examples[211].
However, to the extent their discussion relies upon the model of the auditory system related to

---

[208]Weingerger, N. (2004) Music and the brain *Scientific American* November pp 88-95
[209]Eggermont, J. & Ponton, C.(2002) The neurophysiology of auditory perception: from single units to evoked potentials *Audiol Neurootol* pp 71-99
[210]Harrison, R. (1988) The Biology of Hearing and Deafness. Springfield, Il: Charles C. pg 79
[211]Diehl, R. & Lindblom, B. (2004) Explaining the structure of feature and phoneme inventories: the role of auditory distinctiveness *In* Greenberg, S. Ainsworth, W. Popper, A. & Fay, R. *eds*. Speech Processing in the Auditory System. NY: Springer Chap 3

speech shown in Greenberg and Ainsworth (page 4 of the same source), and in the first two chapters of their text, their task is unmanageable.

Allen and his team have collected thousands of spectrograms from hundreds of informed speakers without obtaining systematic results explaining the speech interpretation task in hearing[212]. A near simultaneous paper provides additional background[213]. His assertions describe his position in 2008, "It has not proved to be possible to generalize from copious examples, or the problem would have proved to be easy. Easy is not a word we may associate with this decoding problem" (page 93). In the colloquial words of the magician, "It is easy if you know the trick.' He has called for broader participation in this type of research to overcome the impasse (page 501). He starts from farther back in the pack than most because, in 2008, he did not accept the fact the OHC, much less the IHC, where neurons (personal conversation).

The fundamental baseline of hearing is developed in;

Section 2.4.2– gross anatomy showing the LOC and MOC paths,
Section 4.3.2– unfolded cochlea showing independent temporal and tonal group paths,
Section 4.5.1– a caricature showing the frequency response of the individual paths,
Section 6.1.1– the two binaural paths combined into single LOC and MOC paths,
Section 8.2.3– details of the creation of the OHC and IHC signal paths &
Section 8.4.2– showing the correlation of multiple OHC signals in thePGN of the thalamus.

The development of two distinct sets of auditory channels is clear from those discussions.

In the case of the temporal channels, it is important to recognize the high degree of relative time dispersal between the individual channels due to the limited (slow) propagation of acoustic energy along the tectorial membrane. Before processing in the higher centers of information extraction, compensation for this dispersion must be removed. As in the visual system (Meyer's loop), the correction is easily accomplished by computational anatomy, the introduction into the temporal channels of a compensating dispersion. This dispersion is easily accomplished due to the similarly slow propagation velocity of stage 3 action potentials and the variation in geometrical path length.

Shannon's model of information transmission, as illustrated by Allen et al. (page 94), was applicable to a single strand serial transmission medium (a telegraph). His receiver (the cochlea by analogy) in fact generates two sets of multiple path signals (roughly 20,000 parallel signal channels as opposed to Shannon's one).

The terminology used in this field is highly tailored to the subject matter. Allen defines the critical terms in his introduction.

Allen et al. describe two methods of describing speech *events* (perceptual features) precisely;

the Speech-Confusion pattern, defined as consonant-vowel confusion as a function of the

---

[212]Allen, J. Regnier, M. Phatak, S. & Li, F.(2009) Nonlinear cochlear signal processing and phoneme perception *In* Cooper, N. & Kemp, D. *eds*. Concepts and Challenges in the Biophysics of Hearing. Singapore: World Scientific

[213]Regnier, M. & Allen, J. (2008) A method to identify noise-robust perceptual features: Application for consonant /t/ *J Acoust Soc Am* vo 123(5) pp 2801-2814

speech-to-noise ratio, and a model acoustic feature (AF) representation called the AI gram, defined as the articulation index density in the spectrotemporal domain.

the Speech-Plosive events,

Allen et al. also introduce the "basic paradox" related to the apparent and instantaneous dynamic range of the human hearing modality. This subject is addressed in **Sections 9.6.4 & xxx** of this work. It will not be addressed here. They make clear their baseline assumption (page 100), "The IHC excitation signal is narrow band with a center frequency that depends on the inner hair cell's location along the basilar membrane." In fact, these are the characteristics of the neurons labeled the OHC, the neurons labeled the IHC exhibit broadband signals with an upper limit determined by their location along the tectorial membrane (with the basilar membrane acting as an inertial mass). Based on their model, they are forced to consider nonlinear (NL) signal processing to attempt to extract the important features from voice patterns. Their summary illustrates their frustration.

9.8.4.2.1 Examples of similar sonorant sounds

Allen and his group prefer to plot their spectrograms as what they describe as Articulation Index spectrograms or AI-grams. The AI-gram plots frequency on the vertical axis and time on the horizontal axis. It also differs from the spectrograms of others by at least four features they describe (page 97);

"First, the AI-gram is normalized to the noise floor."
"Second, unlike a fixed-bandwidth spectrogram, the AI-gram uses a cochlear filter bank with bandwidths given by Fletcher critical bands."
"Finally, the intensity scale in the plot is proportional to the signal-to-noise ratio, in dB, in each critical band, . . ."

Outside of their immediate community, the fundamental presentation is usually described as an auditory or voice spectrogram without these idealizations.

Diehl & Lindblom have published a selection of these spectrograms (pages 109-117) as part of their broader article, including a historical review of the field. The selected spectrograms illustrate some of the critical features they identify as,

1. the sonorant feature,
2. the continuant feature,
3. the nasal feature
4. the feature related to articulation,
5. the voice feature,
6. the strident feature and
7 the vowel feature.

They present a figure similar to Figure 2.4.2-1 (their figure 4.1 based on the morphology of Brodal, 1981). They describe the output of octopus cells as predominantly responding to the onset of sounds but do not recognize distinct temporal and tonal channels projecting to the brain.

Diehl & Lindblom develop the historical development of various terms in speech analysis and focus

on the term “feature theory” to encompass the field. They define a phoneme as a feature bundle. They then ask the question, “What is the explanatory basis of phoneme and feature inventories?”

9.8.4.2.2 An alternate interpretation to the Articulation Index

The Electrolytic Theory of the Neuron provides an alternate to the model used by Allen et al. The main differences are concerned with the adaptation mechanism, the filter bank based on the putative critical bands of Fletcher (or others), and the replacement of the noise floor with the stage 3 threshold for action potential generation. **Figure 9.8.5-1** shows the modified situation based on the earlier figure describing the electrolytic circuitry of the auditory sensory neuron, the arrangement of the auditory signals into two distinct channels based on the OHC and IHC sources, and the stage 3 conversion of the analog signals to action potentials.

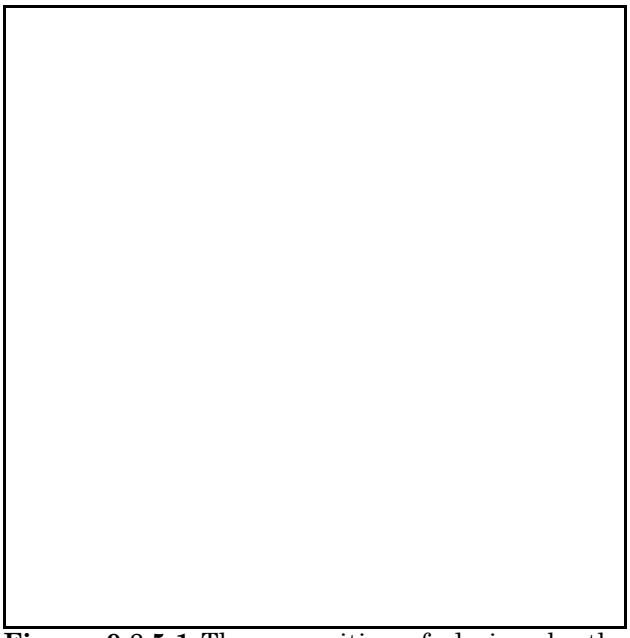


Figure 9.8.5-1 The recognition of plosives by the circuitry of hearing.

The auditory system appears unique in the sense that action potentials are generated at the first Node of Ranvier of the neurons immediately following the IHC sensory neurons. That is the stage 1 sensory neurons interface directly with stage 3 signal projection neurons without intermediary stage 2 signal processing neurons. The stage 2 neurons are in the remote cochlear nucleus.

This initial analog to phasic signal conversion involves a threshold. The threshold is well characterized and will be taken here as a nominal 12 mV relative to the base terminal of the Activa forming the NoR. It is this threshold that is of critical importance in hearing, not a presumed minimum signal-to-noise ratio. For signals below this threshold value at the first NoR, no action potential will be generated in that signaling channel and no distinction between two similar plosives will be reported from the subsequent signal processing (stage 2) and signal extraction (stage 4) circuits.

> While Allen et al. have demonstrated that adding external noise to the acoustic signal provided his subjects will cause loss of a specific plosive percept (page 98), this procedure has little to do with the operation of the hearing modality.

Whether a threshold is exceeded at the first NoR is strongly determined by the output impedance of the sensory neuron and the input circuitry of the NoR. The time constants of the pairs of IHC and NoR described below are well represented in the data of Allen et al. (page 99).

While Allen et al. have continued to conceptualize the adjacent bandpass filters of Fletcher, most investigators confirm and agree there are no frequency domain or temporal domain bandpass filters in the hearing system of the type Allen et al. describe. There is however, a set of spatial domain bandpass filters. The frequency selective mechanism extending across the auditory

spectrum in hearing is the frequency dispersive filter formed by the curvature of the tectorial membrane. as shown in caricature in [**Section 4.5.1**], the result is a series of progressively narrower bandpass filters connected to individual (or adjacent individual groups of) IHC. These channels all exhibit a common low frequency rolloff and progressively lower high frequency rolloff. This arrangement is key to understanding how the system identifies plosive and fricative sounds. The filter arrangement used is illustrated in **Figure 9.8.5-2** combined with one AI-gram from Allen et al.. The high frequency cutoff of each IHC channel filter in frame (A) is determined by its position along the tectorial membrane. The low frequency cutoff of all of the filters is determined by the low frequency rolloff of the stage 0 outer ear (created by the eustachian tube in humans). The filter channels are shown aligned with the frequency scale of the Ai-gram.

The filtering protocol of hearing is fundamentally different from that assumed by Allen et al. although it achieves a similar result. The hearing modality does use the difference between signals captured by low pass filters of different bandwidth. There are no high pass filters used in signal extraction or information extraction related to plosive and fricative speech elements.

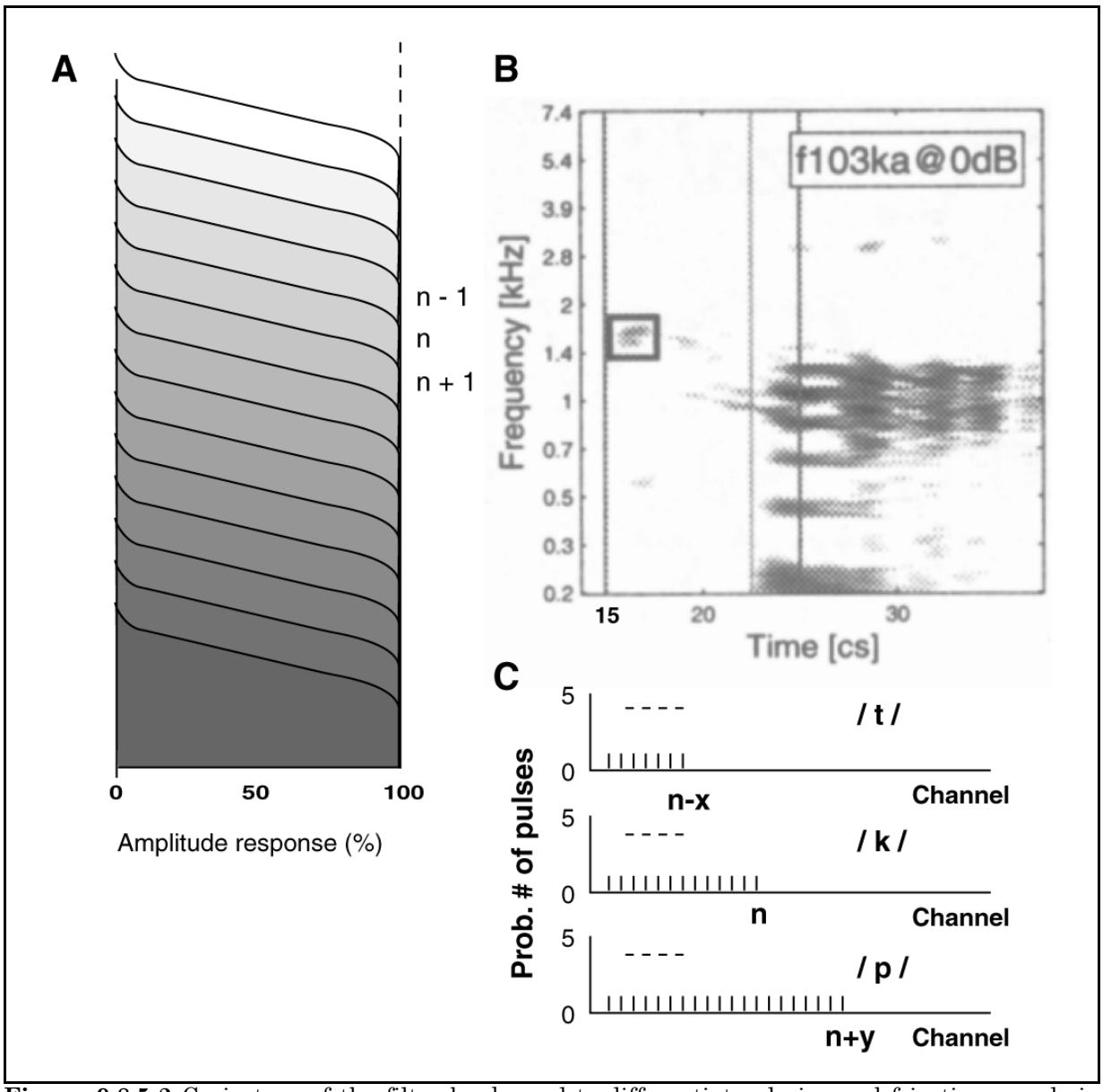


Figure 9.8.5-2 Caricature of the filter bank used to differentiate plosive and fricative sounds in hearing. All of the filters in frame A have the same low frequency cutoff. The AI-gram represents the unvoiced consonant or plosive /k/. The open box in B shows the initial event associated with the start of the plosive. C; several examples of the bit parallel serial words sent to the cochlear nucleus. See text. AI-gram on right from Allen et al., 2009.

In the illustration, the IHC channel n has a cutoff frequency of about 1.8 kHz. Channel n+1 has a high frequency cutoff near 1.4 kHz. As developed in **Section xxx**, the cutoff characteristics are much sharper than that due to single RC filter stages.

All channels numbered less than n pass the signal due to the event within the time interval of 15 and 18 ms after the reference time, C. However, channel n + 1 and all higher numbered channels suppress the signal. Thus, a sufficiently intense acoustic signal in the frequency range passed by the filters numbered less than n exceeds the threshold at the stage 3 encoder in those channels. The result is an action potential will be propagated to the cochlear nucleus in each of those channels. The acoustic signal of the same intensity applied to channels n + 1 and higher will not exceed their stage 3 encoder thresholds and no action potentials will be propagated to the cochlear nucleus by these channels. Therefore the highest numbered channel containing an action potential in the time interval of 15-18 ms represents the event preceding the vowel in the plosive /k/ within the neural system of hearing.

The precise time the action potential is generated in channel n is determined by the intensity profile of the event leading the vowel in the plosive stimulus combined with the temporal response of the channel.

Frame (C) shows the bit-parallel serial words generated by the three unvoiced plosives addressed by Allen et al., / t /, / k / and / p / in the stage 3 channels following the sensory neurons. The events preceding the vowels in the unvoiced consonants /t/, /k/ and /p/ are identified within the neural system by the respective highest numbered IHC channel (starting from the base of the tectorial membrane) issuing an action potential within the event time interval.

The minimum values of x and y in this figure are undetermined. However, Allen et al. have a protocol for generating acoustic signals with these values set arbitrarily. this protocol can be used for establishing the recognizability of the sounds versus the values for these parameters. Such tests may indicate some languages uses smaller intervals than others.

A nominal integration interval of 10 ms is chosen for illustration following the start of the leading event These channels have a pulse bandwidth of 500-1000 pps. Therefore, the number of pulses generated by a perceived initial event in a unvoiced plosive could be not less than one and as many as 5-10. The absence of a count indicated no sensed, and therefore no perceived, signal in that channel. A value in the range of 1-4 could be associated with normal speech and processed in the analytical mode of the CNS. Values above 5 (the dotted lines in the graphs) can be associated with the alarm mode within the CNS. These high values could be due to strong emphasis by the nominal speaker, or a significant distracting influence such as a nearby explosion.

Insufficient information is available concerning the cochlear nucleus (**Section 6.2.3** & **6.5**) to describe the actual signal processing performed relative to the parallel bit serial words associated with the IHC channels. It may involve processing in both the phasic signal domain and recovery of the analog signals before processing in the analog signal domain.

The ability of the system to develop correct percepts of unvoiced consonants will deteriorate precipitously as the stimulus level is decreased below the required threshold level.

While noise masking can also cause a loss in percept development, the signal-to-noise ratio associated with the stimulus is not directly associated with the percept generation process in hearing.

A conventional spectrogram can be used in the above figure just as effectively as the AI-gram. The processing of a spectrogram into an AI-gram is largely unnecessary except for improving the legibility of the printed graph where the black to white ratio of printing is limited to less than 25:1.

There is no technical justification for adjusting the density of the printed bands, as Allen et al. have done, based on either the Fletcher or Zwicker critical bands (**Sections 9.2.3** & **9.5.3**). These bands were developed based on continuous tone tests and are pertinent when discussing the signal manipulation within the peri-geneculate nuclei of the auditory CNS. They have not been shown to be a feature of, or in any way involved in, the temporal IHC channels. The actual adaptation mechanism of hearing is associated with the rate of bond breaking versus the replacement rate of the piezoelectric material in the cilia of the individual sensory neurons.

The spacing of the individual IHC styli (groups of cilia of a single IHC bound as a group) corresponds to the spacing of the OHC styli. This spacing defined in frequency space also defines the channels in the above figure. This spacing (typically 0.5% of center frequency) is much finer than range of a single critical band (typically 8% of its center frequency).

Allen et al. noted the immaturity of their AI-gram (page 97) and suggest additional features. However, there is little reason to incorporate additional masking features into the AI-gram when trying to understand the auditory modality. Such additions could be useful in understanding the effects of noise in the workplace. Nonlinear signal processing (beyond that intrinsic to the sensory neurons and the stage 3 encoding and decoding circuits) is not needed to explain the interpretation and perception of human speech. The stage 2, stage 4 & stage 5 circuitry appears to be remarkably linear (with some thresholding in the decision processes).

9.9 Passive source characterization, including source location as a subelement

[xxx need good copy of early Jeffress in this section]
[xxx see Tobias 1972 including chapter by Jeffress]

Source location is only a portion of the overall task of source characterization. Source characterization encompasses all aspects of data extraction related to an acoustic stimulus. A majority of the data and concepts associated with source characterization come from psychophysical experiments with humans. Most of the data for the narrower subject of source location comes from a combination of both human and animal psychophysics, and electrophysical experiments on animals.

The effectiveness of passive location in hearing depends on the anatomy of the animal, which is closely correlated with its mode of feeding. Humans have their ears mounted on the sides of their heads like the majority of grazing animals. Their direction finding in the forward direction is relatively poor. Animals such as the barn owl that rely on hearing much more than humans to support their predatory tasks have their ears both facing forward. With sufficient separation, their ears operate as an effective antenna array.

Two decidedly different mechanisms are used in passive source location. Both are based on time differences. The first involves the IHC sensory channels and relies upon absolute time of arrival differences of the signal at the two ears. This is the mechanism used to respond to a hand, or nearby thunder, clap. The second involves the OHC sensory channels and relies upon relative time differences as expressed by the phase difference between the received signals at individual frequencies within the stimulus. The two systems compliment each other.

Both the IHC and OHC based systems of source location exhibit a just noticeable interaural delay

(or differential time delay) of about 50 microseconds. This is a much smaller number than that associated with the total delay between acoustic stimulus generation and Stage 4 signal manipulation leading to cognition. Fortunately, the auditory system consists of fixed length acoustic and neural paths that grow very slowly. This allows the subject to maintain precision in his source location capability over time through simple adjustments to his lookup tables associated with source location.

Moore gives some more specific, and in some cases much smaller, time intervals of interest (page 175). "For sounds with ongoing transient disparities, such as bursts of noise, our ability to detect interaural time difference improves with duration of the burst for duration up to about 700 ms, when the threshold disparity (the smallest detectable time difference at the two ears) reaches an asymptotic level of about 6 μs (Tobias and Zerlin, 1959)." This relationship between duration and precision is a clear suggestion of a correlation mechanism operating over a considerable length of time.

McAlpine & Grothe relied upon the conventional wisdom and found problems describing the source location performance of the barn owl and guinea pig[214]. They relied heavily upon the conventional wisdom; they looked for distinct delay lines rather than the delays inherent in the cochlear partition and the axons of neurons, they assumed glycine directly inhibited neural action, they only conceived of the tonal channels of hearing and they conceived of the Jeffress coincidence model in terms of those tonal channels. Their conclusion, based on their concept, was vigorous. ". . .mammalian sound localization must be achieved by some means other than the local-coding strategy suggested by the Jeffress model and, apparently, adopted by the barn owl." In their concept, local-coding referred to the tonal channels compatible with the von Belesky cochlea. Their next sentence leads to the broader model used here. "One possibility is that mammals localize sound by means of a population code, in which the lateral position of a sound source is determined by the relative activation of just two broadly tuned binaural channels, one in each brain hemisphere, beginning in the left and the right MSO. These channels are the first channels of the IHC according to this work. They begin in the cochlea instead of the MSO. The new suggested model of source location of McAlpine & Grothe (last paragraph of their paper, pg 350) is in line with system defined in this section but uses words that will only confuse if presented here.

Figure 9.9.1-1, modified from Stevens & Newman[215], shows the fundamental division between the two types of source location based on measurements. The data set has been overlaid with two construction lines. The dashed lines show the limit in absolute error in localization due to the low-pass filter associated with the adaptation amplifier in each sensory neuron. The dash-dot line shows a similar limitation (shown as a high-pass filter characteristic) due to the limited performance provided by the coincidence detector circuits processing the wideband data provided by the IHC neurons. The slope of the leading edge of the envelope of the tone burst used decreases as the frequency of the tone is lowered. This reduction in slope results in a lower precision in the coincidence measurement circuit. It can be assumed the envelope of the tone burst was not synchronized with the tone oscillator in these early tests. This lack of synchronization can also add a statistical variable to the slope of the leading edge of the tone burst which will be reflected in the measurements. [xxx why such large errors here?]

---

[214]McAlpine, D. & Grothe, B. (2003) sound localization and delay lines–do mammals fit the model? *TINS* vol. 26(7), pp 347-350

[215]Stevens S. & Newman, E. (1936) The localization of actual sources of sound *Am J Psychol* vol. 48, pp 297-306

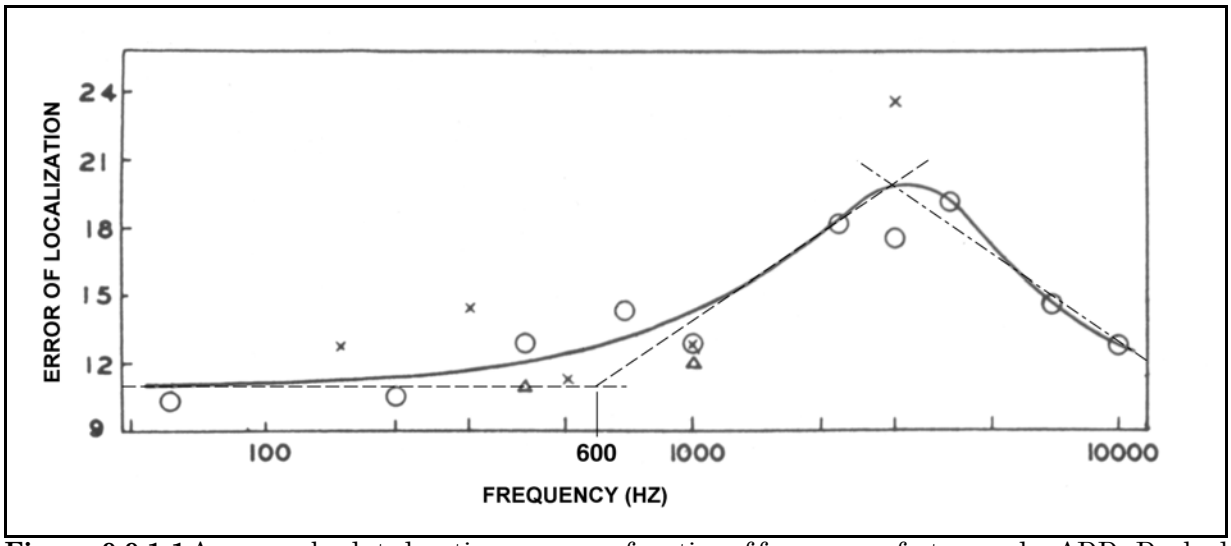


Figure 9.9.1-1 Average absolute location error as a function of frequency of a tone pulse ADD. Dashed line, envelope of error due to low-pass filter in the adaptation amplifier of the OHC. Dash-dot line; error due to coincidence circuit performance as the slope of the leading edge of the pulse envelope decreases in the IHC based circuits. Data from Stevens & Newman, 1936.

[xxx add words concerning differences in impulse responses from IHC along the cochlear partition as well as the amplitude and frequency change during a single formant.]

The figure suggests that the IHC are not used for source location at frequencies below about 3000 Hz. This is consistent with the change in the wiring harness supporting the interface of the IHC sensory neurons and the Stage 2 signal processing that occurs in the vicinity of 16 mm from the base of the cochlear partition in humans.

Passive source location is frequently discussed without differentiating adequately between the cues associated with constant tones, the envelope associated with intermittent tones and impulse stimuli. This leads to difficulty in interpreting the accompanying data properly. The 1990[216] and 1994[217] papers of Heffner & Heffner provide good data without differentiating between these areas carefully.

9.9.2 The direction finding performance of the auditory system NO COUNTER

The direction finding capability of the human is considerably less refined than that of many mammals. This is due primarily to the loss of maneuverability of the pinna of the outer ear. Direction finding in the absence of such maneuverability is inherently ambiguous as easily demonstrated using a set of binaural earphones and a signal source. {xxx Museum of Industry demo] The minibrain associated with the cochlear nuclei is able to provide a signal (an interp) defining the difference in arrival time for signals. However, this signal is not monotonic. To

[216]Heffner, R. & Heffner, H. (1990) Evolution of sound localization in mammals *In* Webster, D. Fay, R. & Popper, A. eds. The Evolutionary Biology of Hearing. NY: Springer-Verlag. Chapter 34, pp 691-715

[217]Heffner, R. & Heffner, H. (1994) xxx

determine the actual direction of the signal, animals with mobile outer ears will make a second determination with the ears rotated to a different position. By combining the computations from these two measurements, a monotonic direction determination (a monotonic interp) can usually be obtained. Even this determination is frequently confirmed by combining information from the visual system to produce a definitive (and much broader) percept related to the signal source.

The auditory system uses a variety of cues related to a stimulus for source location purposes. To maximize the potential for source location, both impulse related (LOC) and tonal based (MOC) channels are used. These channel capacities are different. The impulse channel signals appear dedicated to the alarm mode of operation while the tonal channel signals are dedicated primarily to the awareness and analytical mode of operation. The latter play a major role in the enjoyment of music in humans. This unique human capability is probably associated with the large temporal lobes of the cerebral cortex.

The impulse channel provides the most immediate alarm mode signals, as exemplified by our response to a report of a gun shot, or even a door slamming. The time delays associated with these events are easily characterized.

The simplest cues to analyze are those based on one or a few specific tones. Estimates of the source are provided for single tones based on both amplitude and phase differences between the stimuli arriving at the two ears. However, as will be shown below, these estimates are frequency dependent and ambiguous. Other cues must be sought in order to obtain precise source location estimates.

In most cases, the requirement is for the auditory system to provide an angular precision, at least in the horizontal plane containing the pinna of the ears, sufficient to allow the most sensitive portions of the visual retina to be directed toward the source. This value is about +/– 0.6 degrees for humans, the size of the foveola of the visual system. In some animals, the retinas contain more than one area of peak sensitivity. In these animals, other cues are needed to direct the eyes appropriately.

The Heffner & Heffner team have explored the sound location capability of a large variety of animal in detail[218,219]. They arrived at the same proposition regarding the localization requirement of the auditory system based on their empirical research.

Figure 9.9.1-2 from Heffner & Heffner (1990) suggest this criteria is usually met in a wide variety of animals. However, they did not describe the precise sounds that provide this level of performance. The level of performance is highly dependent on the character of the test stimulus used. The difference in performance between a true click, a "plop" (a poor quality click), and a long tone burst are significant. The 1994 papers described the sound stimuli they used. It did not include a short duration click (qualifying as an impulse). Additional data is provided comparing the performance of many species.

---

[218] Heffner,R. & Heffner, H. (1990) Evolution of sound localization in mammals *In* Webster, D. Fay, R. & Popper, A. eds. The Evolutionary Biology of Hearing. NY: Springer-Verlag. Chapter 34, pp 691-715

[219]Heffner, R. et. al. (1994) Sound localization in chinchillas. I: left/right discrimination *Hear Res* vol. 80, pp 247-257

The data of Heuremann & Colonius is indicative of the performance that can be obtained using a one second duration "click" at 65 dB SPL[220]. This click is a truly poor plop and hardly more than a tone burst. Their results suggest an azimuth precision of about +/– 6.4 to 7.6 degrees. Their data included some tests using an artificial head-related transfer function (HRTF).

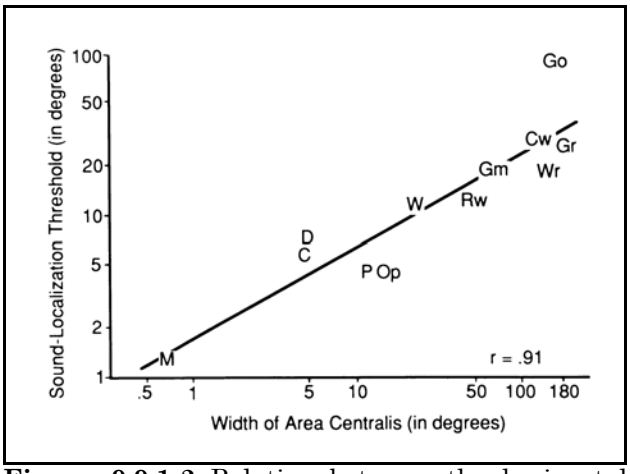


Figure 9.9.1-2 Relation between the horizontal sound localization acuity and the area centralis of vision. M; man. C; domestic cat. D; domestic dog. P; domestic pig. Op; opposum. W; least weasel. Rw; wild rat. Go; gopher. From Heffner & Heffner, 1990.

9.9.1 Coordinates and parameters used in source location research

9.9.1.1 Coordinates used in source location research

Investigators have frequently defined coordinate systems for use in their experiments. However, their notations have lacked consistency. When speaking in terms of inertial coordinates related to a subject, it is important to separate the six degrees of motion properly. There are three *linear coordinates* associated with translational motions of the subject, up–down, left–right and forward–backward. Similarly, there are three *angular coordinates* describing the location of a remote point relative to the subject, its azimuth, its elevation and its radial distance. It is inappropriate to use translational terms to describe angular coordinates.

9.9.1.2 Stimulus parameters as measured at the ears

Obtaining consistent results when measuring the interaural acoustic time difference of humans is difficult. Many investigators have provided data based on different head geometries. Fedderson, et. al. have presented calculations of interaural time delay (ITD) based on a spherical head that match their data well[221], **Figure 9.9.1-3**. This data only applies to the acoustic element of the overall ITD. Additional delay related to the middle ear (small), inner ear and cochlear nerves must be considered when addressing the entire signaling problem. The linearity of this function makes neural computations involving this function simple.

---

[220]Heuermann, H. & Colonius, H. (1999) Localization experiments with saccadic responses in virtual auditory environments *In* Dau, T. et. al. *eds*. Psychophysics, Physiology and Models of Hearing pp 89-92

[221]Fedderson, W et. al. (1957) Localization of high frequency tones *J Acoust Soc Am* vol. 29, pp 988-991

Moller noted (page 10), "When an ear is artificially stimulated, such as by an earphone, the physical basis for directional hearing is naturally lost because the head does not produce an interaural time and intensity difference. In addition to that, and sometimes more importantly, the earphone modifies the resonance effect of the outer ear (canal and concha). An earphone, such as those used in hearing aids, inserted in the ear eliminates the efffect of the concha and modifies the effect of the ear canal. The result is a loss in the acoustic gain at high frequencies normally obtained by the external ear. An earphone with a cushion, such as those used in clinical audiometry and stereo listening, compresses the concha and modifies the ear canal resonance."

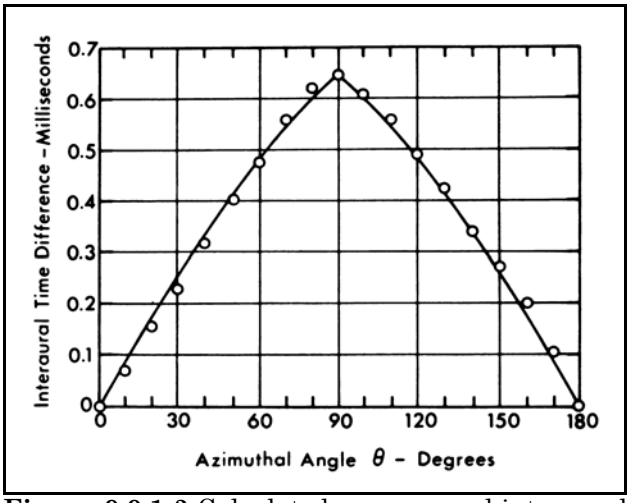


Figure 9.9.1-3 Calculated vs measured interaural time differences as a function of azimuth for a spherical head of 8.75 cm. Only delays associated with the signals in air are considered. Open circles, measured values for the human. From Feddersen, et. al., 1957

Some investigators have defined a set of head-related transfer functions (HRTFs). In the form introduced by Wightman & Kistler and used by Yost, this term appears inappropriate[222]. Their data represented signals at a point and used relative scales. A better display would use an absolute amplitude scale relative to the source intensity (thus introducing a distance factor). **Figure 9.9.1-4** expands on the Wightman & Kistler data set. The original data set was a mixed set, the response of each outer ear when coupled to the environment in the frequency domain and the time delay encountered by each ear in the temporal domain. Not shown was the phase difference associated with the signal in the frequency domain. The data is best interpreted by breaking it into two sets. First, consider the amplitude and time delay related to any click (or envelope of the test signal) in the time domain. Wightman & Kistler noted that the time delay remained relatively constant versus frequency for this fixed angle of stimulation. However, the phase, as a function of frequency clearly does not. Second, consider the amplitude and phase related to any continuous tone in the frequency domain. Their figure has been overlaid with a set of curves representing the phase delay, as a function of frequency, for tones. Looking at the phase information, it is immediately obvious why the sensory neurons of the hearing system in humans was not designed to maintain phase integrity above about 1200 Hz. On a single channel basis, the available phase data is ambiguous for higher frequencies. However, if the hearing system was able to process the frequency and phase information in multiple parallel channels, the information could be assembled in a vernier system. Such a system would provide excellent angular location information based on broadband input signals. This would require selected circuits of Stages 1 through 4 have much broader bandwidth. The limited bandwidth of the sensory neurons precludes this possibility. The experimental data does not suggest the system is able to perform such multi-channel calculations. [xxx confirm this, simplify the text]

[222]Wightman, xxx & Kistler, xxx (1989) xxx

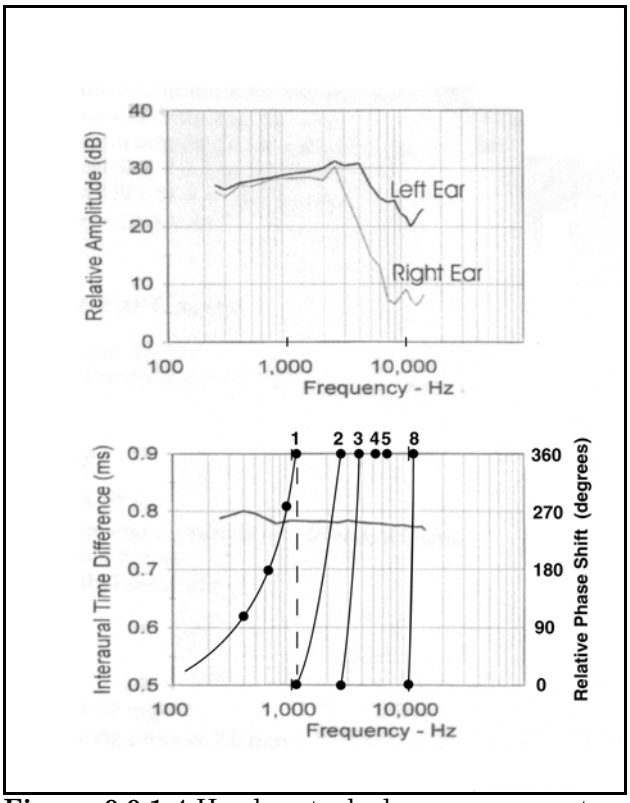


Figure 9.9.1-4 Head perturbed source parameters for a human when presented a wideband noise source directly opposite the left ear. Upper frame, amplitude spectrum for each ear (measured within the ear canal). Lower frame, interaural time difference between the sound arriving at the two ears (approx. horiz. line) and the phase difference for individual frequencies radiated by the source (family of sloping lines). Based on Wightman & Kistler, 1989)

Shaw has provided a useful figure showing the combined effects of resonance in the ear canal and in the outer ear, as well as the affects of diffraction of the head for a source located directly in front of the head[223]. It is reproduced in **Figure 9.9.1-5** and emphasizes the considerable gain introduced into the signaling channel as a function of frequency. The values are considerably higher than those of Wightman & Kistler.

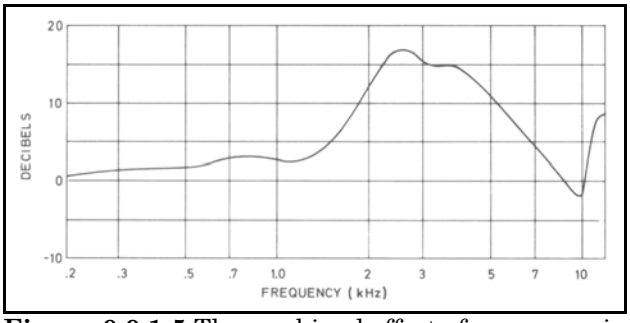


Figure 9.9.1-5 The combined effect of resonance in the ear canal and the outer ear, and diffraction of the head. The curve shows the relative increase in sound pressure at the tympanic membrane with a person located in a free sound field with the sound source placed in front of the person, relative to the sound pressure at the same place but without the person being present. From Shaw, 1974.

[223]Shaw, E. (1974a) Transformation of sound pressure level from the free field to the eardrum in the horizontal plane J Acous Soc Am vol. 56, pp 1848-1861

9.9.1.3 Comparison of the cochlear microphonic, the complete cochlear microphonic and the ABR

Figure 9.9.1-6 presents three frames describing the cochlear microphonic, the complete cochlear microphonic (also known erroneously as the compound action potential) and the auditory brainstem response. See [xxx Moller83 pg 98]

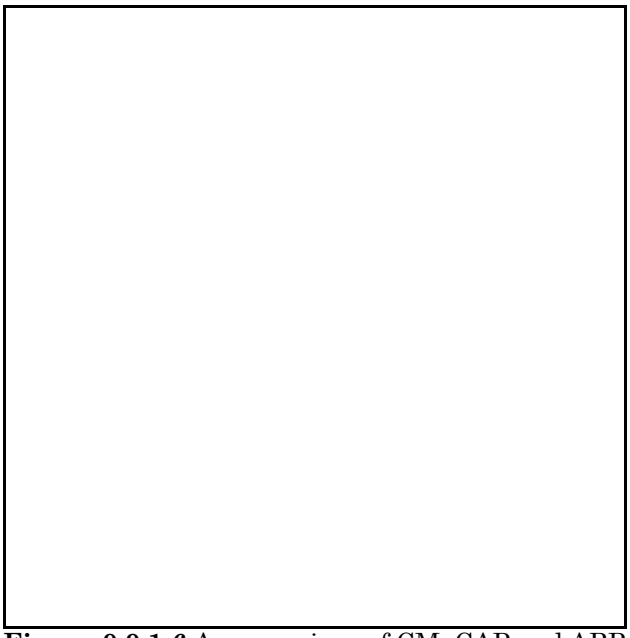


Figure 9.9.1-6 A comparison of CM, CAP and ABR waveforms EMPTY ADD.

9.9.2 Transient- based (pulse) source location processes EMPTY

9.9.3 Tone-based source location processes

Tone based source location involves both differences in phase and differences in amplitude of the tonal signals. The difference in phase can be large for animals with ears separated by more than one-half wavelength of the frequency involved. **Figure 9.9.2-1** shows the critical distance required to achieve a one-half wavelength distance between the ears of xxx. With adequate separation, the performance of the system can be explored using conventional antenna theory. [xxx show quarter wave also?]

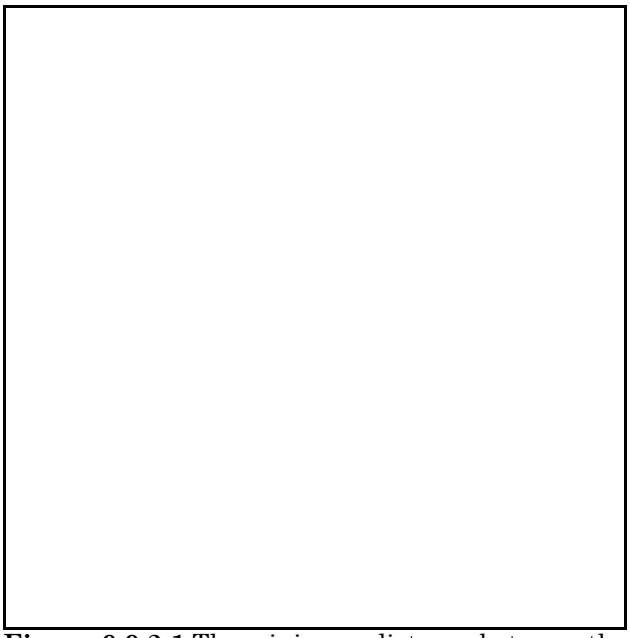


Figure 9.9.2-1 The minimum distance between the ears to achieve one-half wavelength separation EMPTY.

McAlpine & Grothe have provided **Figure 9.9.2-2** illustrating the significance of this difference in predatory and non-predatory animals. The barn owl has the ears located to largely eliminate the interfering effect of the head and uses much higher frequency signaling channels.

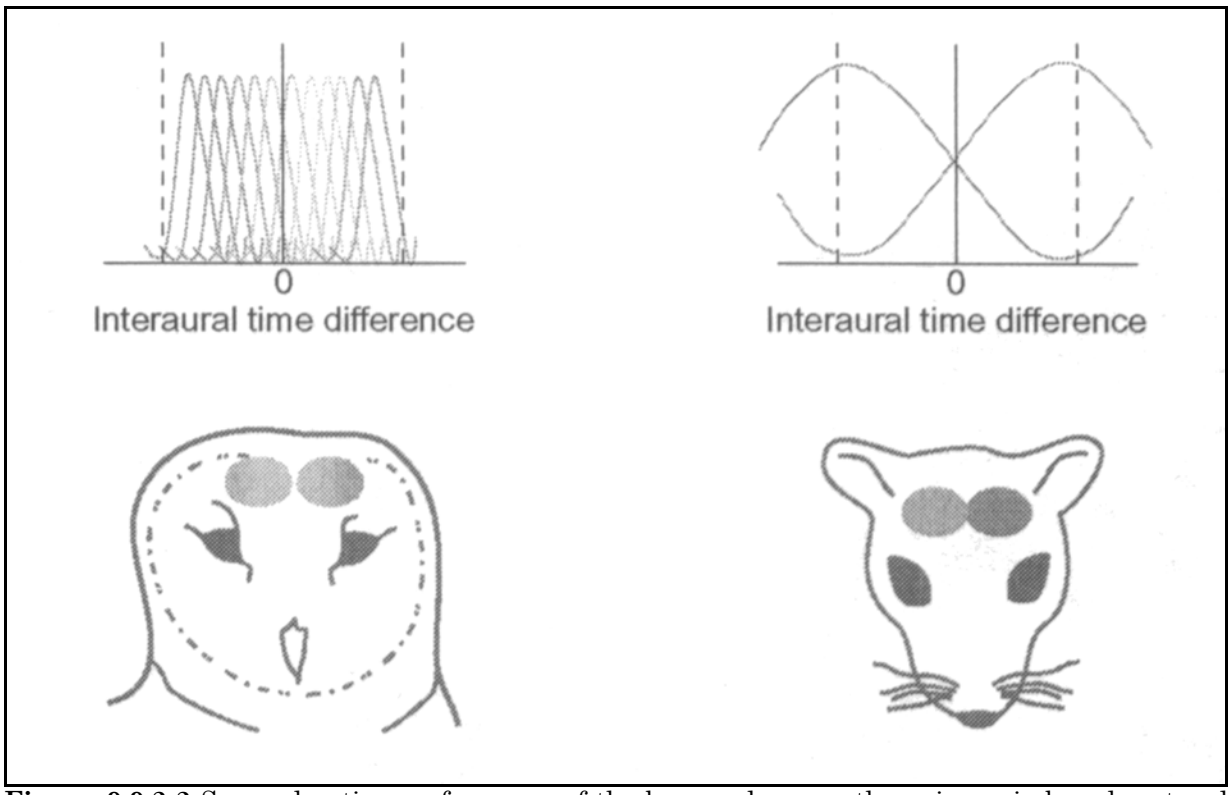


Figure 9.9.2-2 Source location performance of the barn owl versus the guinea pig based on tonal channel differences. Vertical dashed lines represent the relative width of the head compared to the angular accuracy of the source location function. The predator has much higher source location capability. Modified from McAlpine & Grothe, 2003.

9.9.3.1 Psychophysical performance using a single tone

Binaural source location has been studied when the signals delivered to the two ears are at the same frequency but differ in either amplitude or phase.

9.9.3.1.1 Performance varying the phase of a single tone

Fletcher has provided a particularly clear description of binaural source location based on phase, although he did not describe the purpose of the phenomenon. Nor did he introduce the participation of the visual system in resolving the ambiguity as to whether the source is in the front hemisphere or the rear hemisphere relative to the axis of the ears. [xxx chapter 12]

When instrumenting to vary the phase of a pure tone received by the two ears from a synthetic source, Fletcher describes three first order circumstances:

1. "When the difference in phase between the two paths conducting the sound to the ear in zero, then the source appears to be in the median plane, that is, directly *in front* of the individual." In fact, in the absence of visible cues, the subject cannot discern whether the sound is directly in front of the individual or directly behind him.

2. "As the difference in phase increases, the sound image travels, apparently, along a circle toward the ear which is leading in phase." Leading in phase refers to the peak intensity occurring first in that channel. Here again, in the absence of visual cues, the sensation is that the sound moves from the medial plane (front or rear) toward the axis of the ears on the side where the stimulus is leading in phase.

3. "When the image appears to be opposite the ear, it suddenly seems to jump through the head to the other side, finally coming back again to the median plane." It approaches the front medial plane based on visual cues. In the absence of visual cues, the image appears to move ambiguously toward the median plane (either front or rear).

While not stated, the subject will assume the artificial source moved in the horizontal plane unless told otherwise.

Harper & McAlpine have described this phenomenon as phase wrapping in their electrophysical experiments with gerbils and calculations related to a variety of species. They particularly noted the importance of the size of the head with respect to the frequency where phase wrapping is encountered[224].

Rearranging Fletcher's equation, the approximate perceived source angle, θ, from the medial plane is given by, $\theta = \dfrac{\phi}{0.0034 f + 0.8}$ where f is the frequency and ϕ is the phase difference.

Fletcher did not define the limits of the frequency. The equation provides an interesting feature of source binaural source location. At a frequency of 60 Hz, the denominator approximates 1.00 and the spatial angle from the median is approximately equal to the phase difference. However, at higher frequencies, the perceived spatial angle is less than the phase angle. At 500 Hz, the spatial angle approaches +70 (or 110) degrees at 90 degrees phase angle. At this point, the perceived location of the source suddenly jumps to –70 (or –110) degrees in the absence of visual cues. Thus binaural source location is limited at frequencies greater than 60 degrees.

The above equation can be differentiated with respect to both frequency and phase angle. The result gives the spatial angular sensitivity with respect to the angle.

$$\Delta\theta = \frac{\Delta\phi}{0.0034 f + 0.8} - \frac{0.0034 \cdot \phi \cdot \Delta f}{\left(0.0034 f + 0.8\right)^2}$$

To the extent the original equation is accurate, this relationship shows the precision that can be achieved for a differential change in phase angle.

---

[224]Harper, N. & McAlpine, D. (2004) Optimal neural population coding of an auditory spatial cue *Nature* vol. 430, pp 682-685

9.9.3.1.2 Performance varying the amplitude of a single tone EMPTY

[xxx imbalance between the two ears of the typical subject, both in amplitude and possibly phase.

9.9.3.2 Psychophysical performance using multiple tones

9.9.3.2.1 Performance varying the phase of multiple tones

The expressions shown in **Section 9.9.1.1** are complicated functions of the frequency and angle. The expressions shown in **Section 9.9.1.2** based on amplitude differences are also complicated. However, by using both the frequency data and amplitude of multiple individual frequency channels and referring to a lookup table, the auditory system can arrive at a unique best estimate of both the angle and the change in angle of the source. If augmented by cues from the awareness mode of the visual system, no ambiguity remains.

[xxx complete this section]
[xxx tie the above to the portion of the minibrain acting as a lookup table.

9.9.3.2.2 Performance varying the phase of multiple tones due to multiple paths

{xxx fletcher page 213]

The above multiple tone methodology can be extended further. If the individual tones emanating from the source are reflected from at least one additional surface, the auditory system is provided additional cues that may vary significantly with frequency component. This additional variation can be incorporated into the lookup table based on training. It is an effective method of increasing the capability of the system.

Reflections as described above are very useful. They frequently allow a subject deaf in one ear to achieve significant source location capability.

9.9.3.3 Perception of more complex tonal patterns EMPTY

[xxx chapter 13 in Fletcher]

Witton, et. al. have provided data on source location, under diotic and dichotic conditions, where the source energy was modulated[225]. The data appears useful but its application to the overall model or to hearing in the real world appears limited.

9.9.4 Perceiving sound arriving via different paths EMPTY

---

[225]Witton, C. Simpson, M. Henning, G. Rees, A. & Green, G. (2003) Detection and direction-discrimination of diotic and dichotic ramp modulations in amplitude and phase *J Acous Soc Am* vol. 113(1), pp 468-477

[xxx Fletcher 214]

[xxx Both applied to one ear. Tones applied to different ears.
Algazi, et. al. have provided a recent discussion and some measurements related to different shaped torsos and heads in source location[226].

9.10 Binaural hearing

9.10.1 Beats and cross-modulation related to binaural hearing

While discussing pure tone interference, Fletcher raised and answered a very important question. "The question arises: 'Does the same interfering effect exist when the two tones are introduced into opposite ears instead of both being introduced into the same ear?' The answer is 'No.'" His discourse centered around the fact that over a wide range of frequencies, the requirement under binaural conditions was 50 dB higher than under monaural conditions for the same level of interference. He attributed this to the requirement that the two sounds be processed within the same PNS apparatus and that the bone conduction path loss between the two ears in human was about 50 dB. Only at frequencies below 900 Hz was the threshold shift less than 50 dB (page 158). This conclusion is supported by the theory of this work. At frequencies considerably above 600 Hz, the CNS does not process signals related directly to the stimuli. Only the place addresses, and codes associated with the temporal characteristics, of the stimuli are processed. Interference does not occur between such abstract addresses. Beat formation and cross-modulation occur primarily within the sensory neurons, not within the CNS.

Scharf provided a comprehensive set of data in 1969 showing that the two auditory channels operated largely independent of each other[227]. He established that the ultimate summary perception under dichotic conditions was a linear addition of the two perceived signals regardless of the frequency of the two signals. This finding is consistent with the theory presented here that only abstract percepts are summed in signal manipulation of Stage 4 for presentation to the cognitive functions of Stage 5. He also showed that the critical band concept played no role in dichotic summation.

9.10.2 Active source-location techniques in other species

[xxx Pollak, chap 36 in Webster, Fay & Popper 1990]

9.10.2.1 Active location in air–the bats, Order *Chiroptera* EMPTY

[xxx Suga[228] in 1988]

9.10.2.1.1 Potential active jamming of bat sonar

---

[226]Algazi, V. Duda, R. Duraiswami, R. Gumerov, N. & Tang, Z. (2002) Approximating the head-related transfer function using simple geometric models of the head and torso *J Acoust Soc Am* vol. 112(5) pt 1, pp 2053-2064
[227]Scharf, B. (1969) Dichotic Summation of Loudness *J Acoust Soc Am* vol. 45(5) pp 1193-1205
[228]Suga, N. (1988) Auditory neuroetholoogy and speech processing: . . . *In* Edelman, G. Gall, W. & Cowan, W. *eds*. Auditory function. NY: John Wiley & Sons pp 679+

[xxx Alexander has made a useful observation[229]. "One group of moths, the arctiid, or tiger moths, do something remarkably similar to jamming. Not only can they hear bat echolocation calls, but they also have a structure near the base of each wing that produces ultrasonic clicks in approximately the same frequency range as bat sonar. When an echolocating bat closes in on a tiger moth, the moth begins emitting its own ultrasonic sound, and in most cases the bat breaks of the attack." This capability emphasizes the phase-insensitive character of the broadband bat sonar system. The added in-band noise encountered by the bat probably interferes primarily with his ability to determine the velocity of the target (range rate information) using Doppler frequency determination.

9.10.2.2 Active location in water–marine mammals, Order *Cetacea* EMPTY

9.10.2.2.1 Non-cooperative active location in water EMPTY

9.10.2.2.2 Cooperative active location in water, by communications EMPTY

9.10.3 Active source-location by humans

A recent rash of publicity has focused on the ability of blind humans to perform echolocation with reasonable precision. Recognition of this capability is spawning an increased level of research in this phenomenon and its underlying mechanisms.

The phenomenon of active ranging has been achieved using an external source of sound as well as the use of the vocal system to create click sounds, both forms of multistatic sonar. For the vocal source as well as the use of a handheld cricket-type clicker, the sound source is controlled by the brain. The use of uncontrolled external sources is another case of binaural sound location involving receipt of sound along multiple paths, some due to reflections.

Ben Underwood has achieved a remarkable and well documented capability to interpret his world based on his own vocal emanations. While the capability has been described as unique and something he has learned, it appears to be an innate capability that is seldom used to its potential by humans. Kish has documented the capability based on his wide experience as a Orientation and Mobility instructor in "FLASH SONAR PROGRAM– Helping blind people learn to see.[230]" He coined the term flash sonar as a simpler term to use in conversation than echolocation (particularly with the young). He also tries to associate with this term the greater capabilities in echolocation available to one trained in the subject.

As shown in **Section 2.4.1**, the architecture of the human hearing modality is the same as that of the dolphins and the bats, although both the vocal and hearing systems exhibit significantly narrower frequency ranges. The binaural system has an innate capability to perform ranging using

---

[229]Alexander, D (2009) Why Don't Jumbo Jets Flap Their Wings? Brunswick, NJ: Rutgers Univ. Press
[230]Kish, D. (2006) FLASH SONAR– Helping blind people see,
http://www.worldaccessfortheblind.org/snr-curiculm1106.pdf

the intrinsic capabilities of the inferior colliculus. While this organ is frequently described as rudimentary or even atrophied in humans, it exists. These descriptions are based on very limited autopsy statistics and do not reflect any effort by the owner to develop his echolocation skills. If developed at an early age, it is quite possible this organ could provide a significant human capability in totally natural (no external aids) echolocation.

The use of clicking sounds as part of the human vocal capability arose early in human evolution. Its capability was closely aligned to the ability to echolocate based on sound. However, this auditory capability was inferior to the capabilities of the visual system and it was soon abandoned as a utility. The ability to generate and interpret clicking sounds also found a place in human communications. This capability was also found to be of marginal utility and was largely abandoned as language skills developed with the expansion of language out of Africa. The capability lives on today only in the Khoisan languages. Quoting Wikipedia, "The Khoisan languages (also Khoesaan languages) are the indigenous languages of southern and eastern Africa; in southern Africa their speakers are the Khoi and Bushmen (Saan), in east Africa the Sandawe and Hadza. They are famous for their clicks."

Khoisan languages are best known for their use of click consonants as phonemes. These are written with letters such as ! (a post alveolar click) and ‡ (a palatal click). The Ju|'hoan language has some 30 click consonants, not counting clusters, among perhaps 90 phonemes, which include strident and pharyngealized vowels and four tones.

The following material is verbatim from Kish:

III. PRELIMINARY CONSIDERATIONS: There are 3 primary considerations to teaching, using, and evaluating flash sonar. These are target distinction (how detectable are the targets), environmental variables (noise and clutter), and the perceptual factors in the student (hearing issues, presence of vision, attention caapacity).
A. Target Distinction: The maximum resolution of sonic, unaided, human flash based sonar is about 9 square inches (a circular or squarish target) at about 18 inches distance from the listener with a solid target presented alone in open space under quiet conditions using an echo signal with primary frequency at about 3 kHz. This figure is general, and drawn from a synthesis of the literature and my experiences as a user and teacher. Targets that are very narrow, such as a pole, may not bounce back as much sound, and may be more difficult to detect. The more sparse (less dense or solid), the target, such as a fence, the larger it will probably need to be to bounce enough acoustic energy back to a human listener to be detectable. Of particular concern with human sonar is figure ground. This concept acoustically is very similar to this same concept as it relates to the visual system. It has to do with the extent to which the target can be distinguished from its surroundings. Acoustically, we are talking about physical geometry and texture of the target relative to its surroundings. These need to be quite distinct for a target to register to the human audible system, but experience and concentration can narrow this gap. It's hard for me to quantify this distinction, but we get an idea of it from the resolution data above. If a target needs to be about 9 square inches to be perceived at a distance of 18 inches from the listener, this gives us a general idea of how distinct a target needs to be in order to be registered, let alone identified. Objects that are too close to each other tend to blur together, with larger, more dense objects predominating. For example, while a person of average man size may be detectable at about 7 feet, that same person may completely vanish if standing next to a wall or large column. However, he might still be detectable against a chainlink fence. I'm sure we could come up with some

clever formula to approximately represent the extent of distinction necessary, but this is beyond my math, and it would depend very much on target characteristics. I always gage this by ear
when working with students. Ground level targets are also an issue, as the presence of the ground, together with the distance from the ears and the relatively poor angle of perspective all tend to blur ground level targets, unless they present a large surface area,

or are otherwise quite distinct. A 4 inch high curb may be detectable from 9 or 10 feet, but a park bench might only register at 5 or 6, and a coffee table near a couch might not register at all. Incidentally, this is where children have a huge advantage. Their reduced height has the effect of literally making the whole world larger from the auditory perspective. They can detect shorter objects much more easily than adults whose heads are above these same objects.

B. Environmental Variables: Basically these include factors that increase or decrease target distinction. Noise in the environment will make echo signals harder to hear, so targets generally need to be bigger or more solid to register. Reverberation can have a similar effect. Clutter or congestion can obscure a target by causing it to blur with other targets that are too close. When teaching students, I try to choose quiet, open spaces, or focus on highly distinct targets until students become more advanced.

C. Perceptual Variables: Here, we're talking about things like attention, visual functioning, auditory functioning, familiarity with the environment, and self confidence. We should remember that the two primary determinants for success are motivation, and frequent, regular, self directed practice and application under challenging circumstances. Echoes are subtle and require one to be able to attend or at least be motivated to hear them. They are easily masked by noise. Visual functioning generally interferes with echo information as it tends to dominate the attention (for better or worse). Familiarity usually increases registration. It is always easier to find a target when you know what you're looking for.

IV. SONAR SIGNALING

A. Introduction: There are two kinds of sonar - passive and active. Passive sonar is reliant on incidental sounds in the environment which elicit incidental reflections. One can use this to gain information about large features or general layout, but one is reliant on incidental noises that will not be ideal for detection of small features or fine discrimination. Active sonar allows the observer to direct actively a self generated signal into the environment. It is like the difference between taking a picture in strictly ambient lighting, or controlling the lighting by use of a "flash" or strategically placed lighting. While esthetic appreciation may favor the natural look at times, no one can argue that photos and video are always clearer and crisper with sharper detail when the scene is brought under the control of the photographer.

B. In order for sonar to be optimized, three signal characteristics must be present - user control over signal type and directionality, good alignment between signal and ears, minimal masking of the echo, and that the signal is well familiar to the observer. Only an active, self generated or specially designed signal can ensure that these three criteria are met. Cane types and footsteps may give some information, but they are reliant on the travel surface, their directionality cannot be controlled, and such signals may not possess ideal spectral characteristics. An active signal that is self generated and whose characteristics are strictly under user control present the additional advantage of allowing the brain to develop a familiarity with the signal. It is always easier to register something that we recognize. With familiarity, the brain can tune to the signal, and can therefore

register it with less effort under broader conditions. It locks in most readily on signals that it recognizes. Also, since the signal is under the strict control of the user, the brain is

always most sensitive to its effect. Bats always use active signals, and submarine technicians greatly prefer them.
C. The active sonar signal is the basis for the flash sonar approach. Until technology allows us to produce a more ideal signal, we recommend certain types of tongue clicks as the ideal signal. These are intended to be unabtrusive, hands free, and completely under user control without need for reliance on external elements or circumstances.
D. Tongue Clicks: Phoneticists have classified and analyzed five distinct types of tongue clicks. Their names are not important. What we want is a sharp, solid snap, click, or popping sound that the user can control to soft or loud volumes. This is usually produced by pressing the blade of the tongue (flat, middle part) firmly against the roof of the mouth, then pulling downward to break the vacuum. The tip of the tongue should stay more or less stationary and NOT flop down to the bottom of the mouth to form a second "pop." I call this the "cluck click." A tongue click should produce a single, sharp signal, not a double click or clucking sound. Failing the sound being produced by the blade, a respectable sound may be produced by the sides of the tongue against the mollers. This produces the "getiup" click. Another click suitable for temporary purposes is the "tsk tsk" click, the kind we often make to express disapproval. This is produced by the tip of the tongue against the top teeth. Whatever the click, it should ideally not cause odd facial expression, or be used too often or too loudly without cause. Soft clicks should generally be used to detect targets that are close or in quiet environments.
E. Tips for Teaching: Most students can form a suitable click without much training. Students can often learn by modeling. If students hear it enough, they eventually come to do it. For kids, I teach parents and siblings how to make the sound if the student can't do it. Usually, they can. It may help to use a popcycle stick or tongue depressor to show the student where to place their tongue. (I favor tongue depressors, because they are sure to come sterilized and individually wrapped. They are available from any medical supply store; some pharmacists may carry them. A doctor's office may provide a few as samples.) With kids, dipping the tongue depressor in a little honey or jelly helps. Enlisting the help of a speech and language therapist may also help, if one is available. (They may also have some tongue depressors.) If the student just can't make any suitable click at first, have them produce a "ch ch" sound. F. Handheld clickers may be used for certain circumstances. The clicker should be cupped in the hand, button facing outward, and activated by the thumb. Clickers should be sounded either at waste level or above the head, never near the ears. At least 1 second should span between press and release of the button. Clickers should never be activated rapidly, and should only be used out of doors or in open environments.
G. For very beginning stimulus sensitization, some students may benefit from making a continuous "shshshsh" sound, or just using their voice. Some students may not have the breath support to make long "shshshsh" sounds. If this is the case, the transistor radio tuned off station (see II-J) may help. This should be held just below the student's chin.

9.10.3.1 Reserved

9.10.3.2 Sound sources in human echolocation

Kish has described several forms of clicks for the benefit of English speakers and use in his teaching program.

> "**Tongue Clicks**: Phoneticists have classified and analyzed five distinct types of tongue clicks. Their names are not important. What we want is a sharp, solid snap, click, or popping sound that the user can control to soft or loud volumes. This is usually produced by pressing the blade of the tongue (flat, middle part) firmly against the roof of the mouth, then pulling downward to break the vacuum. The tip of the tongue should stay more or less stationary and NOT flop down to the bottom of the mouth to form a second "pop." I call this the "cluck click." A tongue click should produce a single, sharp signal, not a double click or clucking sound. Failing the sound being produced by the blade, a respectable sound may be produced by the sides of the tongue against the mollers. This produces the "getiup" click. Another click suitable for temporary purposes is the "tsk tsk" click, the kind we often make to express disapproval. This is produced by the tip of the tongue against the top teeth. Whatever the click, it should ideally not cause odd facial expression, or be used too often or too loudly without cause. Soft clicks should generally be used to detect targets that are close or in quiet environments."

The above sounds are all unvoiced clicks. Other voiced variants may also be useful.

9.10.3.3 Reserved

9.10.3.4 Achievable echolocation performance by humans EMPTY

9.11 Performance abnormalities in hearing EMPTY

Hearing is a key sensory mode in mammalian physiology. It contributes significantly to awareness of the environment, communications, and enjoyment (at least in humans). The hearing modality can fail or incur significant degradation in a variety of ways.

The hearing modality is uniquely configured to avoid significant erroneous perceptions by employing a variety of mechanism. Some of these mechanisms have only recently enterred the repertoire of modern science. The Marcatili Effect arising in connection with the spiral character of the Organ of Corti is one of these mechanisms. This effect assures that many sources of erroneous acoustic information can not be propagated to the wrong IHC, or more importantly, the wrong OHC.

9.11.1 Gross losses in auditory performance

Clinically observed losses in auditory performance are usually divided into one of three categories;

- Conductive deafness, an impairment of the mechanical system for the transmission of vibrations to the inner ear,
- Sensorineural hearing loss,
- Nerve deafness, deafness due to neural circuit weakness or failure subsequent to the sensory neurons.

The gross degree of deafness is usually described as mild or profound.

9.11.2 Spontaneous and distortion product oto-acoustic emissions

This subject can be addressed from both a normal and a diseased operating condition. There is a great deal of data available related to stimulus frequency oto-acoustic emissions (SOAE's, SFOAE or StOAE's), frequently involving the introduction of an artificial impediment to acoustic signal propagation along the Organ of Corti (OC) . Data related to the spontaneous generation of oto-acoustic emission (SpOAE's), such as that observed in the ear canal of some individuals is less well documented or understood.

Another category of oto-acoustic emission relates to the generation of distortion products (DP) after stimulation by two signals. There are papers reporting the generation of DP's at excessively high intensities as well as at nominal levels. These distortion product oto-acoustic emissions (DPOAE) have been the subject of intense laboratory investigation for several decades.

Both the SOAE and DPOAE areas have suffered from a lack of complete understanding of how the OC operates. Many of the major papers on both SOAE and DPOAE involve mathematical models that have little, and in many cases no, relation to a physical model. They are most often major thought experiments usually employing an isolated, or floating, conceptual model. The scope and variety of the data generated also suggests the appropriate physiological model has not been discovered to date.

9.11.2.1 The physiology of the OC related to OAE

Laboratory investigators have frequently adopted a discretized, or sectional, model of an electrical filter as an analog of the Organ of Corti. Often the model has been treated as a sectionalized model of a coaxial cable. These models have been adopted without detailing the properties of these circuit configurations.

Knight & Kemp have conceptualized the Organ of Corti as a "transmission line" without defining such a device, and then drifted toward a coaxial cable where they also fail to define the characteristics of such a device. Transmission line is an umbrella label in engineering for a broad range of technologies; a coaxial cable is one of the less sophisticated forms. Most of the literature has assumed the OC operates similarly to a coaxial cable. A coaxial cable is typically built of a balanced capacitance and inductance per unit length and negligible series resistance or shunt conductance. However, in the literature, a significant resistance (and often a nonlinear resistance) is generally introduced into the model[231].

The comparison of an OC to a coaxial cable is only possible in a very large scale image. At a finer level of detail, there is little similarity between them. A better analogy is between the OC and strip-line technology. A still better analogy is between the OC and a topographic waveguide.

There is a major difference between a generalized electrical filter section and a coaxial cable. A coaxial cable is designed to have a "real" or resistive, input impedance. Filter sections seldom have a resistive input impedance, unless they are specially designed "end sections." Patuzzi has noted with some stated amazement (page 244) that the input impedance of the Cochlea at the oval window is resistive. This situation may be due to the coaxial cable like properties of the Organ of Corti (OC). However, it is more likely due to the slow wave structure of the OC developed below that prevents any energy from being reflected back to the oval window within the time frame of the

---

[231]Zhang, X. & Mountain, D. (2009) Distortion product emissions: where do they come from? *In* Cooper, N. & Kemp, D. *eds*. Concepts and Challenges in the Biophysics of Hearing. NJ: World Scientific pp 48–54

impedance measurement.

To achieve a resistive input impedance, a coaxial cable is carefully designed to eliminate the roles of its internal series and shunt resistances. Following that optimization, the inductance and capacitance per unit length are matched to give the desired resistive impedance. As a result, the parameters of a coaxial cable are not easily varied along the length of a coaxial cable. On the other hand, the structure of the typical axon can be described as a coaxial cable with special "end sections" designed to insure an appropriate impedance match to the associated Node of Ranvier (Chapter 7 of "Hearing: A 21st Century Paradigm," available at http://neuronresearch.net/hearing/pdf/7Projection.pdf)

Figure 9.11.2-1 shows three potential analogs of the Organ of Corti. Frame A shows a common circuit configuration associated with a section of a generalized electrical filter (but not generally a coaxial cable).

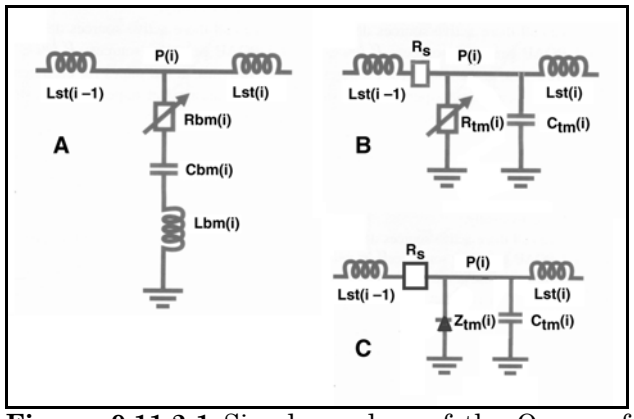


Figure 9.11.2-1 Simple analogs of the Organ of Corti. A; frequent analog of hearing literature. B; analog more appropriate to a coaxial cable. C; analog assigning the dielectric impedance to a diode of limited breakdown potential.

Frame B shows an analog more appropriate to the description of a coaxial cable where the only inductance is the series elements associated with the cylindrical character of the outer cylinder. A series resistance is also shown, although it is typically negligible. The shunt characteristics are those of the shunt capacitance and the shunt resistance (frequently labeled a conductance in electrical analogs. In some discussion in the hearing literature, the two variable resistors are actually nonlinear rather than adjustable.

Frame C shows a configuration more appropriate to the Organ of Corti. The series impedance remains negligible. The shunt conductance is shown as a perfect diode representing the insulation of the shunt capacitance, but exhibiting a finite reverse breakdown potential.

Frame A is labeled to represent the conventional wisdom of the basilar membrane as the principle circuit element. Frames B & C are labeled in conjunction with this work where the surface of the tectorial membrane is the principle circuit element.

As described in **Sections 4.3** & **4.4** of this work, a different technology is more appropriate than the coaxial cable frequently used to describe the OC. Strip-line technology, a specific variant of topographic waveguide technology, is the fundamental substrate used in the OC. It allows the lossless transmission of acoustic energy along a curved path (coiled OC) without involving any reflections at boundaries. In that sense, it is similar to the Topomac of nuclear engineering, a device that attempts to confine a nuclear plasma at very high temperature, without employing any material walls. A topographic waveguide can be modeled using frame C. Properly configured, such a waveguide supports surface acoustic waves and exhibits negligible attenuation. The application of this technology to OAE in the OC will be addressed below.

Virtually all of the data in the literature has been collected under the assumption of basilar

membrane vibration as the mechanism of acoustic energy transport. A totally different picture appears if the energy transport is considered a mechanism involving Hensen's stripe. The tri-annual volumes issued by the Workshop on the Mechanics of Hearing are a wealth of information related to both measured data and a wide variety of models attempting to explain the data. The challenge is to digest all of the data in the context of a more precise model of the OC within the cochlea.

The title of the Zhang & Mountain paper is indicative of the situation after several decades of investigation (the comments at the end of that paper include the suggestive term *paradox*). They give a table of parameter values for their equivalent cable, but only relative to an unspecified parameter $A_{BM}(i)$.

Frosch has provided a similar paper where the parameters of a nominal coaxial cable analog were not specified but the BM stiffness, S, was assumed to vary as a simple exponential over the length of the OC[232]. He then optimizes the product of the stiffness (S) divided by the mass (M) of the BM so that the square root of this ratio equals the frequency of resonance. This optimization incorporates a very difficult computation that has never been found to be addressed within a large ensemble of reviewed papers.. For a acoustic frequency range of 20 to 20,000 cycles, the square root of the stiffness/mass must vary by 1000. The ratio of the stiffness to the mass must vary over a range of one million. For a mass that remains near that of soft body tissue, the stiffness must vary from that of gelatin (think Jello) to that of diamond. This condition cannot be met within the OC. This impossibility is suffered by all of the models incorporating the concept of resonance at a point (in fact at multiple closely spaced but uncoupled resonance points) along the OC. *Any model incorporating the square root of the ratio of S to M to determine a resonant frequency suffers this problem*. The dispersion of acoustic energy by the Marcatili Effect avoids this problem completely.

Many of the models developed to explain DPOAE rely upon the one-dimensional water wave under air (gas) model. This model is discussed in some detail in Appendix W (http://neuronresearch.net/hearing/pdf/AppendW_1D_waves.pdf). The model does not apply to a fluid/fluid or fluid/liquid crystalline interface as found in the OC.

The diversity of opinion within the community concerning a conventional model is addressed in the paper and following discussion from Ren & He[233].

Ren's paradox predates this paper. It probably appears in a 2008 paper by He et al. from the same group[234].

9.11.2.2 Overview of OAE laboratory results

Figure 9.11.2-2 from Siegel, Cerka et al. is particularly useful in understanding how the raw data

---

[232]Frosch, R. (2009) DP phases in mammalian cochleae, predicted from liquid-surface-wave formulas *In* Cooper, N. & Kemp, D. *eds*. Concepts and Challenges in the Biophysics of Hearing. NJ: World Scientific pp 41-47

[233]Ren, T. & He, W. (2009) Differential measurement of basilar membrane vibration in sensitive gerbil cochleae *In* Cooper, N. & Kemp, D. *eds*. Concepts and Challenges in the Biophysics of Hearing. NJ: World Scientific pp 113-121

[234]He, W. Fridberger, A. Porsov, E. Grosh, K. & Ren, T. (2008) Reverse wave propagation in the cochlea *PNAS* vol 105(7), pp 2729-2733

looks when discussing SOAE and DPOAE analyses.

The nearly chaotic nature of the individual records in (a) suggest that employing only one specimen in any laboratory investigation of these phenomenon is of limited value. On the other hand, averaging a set of data may obscure significant features. The signal level versus the noise level shows that averaging multiple sequential records over long intervals is not necessary for individual waveforms. The individual waveforms show "prominent notches and other stable features of individual records appeared at different frequencies in different subjects, causing substantial variability of SFOAE amplitudes at any given frequency across the subject population." The notches are relatively narrow and are less prominent in the phase responses, (b). The group delays shown in (c) have a high degree of scatter but converge around an average of 1 ms at all frequencies above 1.5 kHz in chinchilla. Figure 3 of the Seigel paper provides a similar paper isolating the data from just one subject.

9.11.2.3 Previous SOAE research results

Siegel, Cerka et al. have provided a major study of the SOAE/SFOAE phenomenon without any graphical description of their

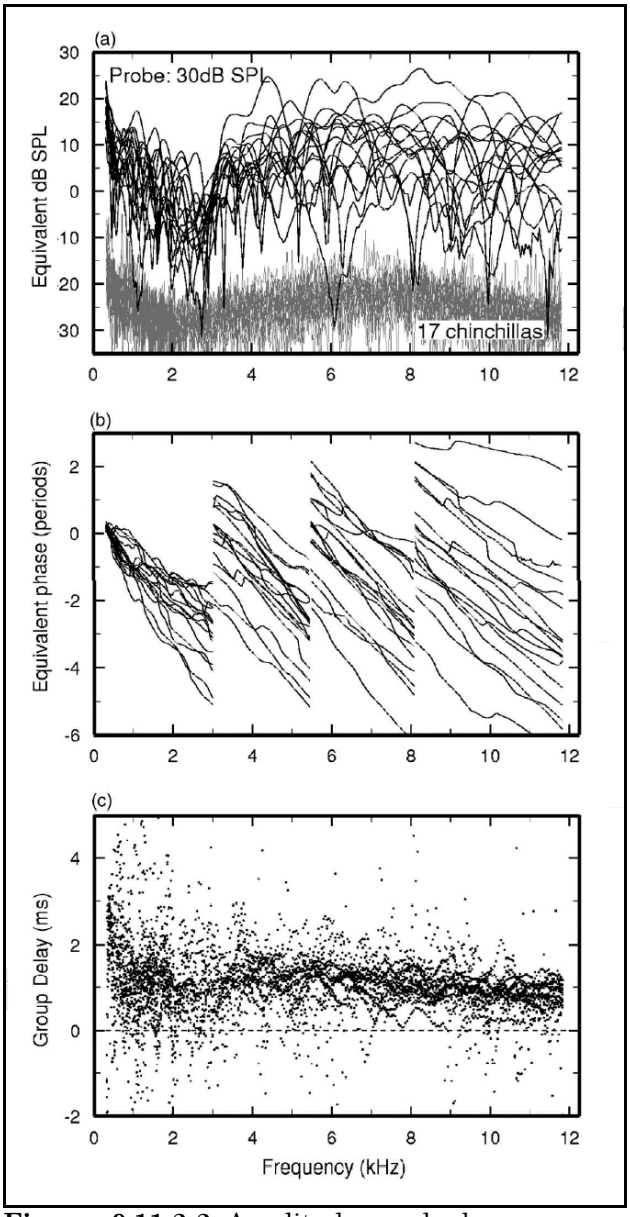


Figure 9.11.2-2 Amplitudes and phases versus frequencies of SFOAE's in 17 chinchillas. (a) Amplitudes of SFOAE's evoked by tones presented at 30 dB SPL (thick traces). Magnitude of baseline noise (thin traces). (b) Phase versus frequency functions for the same measurements. (c) Group delays for the same measurements. Some data points with extreme group-delay values are not shown. From Siegel, Cerka et al., 2005.

physiological model[235]. They do provide a broad bibliography, and indicate they are depending on the basilar membrane as a principle element in the acoustic transduction process, as opposed to being a seismic block supporting the sensing of the tectorial membrane by the hair cells. While they provide data in conflict with the earlier "theory of coherent reflection filtering" (CRF) of Zweig & Shera[236], they do not provide a replacement physical model (section IV-E). They do include a waveform from an auditory-nerve fiber (ANF) in analog form from a very comprehensive paper by Recio-Spinosa et al[237]. but they do not indicate whether it is actually from a stage 1 sensory neuron, or the recovered waveform after integrating an action potential pulse stream. The Recio-Spinosa paper makes it clearer that the waveforms were recovered from the integration of stage 3 action potentials recorded from the auditory nerve.

The 1995 Zweig & Shera paper also surfaces a periodicity with respect to location along the Organ of Corti and not with regard to frequency (note the logarithmic frequency scale in the figure) when observed with a fixed location microphone within the ear canal. This periodicity is frequently found in the data reduced from these experiments. They speculate on potential sources of this periodicity. Shera readdressed this subject in considerable detail in 1999, using data from the 1995 paper but with a figure redrawn to fit a linear frequency scale over a limited range. **Figure 9.11.2-3** illustrates this new figure on how this periodicity appears in the ear canal at very low stimulus levels. The approximate sensation level of the stimulus tone at 1300 Hz is indicated on the right. At the highest level the pressure amplitude varies relatively smoothly with frequency. As the stimulus level is lowered, sound generated within the cochlea combines with the stimulus tone to create an oscillatory acoustic interference pattern that appears superposed on the smoothly varying background seen at high levels. Near 1500 Hz, the frequency spacing Δf_{OAE} between adjacent spectral maxima is approximately 100 Hz. The appearance of this periodicity at low intensity levels is most unexpected. As Shera noted, it suggests its source is not a nonlinearity.

The original figure, but not the recent one, suggests some type of phase inversion between the signal at frequencies below 2000 Hz and those above 2000 Hz. The 1995 paper was a mathematical model with no physical model offered. After considering some sort of corrugation of the basilar membrane, their null hypothesis was, "Although measurements of primate cochlear anatomy find no such corrugation, they do indicate a considerable irregularity in the arrangement of outer hair cells. It is suggested that evoked emissions originate through a novel reflection mechanism, representing an analogue of Bragg scattering in nonuniform, disordered media.". The assumption was the Organ of Corti was operating as a one-dimensional fluid/gas interface device, without regard to whether it was coiled or uncoiled. With a repetition interval on the order of 100 Hz over the low frequency region, it was assumed the signal was due to the reflection of a slow moving traveling wave that existed within the OC and interfered with the similar slow forward traveling wave. Why this interference should be observed in the ear canal was not explained. The comparisons in Table 1 of Zweig & Shera show significant differences between the normal Bragg mechanism and that proposed.

The 1999 paper provides an alternate explanation for the observed phenomenon. It is discussed in a separate section below (**Section 9.11.2.4.2**)

---

[235]Siegel, J. Cerka, A. Recio-Spinosa, A. et al. (2005) Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering *J Acoust Soc Am* vol 118(4), pp 2434-2443

[236]Zweig, G. Shera, C. (1995) The origin of periodicity in the spectrum of evoked otoacoustic emissions *J Acoust Soc Am* vol 98, pp 2018+

[237]Recio-Spinosa, A. Temchin, A. van Dijk, P. et al. (2005) Weiner-Kernal analysis of responses to noise of chinchilla auditory-nerve fibers. *J Neurophysiol* vol 93, pp 3615-3634

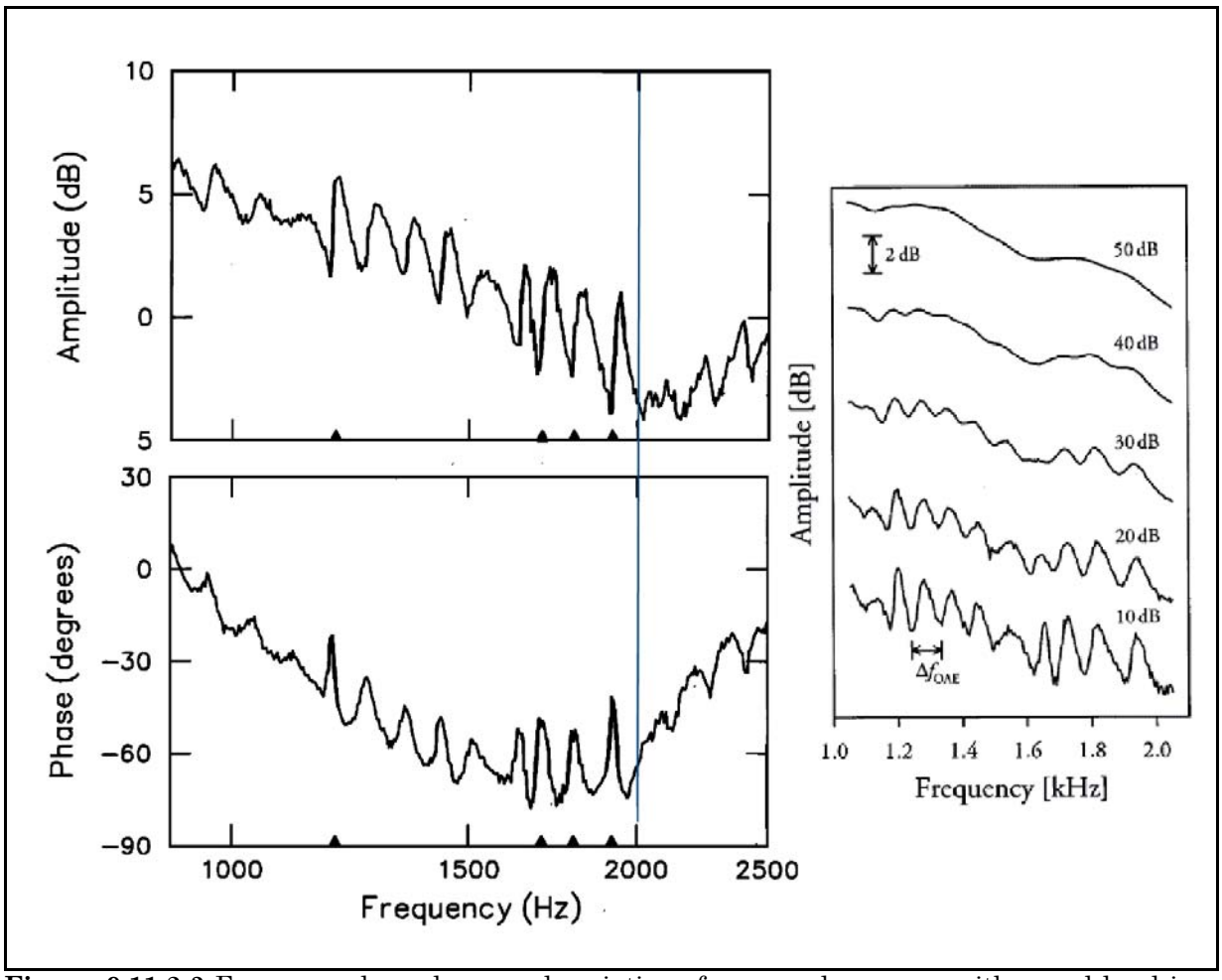


Figure 9.11.2-3 Frequency dependence and variation of ear-canal pressure with sound level in a normal human ear. Left, amplitude and phase information at a fixed (10 dB) stimulus level. Amplitude curves truncated at 1000 Hz and 2000 Hz, and replotted on a linear abscissa on right. From Zweig & Shera, 1995. Right, the curves, offset vertically from one another for clarity, represent the normalized amplitude of the ear-canal pressure, Pec, measured with a constant voltage applied to the earphone driver From Shera & Quinan, 1999.

9.11.2.4 Previous DPOAE research results

Several groups have been involved in DPOAE research for long periods with limited results. Their models are typically floating models based on a longitudinal wave propagating along the axis of an uncoiled basilar membrane acting as the principle structural element. The mechanism of BM excitation at the basal end remains undefined in these models, but is generally assumed to involve some form of compression wave in the scala vestibuli, possibly with a pressure wave in the scala tympani as an antagonist.

9.11.2.4.1 The thought experiment of Knight & Kemp (2001)

Knight & Kemp have represented one of those groups. Their 1995 paper presents data based on an

undefined model, that presumably can be assembled from previous papers of the Kemp group. That paper only addressed subsets of the total distortion product space.

In 2001, they present a paper based on only two subjects. They state categorically, "The simple model used here is not a physical model" (page 1522, summary) and re-emphasize that message in their discussion (page 1521). They then elaborate a very large scale "thought experiment" and site a significant amount of data from, and assertions by, a range of investigators. They attempt to correlate the terms defined by multiple investigators, some of which appear to lack semantic appropriateness, e.g., latency being associated directly with, rather than a result of, delay. Their discussion is based on the basilar membrane as the principle functional element as opposed to the tectorial membrane. The paper contains the normal or larger number of "it must" assertions. At one point, they assert "A simple transmission line model is presented to illustrate how the observed DPOAE maps can arise on the basis of this hypothesis." However, they do not present such a transmission line graphically or describe its technical characteristics, especially as opposed to a simple coaxial cable model. They introduce a gain mechanism that is variable with location and sensitive to the direction of the acoustic waves passing that location. Such performance is suggestive of a much more complex transversal filter rather than a simple transmission line. No discussion of how such a mechanism could be realized in physiology was offered.

They define a,

PATH 1 as involving a DP traveling directly to the base from the point of generation.

- •PATH 1 is associated with a short latency between the stimulus and the DP as measured in the ear canal.
- •They describe the PATH 1 DP generator location as "wave fixed" to indicate their model required it to change as a direct function of the frequency, f_2 ,

PATH 2 as involving a DP traveling initially toward the apex (a forward wave) until it is reflected back toward the base (as a backward wave).

- •PATH 2 is associated with a long latency because of the longer travel time in its round trip path before arriving back at the ear canal.
- •They do not define the PATH 2 DP generator location but define the point of DP reflection as as "place fixed" to indicate their model called for the reflection to be caused by reflectors at fixed locations along the BM, irrespective of f_2 or f_1 ,

They note their analysis does not support the concept of two DP generator locations for the same set of primary frequencies. In section IV(C), they do note the possibility of a combined "wave fixed" emission regime because of anomalies between their two subjects. They occasionally introduce "self cancellation" rather than attenuation. to account for the absence of a particular backward traveling frequency component.

Knight & Kemp do discuss in detail the variation in the performance of DP's based on the f_2/f_1 ratio but do not provide an explanation for this phenomenon or define the transition conditions for a specific DP. Their discussion section is broken into a series of "Possible Explanations."

They only note indirectly that making measurements at $f_2/f_1 < 1.05$ is extremely difficult.

Their musings can be further supported and elaborated based on the physical model of this work and the presentation in **Section 9.11.2.3** below. The figures in that section can be used to explain the discussions of Knight & Kemp in more specific terms related to the physical model presented here and provide a nomograph that can be used by any investigator to aid in interpreting their data. In a broader context these nomographs lead to the inevitable recognition of the role of the

tectorial membrane, the importance of the Marcatili Effect and determining the bandwidth characteristics of the acoustic channel associated with Hensen's stripe.

9.11.2.4.2 Data acquired in the ear canal

In their 1999paper[238], Shera & Quinan assemble a common model based on the literature to that date (with multiple citations). They then assert in the Introduction, "we argue that the common view cannot be correct." They summarize the common wisdom as;

"Common view of evoked emissions

• Evoked otoacoustic emissions arise through nonlinear stimulus "re-emission" by cellular "emission generators" (e.g., outer hair cells).

• Stimulus re-emission occurs because electromechanical nonlinearities—principally those near the peak of the traveling-wave envelope—act as "sources" of backward-traveling waves.

• SFOAEs and DPOAEs both arise through nonlinear distortion (e.g., SFOAEs can be thought of as "zeroth-order" DPOAEs arising from distortion at the first harmonic).

(i) Echo emissions comprise stimulus-frequency and transiently evoked emissions (SFOAEs and
TEOAEs), grouped together because they occur at the frequency (or frequencies) of stimulation, whether the stimulus is a pure tone or an acoustic transient.
(ii) Distortion-product emissions (DPOAEs) occur at frequencies not present in the evoking stimulus (e.g., at the frequency $2f_1 - f_2$ in response to stimulus tones at frequencies f_1 and f_2).

Our use of the distortion-product nomenclature is standard, and we introduce the term "echo emissions" simply as a convenient shorthand. In the common view these two classes of evoked emission both arise from nonlinear electromechanical distortion within the cochlea. Preparatory to further analysis, we review basic properties of echo emissions at low sound levels."

They then define two distinct mechanisms for SFOAE and DPOAE generation. They end the introduction with, "We therefore distinguish two classes of emissions—reflection and distortion-source emissions—based on the mechanisms of their generation. Our mechanism-based taxonomy has important implications for the measurement, interpretation, and use of otoacoustic emissions as noninvasive probes of cochlear function.

Their figure 1 indicates they are talking about phenomena observed with a microphone in the ear canal. Their OAE's appear most prominently at low stimulus levels and appear to represent standing waves in the ear canal. They note, "Near 1500 Hz, the frequency spacing Δf_{OAE} between adjacent spectral maxima is approximately 100 Hz."

---

[238]Shera, C. & Quinan, J. (1999) Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs *J Acoust Soc Am.* vol 105 (2), Pt. 1, pp 782-798

The higher the acoustic level in the ear canal, the smoother their recorded waveforms. They go on;

> "Whereas the shape of the background depends on the acoustics of the ear canal and middle ear, the oscillatory component represents an acoustic interference pattern created by the superposition of the stimulus tone and a second tone at the same frequency originating from within the cochlea (Kemp, 1978, 1979a,b)."

> "We interpret evoked emissions as indicating the presence of backward-traveling waves within the cochlea."

After introducing some mathematics; they further deduce;

> "The linear variation of the phase angle with frequency suggests the presence of a delay (in this case, of about 10 ms)."

Their interpretation continues based on a resonance condition associated with the BM. To interpret their null hypothesis, they introduce one or more "demons" ala the tendency of de Boer and others to introduce allegorical characters to rationalize their conceptual models. Later, they introduce data;

> "Figure 9 provides a test of this prediction in the human ear. The figure shows the phase of the cubic distortion product $2f_1 - f_2$ measured at fixed $f_2/f_1 = 1.2$."

The relevance of using $f_2/f_1 = 1.2$ will be developed in **Section 9.11.3**. The experimental data was apparently all acquired by non-invasive means on willing subjects. No further analysis of this work will be presented.

Their figure 10 provides a putative roadmap through this area of auditory research based on two distinct mechanisms for generating OAE.. They then discuss a "mechanism-based taxonomy" without developing a physical model.

> "Our mechanism-based taxonomy divides evoked otoacoustic emissions into two broad classes:
> (i) Reflection-source emissions, in which backward traveling waves arise through the linear ~coherent! reflection of forward-traveling waves by pre-existing perturbations in the mechanics (e.g., echo emissions such as SFOAEs and TEOAEs measured at low sound levels); and
> (ii) Distortion-source emissions, in which backward traveling waves arise through sources induced by nonlinear distortion (e.g., pure DPOAEs).
>
> Whereas distortion-source emissions would not occur in the absence of cochlear nonlinearities, the coherent reflection responsible for reflection-source emissions is a linear process."

The source and mechanism of their periodic structure that appears at low stimulus level in the ear canal of humans and other primates appears unresolved.

> "The relative amplitude of the oscillatory component grows as the stimulus level is reduced, until, at the sound levels near threshold shown in Fig. 2 ($\leq$5 dB SL), the interference pattern becomes approximately independent of sound level, indicating that stimulus and response are linearly related."

It is reasonable to assume the phenomenon is associated with a SAW type traveling wave within the OC. The displayed waveforms may involve a low level nonlinearity that introduces a hysteresis into the system. A potential source would be the individual OHC (or IHC) near the best frequency that are not excited sufficiently to cause the breakage of any bonds in their piezoelectric transducers. As a result, the energy could be reflected back along the SAW signal path and cause a low level standing wave pattern (of nominally constant amplitude) at least within the OC. Figure 2 in the 1999 paper suggests the amplitude of the standing wave pattern, measured within the ear canal, is only 2 dB with a phase shift of about 10 degrees under these conditions. The shape of the

waveforms suggests the system is not linear at low levels. If it were the sum of two sinusoids remains a sinusoid. The shape of the waveforms suggest there may be a third harmonic that is being reflected, with a different phase shift above and below 2.0 kHz.

9.11.2.4.3 Ear canal data acquired by He et al.

The He et al. paper of 2008 contains a considerable amount of experimental data, many assertions where the nouns appear to lack an adequate number of modifying adjectives and extensive computations, many in simple algebraic form.

> The data from this paper will be used to define the new model of the OC applicable to OAE studies in **Section 9.11.3**.

There is no schematic or block diagram defining the null thesis of this paper; figure 5 is a very simplified conceptual sketch of their basic premise/argument. The figure lacks calibrated scales. There is no definition of where the distortion products of a high intensity two tone stimulus appear along the cochlea, and only two potential (and arbitrary) locations are briefly addressed in the text. Their preferred mechanism of DP generation and projection would suggest the acoustic energy within the system would pass their measurement location twice, at different times along a simple time line. However, their data does not reflect such a situation. They have employed a set of stimulus frequencies (9.00 & 9.45 kHz) that would generate a closely grouped set of cubic DP overlaying the frequencies of the original signals. However, they only address the presence of one of the cubic DP products (8.55 kHz) while ignoring the other (9.9 kHz). The stimuli consisted of simultaneous tone bursts of either 23 or 43 ms. No mention was made of the sidebands generated by the duration of these tone bursts. Their two test points, employing beads interrogated by a laser interferometer, gave significantly different signal amplitudes from their positions only defined in terms of best frequency locations estimated from tone maps available for gerbils.. They acknowledged the use of only a single specimen and the averaging of their data across 10 to 40 acquisition cycles. This paper is in many ways typical of the highly laboratory oriented papers found in the reports of the Mechanics of Hearing organization and the field of hearing research in general.

9.11.2.4.4 Data acquired from neural signals

Kim, Molnar & Matthews[239]. provided a well defined model and accompanying block diagram of the system they were studying in 1980. Their extensive experimental work involved cats and chinchillas. They provide an extensive bibliography through 1980, provided a wealth of data, and use the convention $f_2 > f_1$. While interpreting their data based on the basilar membrane as a principle functional component of the Organ of Corti, their data appears quite compatible with the model proposed here and based on the tectorial membrane as the principle functional component.

---

[239]Kim, D. Molnar, C. & Matthews, J. (1980)Cochlear mechanics: Nonlinear behavior in two-tone responses as reflected in Cochlear-nerve-fiber responses and in ear-canal sound pressure *J Acoust Soc Am* vol 67(5), pp 1704–1721

> "physiologically vulnerable acoustic distortion signals ($2f_1 - f_2$) and ($f_2 - f_1$) in the sound pressure near the eardrums of animals were measured in our recent experiments, using the (Hewlett-Packard 302A, a state of the art at the time) wave analyzer. For neural data, as many as 418 cochlear nerve fibers in each animal were isolated sequentially over two to six days."
>
> They did not quantify the purity of their synthesized stimuli but refer to their instrumentation described in Talmade, Tubis et al[240]. The equipment was developed in their laboratory. It consisted of a signal synthesizer based on a novel but ultimately unsuccessful computer platform, the NeXT/64 (a replacement for the XT family of desktop computers). Several asynchronous clocks were used that led to the potential for what they described as bin-hopping in their spectrograms. This type of artifact was developed in their discussion but was not reported in their data.
>
> No effort was made to explicitly prevent second harmonic signal generation at the digital/analog convertor output. Such harmonic signals were observed in their data for units #80, #125 and #127 shown below. Sidebands 60 dB down would have been considered good for this type of equipment at that time. However, in this investigation, the sidebands would be of similar intensity to the DP's being investigated.

Their data included both waveforms and spectra at many different acoustic energy levels. The spectra showed more than just the minimal number of large components. The awkwardness of using individual specimens was noted;

> "Unit #22, with CF = 600 Hz = (f_2-f_i), shows a predominant response at (f_2-f_i) distortion frequency, while unit #86, with CF = 1160 Hz near ($2f_i - f_2$.), shows a predominant response at ($2f_i - f_2$.) distortion frequency. Unit #127, with CF=4000 Hz, shows a waveform grossly dissimilar to that of the stimulus, with a variety of spectral components at primary and distortion frequencies."

As noted in the introduction to this topic, there is little agreement among the various investigators. Kim et al. noted in 1980;

> "These conclusions are in conflict with conclusions drawn from a series of studies on distortion products in cochlear microphonic (CM) responses by Dallos and his coworkers (e.g., Dallos, 1973a and 1973b; Dallos and Cheatham, 1974). However, a recent study in our laboratory by Gibian and Kim (1979) has demonstrated characteristics of distortion products ($f_2 - f_1$) and
> ($2f_1 - f_2$) in CM responses of the chinchilla which are consistent with our conclusions as stated above."

The intrinsic problem is centered on the model, with the BM as the principle player instead of acting as a seismic block associated with the hair cells sensing energy associated with the tectorial membrane. The more general problem is the grossly different experimental protocols, and locations of measurement, frequently employed by different investigators.

---

[240]Talmadge, C. Tubis, A. Long, G. & Piskorski, P. (1998) Modeling otoacoustic emission and hearing threshold fine structures *J Acoust Soc Am* vol 104(3), Pt. 1, pp 1517-1543

Figure 9.11.2-4 shows the complexity of the spectra obtained by Kim et al. along with the corresponding temporal waveforms. With regard to this figure, which includes both temporal waveforms from the Organ of Corti and spike discharge rates from various neurons collected over 40 second periods, they said;

> "We denote the amplitude and phase of each harmonic frequency component of the discrete Fourier transform of *r(k)* by *R(f)* and Θ(*f*), where *f* is the harmonic frequency and *r(k)* is the spike discharge rate. In all cases except for Figs. 1 and 7 in this paper, we use *R(f)*/SR (SR = spontaneous rate of spike discharge) as the measure of response amplitude, since this has the most desirable features among several alternative response measures, as discussed in Kim and Molnar (1979); the data of Fig. 1 showing R(f)/R(O) and the data of Fig. 7 showing R(f) were processed before we found that R(f)/SR is more useful, but the conclusions based on these figures are not affected by this difference."

The figure shows a few frequency components above the stimuli. These are primarily second harmonics of either of the stimuli, rather than distortion products. This is most clearly seen for the CF = 2100 data for unit #80. These harmonics may have been generated within the ear canal and/or middle ear mechanisms. Unit #125 shows a mixture of a major harmonic at $18 \bullet f_0$ and a series of distortion products at frequencies below f_2.
Note the stress on periodic stimulation and the use of discrete Fourier transforms in their discussions. The top waveform and spectrum show the nominal theoretical conditions for two signals of equal amplitude applied simultaneously and continuously to the ear canal, in the absence of any nonlinearity or discontinuity.

The row labeled CF (characteristic frequency) shows the spectrum at 600 Hz in the presence of a nonlinearity resulting in the "distortion product" at $f_2 - f_1$, $(2 \bullet f_0)$ along with a harmonic of this signal at $2(f_2 - f_1)$, $(4 \bullet f_0)$. All frequencies shown are relative to the base frequency of 300 Hz. $f_2 / f_1 = 1.285$

The row labeled 1160 shows a spectrum dominated by the $2f_1 - f_2$ component at $5 \bullet f_0$ = 1500 Hz, but escorted by components at $2 \bullet f_0$ = 600 Hz, $7 \bullet f_0$ = 2100 Hz, $10 \bullet f_0$ = 3000 Hz & $12 \bullet f_0$ = 3600 Hz. $f_1 / f_2 = 1.33$. These are all expected distortion products as developed in more detail in **Section 9.11.3.2**. The ratio f_2 / f_1 is inverted in this instance to get a value greater than 1.0.

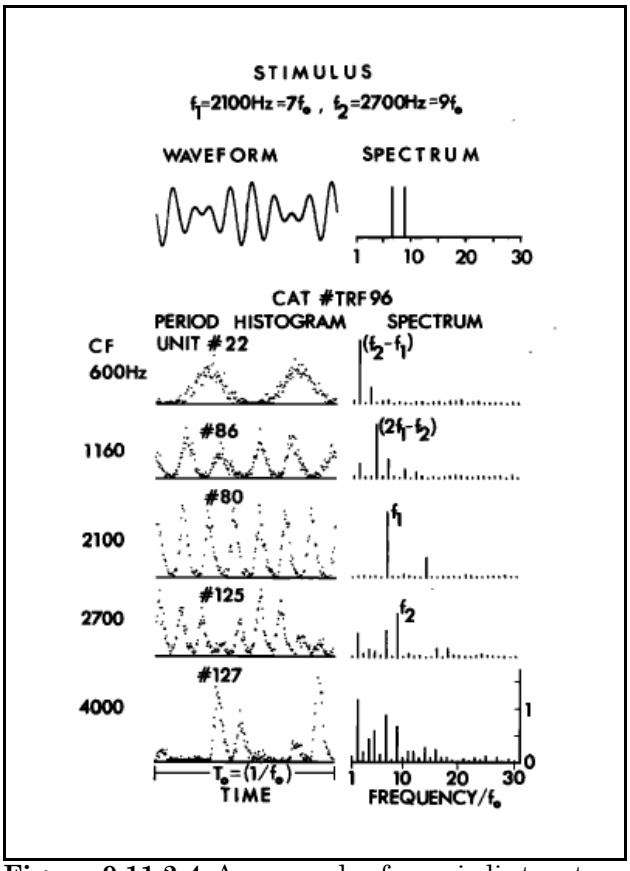


Figure 9.11.2-4 An example of a periodic two-tone stimulus waveform and its spectrum ($f_1 = 7 \bullet f_0$, $f_2 = 9 \bullet f_0$, f_0 = 300 Hz in the top row) and five period histograms and their corresponding spectra for responses to this stimulus of five different cochlear nerve fibers in one cat. The sound pressure levels L_1 and L_2. at f_1 and f_2 frequencies, respectively, were both 64 dB. The spectrum of the period histogram shows the response-amplitude measure *R(f)/R(0)* for $f = f_0$, $2f_0$..., $30f_0$. From Kim, Molnar & Matthews, 1980.

The row labeled 2100 only exhibits spectral components at $f_1 = 7 \bullet f_0 = 2100$ Hz, and the second harmonic of this component at $14 \bullet f_0 = 4200$ Hz.

The row labeled 2700 exhibits primary components at and $f_1 = 7 \bullet f_0 = 2100$ Hz and $f_2 = 9 \bullet f_0 = 2700$ Hz, along with distortion products at $2 \bullet f_0 = 600$ Hz, $4 \bullet f_0 = 1200$ Hz, $5 \bullet f_0 = 1500$ Hz, along with two higher frequency components at $f_1 + f_2 = 16 \bullet f_0 = 4800$ Hz & $2 \bullet f_2 = 18 \bullet f_0 = 5400$ Hz.

Unit #127 with a CF = 4000 Hz (or about $13 \bullet f_0$) also shows a significant second harmonic of f_1, a small second harmonic of f_2 and a third harmonic of f_2, as well as the sum of f_1 & f_2 probably generated prior to the OC. The frequencies at $11 \bullet f_0 = 2 \bullet f_2 - f1$, $12 \bullet f_0 = 3 \bullet f_1 - f_2$, $23 \bullet f_0 = 2 \bullet f_1 + f_2$ and $25 \bullet f_0 = 2 \bullet f_2 + f_1$ appear to be legitimate DP's above f_2 and sensed by neurons within the OC. As developed in **Section 9.11.3**, these data suggests the ability of the Marcatili Effect to suppress DP's above f_2 may be reduced, or a different DP generating mechanism is involved, in this specimen.

These data are all the result of forward SAW waves traveling along the OC. The results are quite compatible with the proposed operation of the OC in **Section 9.11.3**.

> As normally found in the laboratory, no distortion products were measured at sensory neuron fibers with BF's above f_2. An expected selection of lower sideband DP's were reported. In #80, a significant second harmonic of f_1 was measured. #125 showed a significant second harmonic of f_2 and a remnant near f_1 & f_2. It is suggested these two signals are generated either in the synthesizer or within the middle ear or in connection with the movement of the oval window. #127 shows a second harmonic of f_1 and a remnant near f_1 & f_2.

Some of the distortion products may be associated with the stage 3 neural encoding process preceding the recording of pulse signals from the axons of individual stage 3 neurons. This is an aperiodic (not clock driven) encoding process. The process also reflects a stage 1 adaptation mechanism affecting the initial pulse rate in stage 3 neurons.

9.11.2.4.5 The archaic "Classical" mathematical models

Figure 9.11.2-5, from de Boer with a frequency line added, presents his interpretation of DPOAE data based on a quasi-linear method based on a variation of the Neely–Kim model of the same time period, 1996. No data points are included from laboratory experiments. A frequency line has been added showing both the potential real quadratic and cubic DP's. The frequency line shows the location of a 4 kHz DP that is not incorporated in the model even though it is capable of forward wave projection along Hensen's stripe. The frequency line assumes a logarithmic relationship between frequency and position along the OC. The DP at 18 kHz may be influencing the data recorded from the BM and would appear as a backward wave along Hensen's stripe. The DP at 34 kHz would not be expected to propagate as a backward wave due to the curvature of Hensen's stripe near the point of DP generation.

The full caption of the de Boer figure is; "Response of a nonlinear cochlear *model* to a two-component stimulus. A; BM velocity patterns (re 10^{-3} mm/s) produced by presenting two sinusoidal signals at 60 dB SPL. For this figure, f_1 and f_2 are 10 and 14 khz, respectively. *Vertical arrows* indicate response maxima for these components. *Thick lines* show primary components (magnitude only). The dashed line shows the $2f_1–f_2$ DP plotted at a 40-dB higher level. The cross marks the source of this component (see text). B; Phase response of the DP component. At x = 0, the phase is normalized to zero. Model used: see Kanis and de Boer (1993a); a most simplified model of the middle ear is included." ["model" in first sentence emphasized.]

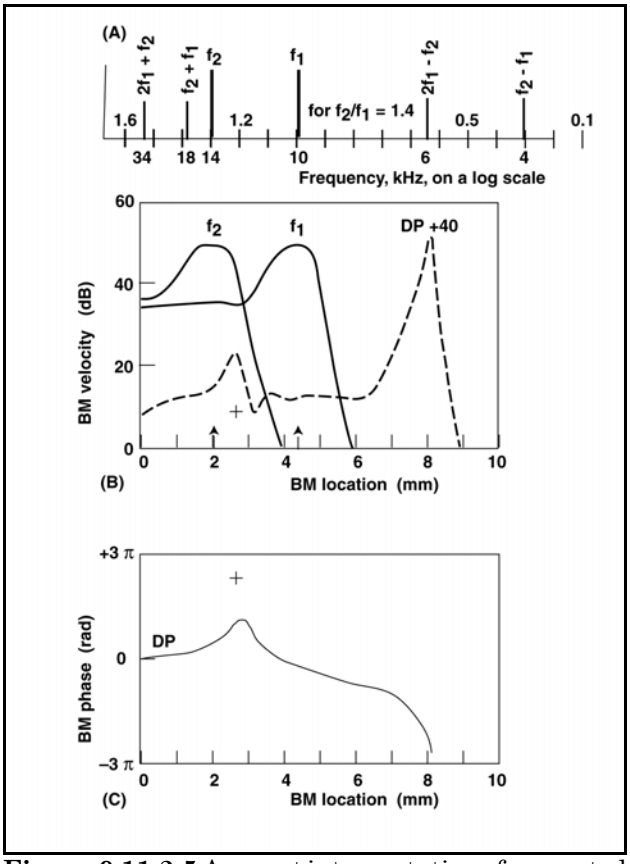


Figure 9.11.2-5 A recent interpretation of computed DPOAE based on the conventional wisdom. The arrows near 2 and 4.3 mm in the middle frame represent stimulation by narrow band sound at 14 and 10 kHz respectively. The peak labeled DP is at 6 kHz. Note the response of the BM is much broader than that associated with that of the TM at the OHC and reflected in the output bandwidth of the respective sensory neurons. The apex is to the right. See text. Modified from de Boer, 1996.

de Boer asserts "The effective source of the DP component (this is the point where the OHC-generated pressure is maximal) is indicated with a *cross*." There is no data supporting the assertion within the parentheses. The next section will present an alternative interpretation.

The dimensions given along the BM are based on the assumption that the energy is only traveling along the longitudinal axis of the BM (an unwound one-dimensional model). If the energy were to be traveling along the TM until their dispersion point, and were then traveling nearly orthogonal to the local BM axis (a two-dimensional model with added features), different values would be expected.

The width of the energy profile captured by the sensory neurons in this configuration would

approximate 0.03% of the nominal position along the BM under the Marcatili Effect based dispersion approach (**Figure 4.5.3-1**). Such narrow profiles, 3 Hz at 10 kHz, would be narrower than the width of the stem of the two arrows shown along the BM axis. These two arrows are identified as the points of peak BM response.

The de Boer model is compatible with energy traveling along the OC at a constant velocity regardless of its location. It does not show any slowing as the energy approaches a BF condition.

- - - - -

Tubis et al[241]. have provided a mathematical tour de force related to DPOAE generation under both steady-state and pulsed conditions. They note;

> "The basic formulas for various types of otoacoustic emissions, and details on the underlying model and approximations used to obtain them, have been extensively discussed in Talmadge et al. (1998), to which the reader is directed for the precise definitions and meanings of the symbols used in this paper. The model used was a one-dimensional macromechanical model with a time-delayed stiffness component in the basilar membrane mechanics of the form suggested by Zweig (1991). The model also includes simplified models of the middle and outer cars."

They do not specifically define their model graphically, but based on their citation to specific pages in Lighthill, it appears they are speaking of water waves at a liquid/gas interface (a condition not found in the Organ of Corti).

The reader should be aware of their statement that seriously undermines the utility of their findings.

> "The time-delayed stiffness terms in [Eq. (13)] greatly complicate the evaluation of the integral. Therefore, for the purpose of obtaining a simple result for the transient build up of the signal, which should at least have qualitative validity, these terms will be neglected."

9.11.2.4.6 A 21st Century model of the OC

Talmadge, Tubis et al. in 1998 (and the other authors of the previous section) were using a "classical" model of the cochlea that is truly irrelevant in the 21st Century[242]. To be ***very, very clear***, their model of **Figure 9.11.2-6(top)** does not represent a realistic model of the mechanisms within the labyrinth. The actual morphological situation is developed in detail in **Section 4.2.2.**

The bottom portion of the figure shows a more accurate representation of the same area. It uses a majority of the terminology found in Talmadge, Tubis et al. The outer ear and middle ear portions remain unchanged. It is important to note the vestibule of the labyrinth is not the scala vestibule. The notation has been changed to highlight the acoustic pressure in the vestibule (P_V) is ***not*** the same as the acoustic pressure in the scala vestibuli (P_{SV}). The acoustic energy in the outer ear

---

[241]Tubis, A. Talmadge, C. & Tong, C. (2000) Modeling the temporal behavior of distortion product otoacoustic emissions *J Acoust Soc Am* vol 107(4), pp 2112–2127

[242]Talmadge, C. Tubis, A. Long, G. & Piskorski, P. (1998) Op. Cit.

travels at a nominal 340 m/s. The energy introduced by the oval window is delivered to the vestibule of the labyrinth as a compression wave traveling at 1500 m/s in the perilymph-like fluid of the vestibule. The energy crosses the vestibule and impinges upon the vestibule caecum (VC) of the tectorial membrane at a very specific angle. The energy passes through the VC to its opposite surface and the angle induces the energy to be transformed into a slowly traveling SAW moving along what becomes Hensen's stripe after some path changing to accommodate packaging. As Hensen's stripe becomes morphologically defined, it enters the cochlear duct, with its active surface facing the tectorial membrane across the scala media (SM) filled with endolymph. As a result, the energy applied to the cochlea is never found in the perilymph of the scala vestibule or scala tympani in its original form. The energy in these media is at very reduced intensity levels and are residuals of the operation of the Organ of Corti.

The new model is compatible with a wide variety of current findings, most of which are

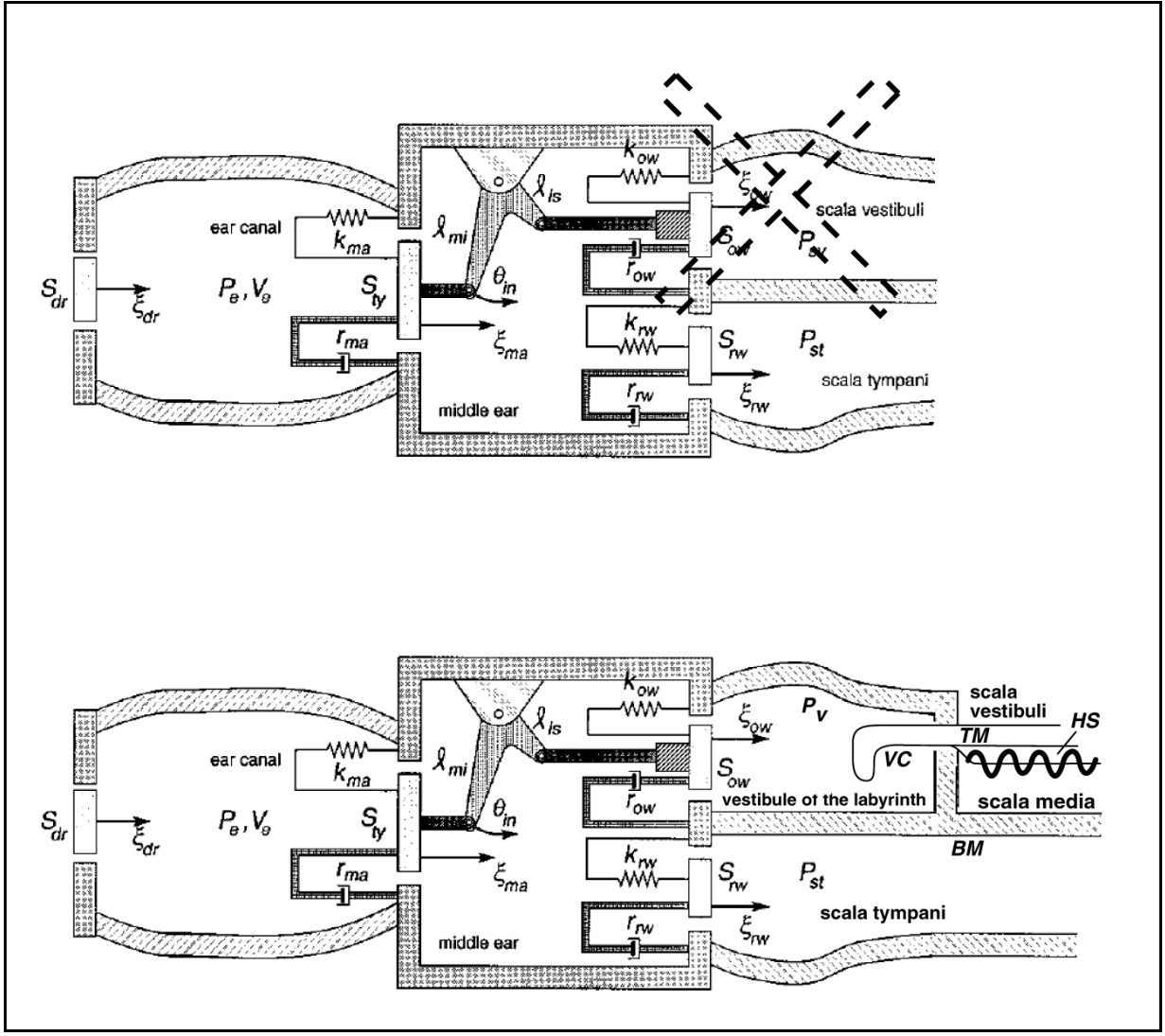


Figure 9.11.2-6 Models of the acoustic elements of the labyrinth. Top; the historical model based on inadequate morphological investigation and inappropriate modeling of the structures. From Talmadge, Tubis et al. 1998. Bottom; a more realistic model of the acoustic portion of the labyrinth based on recent more detailed morphological studies. The acoustic energy crosses the vestibule of the labyrinth and impinges upon the vestibule caecum (VC) of the tectorial membrane (TM) as a compression wave and is converted to a slowly traveling surface acoustic wave (SAW) traveling along Hensen's stripe (HS). See text.

summarized in the recent review by Gavara et al[243]. An exception is their overlooking the acellular Kimura's membrane that incorporates Hensen's stripe and is liquid crystalline in character. The acellular liquid crystalline character of Kimura's membrane is critical to the near lossless propagation of the SAW at velocities on the order of 6 m/s (1/60th the velocity in air and 1/250th the velocity in lymph).

The new model dispenses with the concept of a time-delayed stiffness along the basilar membrane as irrelevant; all the relevant acoustic energy travels along Hensen's stripe and Kimura's membrane. It travels at a nominally constant velocity of 6 m/s, except possibly near the last fractional turn of Hensen's stripe near the helicotrema. (**See Section 4.3.3**).

9.11.3 A paradigm shift in modeling OAE's

[243]Gavara, N. Manoussaki, D. & Chadwick, R. (2011) Auditory mechanics of the tectorial membrane and the cochlear spiral *Curr Opin Otolaryngol Head Neck Surg* vol 19, pp 382–387

[xxx develop the two approaches, simple BM and more complex curved TM to IHC & OHC]

9.11.3.1 The OC based on the tectorial membrane and Marcatili Effect

The Zhang & Mountain calculations and figure can be interpreted using **Figure 9.11.3-1**, modified from **Figure 4.6.1-1**. The acoustic energy is propagating along Hensen's stripe near the inner edge of the tectorial membrane, and generally near, but not congruent with, the inner edge of the BM. The darkest oval represents the boundary of the tectorial membrane most subject to breakdown due to the 40dB SPL energy level in the previous figure. The lighter shaded oval represents the broader area subject to breakdown when energy is applied at the 65 dB SPL level (arbitrarily shown as symmetrical with the first oval). The caricature suggests the energy causes an initial breakdown at a point C and the location of this breakdown at the skirt level may be somewhat different than the frequency associated with that frequency based on BM mapping. The interference to the frequencies remaining in Hensen's stripe can result in DP frequencies of $m\boldsymbol{f}_{residual} \pm n\boldsymbol{f}_1$. Those DP frequencies at lower frequencies than $\boldsymbol{f}_1$ are able to continue to propagate along Hensen's stripe until their appropriate point of dispersion. DP frequency components higher than $\boldsymbol{f}_1$ would be unable to propagate toward the apex due to the curvature of Hensen's stripe and the constraint due to the Marcatili Effect. These higher frequencies could propagate as backward SAW waves if their frequency was sufficiently near the frequency $\boldsymbol{f}_1$ to be within the frequencies that can propagate at the curvature of Hensen's stripe. Higher frequency DP's would be unable to cross the gap between the local curvature and the curvature at which they could propagate along Hensen's stripe efficiently. [xxx should point C, along with A & B be removed from this figure?]

At higher intensity levels, the Zhang & Mountain model suggests the basal edge of the envelope describing breakdown of the tectorial membrane and DP formation moves closer to the stapes at lower ratios of $\boldsymbol{f}_2/\boldsymbol{f}_1$. This appears logical as the total energy density at a given position along the skirt of Hensen's stripe increases in this region under these conditions.

The caricature suggests the breakdown does not occur in the region of the OHC. If the OHC were acting as cochlear amplifiers, as frequently proposed, it would be expected that breakdown would occur in the vicinity of the output of those amplifiers rather than near Hensen's stripe. The energy would be generated as a circular wavefront based on Huygen's principle with only a small amount of the energy following a path back to Hensen's stripe that would allow it to return along Hensen's stripe as a backward traveling SAW. The chance of any energy intercepting Hensen's stripe and becoming a forward traveling SAW appears minimal.

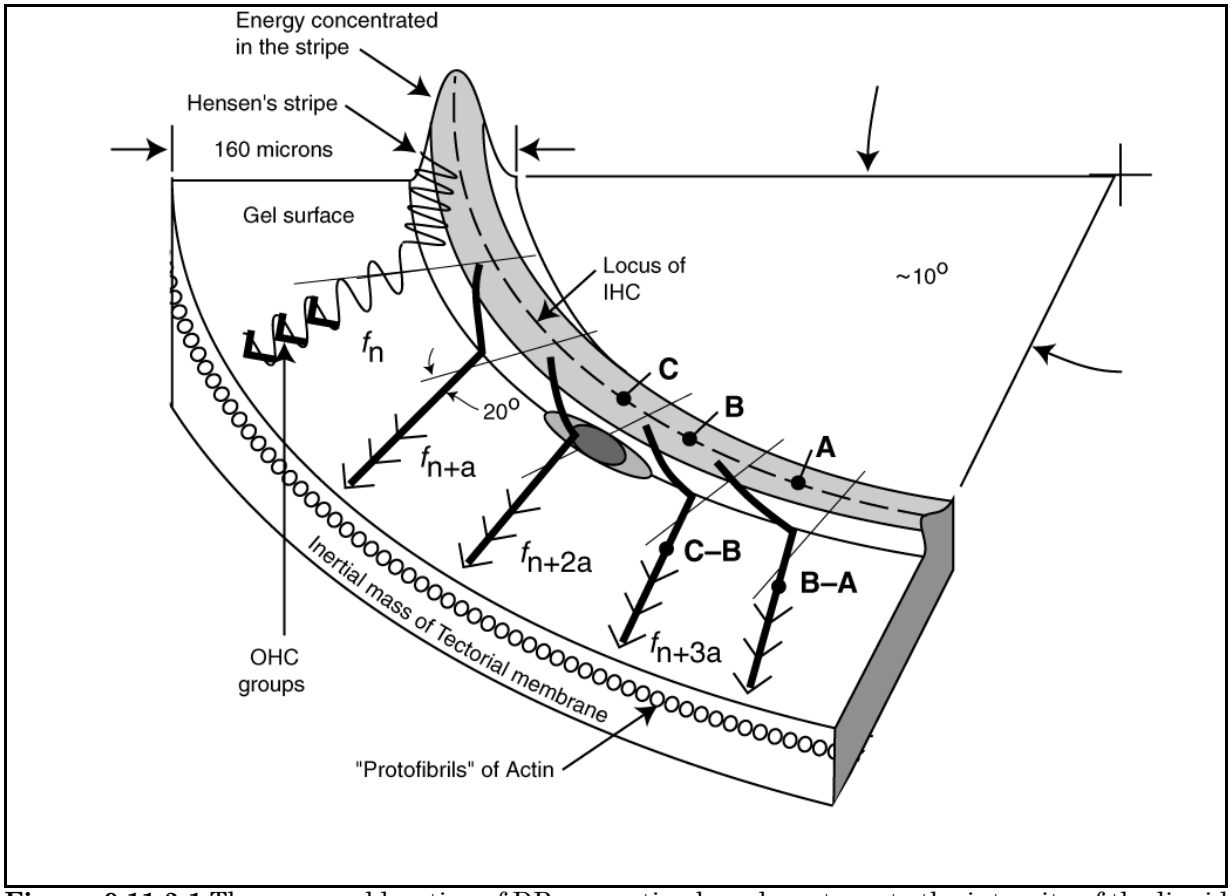


Figure 9.11.3-1 The proposed location of DP generation based on stress to the integrity of the liquid crystalline layer. The darkest oval represents the most likely area at low intensity. The larger oval represents the most likely area as the intensity is raised. See text.

The above interpretation can be readily confirmed by experiment, possibly within the model proposed by Zhang & Mountain. The protocol for such experiments should differentiate between locations along the inner edge of the tectorial membrane near Hensen's stripe and the outer edge near the OHC. They should also accommodate the angle, of about 20 degrees, at which the acoustic energy departs Hensen's stripe on the way to the OHC. In some regions, Hensen's stripe does not overlay the BM at all and measurements of BM motion will not be available for recording data relative to DP formation. The operation proposed here employs a SAW traveling at nominally constant velocity (see **Figure 4.3.3-8** for measured values of this velocity at greater precision), regardless of frequency. This condition is different from the rapidly changing velocity of the conventional interpretation of wave motion near best-frequency points.

9.11.3.2 Existence conditions for reflections and DP's

Figure 9.11.3-2, using the frequencies of He et al, shows the overall potential for the generation of reflections or distortion products within the proposed coiled OC.

First, and briefly, if there is a discontinuity somewhere along Hensen's stripe, any frequency still propagating along the stripe (frequencies below the BP at that location), can be reflected back to

the vestibule, and potentially the stapes. These reflections are defined as stimulus-frequency oto-acoustic emissions (SFOAE) and are observed as a backward propagating SAW.

Second, any frequency propagating along Hensen's stripe at sufficiently high intensity as to introduce a discontinuity due to a liquid crystal substrate breakdown, can cause a reflection of that energy, and potentially any energy still propagating along the stripe (frequencies below the BP at the point of breakdown. These reflections would be by a backward SAW. At the same location, a series of DP's can be expected to appear at the discontinuity due to its "nonlinear" character. Whether these DP's can propagate is the subject of the next paragraphs.

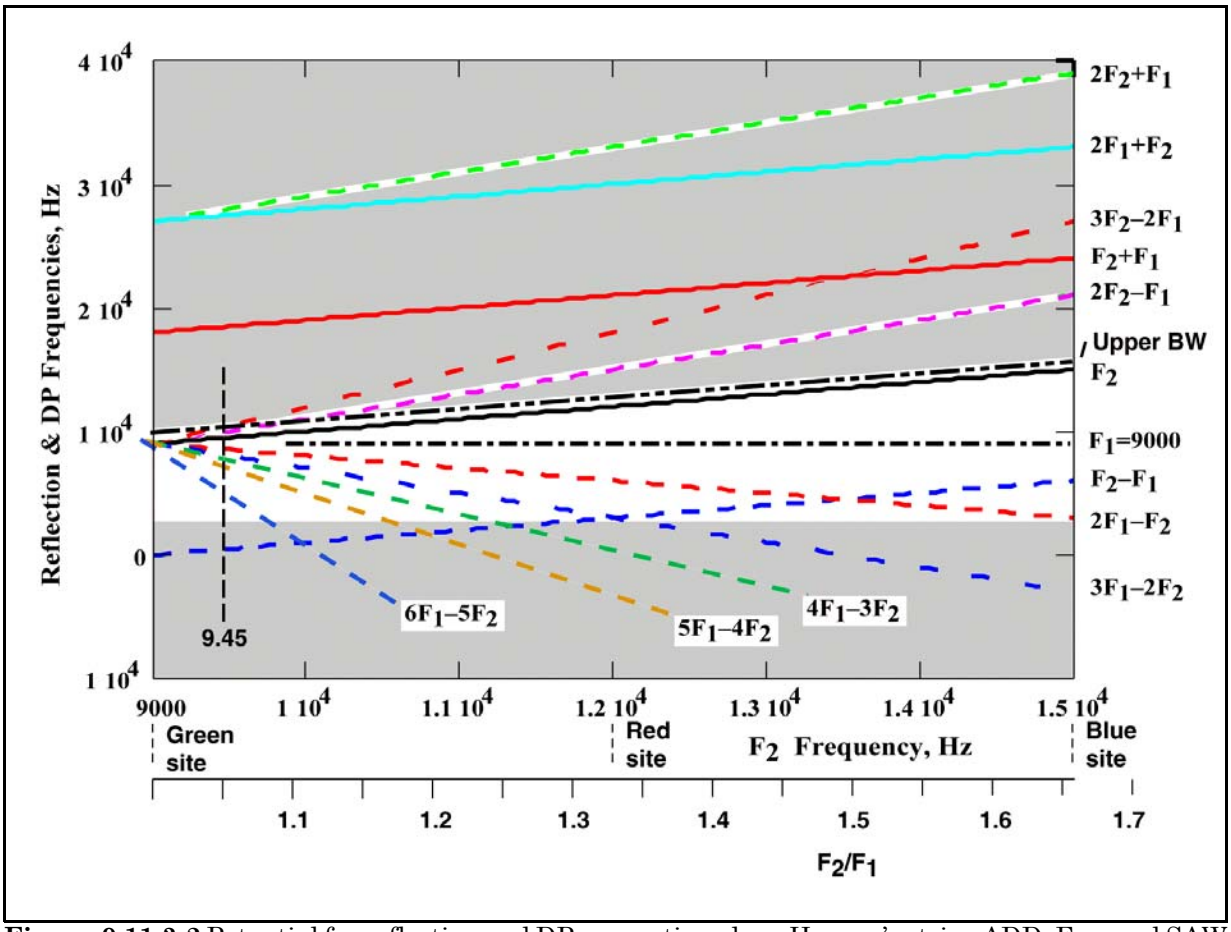


Figure 9.11.3-2 Potential for reflection and DP generation along Hensen's stripe ADD. Forward SAW waves proceed downward in this figure, backward SAW proceed upward. The nominal upper frequency cutoff (Upper BW in figure) of Hensen's stripe was taken as $1.1 \bullet F_2$ in this figure. The blue, red and green sites were used by He et al, 2008. The vertical dashed line at 9.45 kHz indicates the F_2 used by He et al. Their $2F_1–F_2$ DP would have been at 8.55 kHz. See text.

In the general case of a discontinuity, the energy approaching that point as a forward SAW is subject to mixing with other frequencies present and generating distortion products as discussed earlier. At the point of generation, these distortion products will radiate outward with regard to the surface of the liquid crystalline surface in accordance with Huygen's Principle. Some of the energy of those DP's will be able to propagate along Hensen's stripe. The figure illustrates the simple DP's with significant energy due to two-tone mixing. The figure also describes the limits of

this DP propagation under the Electrolytic Theory of the Neuron and this instance of the Marcatili Effect (See the lower frame of **Figure 4.5.1-2**).

Any DP at a higher frequency than F_2 generated within the effective bandwidth of Hensen's stripe at the discontinuity (taken here as $1.1 \bullet F_2$) can be propagated back to the vestibule as a backward wave. As shown, this limits the DP's at frequencies above F_2 to only the DP = $2F_2 - F_1$ where F_2/F_1 is less than 1.1.

Any DP at a frequency below F_2, remains able to propagate in either direction along Hensen's stripe, as either a forward or backward SAW, within the following constraints. For DP's at frequencies below F_2, they are subject to dispersion when they reach the curvature of the stripe corresponding to their BF. If not reflected by that point, they are lost to the oto-acoustic emission process. For frequencies below the effective end point of the dispersion process, they are subject to reflection in the general area of the helicotrema, if not effectively absorbed. If not absorbed, they can reflect and travel as a backward SAW all of the way to the vestibule.

The preference for DP = $2F_1 - F_2$ (8.55 kHz under He et al.) in the laboratory is obvious from this figure. It is likely to be a more intense DP than $3F_1 - 2F_2$ (7.2 kHz under He et al.) and be present, and successfully propagated in either direction over a larger range of F_2. relative to F_1. The DP = $2F_2 – F_1$ (9.90 kHz under He et al) exists in a more limiting situation. **Figure 9.11.3-3** presents an expanded view of the area between 9.0 and 10.0 kHz in the above image. It shows the F_2 test frequency was marginal for generating a $2F_2 – F_1$ DP (9.9 kHz under He et al.) that could propagate as a backward SAW if the limit on backward wave propagation was set by a limit for Hensen's stripe to propagate such a wave of $1.05 \bullet F_2$. The DP at $3F_2 – 2F_1$ (10.35 kHz under He et al.) cannot be propagated as a backward SAW under this condition. If the limit on Hensen's stripe for backward propagation is closer to F_2 than $1.05 \bullet F_2$, (potentially as low as $1.0003 \bullet F_2$) neither of these DP's can propagate as a backward wave.

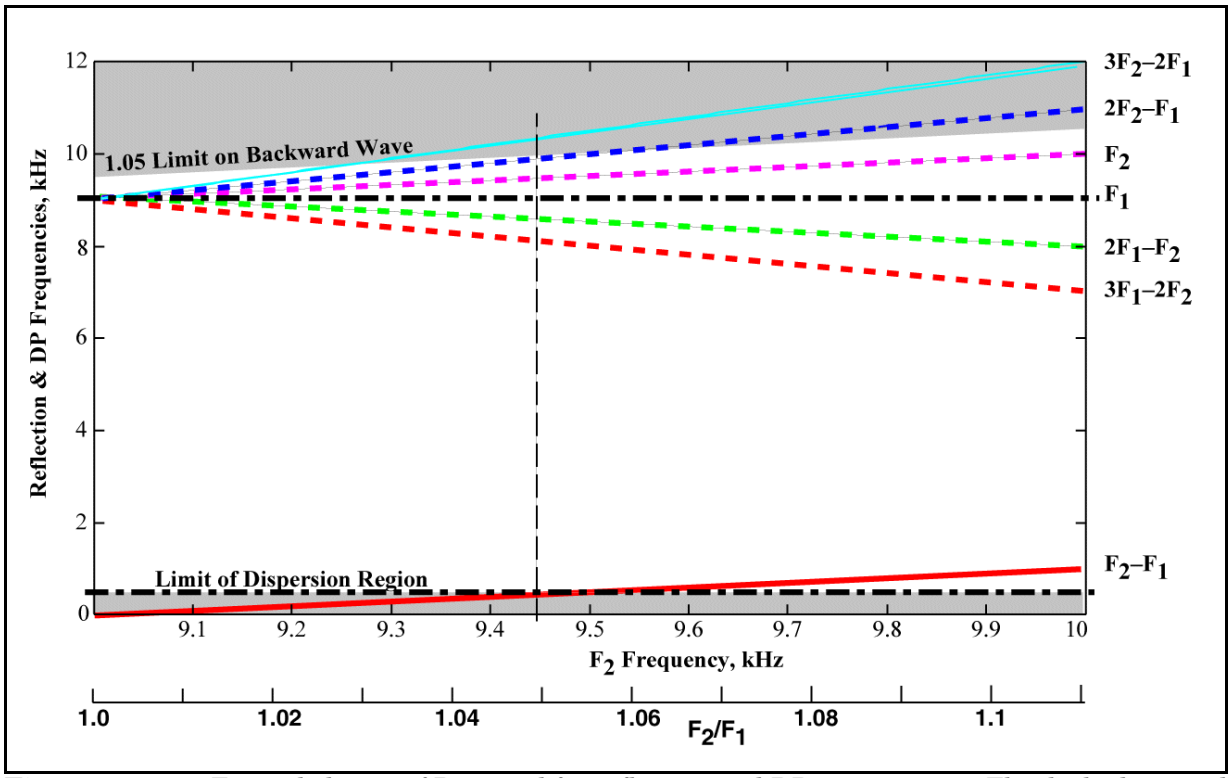


Figure 9.11.3-3 Expanded view of Potential for reflection and DP propagation. The dashed vertical line at 9.45 kHz was the frequency used by He et al. The $2F_2–F_1$ DP is at the margin for backward propagation for a limit on backward propagation of $1.05•F_2$. See text.

For frequencies compatible with forward SAW propagation beyond the discontinuity, these DP's can propagate until they encounter a curvature in Hensen's stripe requiring their dispersion toward the appropriate OHC's. If the DP's frequency is lower than the limit of dispersion associated with that curvature, the energy in the DP is likely to be absorbed in the region of the helicotrema. Potentially, it can be reflected and travel back the length of the Organ of Corti to the vestibule.

This analysis only applies to DP's formed within the Organ of Corti defined in this work and relying upon the Marcatili Effect for its operation. However, it explains the vast majority of the data acquired in the acoustic physiology laboratories over the last few decades, and Ren's paradox in particular.. In this example, DP's traveling as backward SAW waves will not be observed for values of F_2/F_1 exceeding the effective high frequency cutoff of Hensen's stripe, given in this example by $1.1•F_2$.

Ren's group did not observe backward waves for F_2/F_1 of 1.05, indicating the effective bandwidth of Hensen's stripe at that point was below $1.05.•F_2$ (less than the nominal value assumed in the figure). It is quite possible, the effective bandwidth at this point was less than $1.0003•F_2$, in consort with, or approximating, the bandwidth of the channels being dispersed to the OHC. However, it is likely the effective value is between 1.0003 and 1.05. The site of DP generation (green site) in the groups investigations was assumed, not observed or measured. [xxx check this, why no 8.55 kHz backward SAW wave?]

It would be necessary to know more about how they "edited" their Fourier spectrums for the He et

al. paper, and how they performed the inverse transform before making additional comments. The discussion of the Kim, Molnar & Matthew paper in an earlier section (**Section xxx**)suggests the additional details needed for this purpose. The transform procedures found in most mathematical program software packages only take the steady state transform, and deliver only the real part of a complex function. Taking the inverse transform of only the real part of a function can be introduce problems of interpretation. Figure 8 of Kim et al. describe the phase sensitivity of their measurements.

Since He et al. were using pulse stimulation, any reflected principle frequencies should have been accompanied by a pair of sidebands ($F_1 \pm \Delta F_1$ and $F_2 \pm \Delta F_2$). These should have been present along with any DP at 9.9 kHz and its sidebands, at their test locations only a short interval after their stimulus terminated. They asserted they used 23 or 43 ms pulse intervals with 1 ms rise and fall times in their "Experimental Procedures." However, their figure 1*A* shows an interval of 0.8 ms. They also indicated they averaged 10-40 copies of their waveforms to obtain their usable results.

He et al. indicated they collected additional information at F2/F1 ratios of 1.1 and 1.2. Based on this analysis, they should not have expected to see any backward SAW DP's of the $2F_2–F_1$ type.. The potential for observing DP's of the $2F_1–F_2$ as backward SAW's at their observation points, and originating at the discontinuity, remains. Observing the reflections from the helicotrema appears unlikely.

9.11.3.3 A re-interpretation of the early investigators

Smoorenburg performed a major laboratory investigation of perceived OAE during the 1970's[244] ,[245,246,247] following Plomp's earlier work. The last paper includes a discussion of generation of distortion products leading him to define a distinct class of DP's. The distortion products he develops include $f_2 - f_1$ and the distinct class he defines as $f_1–n(f_2–f_1)$ where n is a small integer. These are all frequencies below the frequency f_1 for $f_2>f_1$. They correspond to DP's that continue to forward propagate along Hensen's stripe until they reach the best frequency for each DP where they are dispersed to, and stimulate, the the appropriate OHC. In accordance with the model presented above, the DP's generated at frequencies above f_2 do not backward propagate to the expected best frequency or beyond to middle ear and the ear canal. A paper by Smoorenburg et al. summarizes much of the earlier work[248]. All of this work is consistent with the theoretical model described in the above figures of this section. Smoorenburg notes specifically that only $f_2 - f_1$ increases with f_2. The members of the distinct class decrease in frequency with increasing f_2. Smoorenburg also noted that the higher the integer, n, the lower the maximum frequency of f_2 where the DP is perceived by the subject (in accordance with the fan of **Figure 9.11.3-3**). For frequencies of $0.9x f_1$ up to $f_2 + \delta$, Smoorenburg notes the difficulty of humans to perceive the DP in the presence of the two stimuli. He also notes the variability of the performance of individuals in

---

[244]Smoorenburg, G. (1970) Pitch Perception of Two-Frequency Stimuli, *J Acoust Soc Amer* 48, pp 924-942

[245]Smoorenburg, G. (1972a) Audibility Region of Combination Tones, *J Acoust Soc Amer* 52, Number 2 (Part 2) 603-614

[246]Smoorenburg, G. (1972b) Combination tones and their origin, *J Acoust Soc Amer* 52, Number 2 (Part 2) pp 615-614

[247]Smoorenburg, G. (1974) On the mechanism of combination tone generation and lateral inhibition in hearing *In* Zwicker, E. & Terhardt, E. *eds*. Facts and Models in Hearing. NY: Springer pp 332-343

[248]Smoorenburg, G. Gibson, M. Kitzes, L. et al. (1976) Correlates of combination tones observed in the response of neurons in the anteroventral cochlear nucleus of the cat *J Acoust Soc Am* vol 59(4), pp 945-962

this region. Such variability is probably associated with the musical sophistication of the subject, with “piano tuners” and those with “perfect pitch” probably performing best.

Table of Contents 7 April 2012

Chapter 9 List of Figures 4/7/12

SUBJECT INDEX (using advanced indexing option)