NONLINEAR COCHLEAR SIGNAL PROCESSING AND MASKING IN SPEECH PERCEPTION

Jont B. Allen

University of IL
Urbana IL

1. INTRODUCTION

Auditory masking is critical to our understanding of speech and music processing. There are many classes of masking, but two major classes are easily defined. These two types of masking and their relation to nonlinear (NL) speech processing and coding are the focus of this chapter.

The *first* class of masking, denoted *neural masking*, is due to internal *neural noise*, characterized in terms of the intensity *just noticeable difference*, denoted $\Delta I(I, f, T)$ (abbreviated JND_I) and defined as the "just discriminable change in intensity." The JND_I is a function of intensity I, frequency f and stimulus type T (e.g., noise, tones, speech, music, etc). As an *internal noise*, the JND_I may be modeled in terms of a loudness (i.e., perceptual intensity) noise density along the length of the cochlea[1] ($0 \leq X \leq L$), described in terms of a *partial loudness JND* ($\Delta\mathcal{L}(X, T)$, a.k.a. $\text{JND}_\mathcal{L}$). The loudness JND is a function of the *partial loudness* $\mathcal{L}(X)$, defined as the loudness contribution coming from each cochlear *critical band*, or more generally, along some *tonotopic central auditory representation*. The critical band is a measure of cochlear bandwidth at a given cochlear *place* X. The loudness JND plays a major role in speech and music coding since coding quantization noise may be masked by this internal quantization (i.e., "loudness noise").

The *second* masking class, denoted here as *dynamic-masking*, comes from the NL mechanical action of cochlear *outer hair cell* (OHC) signal processing. It can have two forms, simultaneous and non-simultaneous, also known as *forward masking*, or *post-masking*. Dynamic-masking (i.e., nonlinear OHC signal processing) is well-known (i.e., there is a historical literature on this topic) to be intimately related to questions of cochlear frequency selectivity, sensitivity, dynamic range compression and *loudness recruitment* (the loss of loudness dynamic range). Dynamic masking includes the *upward spread of masking* (USM) effect, or in neural processing parlance, *two-tone suppression* (2TS). It may be underappreciated that NL OHC processing (i.e., dynamic masking) is largely responsible for *forward masking* (FM, or post-stimulus masking), which shows large effects over long time scales. For example OHC effects (FM/USM/2TS) can be as large as 50 dB, with a FM "latency" (return to base line) of up to 200 ms. *Forward masking* (FM) and NL OHC *signal onset enhancement* are important to the detection and identification of perceptual features of a speech signal. Some research has concluded that forward masking is not related to OHC processing (Relkin and Turner, 1988; Hewitt and Meddis, 1991), so the topic remains controversial. Understanding and modeling NL OHC processing is key to many speech processing applications. As a result, a vibrant National Institute of Health driven research effort on OHC biophysics has ensued.

This OHC research effort is paying off at the highest level. Three key examples are notable. *First* is the development of wide dynamic-range multiband compression (WDRC) hearing aids. In the last 10-15 years WDRC signal processing (first proposed in 1937 by Bell Labs researchers Steinberg and Gardner), revolutionized the hearing aid industry. With the introduction of compression signal processing, hearing aids now address the recruitment prob-

[1] The cochlea or inner ear is the organ that converts signals from acoustical to neural signals.

lem, thereby providing speech audibility over a much larger dynamic range, at least in quiet.[2] This powerful circuit (WDRC) is not the only reason hearing aids of today are better. Improved electronics and transducers have made significant strides as well. In the last few years the digital barrier has finally been broken, with digital signal processing hearing aids now becoming common.

A *second* example is the development of otoacoustic emissions (OAE) as a hearing diagnostic tool. Pioneered by David Kemp and Duck Kim, and then developed by many others, this tool allows for cochlear evaluation of neonates. The identification of cochlear hearing loss in the first month has dramatically improved the lives of these children (and their parents). While it is tragic to be born deaf, it is much more tragic for the deafness to go unrecognized until the child is 3 year old, when they fail to learn to talk.[3] With proper and early cochlear implant intervention, these kids can lead nearly normal-hearing lives and even talk on the phone. However they cannot understand speech in noise.[4]

A *third* example of the application of NL OHC processing to speech processing is still an underdeveloped application area. The key open problem here is "How does the auditory system, including the NL cochlea, followed by the auditory cortex, processes human speech?" There are many aspects of this problem including speech coding, speech recognition in noise, hearing aids and language learning and reading disorders in children. If we can solve the *robust phone decoding problem*, we will fundamentally change the effectiveness of human-machine interactions. For example, the ultimate hearing aid is the hearing aid with built in robust speech feature detection and phone recognition. While we have no idea when this will come to be, and it is undoubtedly many years off, when it happens there will be a technology revolution that will change human communications.

Chapter Outline: Several topics will be reviewed. First is the history of cochlear models including extensions that have taken place in recent years. These models include both macromechanics and micromechanics of the tectorial membrane and hair cells. This leads to comparisons of the basilar membrane, hair cell, and neural frequency tuning. Hearing loss, loudness recruitment, as well as other key topics of modern hearing health care, are discussed. The role of NL mechanics and dynamic range are reviewed to help the reader understand the importance of modern wideband dynamic range compression hearing aids as well as the overall impact of NL-OHC processing.

Any reader desiring further knowledge about cochlear anatomy and function or a basic description of hearing, they may consult Pickles (1982); Dallos (1996); Yost (2006).

1.1. Function of the Inner Ear

The goal of cochlear modeling is to refine our understanding of how auditory signals are processed. The two main roles of the cochlea are to separate the input acoustic signal into overlapping frequency bands, and to compress the large acoustic intensity range into the much smaller mechanical and electrical dynamic range of the inner hair cell. This is a basic question of information processing by the ear. The eye plays a similar role as a peripheral organ. It breaks the light image into rod and cone sized pixels, as it compresses the dynamic range of the visual signal. Based on the intensity JND, the corresponding visual dynamic range is about 9 to 10 orders of magnitude of intensity (Hecht, 1934; Gescheider, 1997), while the ear has about 11 to 12. The stimulus has a relatively high information rate. Neurons are low bandwidth neural channels. The eye and the ear must cope with this problem by reducing the stimulus to a large number of low bandwidth signals. It is then the job of the cortex to piece these pixel signals back together, to reconstruct the world as we see and hear it.

The acoustic information coding starts in the cochlea (Fig. 1(a)) which is composed of three major chambers formed by Reissner's membrane and the basilar membrane (BM). Mechanically speaking, there are only two chambers, as Reissner's membrane is only for electrical isolation of the Scala media (SM) (Pickles, 1982; Dallos, 1996). Figure 1(b) shows a blown up view of the organ of Corti

[2] The problems of the impaired ear given speech in noise is poorly understood today, but this problem is likely related to the effects of NL OHC processing.

[3] If you can't hear you don't learn to talk.

[4] It is at least possible that this loss is due to the lack of NL-OHC processing.

where the inner hair cells (IHC) and outer hair cells (OHC) sit between the BM and the tectorial membrane (TM). As the BM moves up and down, the Tectorial membrane (TM) shears against the Reticular Lamina (RL), causing the cilia of the inner and outer hair cells to bend. The afferent auditory nerve fibers which are connected to the inner hair cells carry the signal information into the auditory system. Many fewer efferent fibers bring signals from the auditory system to the base of the outer hair cells. The exact purpose of these efferent fibers, which modulate the neural sensitivity, remains unknown.

Inner Hair Cells: In very general terms, the role of the cochlea is to convert sound at the eardrum into neural pulse patterns along approximately 30,000 neurons of the human auditory (VIIIth) nerve. After being filtered by the cochlea, a low-level pure tone has a narrow spread of excitation which excites the cilia of about 40 contiguous inner hair cells (Allen and Neely, 1992; Allen, 1996b; Dallos, 1996). The IHC excitation signal is narrow band with a center frequency that depends on the inner hair cell's location along the basilar membrane. Each hair cell is about 10 micrometers in diameter while the human basilar membrane is about 35 mm in length (35,000 microns). Thus the neurons of the auditory nerve encode the responses of about 3,500 inner hair cells which form a single row of cells along the length of the BM. Each inner hair cell voltage is a low-pass filtered representation of the detected inner hair cell cilia displacement (Hudspeth and Corey, 1977). Each hair cell is connected to many neurons, having a wide range of spontaneous firing rates and thresholds (Liberman, 1982c). In the cat, for example,[5] approximately 15–20 neurons encode each of these narrow band inner hair cells with a neural timing code. It is widely believed that the neuron information channel between the hair cell and the *cochlear nucleus* is a combination of the mean firing rate and the relative timing between neural pulses (spikes). The mean firing rate is reflected in the loudness coding, while the relative timing carries more subtle cues, including for example pitch information such as speech voicing distinctions.

Outer Hair Cells: As shown in Fig. 1(b) there are typically 3 (occasionally 4) outer hair cells (OHCs) for each inner hair cell (IHCs), leading to approximately 12,000 OHCs in the human cochlea. Outer hair cells are used for intensity dynamic range control. This is a form of NL signal processing, not dissimilar to Dolby sound processing.[6] It is well known (as was first proposed by Lorente de Nó (1937) and Steinberg and Gardner (1937)) that noise damage of "nerve cells" (i.e., OHCs) leads to a reduction of dynamic range, a disorder clinically named *loudness recruitment*[7]

We may describe cochlear processing two ways. First in terms of the signal representation at various points in the system. Second, in terms of models which are our most succinct means of conveying the conclusions of years of detailed and difficult experimental work on cochlear function. The body of experimental knowledge has been very efficiently represented (to the extent that it is understood) in the form of these mathematical models. When no model exists (e.g., because we do not understand the function), a more basic description via the experimental data is necessary. Several good books and review papers are available which make excellent supplemental reading (Littler, 1965; Pickles, 1982; Gescheider, 1997; Hartmann, 1997).

For pedagogical purposes this chapter has been divided into four parts: Besides this Introduction, we have sections on the NL cochlea, Neural masking and finally a brief discussion. Section 2 discusses dynamic masking due to NL aspects of the cochlear outer hair cells. This includes the practical aspects, and theory, of the upward spread of masking (USM) and two-tone suppression. Section 3 discusses neural masking, the JND, loudness recruitment, the loudness-SNR, and the Weber-fraction. Section 4 provides a brief summary.

[5] It is commonly accepted that all mammalian cochleae are similar in function except the frequency range of operation differs between species (e.g., human ≈0.1–20 kHz and cat ≈0.3-50 kHz).

[6] This form of processing was inspired by cochlear function, and was in use long before it was patented by Dolby, in movie sound systems developed by Bell Labs in the 1930's and 1940's. Telephone speech is similarly compressed (Steinberg, 1941) via μ-Law coding.

[7] The word *recruitment*, which describes the abnormal growth of loudness in the impaired ear, is a seriously misleading term, since nothing is being recruited (Neely and Allen, 1997).

1.2. History of cochlear modeling

Typically the cochlea is treated as an uncoiled long thin box, as shown in Fig. 2(a). This represents the starting point for the macromechanical models.

1.2.1. Macromechanics

In his book *On the Sensations of Tone* Helmholtz (1863) likened the cochlea to a bank of highly tuned resonators selective to different frequencies, much like a piano or a harp (Helmholtz, 1857, page 22-58), with each string representing a different place X on the basilar membrane. This model as proposed was quite limited since it leaves out key features, the most important of which is the cochlear fluid coupling between the mechanical resonators. But given the early publication date, the great master of physics and psychophysics Helmloltz shows deep insight and his studies provided many very important contributions.

The next major contribution by Wegel and Lane (1924) stands in a class of its own even today, as a double barreled paper having both deep psychophysical and modeling insight.[8] The paper was the first to quantitatively describe the details of how a high level low frequency tone affects the audibility of a second low-level higher frequency tone (i.e., the *upward spread of masking*). It was also the first publication to propose a "modern" model of the cochlea, as shown in Fig. 2(b). If Wegel and Lane had been able to solve the model equations implied by their circuit (of course they had no computer to do this), they would have predicted cochlear traveling waves. It was their mistake, in my opinion, to make this a single paper. The modeling portion of their paper has been totally overshadowed by their experimental results.[9]

[8] Fletcher published much of the Wegel and Lane data one year earlier (Fletcher, 1923a). It is not clear to me why Wegel and Lane are always quoted for these results rather than Fletcher. In Fletcher's 1930 modeling paper, he mentioned that he was the subject in the Wegel and Lane study. It seems to me that Fletcher deserves some of the credit.

[9] Transmission line theory had been widely exploited by Campbell, the first mathematical research at AT&T research (ca. 1898) with the invention of the wave filter (Campbell, 1903, 1922), which had been used for speech articulation studies (Campbell, 1910; Fletcher, 1922; Fletcher and Steinberg, 1930), and Fletcher and Wegel were fully utilizing Campbell's important discoveries.

It was the experimental observations of G. von Békésy starting in 1928 on human cadaver cochleae which unveiled the physical nature of the basilar membrane traveling wave. What von Békésy found (consistent with the 1924 Wegel and Lane model) was that the cochlea is analogous to a "dispersive" transmission line where the different frequency components which make up the input signal travel at different speeds along the basilar membrane, thereby isolating each frequency component at a different place X along the basilar membrane. He properly identified this dispersive wave a "traveling wave," just as Wegel and Lane had predicted in their 1924 model of the cochlea.

Over the intervening years these experiments have been greatly improved, but von Békésy's fundamental observation of the traveling wave still stands. His original experimental results, however, are *not* characteristic of the responses seen in more recent experiments, in many important ways. These differences are believed to be due to the fact that Békésy's cochleae were dead, and because of the high sound levels his experiments required. He observed the traveling wave using stroboscopic light, in dead human cochleae, at sound levels well above 140 dB SPL.

Today we find that the traveling wave has a more sharply defined location on the basilar membrane for a pure tone input than that observed by von Békésy. In fact, according to measurements made over the last 20 years, the response of the basilar membrane to a pure tone can change in amplitude by more than five orders of magnitude per millimeter of distance along the basilar membrane (e.g., 300 dB/oct is equivalent to 100 dB/mm in the cat cochlea).

1.2.2. The 1-dimensional model of the cochlea

To describe this response it is helpful to call upon the macromechanical *transmission line model* of Wegel and Lane (1924) (Fig. 2(b)) and Fletcher (1930), first quantitatively analyzed by Zwislocki (1948); Ranke (1950); Zwislocki (1950); Peterson and Bogert (1950); Fletcher (1951b,a). This popular transmission line model is now denoted the *one-dimensional* (1-D), or *long-wave* model.

Zwislocki (1948) was first to quantitatively analyze Wegel and Lane's macromechanical cochlear

model, explaining Békésy's traveling wave observations. The stapes input pressure P_1 is at the left, with the input velocity V_1, as shown by the arrow, corresponding to the stapes velocity. This model represents the mass of the fluids of the cochlea as electrical inductors and the BM stiffness as a capacitor. Electrical circuit networks are useful when describing mechanical systems. This is possible because of an electrical to mechanical analog that relates the two systems of equations.[10]

BM impedance. During the following discussion it is necessary to introduce the concept of a *1-port* (two-wire) impedance. *Ohm's Law* defines the impedance as

$$\textit{Impedance} = \frac{\textit{effort}}{\textit{flow}}. \tag{1}$$

In an electrical system the impedance is the ratio of a voltage (effort) over a current (flow). In a mechanical system it is the force (effort) over the velocity (flow).

For *linear time-invariant causal* (LTIC) systems (e.g., an impedance), *phasor* notation is very useful, where the tone is represented as the real part ($\Re$) of the complex exponential[11]

$$e^{i2\pi ft+i\phi} \equiv cos(2\pi ft+\phi) + isin(2\pi ft+\phi). \tag{2}$$

More specifically, impedance is typically defined in the frequency domain using *Laplace transform* notation, in terms of a damped tone

$$Ae^{\sigma t}\cos(2\pi ft+\phi) \equiv A\Re e^{st+i\phi} \tag{3}$$

excitation, characterized by the tone's amplitude A, phase ϕ and *complex Laplace frequency* $s \equiv \sigma + i2\pi f$. When a function such as $Z(s)$ is shown as a function of the complex frequency s, this means that its inverse Laplace transform $z(t) \leftrightarrow Z(s)$ must be *causal*. In the time-domain, the voltage may be found from the current via a convolution with $z(t)$. Three classic examples of such impedances are presented next.

[10] Electrical circuit elements comprise a *de facto* standard for describing such equations. It is possible to write down the equations that describe the system from the circuit of Fig. 2(b), by those trained in the art. Engineers and scientists frequently find it easier to "read" and think in terms of these pictorial circuit diagrams, than to interpret the corresponding equations.

[11] The symbol $\equiv$ denotes "equivalence." It means that the quantity to the left of $\equiv$ is defined by the quantity on the right.

Example 1: The impedance of the tympanic membrane (TM, or eardrum) is defined in terms of a pure tone pressure in the ear canal divided by the resulting TM volume velocity (the velocity times the area of TM motion) (Puria and Allen, 1998; Allen *et al.*, 2005). The pressure (effort) and volume velocity (flow) referred to here are conventionally described using complex numbers, to account for the phase relationship between the two.

Example 2: The impedance of a spring is given by the ratio of the force $F(f)$ to velocity $V(f) = sX(f)$ with displacement X

$$Z(s) \equiv \frac{F}{V} = \frac{K}{s} = \frac{1}{sC}, \tag{4}$$

where the spring constant K is the stiffness, C the compliance and s is the complex radian frequency. The stiffness is represented electrically as a capacitor.[12] Having $s = \sigma + i2\pi f$ in the denominator indicates that the impedance of a spring has a phase of $-\pi/2$ (e.g., -90^o). Such a phase means that when the velocity is $\cos(2\pi ft)$, the force is $\sin(2\pi ft)$. Equation 4 follows from Hooke's Law

$$F = KX = \frac{K}{s}sX = \frac{K}{s}V. \tag{5}$$

Example 3: From *Newton's Law* $F = Ma$ where F is the force, M is the mass, and acceleration $a(s) = sV(s)$ [i.e., the acceleration in the time domain is $dv(t)/dt$]. The electrical element corresponding to a mass is an "inductor," indicated in Fig. 2(b) by a coil. Thus for a mass Z(s) = sM.

From the above relations the magnitude of the impedance of a spring decreases as $1/f$, while the impedance magnitude of a mass is proportional to f. The stiffness with its -90^o phase is called a *lagging* phase, while the mass with its +90^o phase is called a *leading* phase.

Different points along the basilar membrane are represented by the cascaded sections of the lumped transmission line model of Fig. 2(b). The position X along the model is called the *place* variable and corresponds to the longitudinal position along the cochlea. We shall assume that $X = 0$ is at the stapes. The series (horizontal) inductors (coils) denoted L_k represent the fluid mass (inertia) along the

[12] As parallel lines in Fig. 2(b).

length of the cochlea, while the shunt elements represent the mechanical (acoustical) impedance of the corresponding partition (Organ of Corti) impedance, defined as the pressure drop across the partition divided by its volume velocity per unit length

$$Z_p(s,X) = \frac{K_p(X)}{s} + R_p(X) + sM_p, \quad (6)$$

where $K_p(X)$ is the partition stiffness, R_p is the partition resistance. Each inductor going to ground (l_i in Fig. 2(b)) represents the partition plus fluid mass per unit length M_p of the section. Note that sM_p, R_p and K_p/s are impedances, but Mass M_m and stiffness K_p are not. The partition stiffness decreases exponentially along the length of the cochlea, while the mass is frequently approximated as being independent of place.

As shown in Fig. 3(a), for a given input frequency the BM impedance magnitude has a local minimum at the shunt resonant frequency, at which the membrane can move in a relatively unrestricted manner. The shunt *resonance* has special significance because at this resonance frequency $F_{cf}(X)$ the inductor and the capacitor reactance cancel each other, creating an acoustic "hole," where the only impedance element that contributes to the flow resistance is R_p. Solving for $F_{cf}(X)$

$$\frac{K_p(X)}{2\pi i F_{cf}} + 2\pi i F_{cf} M_p = 0. \quad (7)$$

defines the *cochlear map function*, which is a key concept in cochlear modeling:

$$F_{cf}(X) \equiv \frac{1}{2\pi}\sqrt{K_p(X)/M_p}. \quad (8)$$

The inverse of this function specifies the location of the "hole" $X_{cf}(f)$ as shown in Fig. 3(a). In the example of Fig. 3(a) two frequencies are show, at 1 and 8 kHz, with corresponding resonant points shown by $X_{cf}(1)$ and $X_{cf}(8)$.

Basal to $X_{cf}(f)$ in Fig. 3(a), the basilar membrane is increasingly stiff, and apically (to the right of the resonant point), the impedance is mass dominated. In this apical region the impedance has little influence since almost no fluid flows past the low impedance hole. The above description is dependent on the input frequency f since the location of the hole is frequency-dependent. This description is key to our understanding of why the various frequency components of a signal are splayed out along the basilar membrane.

If one puts a pulse of current in at the stapes, the highest frequencies that make up the pulse would be shunted close to the stapes since at high frequencies the hole is near the stapes, while the lower frequencies would continue down the line. As the low-pass pulse travels down the basilar membrane, the higher frequencies are progressively removed, until almost nothing is left when the pulse reaches the right end of the model (the helicotrema end, the apex of the cochlea).

When a single tone is played, the response in the base increases in proportion to the BM compliance (inversely with the stiffness) until there is a local maximum just before the traveling wave reaches the resonant hole, at which point the response plummets, since the fluid flow is shorted by the hole. For a fixed stimulus frequency f there is a maximum along the place axis called the *characteristic place*, denoted $X_{cf}^{(p)}(f)$. Likewise at a given place X as a function of frequency there is a local maximum called the *characteristic frequency*, denoted $F_{cf}^{(p)}(X)$. The relation between the peak in place as a function of frequency or of the peak in frequency as a function of place is also called the *cochlear map*.[13] The cochlear map function $F_{cf}(X)$ plays a key role in cochlear mechanics, has a long history, and is known by many names[14] (Fletcher, 1930; Fletcher and Munson, 1937; Fletcher, 1938, 1940; Steinberg, 1937; Greenwood, 1961a), the most common today being *Greenwood's function*.

The spread of the response around the peak for a fixed frequency is denoted the *critical spread* $\Delta_x(f)$,

[13] There is a serious confusion with conventional terminology here. The resonant frequency of the BM impedance mathematically defines F_{cf} and specifies the frequency on the base of the high-frequency steep portion of the tuning slope, *not* the peak. However the peak is used as the visual cue, *not* the base of the high frequency slope. These two definitions differ by a small factor (that is ignored) that depends directly on the high frequency slope of the response. Over most of the frequency range this slope is huge, resulting in a very small factor, justifying its being ignored. However at very low frequencies the slope is shallow and the factor can then be large. The "droop" in the cochlear map seen in Fig. 3(b) at the apex ($x = L$) may be a result of these conflicting conflicting definitions.

[14] In the speech literature it is called the *Mel-scale*.

while the frequency spread at a given place is called the *critical band* denoted $\Delta_f(X)$. As early as 1933 it was clear that the critical band must exist, as extensively discussed by Fletcher and Munson (1933). At any point along the BM the critical band is proportional to the *critical ratio* $\kappa(X)$, defined as the ratio of pure tone detection intensity at threshold in a background of white noise, to the spectral level of the noise, namely

$$\Delta_f(X) \propto \kappa(X), \tag{9}$$

In the next section we shall show how these various quantities are related via the cochlear map.

Derivation of the cochlear map function: The derivation of the cochlear map is based on "counting" critical bands as shown by Fletcher (Allen, 1996b) and popularized by Greenwood (1961b). The *number of critical bands* N_{cb} may be found by integrating the critical band density over both frequency and place, and equating these two integrals, resulting in the cochlear map $F_{cf}(X)$:

$$N_{cb} \equiv \int_0^{X_{cf}} \frac{dX}{\Delta_f(X)}. = \int_0^{F_{cf}} \frac{df}{\Delta_x(f)} \tag{10}$$

There are approximately 20 pure-tone frequency JNDs per critical band and Fletcher (1938), Fletcher (1953a, page 171)) showed that the *critical ratio* expressed in dB $\kappa_{dB}(X)$ is of the form $aX + b$, where a and b are constants (Allen, 1996b). As verified by Greenwood (1961b, page 1350, Eq. 1) the critical bandwidth in Hz is therefore

$$\Delta_f(X) \propto 10^{\kappa_{dB}(X)/10}. \tag{11}$$

The critical spread $\Delta_x(X)$ is the effective width of the energy spread on the basilar membrane for a pure tone. Based on a suggestion by Fletcher, Allen (1996b) showed that for the cat $\Delta_x(X)$ corresponds to about 2.75 times the basilar membrane width $W_{bm}(X) \propto e^X$. It is reasonable to assume that the same relation would hold in the human case.

The direct observation of the cochlear map in the cat was made by Liberman (1982a) and Liberman and Dodds (1984), and they showed the following empirical formula fit the data

$$F_{cf}(X) = 456\left(10^{2.1(1-X/L)} - 0.8\right), \tag{12}$$

where the length of the cat cochlea is $L = 21$ [mm] and X is measured from the stapes (Liberman, 1982b). The same formula may be used for the human cochlea if $L = 35$ [mm] is used, the 456 is replaced by 165.4, and 0.8 by 0.88. Based on Eq. 12, and as defined in Fig. 3(b), the "slope" of the cochlear map is 3 mm/oct for the cat and 5 mm/oct for the human, as may be determined from the formula $L\log_{10}(2)/2.1$ with $L = 21$ or 35 for cat and human respectively.

For a discussion of work after 1960 on the critical band see Allen (1996b); Hartmann (1997).

2. THE NONLINEAR COCHLEA

In cochlear modeling there are two complex fundamental intertwined problems, *cochlear frequency selectivity* and *cochlear/OHC nonlinearity*. Wegel and Lane's 1924 transmission line wave theory was a most important development, since it was published 26 years prior to the experimental results of von Békésy, and it was based on a simple set of physical principles, conservation of fluid mass, and a spatially variable basilar membrane stiffness. It also gives insight into the NL cochlea, as well as 2-dimensional model wave-transmission effects (mass-loading of the BM).

Over a 15 year period starting in 1971, there was a paradigm shift. Three discoveries rocked the field: (1) nonlinear compressive basilar membrane and inner hair cell measures of neural-like cochlear frequency selectivity (Rhode, 1971; Sellick and Russell, 1978), (2) otoacoustic (ear canal) nonlinear emissions (Kemp, 1978), and (3) motile outer hair cells (Brownell, Bader, Bertran, and de Rabaupierre, 1985). Today we know that these observations are related, and all involve outer hair cells. A theory (e.g., a computational model) is needed to tie these results together. Many groups are presently working out these theories.

On the modeling side during the same period (the 1970's) all the variants of Wegel and Lane 1-D linear theory were becoming dated because: (1) numerical model results became available, which showed that 2-D and 3-D models were more frequency selective than the 1-D model, (2) experimental basilar membrane observations showed that the basilar membrane motion had a nonlinear compressive re-

sponse growth, and (3) improved experimental basilar membrane observations became available which showed increased nonlinear cochlear frequency selectivity.

Because these models and measures are still under development today [the problem has not yet (ca. 2007) been solved], it is necessary to describe the data rather than the models. Data that drives these nonlinear cochlear measures includes:

- The upward spread of masking (USM), first described quantitatively by Wegel and Lane in 1924,
- Distortion components generated by the cochlea and described by Wegel and Lane (1924); Goldstein and Kiang (1968); Smoorenburg (1972); Kemp (1979a); Kim *et al.* (1979); Fahey and Allen (1985) and many others,
- Normal loudness growth and recruitment in the impaired ear (Fletcher and Munson, 1933; Steinberg and Gardner, 1937),
- The frequency dependent neural two–tone suppression observed by Sachs and Kiang (1968); Arthur *et al.* (1971); Kiang and Moxon (1974); Abbas and Sachs (1976); Fahey and Allen (1985); Pang and Guinan (1997) and others,
- The frequency dependent basilar membrane response level compression first described by Rhode (1971, 1978),
- The frequency dependent inner hair cell receptor potential level compression, first described by Sellick and Russell (1978); Russell and Sellick (1978).
- Forward masking data that shows a linear return to baseline after up to 0.2 s (Duifhuis, 1973). There may be compelling evidence that OHCs are the source of forward masking.

We shall discuss each of these, but two related measures are the most important for understanding masking effects, the upward spread of masking (USM) and two tone suppression (2TS).

Basilar membrane nonlinearity: The most basic early and informative of these nonlinear effects was the NL basilar membrane measurements made by Rhode (1971, 1978), as shown in Fig. 4(a), showing the basilar membrane displacement to be a highly NL function of level. For every four dB of pressure level increase on the input, the output displacement (or velocity) only changed one dB. This compressive nonlinearity depends on frequency, and only occurs near the most sensitive region (i.e., the tip of the tuning curve). For other frequencies the system was either linear, namely, one dB of input change gave one dB of output change for frequencies away from the best frequency, or very close to linear. This NL effect was highly dependent on the health of the animal, and would decrease, or would not be present at all, if the animal was not in its physiologically pristine state.

An important and useful measure of cochlear linear and nonlinear response proposed by Rhode (1978, Fig. 8) Fig. 4(b) describes cochlear tuning curves by straight lines on log-log coordinates. Such straight line approximations are called *Bode plots* in the engineering literature. The *slopes* and *break points*, defined as the locations where the straight lines cross, characterize the response.

Otoacoustic Emissions: A few years after Rhode's demonstration of cochlear nonlinearity, David Kemp observed otoacoustic emissions (tonal sound emanating from the cochlea and NL "echos" to clicks and tone bursts) (Kemp, 1978, 1979b,a, 1980, 1986). Kemp's findings were like a jolt to the field, which led to a cottage industry of objective testing of the auditory system, including both cochlear and middle ear tests.

Motile OHCs: Subsequently, Brownell *et al.* (1985) discovered that isolated OHCs change their length when placed in an electric field, thus that the outer hair cell is motile (Brownell, Bader, Bertran, and de Rabaupierre, 1985). This then led to the intuitive and widespread proposal that outer hair cells act as voltage controlled motors that directly drive the basilar membrane on a cycle by cycle basis.[15] OHC NL processing is the basis for both the asymmetry of simultaneous (upward vs. downward spread) and temporal (forward vs. backward) masking.

As summarized in Fig. 5 the OHCs feed back to the BM via the receptor potential, which in turn is modulated by both the position of the basilar mem-

[15] It seems quite clear from a great deal of data that the OHC onset response time is on the order of one cycle or so of the BM impulse response, because the first peak is linear (Recio and Rhode, 2000). The release time must be determined by the OHC membrane properties, the time constant of which must be long relative to the attack.

brane (forming a fast feedback loop), and alternatively by the efferent neurons that are connected to the outer hair cells (forming a slow feedback loop). The details of all this are the topic of a great deal of present research.

The OHCs are the one common element that link all the NL data previously observed, and a missing piece of the puzzle that most needs to be understood before any model can hope to succeed in predicting basilar membrane, hair cell, and neural tuning, and NL compression. Understanding the outer hair cell's two-way mechanical transduction is viewed as the key to solving the problem of the cochlea's dynamic range.

Historically the implication that hair cells might play an important role in cochlear mechanics go back at least to 1936 when loudness recruitment was first reported by Fowler (1936) in a comment by Lorente de Nó (1937), stating that cochlear hair cells are likely to be involved in loudness recruitment.

The same year Steinberg and Gardner (1937) were explicit about the action of recruitment when they concluded

> When someone shouts, such a deafened person suffers practically as much discomfort as a normal hearing person would under the same circumstances. Furthermore for such a case, the effective gain in loudness afforded by amplification depends on the amount of variable type loss present. Owing to the expanding action of this type of loss it would be necessary to introduce a corresponding compression in the amplifier in order to produce the same amplification at all levels.

Therefore as early as 1937 there was a sense that cochlear hair cells were related to dynamic range compression.

In more recent years, theoretical attempts to explain the difference in tuning between normal and damaged cochleae led to the suggestion that OHCs could influence BM mechanics. In 1983 Neely and Kim conclude

> We suggest that the negative damping components in the model may represent the physical action of outer hair cells, functioning in the electrochemical environment of the normal cochlea and serving to boost the sensitivity of the cochlea at low levels of excitation.

In 1999 yet another (a fourth) important discovery was made, that the outer hair cell mechanical stiffness depends on the voltage across its membrane (He and Dallos, 1999, 2000). This change in stiffness, coupled with the naturally occurring internal static pressure, may well account for the voltage dependent accompanying length changes (the cell's voltage dependent motility). This view follows from block diagram feedback model of the organ of Corti shown in Fig. 5 where the excitation to the OHC changes the cell voltage V_{ohc}, which in turn changes the basilar stiffness (Allen, 1997a). It should be noted that this is only one of many possible theories that have been put forth.

This experimental period set the stage for explaining the two most dramatic NL measures of cochlear response, the upward spread of masking and its related neural correlate, two-tone suppression, and may well turn out to be the explanation of the nonlinear forward-masking effect as well (Duifhuis, 1973).

2.0.3. Simultaneous dynamic-masking

The psychophysically measured *upward spread of masking* (USM) and the neurally measured *two-tone suppression* (2TS) are closely related dynamic-masking phenomena. Unfortunately these two measures have traditionally been treated independently in the literature. As will be shown, it is now clear that they are alternative objective measures of the same OHC compressive nonlinearity. Both involve the dynamic suppression of a basal (high frequency) probe due to the simultaneous presentation of an apical (low frequency) suppressor. These two views (USM versus 2TS) nicely complement each other, providing a symbiotic view of cochlear nonlinearity.

Upward Spread of Masking (USM): In a classic paper, Mayer (1876) was the first to describe the asymmetric nature of masking (Titchener, 1923; Duifhuis, 1973). Mayer made his qualitative observations with the use of clocks, organ pipes and tuning forks, and found that that the spread of masking

is a strong function of the probe-to-masker frequency ratio (f_p/f_m).

In 1923, Fletcher published the first quantitative results of tonal masking. In 1924, Wegel and Lane extended Fletcher's experiments[16] using a wider range of tones. Wegel and Lane then discuss the results in terms of their 1-D model described above. As shown in Fig. 6(a), Wegel and Lane's experiments involved presenting listeners with a masker tone at frequency $f_m = 400$ [Hz] and intensity I_m (the abscissa), along with a probe tone at frequency f_p (the parameter used in the figure). At each masker intensity and probe frequency, the threshold probe intensity $I_p^*(I_m)$ is determined, and displayed relative to its threshold *sensation level* (SL) (the ordinate is the probe level at threshold [dB-SL]). The $*$ indicates a threshold measure.

In Fig. 6(a) f_m = 400 Hz, I_m is the abscissa, f_p is the parameter on each curve, in kHz, and the threshold probe intensity $I_p^*(I_m)$ is the ordinate. The dotted line superimposed on the 3 kHz curve $(I_m/10^{60/10})^{2.4}$ represents the suppression threshold at 60 dB-SL which has a slope of 2.4 dB/dB. The dotted line superimposed on the 0.45 kHz curve has a slope of 1 and a threshold of 16 dB SL.

Three regions are clearly evident: the *downward spread* of masking ($f_p < f_m$, dashed curves), *critical band* masking ($f_p \approx f_m$, dashed curve marked 0.45), and the *upward spread* of masking ($f_p > f_m$, solid curves) (Allen, 1997b).

Critical band masking has a slope close to 1 dB/dB (the superimposed dotted line has a slope of 1).[17] The downward spread of masking (the dashed lines in Fig. 6(a)) has a low threshold intensity and a variable slope that is less than one dB/dB, and approaches 1 at high masker intensities. The upward spread of masking (USM), shown by the solid curves, has a threshold near 50 dB re sensation level (e.g., 65 dB SPL), and a growth just less than 2.5 dB/dB. The dotted line superimposed on the f_p=3 kHz curve has a slope of 2.4 dB/dB and a threshold of 60 dB.

The dashed box shows that the upward spread of masking of a probe at 1 kHz can be greater than the masking within a critical band (i.e., f_p = 450 Hz $>$ f_m=400 Hz). As the masker frequency is increased, this "crossover effect" occurs in a small frequency region (i.e., 1/2 octave) above the masker frequency. The crossover is a result of a well documented NL *response migration*, of the excitation pattern with stimulus intensity, described in a wonderful paper by McFadden (1986). Response migration was also observed by Munson and Gardner in a classic paper on forward masking (Munson and Gardner, 1950). This important migration effect is beyond the scope of the present discussion, but is reviewed in (Allen, 1997b; Strope and Alwan, 1997; Allen, 1999b) (see also Fig. 10).

The upward spread of masking is important because it is easily measured psychophysically in normal hearing people, is robust, well documented, and nicely characterizes normal outer hair cell nonlinearities. The psychophysically measured USM has correlates in basilar membrane and hair cell, and is known as two–tone suppression (2TS) in the auditory nerve literature, as shown in Fig. 6(b).

Two–tone suppression: The neural correlate of the psychophysically measured USM is called *two–tone suppression* (2TS). As shown in the insert of Fig. 7(a), first a neural tuning curve is measured. A pure tone probe at intensity $I_P(f_p)$, and frequency f_p, is placed a few dB (e.g., 6 to 10) above threshold at the characteristic (best) frequency of the neuron F_{cf} (i.e., $f_p = F_{cf}$). In 2TS a suppressor tone plays the role of the masker.

There are two possible thresholds. The intensity of the suppressor tone $I_s(f_s)$ at frequency f_s is increased until either the rate response to the probe alone $R(I_P, I_s = 0)$ (a) decreases by a small increment Δ_R, or (b) increases from the undriven spontaneous rate $R(0,0)$ by increment Δ_R. These two criteria are defined in Fig. 7(a) as:

$$R_P(I_P, I_s^*) \equiv R(I_P, 0) - \Delta_R \qquad (13)$$

and

$$R_{spont}(I_P, I_s^*) \equiv R(0,0) + \Delta_R. \qquad (14)$$

Δ_R indicates a fixed small but statistically significant constant change in the rate (e.g., Δ_R = 20 spikes/s is a typical value). The threshold suppressor intensity is defined as $I_s^*(f_s)$.[18] The two threshold definitions R_{spont} and R_P are very different, and both are useful. The more common measure is (a), and the difference in intensity between the two thresholds is quite

[16] Fletcher was the subject (Fletcher, 1930, Page 325).

[17] Four years later in Riesz (1928) shows critical band masking obeys the *near–miss to Weber's Law*, as described in Sec. 3.2.

[18] As before the $*$ indicates the threshold suppressor intensity.

large. The second measure (b) is consistent with neural tuning curve suppression and corresponds to suppression of the probe to threshold.

Neural data of Abbas and Sachs (1976) (their Fig. 8) are reproduced in Fig. 6(b). For this example (see entry in lower-right just below 105), F_{cf} is 17.8 kHz, and the $f_p = F_{cf}$ probe intensity $20 \log 10(|P_1|)$ is 60 dB. The label on the curves is the frequency f_1. The threshold intensity of the associated neural tuning curve has a low spontaneous rate and a 50-55 dB threshold. The left panel of Fig. 6(b) is for apical suppressors that are lower in frequency than the CF probe ($f_s < f_p$). In this case the threshold is just above 65 dB SPL. The suppression effect is relatively strong and independent of frequency. In this example the threshold of the effect is less than 4 dB apart (the maximum shift of the two curves) at suppressor frequencies f_s of 10 and 5 kHz (a one octave separation).

The right panel shows the case $f_s > f_p$. The suppression threshold is close to the neuron's threshold (e.g., 50 dB SPL) for probes at 19 kHz, but increases rapidly with frequency.The strength of the suppression is weak in comparison to the case of the left panel ($f_s < f_p$), as indicated by the slopes of the family of curves.

The importance of the criterion: The data of Fig. 6(b) uses the first suppression threshold definition Eq. 13 R_P (a small drop from the probe driven rate). In this case the F_{cf} probe is well above its detection threshold at the suppression threshold, since according to definition Eq. 13, the probe is just detectably reduced, and thus audible. With the second suppression threshold definition Eq. 14 R_{spont}, the suppression threshold corresponds to the detection threshold of the probe. Thus Eq. 14, *suppression to the spontaneous rate*, is appropriate for Wegel and Lane's masking data where the probe is at its detection threshold $I_P^*(I_m)$. Suppression threshold definition Eq. 14 was used when taking the 2TS data of Fig. 7(b), where the suppression threshold was estimated as a function of suppressor frequency.

To be consistent with a detection threshold criterion, such as the detection criterion used by Wegel and Lane in psychophysical masking, (Eq. 14) must be used. To have a tuning curve pass through the F_{cf} probe intensity of a 2TS experiment (i.e., be at threshold levels), it is necessary to use the suppression to rate criterion given by Eq. 14. This is shown in Fig. 7(b) where a family of tuning curves is taken with different suppressors present. As described by Fahey and Allen (1985), when a probe is placed on a specific tuning curve of Fig. 7(b), corresponding to one of the suppressor level symbols of Fig. 7(b), and a suppression threshold is measured (their Fig. 13). (lower panel), that suppression curve will fall on the corresponding suppression symbol of Fig. 7(b). There is a symmetry between the tuning curve measured in the presents of a suppressor, and a suppression threshold obtained with a given probe. This symmetry holds only for criterion Eq. 14, the detection threshold criterion, which is appropriate for Wegel and Lane's data.

Suppression threshold: Using the criterion Eq. 14, Fahey and Allen (1985) showed (their Fig. 13) that the suppression threshold $I_s^*(I_P)$ in the tails is near 65 dB SPL (0.04 Pa). This is true for suppressors between 0.6 and 4 kHz. A small amount of data are consistent with the threshold being constant to much higher frequencies, but the Fahey and Allen data are insufficient on that point.

Suppression slope: Bertrand Delgutte has written several insightful papers on masking and suppression (Delgutte, 1990a,b, 1995). He first estimated how the intensity growth slope (the ordinate, in dB/dB) of 2TS varies with suppressor frequency (the abscissa) for several probe frequencies (the parameter indicated by the vertical bar) (Delgutte, 1990b). As may be seen in the figure, the suppression growth slope for the case of a low frequency apical suppressor on a high frequency basal neuron (the case of the left panel of Fig. 6(b)), is ≈2.4 dB/dB. This is the same slope as for Wegel and Lane's 400 Hz masker, 3 kHz probe USM data shown in Fig. 6(a). For suppressor frequencies greater than the probe's ($f_s > f_p$), Delgutte reports a slope that is significantly less than 1 dB/dB. Likewise Wegel and Lane's data has slopes much less than 1 for the downward spread of masking.

Summary: The USM and 2TS data show systematic and quantitative correlations between the threshold levels and slopes. The significance of these correlations has special importance because (a) they come from very different measurement methods, and (b) Wegel and Lane's USM are from human, while the 2TS data are from cat, yet they show similar

responses. This implies that the cat and human cochleae may be quite similar in their NL responses.

The USM and 2TS threshold and growth slope (e.g., 50 dB-SL and 2.4 dB/dB) are important features that must be fully understood and modeled before we can claim to understand cochlear function. While there have been several models of 2TS (Kanis and de Boer, 1994; Hall, 1974; Geisler and Nuttall, 1997) as discussed in some detail by Delgutte (Delgutte, 1990b), none are in quantitative agreement with the data. The two–tone suppression model of Hall (Hall, 1974) is an interesting contribution to this problem because it qualitatively explores many of the key issues. Finally forward masking data also show related nonlinear properties that we speculate may turn out to be related to NL OHC function as well (Strope and Alwan, 1997; Régnier and Allen, 2007a,b).

2.1. Outer Hair Cell Transduction

The purpose of this section is to address two intimately intertwined problems *cochlear frequency selectivity* and *cochlear nonlinearity*. The fundamental question in cochlear research today is: *What is the role of the outer hair cell (OHC) in cochlear mechanics?* The OHC is the source of NL effect, and the end product is dynamic masking, including the USM, 2TS and forward masking, all of which include dramatic amounts of gain and tuning variation. The issues are the nature of the NL transformations of the BM, OHC cilia motion, and OHC soma motility, at a given location along the basilar membrane.

The prevailing and popular view is that the OHC provides *cochlear sensitivity* and *frequency selectivity* (Dallos, 1996; Narayan, Temchin, Recio, and Ruggero, 1998; deBoer, 1996; Geisler, 1998). The alternative view, argued here, is that the OHC compresses the excitation to the inner hair cell, thereby providing dynamic range expansion.

There is an important difference between these two views. The *first* view deemphasizes the role of the OHC in providing dynamic range control (the OHC's role is to improve sensitivity and selectivity), and assumes that the NL effects result from OHC saturation.

The *second* view places the dynamic range problem as the top priority. It assumes that the sole purpose of the OHC nonlinearity is to provide dynamic range compression, and that the OHC plays no role in either sensitivity or selectivity, which are treated as important but independent issues.[19]

The dynamic range problem: The question of how the large (up to 120 dB) dynamic range of the auditory system is attained has been a long standing problem which remains fundamentally incomplete. For example, *recruitment*, the most common symptom of neurosensory hearing loss, is best characterized as the loss of dynamic range (Steinberg and Gardner, 1937; Allen, 1996a,b; Neely and Allen, 1997). Recruitment results from outer hair cell damage (Carver, 1978). To successfully design hearing aids that deal with the problem of recruitment, we need models that improve our understanding of *how* the cochlea achieves its dynamic range.

Based on a simple analysis of the IHC voltage, one may prove that the dynamic range of the IHC must be less than 65 dB (Allen, 2001). In fact it is widely accepted that IHC dynamic range is less than 50 dB.

The IHC's transmembrane voltage is limited at the high end by the cell's open circuit (unloaded) membrane voltage, and at the low end by thermal noise. There are two obvious sources of thermal noise, cilia Brownian motion, and Johnson (shot) noise across the cell membrane (Fig. 8).

The obvious question arises: *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*?* The huge amount of indirect evidence has shown that this increased dynamic range results from mechanical NL signal compression provided by outer hair cells. This dynamic range compression shows up in auditory psychophysics and in cochlear physiology in many ways.

This discrepancy in dynamic range forms a basic paradox.

Outer Hair Cell Motility model: A most significant finding in 1985 was of OHC *motility*, namely that the OHC changes its length by up to 5% in response to the cell's membrane voltage (Brownell *et al.*, 1985; Ashmore, 1987; Santos-Sacchi, 1991). This less than 5% change in length must account for

[19] Of course other views besides these two are possible.

a 40 dB (100 times) change in cochlear gain. This observation led to a significant increases in research on the OHC cell's motor properties.

In 1999 it was shown that the cell's longitudinal soma stiffness changes by at least a factor of 2 (>100%), again as a function of cell membrane voltage (He and Dallos, 1999, 2000). A displacement of the cilia in the direction of the tallest cilia, which is called a *depolarizing* stimulus, decreases the magnitude of the membrane voltage $|V_m|$, *decreases* the longitudinal soma stiffness, and *decreases* the cell soma length. A hyper polarizing stimulus increases the stiffness and extends the longitudinal soma length.

Given this much larger relative change in stiffness (a factor of 2) compared to the relative change in length (a factor of 1.05), for a maximum voltage change, it seems possible, or even likely, that the observed length changes (the motility) are simply a result of the voltage dependent stiffness. For example, imagine a spring stretched by applying a constant force (say a weight), and then suppose that the spring's stiffness decreases. It follows from Hooke's Law (Eq. 5) that the spring's length will *increase* when the stiffness decreases.

Each cell is stretched by its internal static pressure $\mathcal{P}$ (Iwasa and Chadwick, 1992), and its stiffness is voltage controlled (He and Dallos, 1999, 2000). The voltage dependent relative stiffness change is much greater than the relative length change. Thus we have the necessary conditions for a stiffness induced motility.

2.2. Micromechanics

Unlike the case of macromechanical models, the physics of every micromechanical model differs significantly. This is in part due to the lack of direct experimental evidence of physical parameters of the cochlea. This is an important and very active area of research [e.g., (Russell *et al.*, 2007)].

To organize our discussion of cochlear micromechanics, we represent each radial cross-section through the cochlear partition as a linear 2-port network. A general formalization in transmission matrix form of the relation between the basilar membrane *input* pressure $P(x,s)$ and velocity $V(x,s)$ and the OHC *output* cilia bundle shear force $f(x,s)$ and shear velocity $v(x,s)$

$$\begin{bmatrix} P \\ V \end{bmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{bmatrix} f \\ v \end{bmatrix}, \qquad (15)$$

where A, B, C, and D are complex functions of place x and radian frequency s.

2.2.1. Passive BM models

The most successful *passive* model of cochlear tuning is the resonant tectorial membrane (RTM) model (Allen, 1980; Allen and Neely, 1992). The RTM model starts from the assumption that the slope S_2 of BM tuning is insufficient to account for the slope S_2 of neural tuning, as seen in Fig. 4(b). This sharpening is accounted for by a reflection in the tectorial membrane, introducing an antiresonance (*spectral zero*) at frequency F_z (see Fig. 4(b)), which is about half an octave below the resonant frequency F_{cf} of the basilar membrane. As described by Allen and Neely (1992), the detailed A, B, C, D elements of Eq. 15 are given by Allen (1980); Allen and Neely (1992).

As described in Allen (1977), the response ratio of IHC cilia bundle displacement to basilar membrane displacement is defined as $H_{ihc}(x,s)$. The parameters of the resonant tectorial membrane (RTM) model may be chosen such that model results fit the experimental neural threshold tuning curves closely, as shown in Fig. 9.

The nonlinear RTM model: The resonant tectorial membrane (RTM) model is made NL by control of the BM stiffness via OHC's stiffness, as modeled in Fig. 10(a). The OHC soma stiffness has been shown to be voltage dependent by Dallos *et al.* (1997) and dependent on Prestin in the membrane wall (Dallos, 2002). If an elastic connection is assumed where the TM attaches to the Limbus, and if this elasticity is similar to that of the cilia of the OHC, then the resulting transfer function between the BM and IHC cilia is strongly filtered at low frequencies (Allen, 1997a, 1999a; Allen and Sen, 1999; Sen and Allen, 2006). Such models are actively under consideration (Russell *et al.*, 2007).

It is postulated that the decrease in OHC stiffness accompanying cilia stimulation results in a decrease of the net BM partition stiffness $K_p(x)$ (i.e. increasing compliance) of Eq. 6. As shown in Fig. 3, this decrease in the local BM stiffness would result in the

partition excitation pattern shifting basally towards the stapes. Such shifts in the BM response patterns are commonly seen. Another way to view this is shown in Fig. 10. This migration of the excitation pattern, combined with the assumption that the TM has a highpass characteristic, means that the cilia excitation gain at CF is nonlinearly compressed as the intensity increases. This compression effect is shown in a cartoon format in Fig. 10(b), while Fig. 10(a) shows the actual calculated model results. Note how the bandwidth $\Delta_f(X))$ remains approximately constant as a function of input intensity.

Sewell (1984) has nicely demonstrated that as the voltage driving the hair cells changes, the neural gain in dB at CF changes proportionally. It is not yet known why the dB gain is proportional to the voltage (1 dB/mv), however this would explain why forward masking decays linearly in dB value with time, after a strong excitation, since the membrane voltage $V_m(t)$ is proportional to e^{t/τ_m}, due to the OHC membrane's $\tau_m = RC$ time constant. In my view, explaining the proportionality between the neural threshold in dB and the linear membrane voltage, is key.

Discussion: Two important advantages of the NL-RTM model include its physically based assumptions (described above), and its simplicity. Given these physical assumptions the NL-RTM model can explain (see the references for the details): a) the basal-ward half-octave traveling wave migration as a function of increasing intensity (McFadden, 1986), b) the upward spread of masking (USM) (Fletcher, 1923a; Wegel and Lane, 1924), two-tone suppression (2TS) (see Sec. 2.0.3), d) distortion product generation (Kemp, 1978; Kim, Siegel, and Molnar, 1979; Allen and Fahey, 1983; Fahey and Allen, 1985; Allen and Lonsbury-Martin, 1993; Fahey and Allen, 1997), e) normal and recruiting loudness growth, and f) hypersensitive tails (Liberman and Dodds, 1984).

From the steep 2.5 dB/dB slope of the USM and 2TS (Fig. 6(a)) it seems necessary that the low frequency suppressor is turning down the high frequency probe even though the growth of the masker at the high frequency's place is linear with masker level, as shown in Fig. 10(b).

2.2.2. Active BM models

One obvious question about active cochlear models is "Are they really necessary?" At least three attempts to answer this question based on detailed comparisons of basilar membrane responses have concluded that the measured responses *cannot* be accounted for by a passive cochlear model (Diependaal, de Boer, and Viergever, 1987; Zweig, 1991; deBoer and Nuttall, 1999, 2000a,b).

The CA hypothesis: The most popular active micromechanical theory is called the *cochlear amplifier* (CA) hypothesis. The concept of the *cochlear amplifier*, originated by Gold, Kemp, Kim and Neely, and named by H. Davis, refers to a hypothetical mechanism within the cochlear partition which increases the *sensitivity* of basilar membrane vibrations to low-level sounds and, at the same time, increases the *frequency selectivity* of these vibrations (Kim, Neely, Molnar, and Matthews, 1980). The CA adds mechanical energy to the cochlear partition at acoustic frequencies by drawing upon the electrical and mechanical energy available from the outer hair cells. In response to a tone, the CA adds mechanical energy to the cochlear traveling wave in the region defined by S_2 [define in Fig. 4(b)] as it approaches the place of maximum response. This energy is reabsorbed at other places along the cochlear partition. The resulting improvement in sensitivity of the ear due to the CA is thought to be 40 dB, or more under certain conditions; however, the details of how this amplification might be accomplished are still unknown. A general discussion of this model is presented in Geisler (1998) and in Allen and Fahey (1992).

It is presumed that this OHC action amplifies the BM signal energy on a cycle-by-cycle basis, increasing the sensitivity (Neely and Kim, 1983, 1986). In some of the models it is assumed that this cycle-by-cycle pressure (force) due to the OHCs causes the sharp BM tuning tip. In most of these models, the CA is equivalent to introducing a frequency dependent negative damping (resistance) into the BM impedance. Nonlinear compression is introduced by assuming that the resistance is signal level dependent. This NL resistance model was first described by Hall (Hall, 1974) for the case of $R > 0$. Thus the CA model is an extension of Hall's model to the case of $R < 0$. In several models NL negative damping is obtained with a nonlinear stiffness and a small delay.

The addition of a small delay introduces a negative real part into the impedance.[20]

Allen and Fahey (1992) developed a method for directly measuring the cochlear amplifier (CA) gain. All of the studies to date using this method have found no gain. However many researchers continue to believe that the CA has gain. Given that the gain is on the order of 40-50 dB, this is difficult to understand. A nice summary of this situation has been recently published in Shera and Guinan (2007). The reasons for the failure to directly measure any CA gain are complex and multifaceted, and many important questions remain open. One possibility that remains open is that the many observed large NL OHC BM effects we see are not due to cycle by cycle power amplification of the BM traveling wave.

2.2.3. Discussion and summary

Discussion: Both active and passive BM models are reasonably successful at simulating the neural threshold response tuning curves. Thus we need to look elsewhere to contrast the difference between these two approaches, such as 2TS/USM. While the passive RTM model is easily made NL with the introduction of $K_{ohc}(V_m)$, differences between *nonlinear* RTM and CA models have not yet been investigated. The CA and RTM models differ in their interpretation of damaged cochlear responses. In CA models, the loss of sensitivity of the cochlea with damage is interpreted as a loss of CA gain while in passive models, the loss of sensitivity has been interpreted as a 2:1 change in the BM stiffness (Allen, 1991).

The discovery of OHC motility demonstrates the existence of a potential source of mechanical energy within the cochlear partition which is suitably positioned to influence vibrations of the basilar membrane. It is still an open question whether this source of energy is sufficient to power a CA at high frequencies.

One possible advantage of the CA is that of improving the signal-to-noise ratio in front of the IHC detector. A weakness of the CA models has been their lack of specificity about the physical realization of the active elements. Until we have a detailed physical representation for the CA, RTM models have the advantage of being simpler and more explicit.

The discovery by He and Dallos that the OHC soma stiffness is voltage dependent is an exciting development for the NL passive RTM model, as it greatly simplifies the implementation of the physical model. The RTM model has been in disfavor because many feel it does not account for basilar membrane tuning. This criticism is largely due to the experimental results of physiologists who have measured the BM-ear canal transfer function, and found the tuning of BM velocity to be similar to neural threshold response data. Much of the experimental BM data, however, are not convincing on this point, with the BM slope S_2 [Fig. 4(b)] generally being much smaller than that of neural responses. The question of whether an active model is required to simulate measured BM responses is still being debated.

Better estimates of the amplitude of cilia bundle displacement at a given sound pressure level directly address the sensitivity questions. If the estimate of Russell of 30 mV/degree is correct (Russell *et al.*, 1986), then the cochlear sensitivity question may be resolved by having very sensitive detectors. Also, better estimates are needed of the ratio of the BM frequency response to the IHC frequency response, both at high and low frequencies. Rhode's approach of using the slopes of Fig. 4(b) rather than traditional *ad hoc* bandwidth measures,[21] might be a useful tool in this regard.

Summary: This section has reviewed what we know about the cochlea. The *Introduction* section reviews the nature of modeling and briefly describes the anatomy of the inner ear, and the function of inner and outer hair cells. In Sec. 1.2 we reviewed the history of cochlear modeling. The Wegel and Lane paper was a key paper that introduced the first detailed view of masking, and in the same paper introduced the first modern cochlear model Fig. 2(b). We presented the basic tools of cochlear modeling, *impedance*, and introduced the *Transmission matrix* method (2-port analysis). We described how these models work in intuitive terms, including how the basilar membrane may be treated as having a fre-

[20] In mathematical physics, NL damping resonators are described by *van der Pol equations*, while NL stiffness resonators are described by *Duffing equations* (Pipes, 1958).

[21] The bandwidth 10 dB down relative to the peak has been popular but arbitrary and thus poor, criterion in cochlear research. A second somewhat better bandwidth measure is Fletcher's *equivalent rectangular bandwidth* discussed in Allen (1996b).

quency dependent acoustic hole. The location of the hole, as a function of frequency, is called the cochlear map. This hole keeps fluid from flowing beyond a certain point, producing the cochlear traveling wave.

We reviewed and summarized the NL measures of cochlear response. Since these data are not fully understood, and have not been adequately modeled, this is the most difficult section. However it is worth the effort to understand these extensive data and to appreciate the various relations between them, such as the close parallel between two–tone suppression and the upward spread of masking, and between loudness recruitment and outer hair cell damage.

We reviewed several models of the hair cell, including forward and reverse transduction. Some of this material is recently published, and the view of these models could easily change over the next few years, as we better understand reverse transduction.

Finally in Sec. 2.2 we reviewed the basics of micromechanics. We have presented the two basic types of models, *Passive* and *Active* models, with a critical review of each.

3. NEURAL MASKING

When modeling human psychophysics we must carefully distinguish the external *physical* variables, which we call Φ variables, from the internal *psychophysical* variables, which we refer to as Ψ variables.[22] Psychophysical modeling seeks a transformation from the Φ domain to the Ψ domain. The Φ–intensity of a sound is easily quantified by direct measurement. The Ψ–intensity is the loudness. The idea that loudness could be quantified was first suggested by Fechner (1966) in 1860, which raised the question of the quantitative transformation between the physical and psychophysical intensity. For a recent review of this problem see Schlauch *et al.* (1995). This section is based on an earlier report by Allen (1999b) and Allen and Neely (1997).

An increment in the intensity of a sound that results in a *just noticeable difference* is called an intensity JND. Fechner suggested quantifying the intensity-loudness growth transformation by counting the number of the *loudness JNDs* between two intensity values. However, after many years of work, the details of the relationship between loudness and the intensity JNDs have remained unclear (Zwislocki and Jordan, 1986; Viemeister, 1988; Plack and Carlyon, 1995).

The contribution of Allen and Neely (1997); Allen (1999b) is that it takes a new view of the problem of the intensity JND and loudness by merging the 1953 Fletcher neural excitation pattern model of loudness (Allen, 1995, 1996b) with auditory signal detection theory (Green and Swets, 1966). It is generally accepted that the intensity JND is the physical correlate of the psychological–domain uncertainty corresponding to the psychological intensity representation of a signal. Along these lines, for long duration pure tones and wideband noise, we assume that the Ψ–domain intensity is the loudness, and that the loudness JND results from loudness "noise" due to its stochastic representation.

To model the intensity JND we must define a *decision variable* associated with loudness and its random fluctuations. We call this loudness random decision variable the *single–trial loudness*. Accordingly we define the loudness and the loudness JND in terms of the first and second moments of the single–trial loudness, that is the mean and variance of the distribution of the single-trial loudness decision variable. We also define the ratio of the mean loudness to the loudness standard deviation as the *loudness signal–to–noise ratio* SNR_L.

Our ultimate goal in this work is to use signal detection theory to unify masking and the JND, following the 1947 outline of this problem by Miller (1947). Tonal data follows the "near–miss to Weber's Law" (thus does not obey Weber's law), while the wideband noise data does obey Weber's law.[23] We will show that the transformation of the Φ–domain (intensity) JND data (both tone and noise) into the Ψ domain (loudness) unifies these two types of JND data, since $\mathrm{SNR}_L(L)$ is the same for both the tone and noise cases. To help understand these results, we introduce the concept of a near–miss to Stevens' law, which we show cancels the near–miss to Weber's law, giving the invariance in SNR_L for the tone case (Allen and Neely, 1997). This work has applications

[22] It may be helpful to note that Φ and Ψ sound similar to the initial syllable of the words *phy*sical and *psy*chological, respectively (Boring, 1929)

[23] Weber's law says that the relative JND is a constant, as discussed in Sec. 3.2.

in speech and audio coding.

For the case of tones, we have chosen to illustrate our theoretical work using the classical intensity modulation measurements of Riesz (1928) who measured the intensity JND using small, low-frequency (3-Hz), sinusoidal modulation of tones. "Modern" methods generally use "pulsed" tones which are turned on and off somewhat abruptly, to make them suitable for a two-alternative, forced-choice (2AFC) paradigm. This transient could trigger cochlear forward masking. Riesz's modulation method has a distinct advantage for characterizing the internal signal detection process, because it maintains a nearly steady-state small-signal condition within the auditory system, minimizing any cochlear forward masking component. The interpretation of intensity JNDs is therefore simplified since underlying stochastic processes are stationary.

An outline of this neural masking section is as follows: After some basic definitions in Sec. 3.1 and a review of historical models (e.g., Weber and Fechner), in Sec. 3.2, we explore issues surrounding the relation between the intensity JND and loudness, for the special case of tones in quiet and for wide–band noise. First, we look at formulae for counting the number of intensity and loudness JNDs and we use these formulae, together with decision-theoretic principles, to relate loudness to the intensity JND. We then review the loudness–JND theory developed by Hellman and Hellman (1990), which provided the inspiration for the present work. Next we empirically estimate the loudness SNR, defined as the loudness divided the loudness variance, and proportional to $L/\Delta L$, as a function of both intensity and loudness, using the tonal JND data of Riesz (1928) and the loudness growth function of Fletcher and Munson (1933). We then repeat this calculation for Miller's wideband noise JND and loudness data. Finally we propose a model of loudness that may be used to compute the JND. This model merges Fletcher's neural excitation pattern model of loudness with signal detection theory.

3.1. Basic definitions

We need a flexible yet clear notation, that accounts for important time fluctuations and modulations that are present in the signals, such as beats and gated signals. We include a definition of *masked threshold* because we view the intensity JND as a special case of the masked threshold (Miller, 1947). We include a definition of *beats* so that we can discuss their influence on Riesz's method for the measurement of intensity JNDs.

Intensity: In the time domain, it is common to define the Φ–*intensity* in terms of the time-integrated squared signal pressure $s(t)$, namely,

$$I_s(t) \equiv \frac{1}{\varrho c T} \int_{t-T}^{t} s^2(t) dt, \quad \text{Watts/m}^2 \qquad (16)$$

where T is the integration time, assumed long compared with the period, and ϱc is the specific acoustic impedance of air. The *intensity level* is defined as I_s/I_{ref}, and the *sound pressure level* as s/s_{ref} where the reference intensity is I_{ref} or 10^{-10} μW/cm^2 and the reference pressure $s_{\text{ref}} = 20$ μPa. These two reference levels are equivalent at only one temperature,[24] but both seem to be in use.

Intensity of masker + probe: The JND is sometimes called "self-masking," to reflect the view that it is determined by the internal noise of the auditory system. To model the JND it is useful to define a more general measure called the *masked threshold*, which is defined in the Φ domain in terms of a pressure scale factor α applied to the probe signal $p(t)$ that is then added to the masking pressure signal $m(t)$. The relative intensity of the probe and masker is varied by changing α. Setting $s(t) = m(t) + \alpha p(t)$, we denote the combined intensity as

$$I_{m+p}(t, \alpha) \equiv \frac{1}{\varrho c T} \int_{t-T}^{t} (m(t) + \alpha\, p(t))^2 dt. \quad (17)$$

The unscaled probe signal $p(t)$ is chosen to have the same long–term average intensity as the masker $m(t)$, defined as I. Let $I_m(t)$ be the intensity of the masker with no probe ($\alpha = 0$), and $I_p(t, \alpha) = \alpha^2 I$ be the intensity of the scaled probe signal with no

[24] Equivalence of the pressure and intensity references requires that ϱc = 40 cgs Rayls. At standard atmospheric pressure, this is only true when the temperature is about 39 °C.

masker. Thus[25]

$$I \equiv I_{m+p}(t,0) = I_m(t) = I_p(t,1).$$

Beats: Rapid fluctuations having frequency components outside the bandwidth of the T-s rectangular integration window are very small and will be ignored. Accordingly we drop the time dependence in terms I_m and I_p. Because of beats between $m(t)$ and $p(t)$ (assuming the spectra of these signals are within a common critical band) one must proceed carefully. Slowly varying correlations between the probe and masker having frequency components within the bandwidth of the integration window may *not* be ignored, as with beats between two tones separated in frequency by a few Hz. Accordingly we keep the time dependence in the term $I_{m+p}(t,\alpha)$ and other slow–beating time dependent terms. In the Φ domain these beats are accounted for with a probe–masker correlation function $\rho_{mt}(t)$ (Green and Swets, 1966, Page 213).

Intensity increment $\delta I(t,\alpha)$: Expanding Eq. 17 and solving for the *intensity increment* δI we find

$$\begin{aligned} \delta I(t,\alpha) &\equiv I_{m+p}(t,\alpha) - I & (18)\\ &= \left(2\alpha\rho_{mp}(t) + \alpha^2\right) I, & (19)\end{aligned}$$

where

$$\rho_{mp}(t) = \frac{1}{\varrho c T I}\int_{t-T}^{t} m(t)p(t)dt \qquad (20)$$

defines a normalized cross correlation function between the masker and the probe. The correlation function must lie between -1 and 1.

Detection threshold: As the probe to masker ratio α is slowly increased from zero, the probe can eventually be detected. We specify the *detection threshold* as α_* where the asterisk indicates the threshold value of α at which a subject can discriminate intensity $I_{m+p}(t,\alpha_*)$ from intensity $I_{m+p}(t,0)$ 50% of the time, corrected for chance [i.e., obtain a 75% correct score in a direct comparison of the two signals (Green and Swets, 1966, Page 129)]. The quantity $\alpha_*(t,I)$ is the probe to masker rms pressure ratio at the detection threshold. It is a function of the masker intensity I and, depending on the experimental setup, time.

Masked threshold intensity: The *masked threshold intensity* is defined in terms of α_* as

$$I_p^*(I) \equiv I_p(\alpha_*) = \alpha_*^2 I,$$

which is the threshold intensity of the probe in the presence of the masker.

The masked threshold intensity is a function of the stimulus modulation parameters. For example, tone maskers and narrow–band noise maskers of equal intensity, and therefore approximately equal loudness, give masked thresholds that are about 20 dB different (Egan and Hake, 1950). As a second example, when using the method of beats (Riesz, 1928), the just–detectable modulation depends on the beat frequency. With "modern" 2AFC methods , the signals are usually gated on and off (100% modulation) (Jesteadt *et al.*, 1977). According to Stevens and Davis (p. 142, 1983)

> A gradual transition, such as the sinusoidal variation used by Riesz, is less easy to detect than an abrupt transition; but, as already suggested, an abrupt transition may involve the production of unwanted transients.

One must conclude that the *relative masked threshold* [i.e., $\alpha_*(t,I)$] is a function of the modulation conditions.

Ψ–domain temporal resolution: When modeling time varying psychological decision variables, the relevant integration time T is not the duration defined by the Φ–intensity Eq. 16, rather the integration time is determined in the Ψ–domain. This important Ψ–domain model parameter is called *loudness temporal integration* (Yost, 1994). It was first explicitly modeled by Munson in 1947.

The Φ–domain temporal resolution (T) is critical to the definition of the JND in Riesz's experiment (see appendix A) because it determines the measured intensity of the beats. The Ψ–domain temporal resolution plays a different role. Beats cannot be heard if they are faster than, and therefore "filtered" out by, the Ψ domain response. The Ψ–domain temporal resolution also impacts results for gated stimuli, such as

[25] Because of small fluctuations in I_m and I_p due to the finite integration time T, this equality cannot be exactly true. We specifically ignore these small rapid fluctuations – when these rapid fluctuations are important, our conclusions and model results must be reevaluated.

in the 2AFC experiment, though its role is poorly understood in this case. To model the JND as measured by Riesz's method of just–detectable beats, one must know the Ψ–domain resolution duration to calculate the probe–masker effective correlation $\rho_{mp}(t)$ in the Ψ domain. It may be more practical to estimate the Ψ domain resolution from experiments that estimate the degree of correlation, as determined by the beat modulation detection threshold as a function of the beat frequency f_b.

In summary, even though Riesz's modulation detection experiment is technically a masking task, we treat it, following Riesz (1928), Miller (1947), and Littler (1965), as characterizing the intensity JND. It follows that the Ψ–domain temporal resolution plays a key role in intensity JND and masking models.

The intensity JND ΔI: The intensity *just-noticeable difference* (JND) is[26]

$$\Delta I(I) \equiv \delta(t, \alpha_*), \tag{21}$$

the intensity increment at the masked threshold, for the special case where the probe signal is equal to the masking signal ($p(t) = m(t)$). From Eq. 19 with α set to threshold α_* and $\rho_{mp}(t)$ = 1 [see Eq. 20]

$$\Delta I(I) = (2\alpha_* + \alpha_*^2)I. \tag{22}$$

An important alternative definition for the special case of the *pure–tone JND* is to let the masker be a pure tone, and let the probe be a pure tone of a slightly different frequency (e.g., a beat frequency difference of f_b = 3 Hz). This was the definition used by Riesz in 1928. Beats are heard at f_b = 3 Hz, and assuming the period of 3 Hz is within the passband of the Ψ–temporal resolution window, $\rho_{mp}(t) = \sin(2\pi f_b t)$ and

$$\Delta I(t, I) = [2\alpha_* \sin(2\pi f_b t) + \alpha_*^2]I. \tag{23}$$

If the beat period is less than the Ψ temporal resolution window, the beats are "filtered" out by the auditory brain and we do not hear the beats. In this case $\Delta I(I) = \alpha_*^2 I$.

Internal noise: It is widely accepted that the pure–tone intensity JND is determined by the *internal noise* of the auditory system (Siebert, 1965; Raab and Goldberg, 1975), and that ΔI is proportional to the standard deviation of the Ψ–domain decision variable that is being discriminated in the intensity detection task, reflected back into the Φ domain. The usual assumption, from signal detection theory, is that $\Delta I = d'\sigma_I$, where d' is defined as the proportionality between the change in intensity and the variance $d' \equiv \Delta I/\sigma_I$. Threshold is typically when $d' = 1$ but can depend on the experimental design, and σ_I is the intensity standard deviation of the Φ–domain intensity due to Ψ–domain auditory noise (Hartmann, 1997, Chapter 4).

Hearing threshold: The *hearing threshold* (or unmasked threshold) *intensity* may be defined as the intensity corresponding to the first (lowest intensity) JND. The hearing threshold is represented as $I_p^*(0)$ to indicate the probe intensity when the masker intensity is small (i.e., $I \to 0$). It is believed that internal noise is responsible for the hearing threshold, however, there is no reason to assume that this noise is the same as the internal noise that produces the JND.

Loudness L: The *loudness* L of a sound is the Ψ intensity. The *loudness growth function* $L(I)$ depends on the stimulus conditions. For example $L(I)$ for a tone and for wideband noise are not the same functions. Likewise the loudness growth function for a 100 ms tone and a 1-s tone differ. When defining a *loudness scale* it is traditional to specify the intensity, frequency, and duration of a tone such that the loudness growth function is one (i.e., $L(I_{\text{ref}}, f_{\text{ref}}, T_{\text{ref}})$ = 1 defines a loudness scale). For the sone scale, the reference signal is a I_{ref} = 40 dB SPL tone at f_{ref} = 1 kHz with duration T_{ref} = 1-s. For Fletcher's LU loudness scale, the reference intensity is the hearing threshold, which means that 1 sone = 975 LU (Fletcher, 1953b) for a "normal" hearing person. In the next section we shall show that Fletcher's LU loudness scale is a more natural scale than the sone scale (the ANSI and ISO standard scales). For a detailed discussion of how loudness is measured see Allen (1996b).

The single–trial loudness: A fundamental postulate of psychophysics is that all decision variables (i.e., Ψ variables) are random variables, drawn from some probability density function (Green and Swets, 1966, Chapter 5). For early discussions of this point see Montgomery (1935) and p. 144 of Stevens and Davis (1983). To clearly indicate the distinction between random and nonrandom variables, a tilde ($\sim$)

[26] It is traditional to define the intensity JND to be a function of I, rather than a function of $\alpha(I)$, as we have done here. We shall treat both notations as equivalent [i.e., $\Delta I(I)$ or $\Delta I(\alpha(I))$].

is used to indicate a random variable.[27]

We define the loudness decision variable as the *single–trial loudness* $\widetilde{L}$, which is the sample–loudness heard on each stimulus presentation. The loudness L is then the expected value of the single–trial loudness $\widetilde{L}$

$$L(I) \equiv \mathcal{E}\widetilde{L}(I). \tag{24}$$

The second moment of the single-trial loudness

$$\sigma_L^2 \equiv \mathcal{E}(\widetilde{L} - L)^2 \tag{25}$$

defines the loudness *variance* σ_L^2 and *standard deviation* σ_L.

3.1.1. Derived definitions

The definitions given above cover the basic variables. However many alternative forms (various normalizations) of these variables are used in the literature. These derived variables were frequently formed with the hope of finding an invariance in the data. This could be viewed as a form of modeling exercise that has largely failed (e.g., the near–miss to Weber's law), and the shear number of combinations has led to serious confusions (Yost, 1994, p. 152). Each normalized variable is usually expressed in dB, adding an additional unnecessary layer of confusion to the picture. For example, *masking* is defined as the masked threshold normalized by the unmasked (quiet) threshold, namely

$$M \equiv I_p^*(I)/I_p^*(0).$$

It is typically quoted in dB re sensation level (dB-SL) The intensity JND is frequently expressed as a *relative JND* called the *Weber fraction* defined by

$$J(I) \equiv \Delta I(I)/I. \tag{26}$$

From the signal detection theory premise that $\Delta I = d'\sigma_I$ (Hartmann, 1997), J is just the reciprocal of an effective signal to noise ratio defined as

$$\mathrm{SNR}_I(I) \equiv I/\sigma_I(I) \tag{27}$$

since

$$J = d'\sigma_I/I = d'/\mathrm{SNR}_I. \tag{28}$$

[27] As a mnemonic, think of the $\sim$ as a "wiggle" associated with randomness.

One conceptual problem with the Weber fraction J is that it is an *effective* noise–to–signal ratio, expressed in the Φ (physical) domain, but determined by a Ψ (psychophysical) domain mechanism (internal noise), as may be seen from Fig. 11.

Loudness JND ΔL: Any suprathreshold Ψ–domain increments may be quantified by corresponding Φ domain increments. The *loudness JND* $\Delta L(I)$ is defined as the change in loudness $L(I)$ corresponding to the intensity JND $\Delta I(I)$. While it is not possible to measure ΔL directly, we assume that we may expand the loudness function in a Taylor series, giving

$$L(I + \Delta I) = L(I) + \Delta I \frac{dL}{dI}\bigg|_I + \mathrm{HOT},$$

where HOT represents *higher–order terms*, which we shall ignore. If we solve for

$$\Delta L \equiv L(I + \Delta I) - L(I) \tag{29}$$

we find

$$\Delta L = \Delta I \frac{dL}{dI}\bigg|_I. \tag{30}$$

We call this expression the *small–JND* approximation. The above shows that the loudness JND $\Delta L(I)$ is related to the intensity JND $\Delta I(I)$ by the slope of the loudness function, evaluated at intensity I. According to the signal detection model, the standard deviation of the single-trial loudness is proportional to the loudness JND, namely

$$\Delta L = d'\sigma_L. \tag{31}$$

A more explicit way of expressing this assumption is

$$\frac{\Delta L}{\Delta I} = \frac{\sigma_L}{\sigma_I} \tag{32}$$

where we have assumed here that d' in both the Φ and Ψ domains is the same and thus cancels.

Loudness SNR: In a manner analogous to the Φ–domain SNR_I, we define the Ψ–domain loudness SNR as $\mathrm{SNR}_L(L) \equiv L/\sigma_L(L)$. Given Eq. 31, it follows that

$$\mathrm{SNR}_I = \nu \mathrm{SNR}_L, \tag{33}$$

where ν is the slope of the log–loudness function with respect to log–intensity. If we express the loudness as a power law

$$L(I) = I^\nu$$

and let $x = \log(I)$ and $y = \log(L)$, then $y = \nu x$. If the change of ν with respect to dB SPL is small, then $dy/dx \approx \Delta y/\Delta x \approx \nu$. Since $d\log(y) = dy/y$ we get

$$\Delta L/L = \nu \Delta I/I. \tag{34}$$

From Eq. 32, Eq. 33 follows.

Equation 33 is important because *(a)* it tells us how to relate the SNRs between the Φ and Ψ domains, *(b)* every term is dimensionless, *(c)* the equation is simple, since ν is approximately constant above 40 dB SL (i.e., Stevens' law), and because *(d)* we are used to seeing and thinking of loudness, intensity, and the SNR, on log scales, and ν as the slope on log-log scales.

Counting JNDs: While the concept of counting JNDs has been frequently discussed in the literature, starting with Fechner, unfortunately the actual counting formula (i.e., the equation) is rarely provided. As a result of a literature search, we found the formula in Nutting (1907), Fletcher (1923a), Wegel and Lane (1924), Riesz (1928), Fletcher (1929), and Miller (1947).

To derive the JND counting formula, Eq. 30 is rewritten as

$$\frac{dI}{\Delta I} = \frac{dL}{\Delta L}. \tag{35}$$

Integrating over an interval gives the total number of intensity JNDs

$$N_{12} \equiv \int_{I_1}^{I_2} \frac{dI}{\Delta I} = \int_{L_1}^{L_2} \frac{dL}{\Delta L}, \tag{36}$$

where $L_1 = L(I_1)$ and $L_2 = L(I_2)$. Each integral counts the total number of JNDs in a different way between I_1 and I_2 (Riesz, 1928; Fletcher, 1929). The number of JNDs must be the same regardless of the domain (i.e., the abscissa variable), Φ or Ψ.

3.2. Empirical models

This section reviews some earlier empirical models of the JND and its relation to loudness relevant to our development.

Weber's Law In 1846 it was suggested by Weber that $J(I)$ is independent of I. According to Eq. 22 and Eq. 26

$$J(I) = 2\alpha_* + \alpha_*^2.$$

If J is constant, then α_* must be constant, which we denote by $\alpha_*(\not{I})$ (we strike out I to indicate that α_* is not a function of intensity). This expectation, which is called Weber's law (Weber, 1988), has been successfully applied to many human perceptions. We refer the reader to the helpful and detailed review of these questions by Viemeister (1988), Johnson *et al.* (1993), and Moore (1982).

Somewhat frustrating is the empirical observation that $J(I)$ is not constant for the most elementary case of a pure tone (Riesz, 1928; Jesteadt *et al.*, 1977). This observation is referred to as *the near–miss to Weber's law* (McGill and Goldberg, 1968b). It remains unexplained why Weber's law holds as well as it does (Green, 1988, 1970, p. 721) (it holds approximately for the case of wide band noise), or even why it holds at all. Given the complex and NL nature of the transformation between the Φ and Ψ domains, coupled with the belief that the noise source is in the Ψ domain, it seems unreasonable that a law as simple as Weber's law, could hold in any general way. A transformation of the JND from the Φ domain to the Ψ domain might clarify the situation.

Weber's law does make one simple prediction that is potentially important. From Eq. 36 along with Weber's law $J_0 \equiv J(\not{I})$ we see that the formula for the number of JNDs is

$$N_{12} = \int_{I_1}^{I_2} \frac{dI}{J_0 I} \tag{37}$$

$$= \frac{1}{J_0} \ln(I_2/I_1). \tag{38}$$

Fechner's postulate In 1860 Fechner postulated that the loudness JND $\Delta L(I)$ is a constant[28] (Stevens, 1951; Fechner, 1966; Luce, 1993; Plack and Carlyon, 1995). We shall indicate such a constancy with respect to I as $\Delta L(\not{I})$ (as before, we strike out the I to indicate that ΔL is *not* a function of intensity). As first reported by Stevens (1961), we shall show that Fechner's postulate is not generally true.

The Weber–Fechner law. It is frequently stated (Luce, 1993) that Fechner's postulate ($\Delta L(\not{I})$) and Weber's law ($J_0 \equiv J(\not{I})$) lead to the conclusion that the difference in loudness between

[28] We are only considering the auditory case of Fechner's more general theory.

any two intensities I_1 and I_2 is proportional to the logarithm of the ratio of the two intensities, namely

$$\frac{L(I_2) - L(I_1)}{\Delta L} = \frac{1}{J_0} \log(I_2/I_1). \qquad (39)$$

This comes from Eq. 36 by assuming Weber's law and Fechner's Hypothesis. This result is called *Fechner's law* (also called the *Weber-Fechner law*). It is not true because of the false assumptions.

3.3. Models of the JND

Starting in 1923, Fletcher and Steinberg studied loudness coding of pure tones, noise, and speech (Fletcher, 1923a,b; Fletcher and Steinberg, 1924; Steinberg, 1925), and proposed that loudness was related to neural spike count (Fletcher and Munson, 1933), and even provided detailed estimates of the relation between the number of spikes and the loudness in sones (Fletcher, 1953b, Page 271). In 1943 De Vries first introduced a photon counting Poisson process model as a theoretical basis for the threshold of vision (De Vries, 1943). Siebert (1965) proposed that Poisson–point–process noise, resulting from the neural rate code, acts as the internal noise that limits the frequency JND (Green, 1970; Jesteadt *et al.*, 1977). A few years later (Siebert, 1968), and independently[29] McGill and Goldberg (1968a) proposed that the Poisson internal noise (PIN) model might account for the intensity JND, but they did not find this to produce a reasonable loudness growth function. Hellman and Hellman (1990) further refined the argument that Poisson noise may be used to relate the loudness growth to the intensity JND, and they found good agreement between the JND and realistic loudness functions.

Given Poisson Noise, the variance is equal to the mean, thus

$$\Delta L(L) \propto \sqrt{L}. \qquad (40)$$

This may be rewritten as $\sigma_L^2 \propto L$. We would expect this to hold if the assumptions of McGill and Goldberg (1968b) (i.e., the PIN model) are valid.

A direct estimate of $\Delta L(L)$ In the following we directly compare the loudness–growth function of Fletcher and Munson to the number of JNDs N_{12} from Riesz (Riesz, 1928; Allen and Neely, 1997). The Fletcher and Munson loudness data (Munson, 1932) were determined for long duration tonal stimuli using the loudness balance method (Fletcher and Munson, 1933), the method of *constant stimuli* (Yost, 1994), and the assumption of additivity of partial loudness. Riesz's data were also determined for long duration stimuli with just–detectable modulation (i.e., they were tone-like sounds). Since the JND depends on the modulation depth, as discussed in the *Definitions* section, Riesz's JND data seem to be ideal for this comparison since both the loudness data and the JND data have minimal (and similar) modulation parameters (Riesz's continuous tonal stimuli, which have just–detectable modulations, are more tone-like than gated 2AFC stimuli).

3.4. The direct estimate of ΔL

The above discussion has

(a) drawn out the fundamental nature of the JND,

(b) shown that the PIN loudness model holds below 5 sones (5,000 LU) (The solid line in the lower right panel of Fig. 11 below 5000 LU obeys the PIN model, and the data for both tones and wide band noise fall close to this line below 5000 LU)[30]

(c) shown that above 5 sones the PIN model fails and the loudness SNR remains constant.

Given its importance, it is reasonable to estimate ΔL directly from its definition Eq. 29, using Riesz's $\Delta I(I)$ and Fletcher and Munson's 1933 estimate of $L(I)$.

Miller's 1947 famous JND paper includes wide–band noise loudness–level results. We transformed these data to loudness using Fletcher and Munson (1933) reference curve (i.e., Fig. 12(b) upper left).

3.4.1. Loudness growth, recruitment and the OHC

In 1924 Fletcher and Steinberg published an important paper on the measurement of the loudness of speech signals (Fletcher and Steinberg, 1924). In this paper, when describing the growth of loudness, the authors state

[29] W. Siebert, personal communication.

[30] One sone is 975 LUs (Allen and Neely, 1997, page 3631), thus 5000 LUs = 5.13 LU. From the loudness scale this corresponds to a 1 kHz pure tone at 60 dB SL.

> the use of the above formula involved a *summation of the cube root of the energy rather than the energy*.

This cube–root dependence had first been described by Fletcher the year before (Fletcher, 1923a).

In 1930 Fletcher postulated that there was a monotonic relationship between central nerve firings rates and loudness. Given a tonal stimulus at the ear drum, Stevens' law says that the loudness is given by

$$L \equiv L(f, x, I) \propto I^{\nu}, \tag{41}$$

where $\{f, x, I\}$ are the frequency, place and intensity of the tone, respectively. The exponent ν has been experimentally established to be in the range between $1/4$ and $1/3$ for long duration pure tones at 1 kHz. Fletcher and Munson (1933) found $\nu \approx 1/4$ at high intensities and approximately 1 near threshold. Although apparently it has not been adequately documented, ν seems to be close to 1 for the *recruiting ear* (Neely and Allen, 1997).

Recruitment: What is the source of Fletcher's cube root loudness growth (i.e., Stevens' Law)? Today we know that cochlear outer hair cells (OHC) are the source of the cube root loudness growth observed by Fletcher.

From noise trauma experiments on animals and humans, we may conclude that recruitment (abnormal loudness growth) occurs in the cochlea (Carver, 1978; Gardner, 1994). Steinberg and Gardner (1937) described such a loss as a "variable loss" (i.e., sensory–neural loss) and partial recruitment as a mixed loss (i.e., having a conductive component) (Steinberg and Gardner, 1937, 1940). They and Fowler verified the conductive component by estimating the air–bone gap. In a comment to Fowler's original presentation on loudness recruitment in 1937, the famous anatomist Lorente de No theorized that recruitment is due to hair cell damage (Lorente de Nó, 1937). Steinberg and Gardner clearly understood recruitment, as is indicated in the following quote (Steinberg and Gardner, 1937, page 20)

> Owing to the expanding action of this type of loss it would be necessary to introduce a corresponding compression in the amplifier in order to produce the same amplification at all levels.

This compression/loss model of hearing and hearing loss, along with the loudness models of Fletcher and Munson (1933), are basic to an eventual quantitative understanding of NL cochlear signal processing and the cochlea's role in detection, masking and loudness in normal and impaired ears. The work by Fletcher (1950) and Steinberg and Gardner (1937), and work on modeling hearing loss and recruitment (Allen, 1991) support this view.

In summary, many studies agree: The cube–root loudness growth starts with the NL compression of basilar membrane motion due to stimulus dependent voltage changes within the OHC.

3.5. Determination of the loudness SNR

The pure-tone and wideband noise JND results may be summarized in terms of the loudness $\text{SNR}_L(I)$ data shown in Fig. 12(a) where we show $\Delta L/L = d'/\text{SNR}_L$, as a function of loudness. As before we separate frequencies into separate panels.

For noise below 55 dB SL the loudness signal–to–noise ratio $\text{SNR}_L \equiv L/\sigma_L$ increases as the cube–root (1 - 2/3 = 1/3) of the loudness; namely the noise increases by a factor of 2 when the loudness increases by a factor of 8. For levels above about 55 dB SL, $\text{SNR}_L(L)$ remains approximately constant with a value between 20 and 60 for both tones and noise.

To the extent that the curves are all approximately the same across frequency, Fig. 12(a) provides a stimulus independent description of the relation between the intensity JND and loudness. This invariance in SNR_L seems significant. Where the high level segment of SNR_L is constant, the intensity resolution of the auditory system has a fixed internal *relative* resolution (Ekman, 1959). The obvious interpretation is that as the intensity is increased from threshold, the neural rate–limited SNR increases until it saturates due to some *other* dynamic range limit, such as that due to some form of central nervous system (CNS) noise.

Near–miss to Stevens' law. In Fig. 12(a) we show a summary of $L(I)$, $\nu(I)$, $J(I)$ and $\Delta L/L = d'/\text{SNR}_L$ for the tone and noise data. For tones the intensity exponent $\nu(I)$ varies systematically between 0.3 and 0.4 above 50 dB SL, as shown by the solid line in the upper–right panel. We have highlighted this change in the power law with intensity

for a 1 kHz tone in the upper–right panel with a light–solid straight line. It is logical to call this effect the *near–miss to Stevens' law*, since it cancels the near–miss to Weber's law, giving a constant relative loudness JND $\Delta L/L$ for tones.

In the lower–right panel we provide a functional summary of $\Delta L/L$ for both tones and noise with the light–solid line described by

$$\frac{\Delta L(L)}{L} = h\left[\min(L, L_0)\right]^{-1/2}, \qquad (42)$$

where $h = \sqrt{2}$ and $L_0 = 5000$ LU (≈ 5 sones). We call this relation the Saturated Poisson Internal Noise (SPIN) model. With these parameter values, Eq. 42 appears to be a lower bound on the relative loudness JND_L for both tones and noise.

In Fig. 12(a) the second top panel shows the exponent $\nu(I)$ [see Fig. 11], for both Fletcher and Munson's and Miller's loudness growth function. In the lower–left panel we see $\Delta I/I$ versus I for Miller's subjects, Miller's equation, and Riesz's JND equation. In second from left bottom panel we show the $\Delta L/L$ versus L for the noise and tones cases. From Eq. 34 $\Delta L/L = \nu(I)J(I)$. Note how the product of $\nu(I)$ and $J(I)$ is close to a constant for tones above 65 dB SL. This invariance justifies calling the variations in the power–law exponent $\nu(I)$ for tones the "near–miss to Stevens' law." For reference, 1 sone is 975 LU. The upper–left panel shows the Fletcher-Munson loudness data from their Table III (Fletcher and Munson, 1933). The upper–right panel is a plot of the slope of the loudness with respect to intensity (LU-cm/W). In the lower–left we show the relation between the SPIN–model (Eq. 45 with $h = 2.4$) relative JND (solid line), calculated from the Fletcher–Munson loudness data, and the measured relative JND obtained by (Riesz, 1928) at 1 kHz. We display both Riesz's formula (dashed line) and Riesz's raw data (circles), which may be found in Fletcher (1953, 1995). In the lower–right we compare the SPIN–model relative JND (Eq. 45, with $h = 3.0$), and the relative JND computed from the Jesteadt *et al.* formula (dashed line) and data from their Table B-I (circles). They measured the JND using pulsed tones for levels between 5 and 80 dB. For reference, 1 sone is 975 LU.

3.6. Weber–fraction formula

In this section we derive the relation between the Weber fraction $J(I)$ given the loudness $L(I)$ starting from the *small–JND approximation*

$$\Delta L = \Delta I L'(I), \qquad (43)$$

where $L'(I) \equiv dL/dI$. If we solve this equation for ΔI and divide by I we find

$$J(I) \equiv \frac{\Delta I}{I} = \frac{\Delta L}{IL'(I)}. \qquad (44)$$

Finally we substitute the SPIN model Eq. 42

$$J(I) = \frac{hL(I)}{IL'(I)}\left[\min(L(I), L_0)\right]^{-1/2} \qquad (45)$$

This formula is the same as that derived by Hellman and Hellman (1990) when $L \leq L_0$. In Fig. 12(b) we plot Eq. 45 in the lower two panels labeled "SPIN model." From the lower–left panel of this figure, $h = 2.4$ and $L_0 = 10,000$ LU. For levels between 0 and 100 dB SL, the SPIN model (solid curve) fit to Riesz's data and Riesz's formula is excellent. Over this 100 dB range the curve defined by the loudness function fits as well as the curve defined by Riesz's formula (Allen and Neely, 1997). The excellent fit gives us further confidence in the basic assumptions of the model.

In the lower–right panel we have superimposed the JND data of Jesteadt *et al.* (1977) with $h = 3$ and $L_0 = 10,000$ LU for comparison to Eq. 45. The Jesteadt *et al.* data were taken with gated stimuli (100% modulation) and 2AFC methods. It is expected that the experimental method would lead to a different value of h than the valued required for Riesz's data set. The discrepancy between 0 and 20 dB may be due to the 100% modulation for these stimuli. The fit from 20 to 80 dB SL is less than a 5% maximum error, and much less in terms of rms error. Note the similarity in slope between the model and the data.

4. DISCUSSION AND SUMMARY

Inspired by the Poisson internal noise (PIN) based theory of Hellman and Hellman (1990), we have developed a theoretical framework that can be used to

explore the relationship between the pure–tone loudness and the intensity JND. The basic idea is to combine Fletcher's neural excitation response pattern model of loudness with signal detection theory. We defined a random decision variable called the single–trial loudness. The *mean* of this random variable is the loudness, while its *standard deviation* is proportional to the loudness JND. We define the loudness signal–to–noise ratio SNR_L as the ratio of loudness (the signal) to standard deviation (a measure of the noise).

4.1. Model validation

To evaluate the model we have compared the loudness data of Fletcher and Munson (1933) with the intensity JND data of Riesz (1928), for tones. A similar comparison was made for noise using loudness and intensity JND data from Miller (1947). We were able to unify the tone and noise data by two equivalent methods. First, since the loudness SNR is proportional to the ratio of the loudness to the JND $L/\Delta L$, the SNR is also a piecewise power–law function we call the SPIN model. All the data are in excellent agreement with the SPIN model, providing support for the validity of this theory.

4.2. The noise model

The SPIN model. Equation 42 summarizes our results on the relative loudness JND for both tones and noise. Using this formula along with Eq. 33, the JND may be estimated for tones and noise once the loudness has been determined, by measurement, or by model. Fechner's postulate, that the loudness JND is constant, is not supported by our analysis, in agreement with Stevens (1961).

The PIN model. The success of the PIN model is consistent with the idea that pure–tone loudness is based on neural discharge rate. This theory should apply between threshold and moderate intensities (e.g., < 60 dB) for "frozen stimuli" where the JND is limited by internal noise.

CNS noise. Above 60 dB SL we find that the loudness signal–to–noise ratio saturated with a constant loudness SNR between 30 and 50 for both the tone and noise conditions, as summarized by Ekman's law (Ekman, 1959). We conclude that the Hellman and Hellman theory must be modified to work at these higher intensities.

Weber's law. It is significant that while both $J(I)$ and $\nu(I)$ vary with intensity, the product is constant above 60 dB SL. Given that $J = d'/\nu\mathrm{SNR}_L$, the saturation in SNR_L explains Weber's law for wideband signals (since ν and SNR_L for that case are constant) as well as the near–miss to Weber's law for tones, where ν is not constant (the near–miss to Stevens' law, Fig. 12(a)).

Generalization to other data. If $\sigma_L(L, \not{I})$ depends on L, and is independent of I, then the $\mathrm{SNR}_L(L)$ should not depend on the nature of the function $L(I)$ (i.e., it should be true for any $L(I)$). This prediction is supported by our analysis summarized by Eq. 42. It will be interesting to see how SNR_L depends on L and I for subjects having a hearing–loss induced recruitment, and how well this theory explains other data in the literature, such as loudness and JNDs with masking induced recruitment (Schlauch, Harvey, and Lanthier, 1995).

Conditions for model validity. To further test the SPIN model, several conditions must be met. First the loudness and the JND must have been measured under the same stimulus conditions. Second, the internal noise must be the dominant factor in determining the JND. This means that the stimuli must be frozen (or have significant duration and bandwidth), and the subjects well trained in the task. As the signal uncertainty begins to dominate the internal noise, as it does in the cases of roving the stimulus, the intensity JND will become independent of the loudness.

As discussed by Stevens and Davis (Stevens and Davis, 1938b, p. 141-143), JND data are quite sensitive to the modulation conditions. The Riesz (1928) and Munson (1932) data make an interesting comparison because they are taken under steady–state conditions and are long duration tonal signals. Both sets of experimental data (i.e., Riesz and Munson) were taken in the same laboratory within a few years of each other.[31] Riesz (1928) states that he used the same methods as Wegel and Lane (1924), and it is likely that Munson (1932) did as well.

Differences in the signal conditions are the most likely explanation for the differences observed in the

[31] In 1928 Wegel, Riesz, and Munson were all members of Fletcher's department.

intensity JND measurements of Riesz and Jesteadt shown in Fig. 12(b). One difference between the data of Riesz (1928) and Jesteadt *et al.* (1977) is that Riesz varied the amplitude of the tones in a sinusoidal manner with a small (i.e., just detectable) modulation index, while Jesteadt *et al.* alternated between two intervals of different amplitude, requiring that the tones be gated on and off (i.e., a 100% modulation index).

The neural response to transient portions of a stimulus is typically larger than the steady-state response (e.g., neural overshoot) and, therefore, may dominate the perception of stimuli with large abrupt changes in amplitude. The fact that the intensity JND is sensitive to the time interval between two tones of different amplitude (Stevens and Davis, 1938b) is another indication that neural overshoot may play a role.

It would be interesting to check the SPIN model on loudness and JND data taken using gated signals, given the observed sensitivity to the modulation. While these JND data are available (Jesteadt, Wier, and Green, 1977), one would need loudness data taken with identical (or at least similar) modulations. We are not aware of such data.

References

Abbas, P. and Sachs, M. (**Jan. 1976**), "Two–tone suppression in auditory–nerve fibers: Extension of a stimulus–response relationship," J. Acoust. Soc. Am. **59**(1), 112–122.

Allen, J. B. (**1977**), "Cochlear micromechanics - A mechanism for transforming mechanical to neural tuning within the cochlea," J. Acoust. Soc. Am. **62**, 930–939.

Allen, J. B. (**1980**), "Cochlear micromechanics: A physical model of transduction," J. Acoust. Soc. Am. **68**(6), 1660–1670.

Allen, J. B. (**1991**), "Modeling the noise damaged cochlea," in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Springer-Verlag, New York), pp. 324–332.

Allen, J. B. (**1995**), "Harvey Fletcher 1884–1981," in *The ASA edition of Speech, Hearing in Communication*, edited by J. B. Allen (Acoustical Society of America, Woodbury, New York), pp. A1–A34.

Allen, J. B. (**1996**a), "DeRecruitment by multiband compression in hearing aids," in *Psychoacoustics, speech, and hearing aids*, edited by B. Kollmeier (World Scientific Press, Singapore), pp. 141–152.

Allen, J. B. (**Apr. 1996**b), "Harvey Fletcher's role in the creation of communication acoustics," J. Acoust. Soc. Am. **99**(4), 1825–1839.

Allen, J. B. (**1997**a), "OHCs shift the excitation pattern via BM tension," in *Diversity in auditory mechanics*, edited by E. Lewis, G. Long, R. Lyon, P. Narins, C. Steele, and E. Hecht-Poinar (World Scientific Press, Singapore), pp. 167–175.

Allen, J. B. (**Sep. 1997**b), "A Short History of Telephone Psychophysics," J. Audio Eng. Soc. **Reprint 4636**, 1–37.

Allen, J. B. (**1999**a), "Derecruitment by Multiband compression in hearing aids," in *The Efferent Auditory System*, edited by C. Berlin (Singular, 401 West A St., Suite 325, San Diego, CA 92101), chap. 4, pp. 73–86, includes a CDROM video talk by J. B. Allen in MP3 format.

Allen, J. B. (**1999**b), "Psychoacoustics," in *Wiley Encyclopedia of Electrical and Electronics Engineering*, edited by J. Webster (John Wiley & Sons, Inc, New York, NY), vol. 17, pp. 422–437.

Allen, J. B. (**2001**), "Nonlinear Cochlear Signal Processing," in *Physiology of the Ear, Second Edition*, edited by A. Jahn and J. Santos-Sacchi (Singular Thomson Learning, 401 West A Street, Suite 325 San Diego, CA 92101), chap. 19, pp. 393–442.

Allen, J. B. and Fahey, P. F. (**1983**), "Nonlinear behavior at threshold determined in the auditory canal on the auditory nerve," in *Hearing – Physiological bases and psychophysics*, edited by R. Klinke and R. Hartmann (Springer-Verlag, Bad Nauheim, Germany), pp. 128–134.

Allen, J. B. and Fahey, P. F. (**Jul. 1992**), "Using acoustic distortion products to measure the cochlear amplifier gain on the basilar membrane," J. Acoust. Soc. Am. **92**(1), 178–188.

Allen, J. B., Jeng, P. S., and Levitt, H. (**Jul 2005**), "Evaluating Human Middle Ear Function via an Acoustic Power Assessment," Jol. of Rehabil. Res. Dev. **42**(4), 63–78.

Allen, J. B. and Lonsbury-Martin, B. L. (**Jan. 1993**), "Otoacoustic emissions," J. Acoust. Soc. Am. **93**(1), 568–569.

Allen, J. B. and Neely, S. (**Jul. 1992**), "Micromechanical models of the cochlea," Physics Today **45**(7), 40–47.

Allen, J. B. and Neely, S. T. (**Dec. 1997**), "Modeling the relation between the intensity JND and loudness for pure tones and wide–band noise," J. Acoust. Soc. Am. **102**(6), 3628–3646.

Allen, J. B. and Sen, D. (**1999**), "Is tectorial membrane filtering required to explain two tone suppression and the upward spread of masking?" in *Recent Developments in Auditory Mechanics*, edited by H. Wada, T. Takasaka, K. Kieda, K. Ohyama, and T. Koike (World Scientific Publishing Co., PO Box 128, Farrer Road, Singapore 912805), pp. 137–143.

Arthur, R., Pfeiffer, R., and Suga, N. (**1971**), "Properties of "two-tone inhibition" in primary auditory neurons," J. Physiology (London) **212**, 593–609.

Ashmore, J. (**1987**), "A fast motile response in guinea-pig outer hair cells: the molecular basis of the cochlear amplifier," Journal of Physiology (London) **388**, 323–347.

Boring, E. G. (**1929**), *History of Psychophysics* (Appleton–Century, New York).

Brownell, W., Bader, C., Bertran, D., and de Rabaupierre, Y. (**1985**), "Evoked mechanical responses of isolated cochlear outer hair cells," Science **227**, 194–196.

Campbell, G. (**1903**), "On Loaded Lines in Telephonic Transmission," Phil. Mag. **5**, 313–331, see Campbell22a (footnote 2): In that discussion, ... it is tacitly assumed that the line is either an actual line with resistance, or of the limit such with R=0.

Campbell, G. (**Nov. 1922**), "Physical Theory of the Electric Wave Filter," Bell System Tech. Jol. **1**(1), 1–32.

Campbell, G. A. (**Jan. 1910**), "Telephonic Intelligibility," Phil. Mag. **19**(6), 152–9.

Carver, W. F. (**1978**), "Loudness balance procedures," in *Handbook of Clinical Audiology, 2^d edition*, edited by J. Katz (Williams and Wilkins, Baltimore MD), chap. 15, pp. 164–178.

Dallos, P. (**1996**), "Cochlear Neurobiology," in *The cochlea*, edited by P. Dallos, A. Popper, and R. Fay (Springer, New York), pp. 186–257.

Dallos, P. (**Feb. 2002**), "Prestin and the electromechanical reponses of outer hair cells," ARO-2002 **25**, 189.

Dallos, P., He, D. Z., Lin, X., Sziklai, I., Mehta, S., and Evans, B. N. (**Mar. 1997**), "Acetylcholine, Outer Hair Cell Electromotility, and the Cochlear Amplifier," Journal of Neuroscience **17**(6), 2212–2226.

De Vries, H. (**1943**), "The quantum character of light and its bearing upon the threshold of vision, the differential sensitivity and the acuity of the eye," Physica **10**, 553–564.

deBoer, E. (**1996**), "Mechanics of the cochlea: modeling efforts," in *The cochlea* (Springer, New York), pp. 258–317.

deBoer, E. and Nuttall, A. (**1999**), "The "inverse problem" solved for a three-dimensional model of the cochlear. III Brushing-up the solution method," J. Acoust. Soc. Am. **105**(6), 3410–3420, need exact reference.

deBoer, E. and Nuttall, A. (**2000**a), "The mechanical waveform of the basilar membrane. II. From data to models–and back," J. Acoust. Soc. Am. **107**(3), 1487–1496.

deBoer, E. and Nuttall, A. (**2000**b), "The mechanical waveform of the basilar membrane. III. Intensity effects," J. Acoust. Soc. Am. **107**(3), 1496–1507.

Delgutte, B. (**1990**a), "Physiological mechanisms of psychophysical masking: Observations from auditory-nerve fibers," J. Acoust. Soc. Am. **87**, 791–809.

Delgutte, B. (**1990**b), "Two-tone suppression in auditory-nerve fibres: Dependence on suppressor frequency and level," Hearing Res. **49**, 225–246.

Delgutte, B. (**1995**), "Physiological models for basic auditory percepts," in *Auditory Computation*, edited by H. Hawkins and T. McMullen (Springer Verlag, New York).

Diependaal, R., de Boer, E., and Viergever, M. (**1987**), "Cochlear power flux as an indicator of mechanical activity," J. Acoust. Soc. Am. **82**, 917–926.

Duifhuis, H. (**1973**), "Consequences of peripheral frequency selectivity for nonsimultaneous masking," J. Acoust. Soc. Am. **54**(6), 1472–1488, nice description of of AM Mayer's work.

Egan, J. and Hake, H. (**1950**), "On the masking pattern of a simple auditory stimulus," J. Acoust. Soc. Am. **22**, 662–630.

Ekman, G. (**1959**), "Weber's law and related functions," Psychology **47**, 343–352.

Fahey, P. F. and Allen, J. B. (**Feb. 1985**), "Nonlinear phenomena as observed in the ear canal, and at the auditory nerve," J. Acoust. Soc. Am. **77**(2), 599–612.

Fahey, P. F. and Allen, J. B. (**Nov. 1997**), "Measurement of distortion product phase in the ear canal of cat," J. Acoust. Soc. Am. **102**(5), 2880–2891.

Fechner, G. (**1966**), "Translation of: Elemente der Psychophysik," in *Elements of Psychophysics, Volume I*, edited by H. Adler (Holt, Rinehart, and Winston (Breitkopf and Hartel, Leipzig, 1860), New York).

Fletcher, H. (**Jun. 1922**), "The nature of speech and its interpretation," J. Franklin Inst. **193**(6), 729–747.

Fletcher, H. (**Sep. 1923**a), "Physical Measurements of Audition and their Bearing on the Theory of Hearing," J. Franklin Inst. **196**(3), 289–326.

Fletcher, H. (**Oct. 1923**b), "Physical Measurements of Audition and their Bearing on the Theory of Hearing," Bell System Tech. Jol. **ii**(4), 145–180.

Fletcher, H. (**1929**), *Speech and Hearing* (D. Van Nostrand Company, Inc., New York).

Fletcher, H. (**Apr. 1930**), "A space-time pattern theory of hearing," J. Acoust. Soc. Am. **1**(1), 311–343.

Fletcher, H. (**Apr. 1938**), "Loudness, Masking and Their Relation to the Hearing Process and the Problem of Noise Measurement," J. Acoust. Soc. Am. **9**, 275–293.

Fletcher, H. (**May 1940**), "Auditory Patterns," Reviews of Modern Physics **12**(1), 47–65.

Fletcher, H. (**Jan. 1950**), "A method of Calculating Hearing Loss for Speech from an Audiogram," J. Acoust. Soc. Am. **22**, 1–5.

Fletcher, H. (**Dec. 1951**a), "Acoustics," Phys Today **4**, 12–18.

Fletcher, H. (**Nov. 1951**b), "On the dynamics of the cochlea," J. Acoust. Soc. Am. **23**, 637–645.

Fletcher, H. (**1953**a), *Speech and Hearing in Communication* (Robert E. Krieger Publishing Company, Huntington, New York).

Fletcher, H. (**1953**b), *Speech and Hearing in Communication* (Robert E. Krieger Publishing Company, Huntington, New York).

Fletcher, H. (**1995**), "Speech and Hearing in Communication," in *The ASA edition of Speech and Hearing in Communication*, edited by J. B. Allen (Acoustical Society of America, Suite 1N01, 2 Huntington Quadrangle, Melville, New York).

Fletcher, H. and Munson, W. (**1933**), "Loudness, Its Definition, Measurement, and Calculation," J. Acoust. Soc. Am. **5**, 82–108.

Fletcher, H. and Munson, W. (**1937**), "Relation Between Loudness and Masking," J. Acoust. Soc. Am. **9**, 1–10.

Fletcher, H. and Steinberg, J. (**Sep. 1924**), "The dependence of the loudness of a complex sound upon the energy in the various frequency regions of the sound," Phy. Rev. **24**(3), 306–317.

Fletcher, H. and Steinberg, J. (**Jan. 1930**), "Articulation testing methods," J. Acoust. Soc. Am. **1**(2.2), 17–113 (1+), intelligibility Lists pages 65-113 (47+).

Fowler, E. (**1936**), "A method for the early detection of otosclerosis," Archives of Otolaryngology **24**(6), 731–741.

Gardner, M. (**1994**), Personal communication.

Geisler, C. D. and Nuttall, A. L. (**jul 1997**), "Two-tone suppression of basilar membrane vibrations in the base of the guinea pig cochlea using "low-side" suppressors," J. Acoust. Soc. Am. **102**(1), 430–440.

Geisler, D. C. (**1998**), *From Sound to Synapse: Physiology of the Mammalian Ear* (Oxford University Press).

Gescheider, G. (**1997**), *Psychophysics: The Fundamentals, 3d edition* (Lawrence Erlbaum Associates, Mahwah, NJ; London).

Goldstein, J. L. and Kiang, N. (**1968**), "Neural correlates of the aural combination tone 2f1-f2," Proc. of the IEEE **56**, 981–992.

Green, D. (**May 1970**), "Application of detection theory in psychophysics," Proceedings of the IEEE **58**(5), 713–723.

Green, D. (**1988**), "Audition: Psychophysics and Perception," in *Stevens' Handbook of Experimental Psychology*, edited by R. Atkinson, R. Herrnstein, G. Lindzey, and R. Luce (John Wiley & Sons, Inc., New York), chap. 6, pp. 327–376.

Green, D. M. and Swets, J. A. (**1966**), *Signal Detection Theory and Psychophysics* (John Wiley and Sons, INC., New York).

Greenwood, D. D. (**April 1961**a), "Auditory Masking and the Critical Band," J. Acoust. Soc. Am. **33**(4), 484–502.

Greenwood, D. D. (**Oct. 1961**b), "Critical bandwidth and the frequency coordinates of the basilar membrane," J. Acoust. Soc. Am. **33**, 1344–1356.

Hall, J. (**1974**), "Two-tone distortion products in a nonlinear model of the basilar membrane," J. Acoust. Soc. Am. **56**, 1818–1828.

Hartmann, W. M. (**1997**), *Signals, Sound, and Sensation* (AIP Press, American Institute of Physics, Woodbury, NY).

He, D. and Dallos, P. (**Jul. 1999**), "Somatic stiffness of cochlear outer hair cells is voltage-dependent," Proc. Nat. Acad. Sci. **96**(14), 8223–8228.

He, D. and Dallos, P. (**Jul. 2000**), "Properties of Voltage-Dependent Somatic Stiffness of Cochlear Outer Hair Cells," J. of the Assoc. for Res. in Otolaryngology **1**(1), 64–81.

Hecht, S. (**1934**), "Vision II. The nature of the photoreceptor process," in *Handbook of General Experimental Psychology*, edited by C. Murchison (Clark University Press, Worcester, MA).

Hellman, W. and Hellman, R. (**Mar. 1990**), "Intensity discrimination as the driving force for loudness. Application to pure tones in quiet," J. Acoust. Soc. Am. **87**(3), 1255–1271.

Helmholtz, H. L. F. (**1857**), *Helmholtz's popular scientific lectures* (Dover (1962), New York).

Helmholtz, H. L. F. (**1863**), *On the sensations of tone* (Dover (1954), New York).

Hewitt, M. and Meddis, R. (**aug 1991**), "An evaluation of eight computer models of mammalian inner hair-cell function," J. Acoust. Soc. Am. **90**(2), 904–917.

Hudspeth, A. and Corey, D. (**Jun. 1977**), "Sensitivity, polarity, and conductance change in the response of vertebrate hair cells to controlled mechanical stimuli," Proc. Nat. Acad. Sci. **74**(6), 2407–2411.

Iwasa, K. and Chadwick, R. (**Dec. 1992**), "Elasticity and Active Force Generation of Cochlear Outer Hair Cells," J. Acoust. Soc. Am. **92**(6), 3169–3173.

Jesteadt, W., Wier, C., and Green, D. (**Jan. 1977**), "Intensity discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**(1), 169–177.

Johnson, J., Turner, C., Zwislocki, J., and Margolis, R. (**Feb. 1993**), "Just noticeable differences for intensity and their relation to loudness," J. Acoust. Soc. Am. **93**(2), 983–991.

Kanis, L. and de Boer, E. (**Oct. 1994**), "Two-tone Suppression in a Locally Active Nonlinear Model of the Cochlea," J. Acoust. Soc. Am. **96**(4), 2156–2165.

Kemp, D. (**1978**), "Stimulated acoustic emissions from within the human auditory system," J. Acoust. Soc. Am. **64**, 1386–1391.

Kemp, D. (**1979**a), "Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea," Archives of Oto-Rhino-Laryngology **224**, 37–45.

Kemp, D. (**1979**b), "The evoked cochlear mechanical response and the auditory microstructure – evidence for a new element in cochlear mechanics," in *Models of the auditory system and related signal processing techniques* (Scandinavian Audiology, Supplementum 9), 9, pp. 35–47.

Kemp, D. (**1980**), "Towards a model for the origin of cochlear echoes," Hearing Resh. **2**, 533–548.

Kemp, D. (**1986**), "Otoacoustic emissions, travelling waves and cochlear mechanisms." J. Acoust. Soc. Am. **22**, 95–104.

Kiang, N.-S. and Moxon, E. (**1974**), "Tails of tuning curves of auditory-nerve fibers," J. Acoust. Soc. Am. **55**, 620–630.

Kim, D., Neely, S., Molnar, C., and Matthews, J. (**1980**), "An active cochlear model with negative damping in the cochlear partition: Comparison with Rhode's ante- and post-mortem results," in *Psychological, Physiological and Behavioral Studies in Hearing*, edited by G. van den Brink and F. Bilsen (Delft Univ. Press, Delft, The Netherlands), pp. 7–14.

Kim, D., Siegel, J., and Molnar, C. (**1979**), "Cochlear Nonlinear Phenomena in Two-Tone Responses," in *Scandinavian Audiology, Supplementum 9*, edited by M. Hoke and E. DeBoer, pp. 63–82.

Liberman, M. (**Nov. 1982**a), "The cochlear frequency map for the cat: Labeling auditory-nerve fibers of known characteristic frequency," J. Acoust. Soc. Am. **72**(5), 1441–1449.

Liberman, M. (**Nov. 1982**b), "The cochlear frequency map for the cat: Labeling auditory-nerve fibers of known characteristic frequency," J. Acoust. Soc. Am. **72**(5), 1441–1449.

Liberman, M. (**1982**c), "Single-neuron labeling in the cat auditory nerve," Science **216**, 163–176.

Liberman, M. and Dodds, L. (**1984**), "Single Neuron Labeling and Chronic Cochlear Pathology III: Stereocilia Damage and Alterations of Threshold Tuning Curves," Hearing Resh. **16**, 55–74.

Littler, T. S. (**1965**), *The physics of the ear* (Pergamon Press, Oxford, England).

Lorente de Nó, R. (**1937**), "The diagnosis of diseases of the neural mechanism of hearing by the aid of sounds well above threshold," Transactions of the American Otological Society **27**, 219–220.

Luce, R. (**1993**), *Sound and Hearing* (Lawrence Earlbaum Associates, Hilldale, NJ).

Mayer, A. M. (**1876**), "Research in acoustics," Philosophy Magazine **2**, 500–507, in Benchmark Papers in Acoustics, vol. 13, Earl D. Schubert, Ed.

McFadden, D. (**1986**), "The curious half–octave shift: Evidence of a basalward migration of the traveling–wave envelope with increasing intensity," in *Applied and Basic Aspects of Noise-Induced Hearing Loss*, edited by R. Salvi, D. Henderson, R. Hamernik, and V. Coletti (Plenum), pp. 295–312.

McGill, W. J. and Goldberg, J. P. (**1968**a), "Pure–tone intensity discrimination as energy detection," J. Acoust. Soc. Am. **44**, 576–581.

McGill, W. J. and Goldberg, J. P. (**1968**b), "A study of the near–miss involving Weber's law and pure tone intensity discrimination," Perception and Psychophysics **4**, 105–109.

Miller, G. A. (**1947**), "Sensitivity to changes in the intensity of white noise and its relation to masking and loudness," J. Acoust. Soc. Am. **19**, 609–619.

Montgomery, H. C. (**1935**), "Influence of experimental technique on the measurement of differential intensty sensitivity of the ear," J. Acoust. Soc. Am. **7**, 39–43.

Moore, B. C. J. (**1982**), *An introduction to the psychology of hearing (second edition)* (Academic Press, London, New York).

Munson, W. (**Jul. 1947**), "The growth of auditory sensation," J. Acoust. Soc. Am. **19**, 584–591.

Munson, W. A. (**1932**), "An Experimental determination of the equivalent loudness of pure tones," J. Acoust. Soc. Am. **4**(7), ABSTRACT.

Munson, W. A. and Gardner, M. B. (**Mar. 1950**), "Loudness patterns–a new approach," J. Acoust. Soc. Am. **22**(2), 177–190.

Narayan, S., Temchin, A., Recio, A., and Ruggero, M. (**Dec. 1998**), "Frequency tuning of basilar membrane and auditory nerve fibers in the same cochleae," Science **282**, 1882–1884, shows BM and neural differ by 3.8 dB/oct.

Neely, S. (**1992**), "A model of cochlear mechanics with outer hair cell motility," J. Acoust. Soc. Am. .

Neely, S. and Kim, D. O. (**1983**), "An active cochlear model showing sharp tuning and high sensitivity," Hearing Res. **9**, 123–130.

Neely, S. and Kim, D. O. (**1986**), "A model for active elements in cochlear biomechanics," J. Acoust. Soc. Am. **79**, 1472–1480.

Neely, S. T. and Allen, J. B. (**1997**), "Relation between the Rate of growth of loudness and the intensity DL," in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt and *et al.* (Lawrence Erlbaum Assoc., Mahwah, NJ), pp. 213–222.

Nutting, P. G. (**1907**), "The complete form of Fechner's Law," Bulletin of the Bureau of Standards **3**(1), 59–64.

Pang, X. and Guinan, J. (**Dec. 1997**), "Growth rate of simultaneous masking in cat auditory-nerve fibers: Relationship to the growth of basilar-membrane motion and the origin of two-tone suppression," J. Acoust. Soc. Am. **102**(6), 3564–3574, beautiful data showing the slope of suppression for low frequency suppressors.

Peterson, L. C. and Bogert, B. P. (**1950**), "A Dynamical Theory of the Cochlea," J. Acoust. Soc. Am. **22**, 369–381.

Pickles, J. O. (**1982**), *An introduction to the physiology of hearing* (Academic Press Inc., London, England).

Pipes, L. A. (**1958**), *Applied Mathematics for Engineers and Physicists* (McGraw Hill, NYC, NY).

Plack, C. and Carlyon, R. (**1995**), "Loudness perception and intensity coding," in *Hearing, Handbook of Perception and Cognition*, edited by B. Moore (Academic Press, San Diego), chap. 4, pp. 123–160.

Puria, S. and Allen, J. B. (**Dec. 1998**), "Measurements and model of the cat middle ear: Evidence for tympanic membrane acoustic delay," J. Acoust. Soc. Am. **104**(6), 3463–3481.

Raab, D. and Goldberg, I. (**Feb. 1975**), "Auditory intensity discrimination with bursts of reproducible noise," J. Acoust. Soc. Am. **57**(2), 473–447.

Ranke, O. (**1950**), "Theory operation of the cochlea: A contribution to the hydrodynamics of the cochlea," J. Acoust. Soc. Am. **22**, 772–777.

Recio, A. and Rhode, W. (**Nov. 2000**), "Basilar membrane responses to broadband stimuli," J. Acoust. Soc. Am. **108**(5), 2281–2298.

Régnier, M. and Allen, J. B. (**2007**a), "The importance of across-frequency timing coincidences in the perception of some English consonants in noise," in *Abst.* (ARO, Denver).

Régnier, M. and Allen, J. B. (**2007**b), "Perceptual cues of some CV sounds studied in noise," in *Abst.* (AAS, Scottsdale).

Relkin, E. and Turner, C. (**Aug. 1988**), "A reexamination of forward masking in the auditory nerve," J. Acoust. Soc. Am. **84**(2), 584–591.

Rhode, W. (**1971**), "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique," J. Acoust. Soc. Am. **49**, 1218–1231.

Rhode, W. (**Jun. 1978**), "Some observations on cochlear mechanics," J. Acoust. Soc. Am. **64**, 158–176.

Riesz, R. R. (**1928**), "Differential intensity sensitivity of the ear for pure tones," Phy. Rev. **31**(2), 867–875.

Russell, I., Legan, P., Lukashkina, V., Lukashkin, A., Goodyear, R., and Richardson, G. (**Jan. 2007**), "Sharpened cochlear tuning in a mouse with a genetically modified tectorial membrane," Nature Neuroscience , 1–9.

Russell, I., Richardson, and Cody (**May 1986**), "Mechanosensitivity of mammalian auditory hair cells *in vitro*," Nature **321**(29), 517–519.

Russell, I. and Sellick, P. (**1978**), ""Intracellular studies of hair cells in the mammalian cochlea"," J. Physiol. **284**, 261–290.

Sachs, M. B. and Kiang, Y. S. (**1968**), "Two-Tone inhibition in auditory-nerve fibers," J. Acoust. Soc. Am. **43**, 1120–1128.

Santos-Sacchi, J. (**Oct. 1991**), "Reversible inhibition of voltage–dependent outer hair cell motility and capacitance," J. of Neurosci. **11**(10), 3096–3110.

Schlauch, R., Harvey, S., and Lanthier, N. (**Oct. 1995**), "Intensity resolution and loudness in broadband noise," J. Acoust. Soc. Am. **98**(4), 1895–1902.

Sellick, P. and Russell, I. (**1978**), "Intracellular studies of cochlear hair cells: Filling the gap between basilar membrane mechanics and neural excitation," in *Evoked electrical activity in the auditory nervous system*, edited by F. Naunton and C. Fernandez (Academic Press, New York), pp. 113–140.

Sen, D. and Allen, J. B. (**Apr. 2006**), "Functionality of cochlear micromechanics–as elucidated by the upward spread of masking and two tone suppression," Acoustics Australia **34**(1), 43–51.

Sewell, W. (**1984**), "The effects of furosemide on the endocochlear potential and auditory-nerve fiber tuning curves in cats," Hearing Res. **14**, 305–314.

Shera, C. and Guinan, J. (**Feb. 2007**), "Cochlear traveling-wave amplification, suppression, and beamforming probed using noninvasive caliration of intracochlear distortion sources," J. Acoust. Soc. Am. **121**(2), 1003–1016.

Siebert, W. (**1965**), "Some implications of the stochastic behavior of primary auditory neurons," Kybernetik **2**, 205–215.

Siebert, W. (**1968**), "Stimulus transformations in the peripheral auditory system," in *Recognizing patterns*, edited by P. Kolers and M. Eden (MIT Press, Cambridge, MA), chap. 4, pp. 104–133.

Smoorenburg, G. (**1972**), "Combination tones and their origin," J. Acoust. Soc. Am. **52**(2), 615–632.

Steinberg, J. (**1925**), "The Loudness of a Sound and Its Physical Stimulus," Physical Review **26**, 507.

Steinberg, J. (**1937**), "Positions of stimulation in the cochlea by pure tones," Journal of the Acoustical Society of America **8**, 176–180, cochlear map estimate; Monograph B-973.

Steinberg, J. (**Oct. 1941**), "Stereophonic sound-film system-pre- and post-equalization of compandor systems," Journal of the Acoustical Society of America **13**, 107–114, b-1327.

Steinberg, J. and Gardner, M. (**Jul. 1937**), "Dependence of hearing impairment on sound intensity," Journal of the Acoustical Society of America **9**, 11–23.

Steinberg, J. C. and Gardner, M. B. (**Jan. 1940**), "On the auditory significance of the term hearing loss," J. Acoust. Soc. Am. **11**, 270–277.

Stevens, S. (**1951**), "Mathematics, Measurement, and Psychophysics," in *Handbook of Experimental Psychology*, edited by S. Stevens (John Wiley & Sons, Inc., New York), chap. 1, pp. 1–49.

Stevens, S. (**1961**), "To honor Fechner and repeal his law," Science .

Stevens, S. and Davis, H. (**1938**a), *Hearing, Its Psychology and Physiology* (Republished by the the Acoustical Society of America in 1983, Woodbury, New York).

Stevens, S. and Davis, H. (**1938**b), *Hearing, Its Psychology and Physiology* (The Acoustical Society of America, Woodbury, New York).

Strope, B. and Alwan, A. (**Sep. 1997**), "A model of Dynamic Auditory Perception and its application to robust word recognition," IEEE Trans. Acoust. Speech and Sig. Processing **5**(5), 451–464.

Titchener, E. (**1923**), *Experimental Psychology, A Manual of Laboratory Practice, Vol. II* (The Macmillan Co., London).

Viemeister, N. F. (**1988**), "Psychophysical aspects of auditory intensity coding," in *Auditory Function*, edited by G. Edelman, W. Gall, and W. Cowan (Wiley, New York), chap. 7, pp. 213–241.

Weber, E. H. (**1988**), "Der Tastsinn und das Gemainfül," in *Handwörterbuch der Physiologie*, edited by R. Wagner (Vieweg, Braunschweig), vol. 3, chap. 7, pp. 481–588.

Wegel, R. L. and Lane, C. (**Feb. 1924**), "The Auditory Masking of One Pure Tone by Another and Its Probable Relation to the Dynamics of the Inner Ear," Physical Review **23**, 266–285.

Yost, W. A. (**1994**), *Fundamentals of Hearing, An Introduction* (Academic Press, San Diego, London).

Yost, W. A. (**2006**), *Fundamentals of Hearing, An Introduction* (Academic Press, San Diego, London).

Zweig, G. (**1991**), "Finding the impedance of the organ of Corti," J. Acoust. Soc. Am. **89**, 1229–1254.

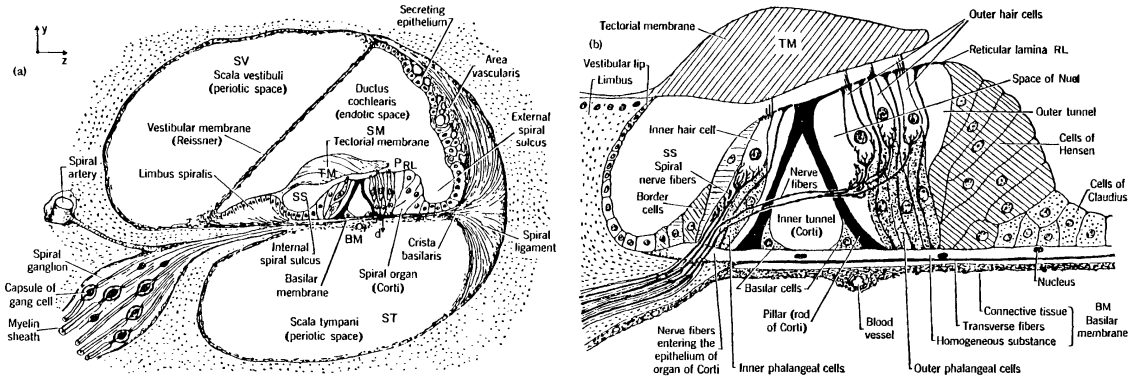
Zwislocki, J. (**1948**), "Theorie der Schneckenmechanik," Acta Otolaryngol. [supl.] , 72.

Zwislocki, J. (**1950**), "Theory of the acoustical action of the cochlea," J. Acoust. Soc. Am. **22**, 779–784.

Zwislocki, J. and Jordan, H. (**1986**), "On the relation of intensity JNDs to loudness and neural noise," J. Acoust. Soc. Am. **79**, 772–780.

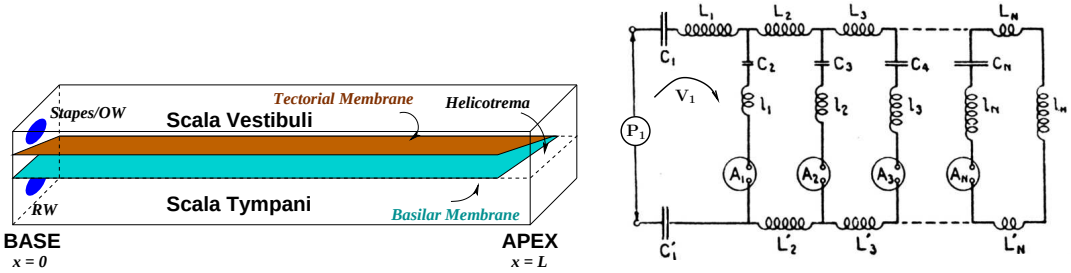
Contents

List of Figures



(a) Cross-section through the cochlear duct (b) Cross-section through the Organ of Corti.

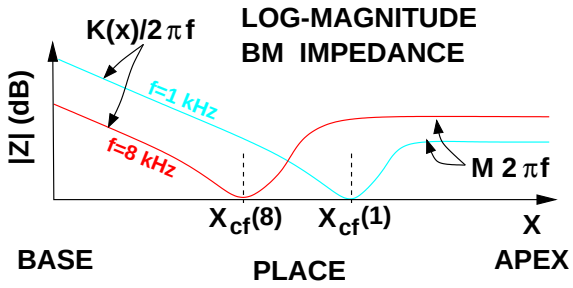
Figure 1: *On the left we see all the major structures of the cochlea. The three chambers are filled with fluid. Reissner's membrane is an electrical barrier and is not believed to play a mechanical role. The right panel shows the inner and outer hair cells, pillar cells and other supporting structures, the basilar membrane (BM), and the tectorial membrane (TM).*



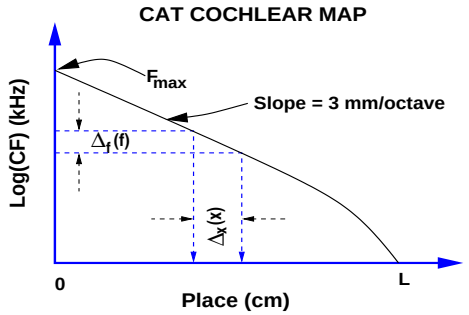
(a) The *Base* (x=0) is the high frequency end of the cochlea while the *Apex* ($x = L$) carries the low frequencies.

(b) The model is built from a cascade of electrical section.

Figure 2: *On the left se see the basic 2-D box-model of the cochlea, and on the right the 1924 Wegel and Lane electrical equivalent circuit.*

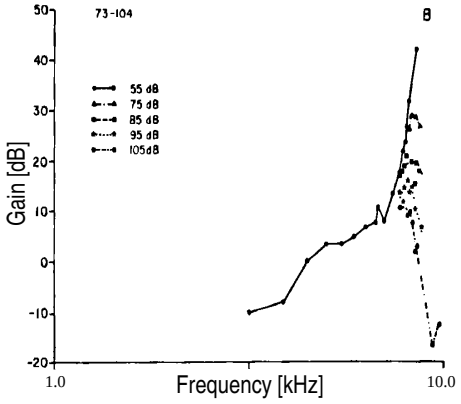


(a) Plot of the log-magnitude of the impedance as a function of place for two different frequencies of 1 and 8 kHz.

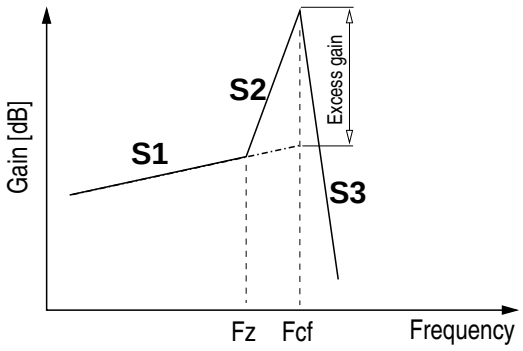


(b) Cochlear map of the cat following Liberman and Dodds.

Figure 3: *On the left showing the impedance, the region labeled $K(X)$ is the region dominated by the stiffness and has impedance $K(X)/s$. The region labeled M is dominated by the mass and has impedance sM. The characteristic places for 1 and 8 kHz are shown as X_{cf}. The resonance frequency depends on place according to the* cochlear map function*, as shown by the plot on the left. A* critical bandwidth $\Delta_f(f)$ *and a* critical spread $\Delta_x(X)$ *area related through the cochlear map.*



(a) This panel shows a reproduction of Figure 9a panel B from Rhode (1978), showing the response of the basilar membrane for his most sensitive animal. The graduals along the abscissa are at 0.1, 1.0 and 10.0 kHz.



(b) Basic definition of the 6 parameters for characterizing a tuning curve: slopes S_1, S_2, S_3, frequencies F_z and F_{cf}, and the Excess-gain.

Figure 4: *There are 6 numbers that characterize every curve, three slopes (S_1, S_2, S_3), in dB/oct, two frequencies (F_z, F_{cf}), and the* Excess-gain *characterizes the amount of gain at F_{cf} relative to the gain defined by S_1. The Excess-gain depends on the input level for the case of a nonlinear response like the cochlea. Rhode found up to ≈35 dB of excess gain at 7.4 kHz and 55 dB SPL, relative to the gain at 105 dB SPL. From of the 55 dB SPL curve of Fig. 4(a) (the most sensitive case), and his Table I, $S_1 = 9$, $S_2 = 86$, and $S_3 = -288$ (dB/oct)), F_z = 5 kHz, F_{cf} = 7.4 kHz, and an* excess gain *of 27 dB. Rhode reported S_1 = 6 dB/oct, but 9 seems to be a better fit to the data, so 9 dB/oct is the value we have used for our comparisons.*

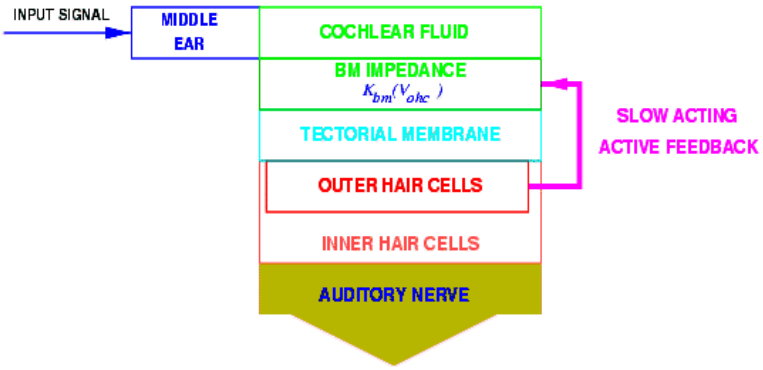
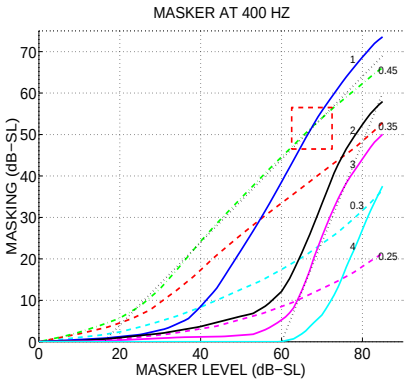
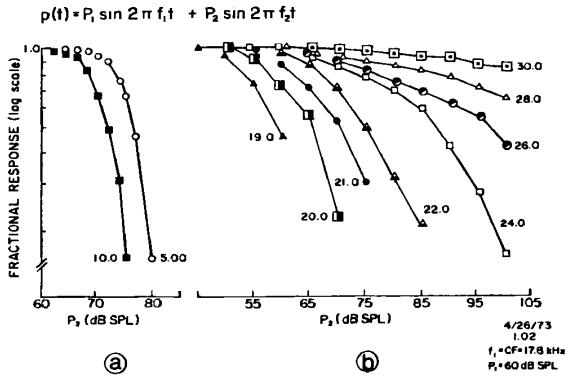


Figure 5: *Block flow diagram of the inner ear (Allen, 1997a).*

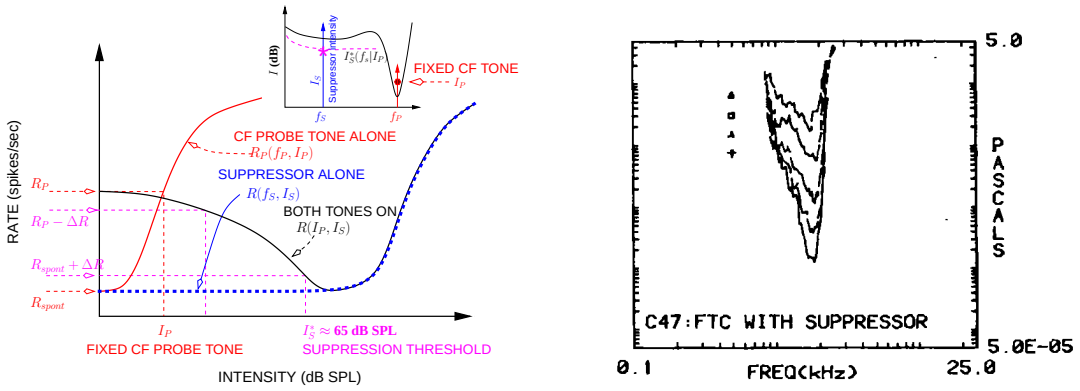


(a) Upward spread of masking as characterized by Wegel and Lane in 1924. The solid lines correspond to the probe being higher than the 400 Hz masker, while the dashed lines correspond to the 400 Hz probe lower than the masker. On the left we see upward spread of masking functions from Wegel and Lane for a 400 Hz low frequency masker. The abscissa is the masker intensity I_m in dB-SL while the ordinate is the threshold probe intensity $I_p^*(I_m)$ in dB-SL. The frequency of the probe f_p, expressed in kHz, is the parameter indicated on each curve. The dashed box shows that the masking due to a 1 kHz tone becomes more than that at 450 Hz, for a 400 Hz probe. This is the first observation of *excitation pattern migration* with input intensity.



(b) Two-tone suppression (2TS) IO functions from Fig. 8 of Abbas and Sachs (1976). On the left is low-side suppression and on the right we see high-side suppression. In 2TS the suppressor plays the role of the masker and the probe the role of the maskee. Note that the threshold of suppression for low-side suppressor (masker) is close to 70 dB SPL, which is similar to human low-side suppressors, the case of the Wegel and Lane USM (left) (60-70 dB-SPL). The onset of suppression for high-side suppressors is close to the neuron's CF threshold of 50 dB, as elaborated further in Fig. 7(a).

Figure 6: *On the left we see the psychoacoustic measure of 2TS, called the upward spread of masking. On the right are related measures taken in the auditory nerve by a procedure called two-tone suppression (2TS). Low-side and high-side masking or suppression have very different thresholds and slopes. These suppression slopes and thresholds are very similar between 2TS and the USM.*



(a) Definitions of 2TS low-side masking procedure (b) Example of 2TS (low-side masking) in the Cat AN.

Figure 7: *On the left are shown the definitions used in 2TS, while on the right we see 2TS in in a cat neural tuning curve. A cat neural tuning curve taken with various "low-side" suppressors present (suppressor below the best frequency), as indicated by the symbols. The tuning curve with the lowest threshold is for no suppressor. When the suppressor changes by 20 dB, the F_{cf} threshold changes by 36 dB. Thus for a 2 kHz neuron, the slope is 36/20, or 1.8. These numbers are similar to those measure by Delgutte (1990b). One Pascal = 94 dB SPL.*

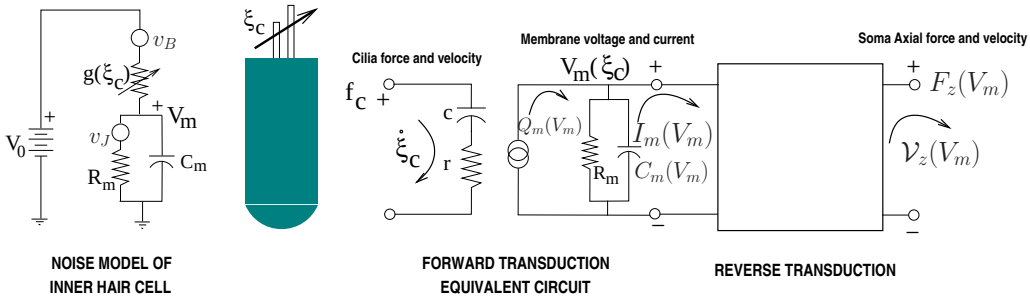
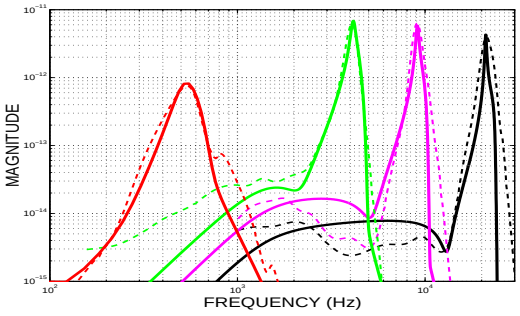
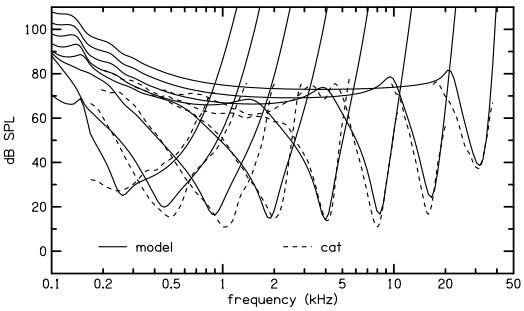


Figure 8: *On the far left is the electrical equivalent circuit model of an OHC with thermal noise sources due to the cell leakage resistance Johnson and shot noise v_J and the Brownian motion of the cilia, represented by the voltage noise source v_B. The cilia force f_c and velocity $\dot{\xi}_c$ are the stimulus (input) variables to the* FORWARD TRANSDUCTION*, and are loaded by the mechanical impedance of the cilia viscous drag r and compliance c. When the cilia move, current flows into the cell charging the membrane capacitance and thus changing the membrane voltage V_m. This membrane capacitance $C_m(V_m)$ is voltage dependent (i.e., it is NL). The membrane voltage has also been shown to control the cell's soma axial stiffness. It follows that the axial force $F_z(V_m)$ the cell can deliver, and the axial velocity $\mathcal{V}_z(V_m)$ of the cell, must also depend on the membrane voltage. The precise details of how all this works is unknown.*

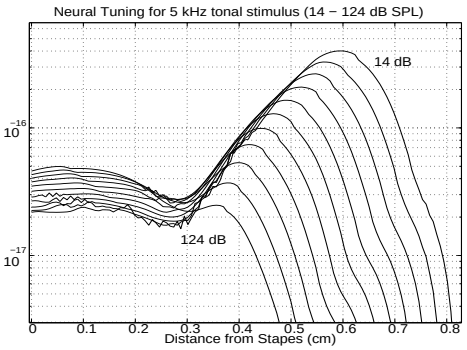


(a) Comparison between neural data and the computed model excitation patterns from Allen's passive RTM model (transfer function format). This CA model assumes an IHC cilia bundle displacement of about 50 pm at the neural rate threshold.

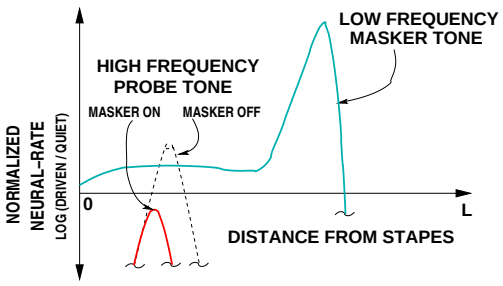


(b) Comparison between neural data computed tuning curves from Neely's active model (Neely, 1992). This CA model assumes an IHC cilia bundle displacement of 300 pm (0.3 nm) at the neural rate threshold.

Figure 9: *The tuning curves shown by the dashed lines are the average of single nerve fiber responses from six cats obtained by M. C. Liberman and B. Delgutte.*



(a) Compression in the NL-RTM model. Note how the response at the peak is reduced as the BM stiffness changes, causing the peak to shift to the base. As this happens the response in the tail region between $0 \leq X \leq 0.3$ cm becomes more sensitive, and thus shows an expansive NL response. All of these effects have been seen in real BM data.



(b) Cartoon showing the effect of a low-side masker on a high frequency tone as a function of position along the basilar membrane. When the suppressor is turned on the CF of the high-frequency probe becomes less sensitive and shifts to higher frequencies. We model this effect in the panel on the left as BM stiffness that depends on level (i.e., $K_p(I_s)$).

Figure 10: *In (a) results of model calculations by Sen and Allen (2006) are shown of a NL BM stiffness model. On the right shows a cartoon of what might happen to the excitation pattern of a low-level probe when a suppressor is turned on given such a nonlinearity. The presence of the suppressor causes the probe to be suppressed and shifted slightly toward the base when the stiffness is decreased with increased level. It may be inferred from Fig. 3(a) that if the BM stiffness is reduced the location of the maximum will shift to the base, as is seen in real data.*

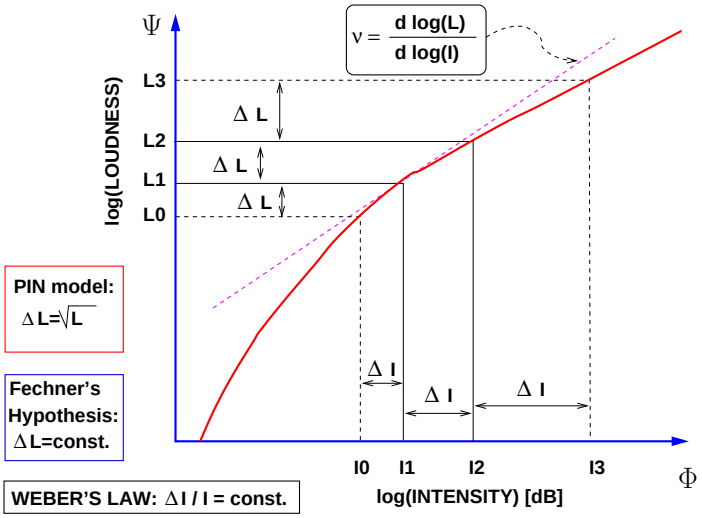
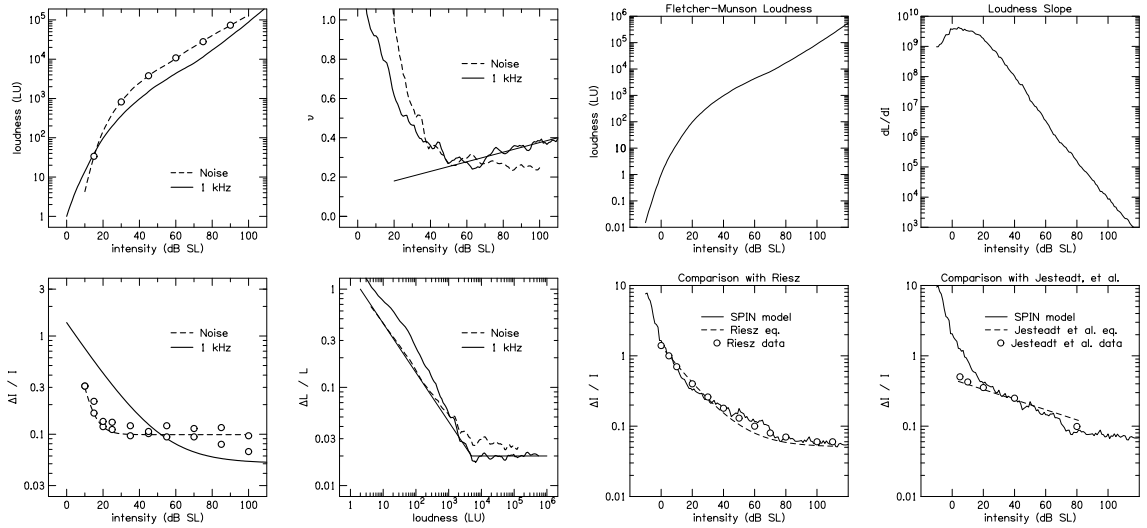


Figure 11: *This figure summarizes all the historical ideas about psychophysics and the relations between the* Φ *and* Ψ *variables. Along the abscissa we have the physical variable, intensity, and along the ordinate, the psychological variable loudness. The curve represents the loudness, on a log-intensity log-loudness set of scales. A JND in loudness is shown as* ΔL *and it depends on loudness, as described by the* Poisson internal noise *(PIN) model shown in the box on the left. Fechner assumed that* ΔL *was constant, which we now know to be incorrect. The loudness JND is reflected back into the physical domain as an intensity JND* ΔI*, which also depends on level. Weber's law, is therefore not true in general (but is approximately true for wide-band noise). Our analysis shows that the loudness SNR and the intensity SNR must be related by the slope of the loudness growth function, as given by Eq. 33. These relations are verified in Fig. 12(a), as discussed in detail in Allen and Neely (1997).*



(a) *The direct derivation of ΔL based on pure tone JND and loudness data data from Miller (1947); Riesz (1928); Fletcher and Munson (1933).*

(b) *Test of the model derived on the left based on a comparison between loudness data and intensity JND data at 1 kHz, using the SPIN model.*

Figure 12: *(a) In 1947 Miller measured the JND_I and the loudness–level for two subjects using wideband modulated noise (0.15–7 kHz) for levels between 3 and 100 dB SL. The noise (dash line) and pure tone (solid line) loudness are shown in the upper–left panel. The similarity between $\Delta L/L$ derived from the loudness curves for pure tones and for noise provide an almost perfect fit to the SPIN model. which results from assuming the noise is neural point-process noise. See the text for a summary of these results. (b) Test of the SPIN model against the classic results of Riesz (1928); Jesteadt* et al. *(1977).*