3. Nonlinear Cochlear Signal Processing and Masking in Speech Perception

J. B. Allen

There are many classes of masking, but two major classes are easily defined: *neural masking* and *dynamic masking*. Neural masking characterizes the internal noise associated with the neural representation of the auditory signal, a form of loudness noise. Dynamic masking is strictly cochlear, and is associated with cochlear outer-hair-cell processing. This form is responsible for dynamic nonlinear cochlear gain changes associated with sensorineural hearing loss, the upward spread of masking, two-tone suppression and forward masking. The impact of these various forms of masking are critical to our understanding of speech and music processing. In this review, the details of what we know about nonlinear cochlear and basilar membrane signal processing is reviewed, and the implications of neural masking is modeled, with a comprehensive historical review of the masking literature. This review is appropriate for a series of graduate lectures on nonlinear cochlear speech and music processing, from an auditory point of view.

3.1 Basics

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 $\mathrm{JND_I}$) and defined as the *just discriminable change in intensity*. The $\mathrm{JND_I}$ is a function of intensity I, frequency f and stimulus type T (e.g., noise, tones, speech, music, etc.). As an *internal noise*, the $\mathrm{JND_I}$ may be modeled in terms of a loudness (i. e., perceptual intensity) noise density along the length of the cochlea ($0 \le X \le L$), described in terms of a *partial loudness JND* ($\Delta\mathcal{L}(X, T)$, a.k.a. $\mathrm{JND}_{\mathcal{L}}$). The cochlea or inner ear is the organ that converts signals from acoustical to neural signals. 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 nonsimultaneous, 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 an 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 [3.1, 2], so the topic remains controversial. Understanding and modeling NL OHC processing is key to many speech processing applications. As a result, a vibrant research effort driven by the National Institute of Health 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 researchers at Bell Labs [3.3]), revolutionized the hearing-aid industry. With the introduction of compression signal processing, hearing aids now address the recruitment problem, thereby providing speech audibility over a much larger dynamic range, at least in quiet. 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. 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 improves 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 three years old, when they fail to learn to talk. If you cannot hear you do not learn to talk. 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. It is at least possible that this loss is due to the lack of NL OHC processing.

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.

In this chapter 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* [3.4], *Dallos* [3.5], *Yost* [3.6].

3.1.1 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 nine to

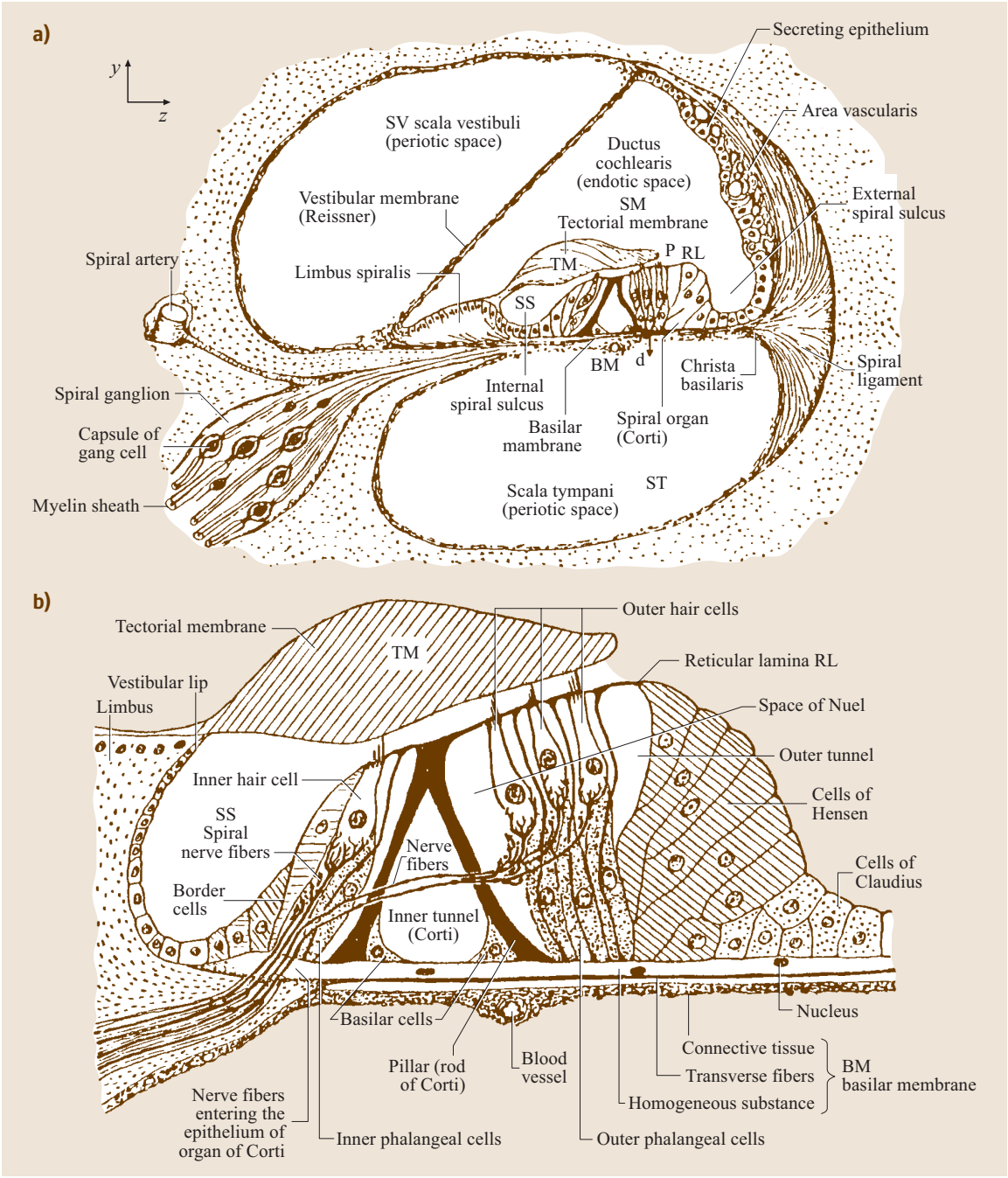


Fig. 3.1a,b On the left we see all the major structures of the cochlea **(a)**. The three chambers are filled with fluid. Reissner's membrane is an electrical barrier and is not believed to play a mechanical role. The right panel **(b)** shows the inner and outer hair cells, pillar cells and other supporting structures, the basilar membrane (BM), and the tectorial membrane (TM)

ten orders of magnitude of intensity [3.7, 8], while the ear has about 11 to 12. The stimulus has a relatively high information rate. Neurons are low-bandwidth 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. 3.1a) 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) [3.4, 5]. Figure 3.1b shows a blown-up view of the organ of Corti 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 TM shears against the reticular lamina (RL), causing the cilia of the inner and outer hair cells to bend. The afferent auditory nerve fibers that 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 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 [3.5, 9, 10]. The IHC excitation signal has a narrow bandwidth and a center frequency that depends on the inner hair cell's location along the basilar membrane. Each hair cell is about 10 μm in diameter while the human basilar membrane is about 35 mm in length (35 000 μm). Thus the neurons of the auditory nerve encode the responses of about 3500 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 [3.11]. Each hair cell is connected to many neurons, having a wide range of spontaneous firing rates and thresholds [3.12]. In the cat, for example, approximately 15–20 neurons encode each of these narrow band inner hair cells with a neural timing code. 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). 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. 3.1b there are typically three (occasionally four) 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. 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 1930s and 1940s. Telephone speech is similarly compressed [3.13] via μ-law coding. It is well known (as was first proposed by *Lorente de Nó* [3.14] and *Steinberg* [3.3]) that noise damage of *nerve cells* (i. e., OHCs) leads to a reduction of dynamic range, a disorder clinically named *loudness recruitment*. The word *recruitment*, which describes the abnormal growth of loudness in the impaired ear, is a seriously misleading term, since nothing is being recruited [3.15].

We may describe cochlear processing two ways: first in terms of the signal representation at various points in the system; and 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 that make excellent supplemental reading are available [3.4, 8, 16, 17].

For pedagogical purposes this chapter has been divided into four parts. Besides this introduction, we include sections on the NL cochlea, neural masking, and finally a brief discussion. Section 3.2 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 3.3 discusses neural masking, the JND, loudness recruitment, the loudness signal-to-noise ratio

(SNR), and the Weber fraction. Section 3.4 provides a brief summary.

3.1.2 History of Cochlear Modeling

Typically the cochlea is treated as an uncoiled long thin box, as shown in Fig. 3.2a. This represents the starting point for the macromechanical models.

Macromechanics

In his book *On the Sensations of Tone Helmholtz* [3.18] likened the cochlea to a bank of highly tuned resonators selective to different frequencies, much like a piano or a harp [3.19, p. 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 Helmholtz shows deep insight and his studies provided many very important contributions.

The next major contribution by *Wegel* and *Lane* [3.20] stands in a class of its own even today, as a double-barreled paper having both deep psychophysical and modeling insight. Fletcher published much of the Wegel and Lane data one year earlier [3.21]. 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. 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. 3.2b. 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. 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 [3.22, 23], which had been used for speech articulation studies [3.24–26], and Fletcher and Wegel were fully utilizing Campbell's important discoveries.

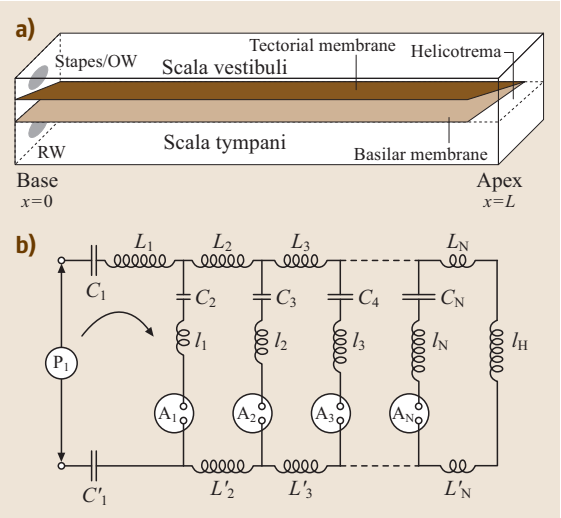


Fig. 3.2a,b On the left (a) see the basic 2-D box model of the cochlea. The *Base* ($x = 0$) is the high-frequency end of the cochlea while the *Apex* ($x = L$) carries the low frequencies. On the right (b) the 1924 Wegel and Lane electrical equivalent circuit. The model is built from a cascade of electrical sections

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 for a pure tone input the traveling wave has a more sharply defined location on the basilar membrane 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).

The One-Dimensional Model of the Cochlea

To describe this response it is helpful to call upon the macromechanical *transmission line models* of *Wegel* [3.20] (Fig. 3.2b) and *Fletcher* [3.27], first quantitatively analyzed by *Zwislocki* [3.28, 29], *Ranke* [3.30], *Peterson* and *Bogert* [3.31], *Fletcher* [3.32, 33]. This popular transmission line model is now denoted the *one-dimensional* (1-D), or *long-wave* model.

Zwislocki [3.28] 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 capacitors. 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. 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. 3.2b, 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.

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

$$\text{Impedance} = \frac{\text{effort}}{\text{flow}} \,. \tag{3.1}$$

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 (i. e., an impedance), *phasor* notation is very useful, where the tone is represented as the real part (Re) of the complex exponential

$$\mathrm{e}^{\mathrm{i}2\pi ft+\mathrm{i}\phi} \equiv \cos(2\pi ft+\phi)+\mathrm{i}\sin(2\pi ft+\phi) \,. \tag{3.2}$$

The symbol $\equiv$ denotes *equivalence*. It means that the quantity to the left of $\equiv$ is defined by the quantity on the right. More specifically, impedance is typically defined in the frequency domain using *Laplace transform* notation, in terms of a damped tone

$$A\,\mathrm{e}^{\sigma t}\cos(2\pi ft+\phi) \equiv A\,\mathrm{Re}\,\mathrm{e}^{st+\mathrm{i}\phi} \tag{3.3}$$

excitation, characterized by the tone's amplitude A, phase ϕ and *complex Laplace frequency* $s \equiv \sigma + \mathrm{i}2\pi f$. 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.

Example 3.1: 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) [3.34, 35]. 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 3.2: The impedance of a spring is given by the ratio of the force $F(f)$ to velocity $V(f) = sX(f)$ with displacement X

$$Z(s) \equiv \frac{F}{V} = \frac{K}{s} = \frac{1}{sC} \,, \tag{3.4}$$

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 (as parallel lines in Fig. 3.2b). Having $s = \sigma + \mathrm{i}2\pi f$ in the denominator indicates that the impedance of a spring has a phase of $-\pi/2$ (e.g., $-90°$). Such a phase means that when the velocity is $\cos(2\pi ft)$, the force is $\sin(2\pi ft)$. This follows from Hooke's law

$$F = KX = \frac{K}{s}sX = \frac{K}{s}V \,. \tag{3.5}$$

Example 3.3: 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 $\mathrm{d}v(t)/\mathrm{d}t$). The electrical element corresponding to a mass is an *inductor,* indicated in Fig. 3.2b by a coil. Thus for a mass $Z(s) = sM$.

From these 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° phase is called a *lagging* phase, while the mass with its +90° 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. 3.2b. The position X along the model is called the *place* variable and corresponds to the longitudinal position along the cochlea. The series (horizontal) inductors (coils) denoted by L_k represent the fluid mass (inertia) along the 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

$$Z_{\mathrm{p}}(s, X) = \frac{K_{\mathrm{p}}(X)}{s} + R_{\mathrm{p}}(X) + sM_{\mathrm{p}} \,, \tag{3.6}$$

where $K(X)$ is the partition stiffness, and R_{p} is the partition resistance. Each inductor going to ground (l_i in Fig. 3.2b) represents the partition plus fluid mass per unit length M_{p} of the section. Note that sM, R_{p} and K/s are impedances, but the mass M and stiffness K 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. 3.3a, for a given input frequency the BM impedance magnitude has a local minimum at the shunt resonant frequency, where the membrane that can move in a relatively unrestricted manner. The shunt *resonance* has special significance because at this resonance frequency $F_{\mathrm{cf}}(X)$ 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 $F_{\mathrm{cf}}(X)$

$$\frac{K_{\mathrm{p}}(X)}{2\pi \mathrm{i} F_{\mathrm{cf}}} + 2\pi \mathrm{i} F_{\mathrm{cf}} M_{\mathrm{p}} = 0 \tag{3.7}$$

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

$$F_{\mathrm{cf}}(X) \equiv \frac{1}{2\pi} \sqrt{\frac{K_{\mathrm{p}}(X)}{M_{\mathrm{p}}}} \,. \tag{3.8}$$

The inverse of this function specifies the location of the *hole* $X_{\mathrm{cf}}(f)$ as shown in Fig. 3.3a. In the example of Fig. 3.3a two frequencies are show, at 1 and 8 kHz, with corresponding resonant points shown by $X_{\mathrm{cf}}(1)$ and $X_{\mathrm{cf}}(8)$.

Basal to $X_{\mathrm{cf}}(f)$ in Fig. 3.3a, the basilar membrane is increasingly stiff, and apically (to the right of the resonant point), the impedance is mass dominated. The above description is dependent on the input frequency f since the location of the hole is frequency dependent. In this apical region the impedance has little influence since almost no fluid flows past the low-impedance hole. 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 end of the model (the helicotrema end, the apex of the cochlea).

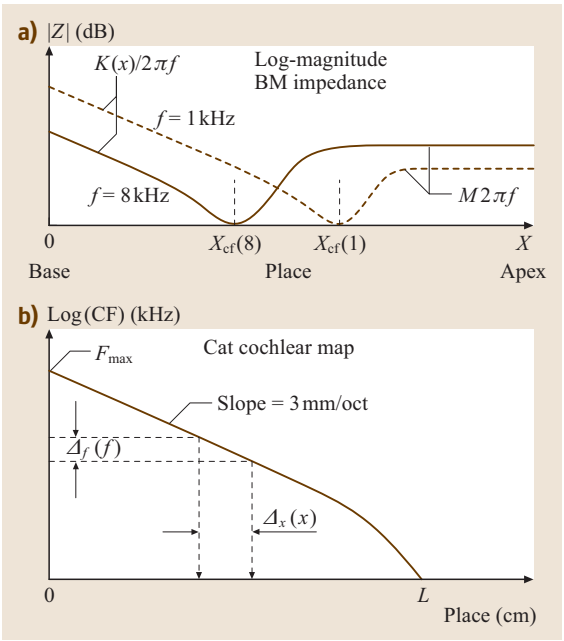


Fig. 3.3 **(a)** Plot of the log-magnitude of the impedance as a function of place for two different frequencies of 1 and 8 kHz showing the impedance; the region labeled $K(X)$ is the region dominated by the stiffness and has impedance $K(X)/s$. The region labeled M is dominated by the mass and has impedance sM. The characteristic places for 1 and 8 kHz are shown as X_{cf}. **(b)** Cochlear map of the cat following Liberman and Dodds. The resonance frequency depends on place according to the *cochlear map function* **(b)**. A *critical bandwidth* $\Delta_f(f)$ and a *critical spread* $\Delta_x(X)$ area related through the cochlear map

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 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 by $X_{\mathrm{cf}}^{(p)}(f)$. Likewise at a given place X as a function of frequency there is a local maximum called the *characteristic frequency*, denoted by $F_{\mathrm{cf}}^{(p)}(X)$. 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*. There is 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. 3.3b at the apex ($x = L$) may be a result of these conflicting definitions. The cochlear map function $F_{\mathrm{cf}}(X)$ plays a key role in cochlear mechanics, has a long history, and is known by many names [3.27, 36–40], the most common today being *Greenwood's function*. In the speech literature it is called the *Mel scale*.

The spread of the response around the peak for a fixed frequency is denoted the *critical spread* $\Delta_x(f)$, 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* [3.41]. 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 [3.42], namely

$$\Delta_f(X) \propto \kappa(X) \,. \tag{3.9}$$

In the next section we shall show how the the relations between 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* [3.10] and popularized by *Greenwood* [3.43]. 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 $F_{\mathrm{cf}}(X)$:

$$N_{\mathrm{cb}} \equiv \int_0^{X_{\mathrm{cf}}} \frac{\mathrm{d}X}{\Delta_x(X)} = \int_0^{F_{\mathrm{cf}}} \frac{\mathrm{d}f}{\Delta_f(f)} \,. \tag{3.10}$$

There are approximately 20 pure-tone frequency JNDs per critical band [3.37], [3.42, p. 171], and *Fletcher* showed that the *critical ratio* expressed in dB $\kappa_{\mathrm{dB}}(X)$ is of the form $aX + b$, where a and b are constants [3.10]. As verified by *Greenwood* [3.43, p. 1350, (1)] the critical bandwidth in Hz is therefore

$$\Delta_f(X) \propto 10^{\kappa_{\mathrm{dB}}(X)/10} \,. \tag{3.11}$$

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* showed that for the cat, $\Delta_x(X)$ corresponds to about 2.75 times the basilar membrane width with $W_{\mathrm{bm}}(X) \propto \mathrm{e}^X$ [3.10]. 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* [3.44] and *Liberman* and *Dodds* [3.45], and they showed the following empirical formula fit the data

$$F_{\mathrm{cf}}(X) = 456(10^{2.1(1-X/L)} - 0.8) \,, \tag{3.12}$$

where the length of the cat cochlea is $L = 21$ mm, and X is measured from the stapes [3.44]. 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 (3.12), and as defined in Fig. 3.3b, 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* [3.10] and *Hartmann* [3.17].

3.2 The Nonlinear Cochlea

3.2.1 Cochlear Modeling

In cochlear modeling there are two fundamental intertwined complex 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 gives insight into both the NL cochlea, as well has two-dimensional (2-D) model frequency-selective 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 [3.47, 48],
2. otoacoustic (ear canal) nonlinear emissions [3.49], and
3. motile outer hair cells [3.50].

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 such 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- and three-dimensional (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 response growth, and
3. improved experimental basilar membrane observations became available which showed increased nonlinear cochlear frequency selectivity.

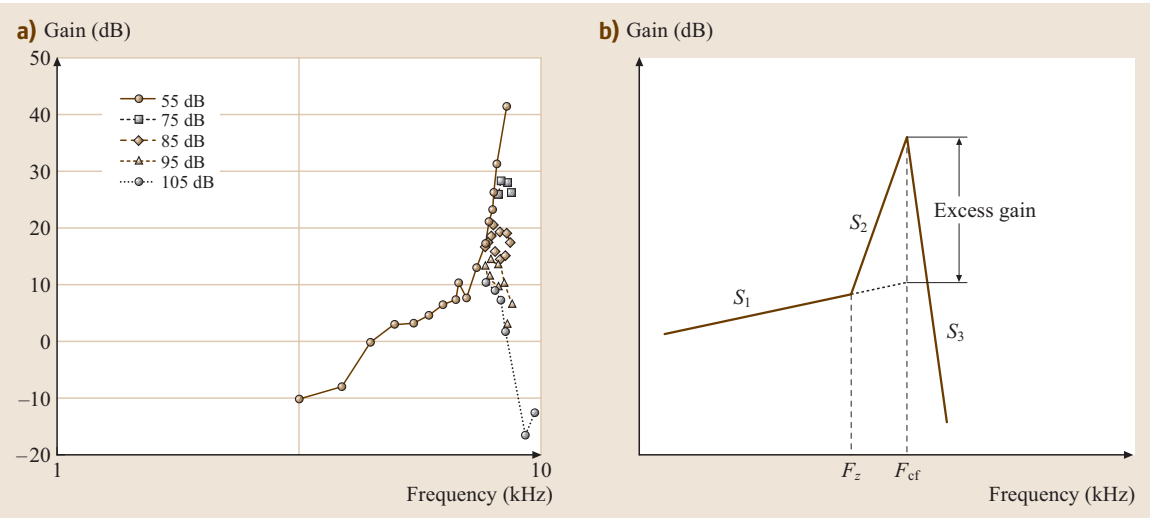


Fig. 3.4a,b There are six numbers that characterize every curve, three slopes (S_1, S_2, S_3), in dB/oct, two frequencies (F_z, F_{cf}), and the *excess gain* characterizes the amount of gain at F_{cf} relative to the gain defined by S_1. The excess gain depends on the input level for the case of a nonlinear 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 **(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. **(a)** Response of the basilar membrane for his most sensitive animal. The graduations along the abscissa are at 0.1, 1.0 and 10.0 kHz (after [3.46, Fig. 9a]) **(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

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 include:

- 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* [3.20], *Goldstein* and *Kiang* [3.52], *Smoorenburg* [3.53], *Kemp* [3.54], *Kim* et al. [3.55], *Fahey* and *Allen* [3.56] and many others,
- Normal loudness growth and recruitment in the impaired ear [3.3, 41],
- The frequency dependent neural two-tone suppression observed by *Sachs* and *Kiang* [3.57], *Arthur* et al. [3.58], *Kiang* and *Moxon* [3.59], *Abbas* and *Sachs* [3.60], *Fahey* and *Allen* [3.56], *Pang* and *Guinan* [3.61], and others,
- The frequency-dependent basilar membrane response-level compression first described by *Rhode* [3.46, 47],
- The frequency-dependent inner-hair-cell receptor potential level compression, first described by *Sellick* and *Russell* [3.48], *Russell* and *Sellick* [3.62].
- Forward masking data that shows a linear return to baseline after up to 0.2 s [3.63]. 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 these NL 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* [3.46, 47], as shown in Fig. 3.4a, showing that 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 (e.g., 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, when the animal was not in its physiologically pristine state.

An important and useful measure of cochlear linear and nonlinear response first proposed by *Rhode* [3.46, Fig. 8], is shown in Fig. 3.4b which 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) [3.49,54,64–66]. 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. [3.50] discovered that isolated OHCs change their length when placed in an electric field, thus that the outer hair cell is motile. 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. 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 [3.67]. The release time must be determined by the OHC membrane properties, which is slow relative to the attack. Thus OHC NL processing is the basis for both the frequency asymmetry of simultaneous (upward versus downward spread) and temporal (forward versus backward) masking.

As summarized in Fig. 3.5, OHCs provide feedback to the BM via the OHC receptor potential, which in turn is modulated by both the position of the basilar

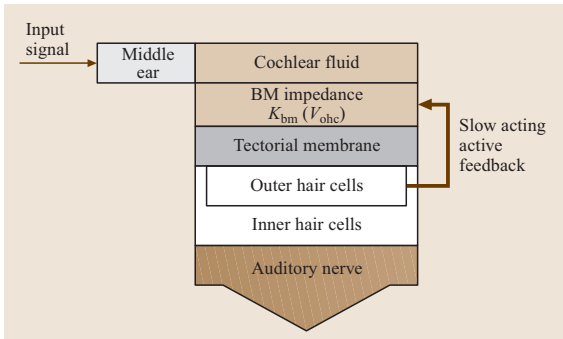


Fig. 3.5 Block flow diagram of the inner ear (after *Allen* [3.51])

membrane (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.

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, or 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* [3.68] in a comment by *R. Lorente de Nó* [3.14] stating that cochlear hair cells are likely to be involved in loudness recruitment.

The same year *Steinberg* and *Gardner* [3.3] 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 clear sense that cochlear hair cells were related to dynamic range compression.

More recently, 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* [3.69] concluded:

> *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 [3.70, 71]. 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 the block diagram feedback model of the organ of Corti shown in Fig. 3.5 where the excitation to the OHC changes the cell voltage V_{ohc}, which in turn changes the basilar stiffness [3.51]. This is one of several 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 [3.63].

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. Historically these two measures have 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* [3.72] was the first to describe the asymmetric nature of masking [3.63, 73]. 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) [3.63].

In 1923, Fletcher published the first quantitative results of tonal masking. In 1924, Wegel and Lane extended Fletcher's experiments (Fletcher was the subject [3.27, p. 325]) 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. 3.6a, Wegel and Lane's experiments involved presenting listeners with a masker tone at frequency $f_m = 400\,\text{Hz}$ 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

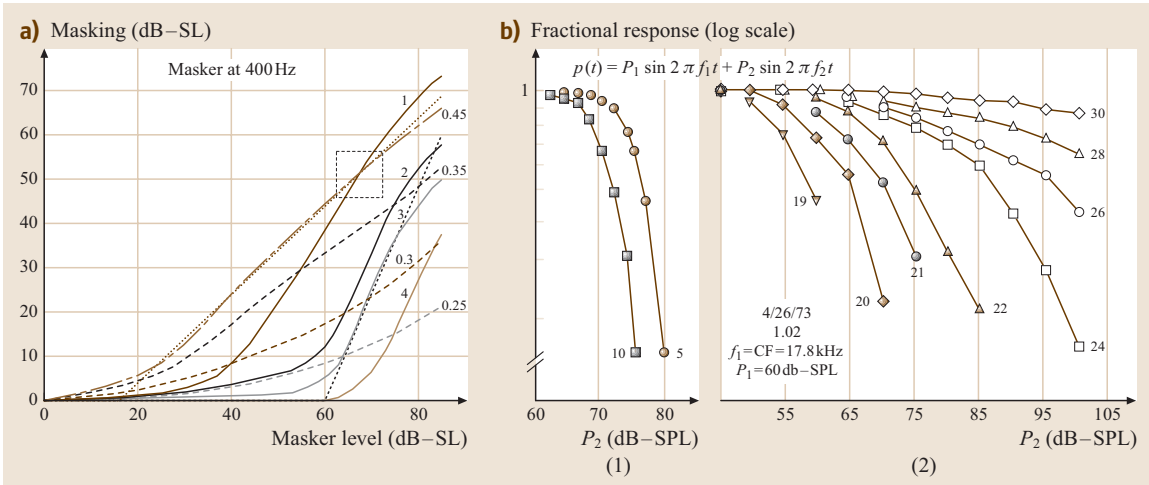


Fig. 3.6a,b On the left **(a)** we see the psychoacoustic measure of 2TS, called the upward spread of masking. On the right **(b)** are related measures taken in the auditory nerve by a procedure called two-tone suppression (2TS). Low- and high-side masking or suppression have very different thresholds and slopes. These suppression slopes and thresholds are very similar between 2TS and the USM. **(a)** 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) input–output (IO) functions from *Abbas* and *Sachs* [3.60, Fig. 8]. On the *left* (1) is low-side suppression and on the *right* (2) 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 (1) (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. 3.7a

is the probe level at threshold [dB − SL]). The asterisk indicates a threshold measure.

In Fig. 3.6a $f_m = 400$ 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) [3.74].

Critical band masking has a slope close to 1 dB/dB (the superimposed dotted line has a slope of 1). Four years later *Riesz* [3.75] shows critical band masking obeys the *near miss to Weber's law*, as described in Sect. 3.3.2. The downward spread of masking (the dashed lines in Fig. 3.6a) 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 − SL.

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

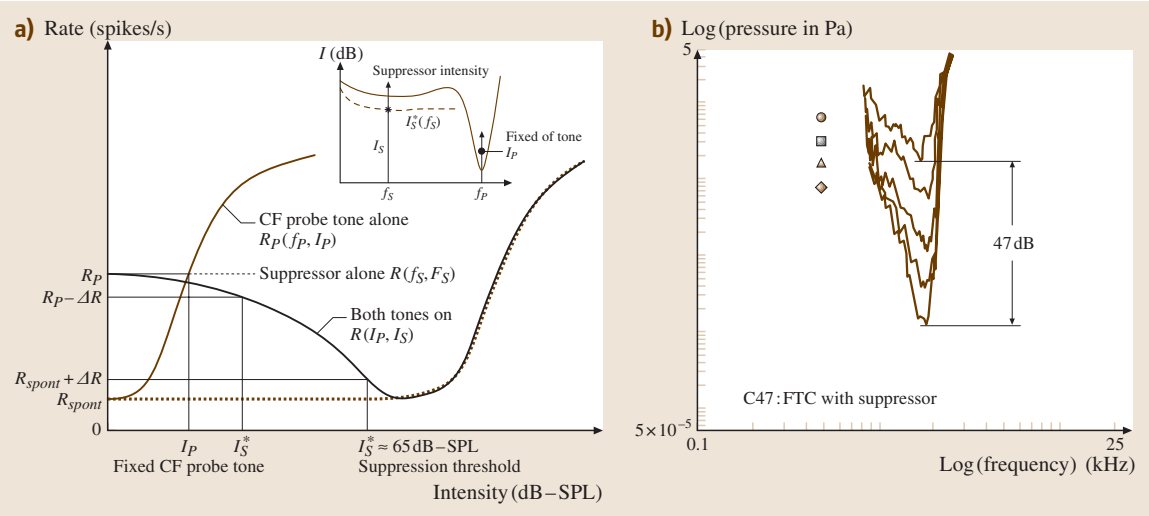


Fig. 3.7 **(a)** Definitions of 2TS low-side masking procedure (see (3.13) and (3.14)). **(b)** Example of 2TS (low-side masking) in the cat auditory nerve (AN). A cat neural tuning curve taken with various *low-side* suppressors present (suppressor below the best frequency), as indicated by the symbols. 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* [3.80]. One Pa = 94 dB − SPL

stimulus intensity, described in a wonderful paper by *McFadden* [3.76]. Response migration was also observed by *Munson* and *Gardner* in a classic paper on forward masking [3.77]. This important migration effect is beyond the scope of the present discussion, but is reviewed in [3.74, 78, 79] discussed in the caption of Fig. 3.10.

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 signals, and is known as two-tone suppression (2TS) in the auditory nerve literature, as discussed in the caption of Fig. 3.6b.

Two-Tone Suppression. The neural correlate of the psychophysically measured USM is called *two-tone suppression* (2TS). As shown in the insert of Fig. 3.7a, 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

1. the rate response to either the probe alone $R(I_p, I_s = 0)$ decreases by a small increment Δ_R, or
2. drops to the small increment Δ_R, just above the undriven spontaneous rate $R(0, 0)$.

These two criteria are defined in Fig. 3.6b and may be written

$$R_p\left(I_p, I_s^*\right) \equiv R(I_p, 0) - \Delta_R \tag{3.13}$$

and

$$R_{spont}\left(I_p, I_s^*\right) \equiv R(0, 0) + \Delta_R\,; \tag{3.14}$$

Δ_R indicates a fixed small but statistically significant constant change in the rate (e.g., $\Delta_R = 20$ spikes/s is a typical value). The threshold suppressor intensity is defined as $I_s^*(f_s)$, and as before the $*$ indicates the threshold suppressor intensity. The two threshold definitions (3.13) and (3.14) are very different, and both are useful. The difference in intensity between the two thresholds is quite large, and the more common measure used by *Abbas* and *Sachs* [3.60] is (3.13). The second measure (3.14) is consistent with neural tuning curve suppression, and is therefore the more interesting of the two. It corresponds to suppression of the probe to threshold.

Neural data of *Abbas* and *Sachs* [3.60, Fig. 8] are reproduced in Fig. 3.6b. 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 is has a low spontaneous rate and a 50–55 dB threshold. The left panel of Fig. 3.6b is for apical suppressors that are lower in frequency than the characteristic frequency (CF) probe ($f_s < f_p$). In this case the threshold is just above 65 dB − SPL. The suppression effect is relatively strong and almost 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. 3.6b uses the first suppression threshold definition (3.13) 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 (3.13), the probe is just detectably reduced, and thus audible. With the second suppression threshold definition (3.14), the suppression threshold corresponds to the detection threshold of the probe. Thus (3.14), *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 (3.14) was used when taking the 2TS data of Fig. 3.7b, 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, (3.14) 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 (3.14). This is shown in Fig. 3.7b where a family of tuning curves is taken with different suppressors present. As described by *Fahey* and *Allen* [3.56, Fig. 13], when a probe is placed on a specific tuning curve of Fig. 3.7b, corresponding to one of the suppressor level symbols of Fig. 3.7b, and a suppression threshold is measured, that suppression curve will fall on the corresponding suppression symbol of Fig. 3.7b. 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 only holds for criterion (3.14), the detection threshold criterion, which is appropriate for Wegel and Lane's data. If one uses (3.13) as in [3.60] they will not see this symmetry as cleary.

Suppression Threshold. Using the criterion (3.14), *Fahey* and *Allen* [3.56, Fig. 13] showed 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. *Delgutte* has written several insightful papers on masking and suppression [3.80–82]. He estimated how the intensity growth slope (in dB/dB) of 2TS varies with suppressor frequency for several probe frequencies [3.80]. As may be seen in his 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. 3.6b), 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. 3.6a. 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.

One may conclude that USM and 2TS data show systematic and quantitative correlations between the threshold levels and slopes. The significance of these correlations has special importance because

1. they come from very different measurement methods, and
2. 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 [3.83–85] as discussed in some detail by *Delgutte* [3.80], none are in quantitative agreement with the data. The two-tone suppression model of *Hall* [3.84] 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 [3.78, 86, 87].

3.2.2 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 the 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 *cochlear-amplifier* view is that the OHC provides *cochlear sensitivity* and *frequency selectivity* [3.5, 88–94]. 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. Of course other views besides these two are possible.

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 [3.3, 10, 15, 95]. Recruitment results from outer-hair-cell damage [3.96]. 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 [3.97]. 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. 3.8).

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

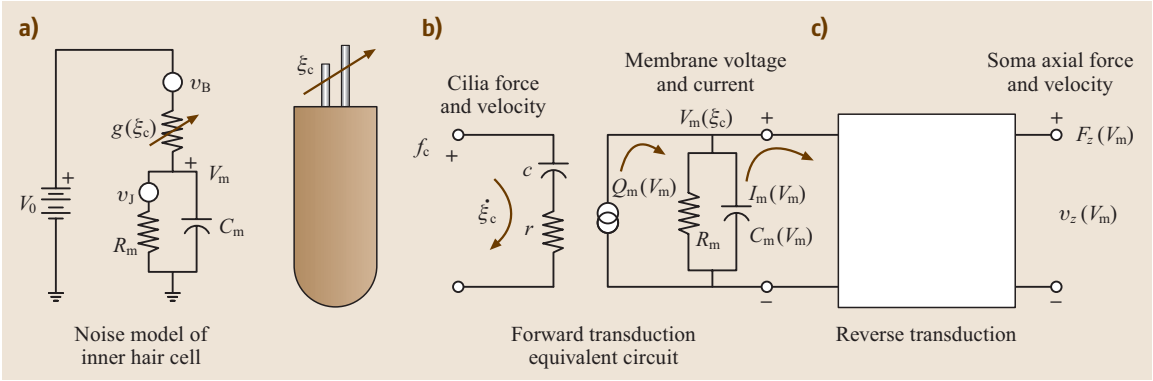


Fig. 3.8a–c On the far left **(a)** is the electrical equivalent circuit model of an IHC with thermal noise sources due to the cell leakage resistance Johnson and shot noise v_J and the Brownian motion of the cilia, represented by the voltage noise source v_B. The cilia force f_c and velocity $\dot{\xi}_c$ are the stimulus (input) variables to the forward transduction **(b)**, and are loaded by the mechanical impedance of the cilia viscous drag r and compliance c. **(c)** For OHCs, when the cilia move, current flows into the cell charging the membrane capacitance, 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 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 thus forms the basic dynamic-range dilemma.

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 [3.50, 99, 100]. This less than 5% change in length must account for a 40 dB (100 times) change in cochlear sensitivity. 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 [3.70, 71]. 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 hyperpolarizing 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 (3.5) that the spring's length will *increase* when the stiffness decreases.

Each cell is stretched by its internal static pressure $\mathcal{P}$ [3.101], and its stiffness is voltage controlled [3.70, 71]. The voltage dependent relative stiffness change is much greater than the relative motility change. Thus we have the necessary conditions for stiffness-induced motility.

3.2.3 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., [3.102]).

To organize our discussion of cochlear micromechanics, we represent each radial cross-section through the cochlear partition (Fig. 3.1b) as a linear two-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

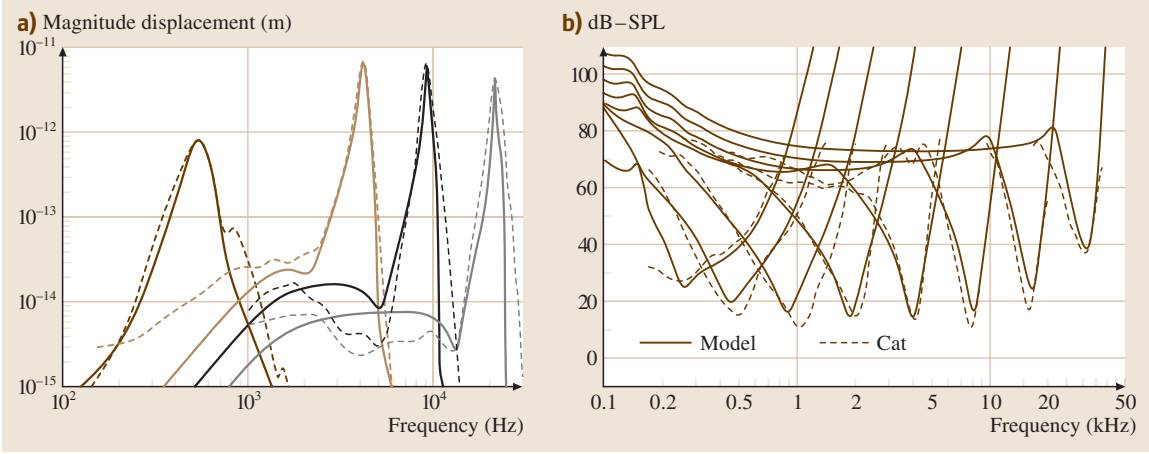


Fig. 3.9a,b The tuning curves shown by the dashed lines are the average of single nerve fiber responses from six cats obtained by M. C. Liberman and B. Delgutte. **(a)** Comparison between neural data and the computed model excitation patterns from Allen's passive RTM model (transfer function format). 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 [3.98]. This CA model assumes an IHC cilia bundle displacement of 300 pm (0.3 nm) at the neural rate threshold

$v(x, s)$

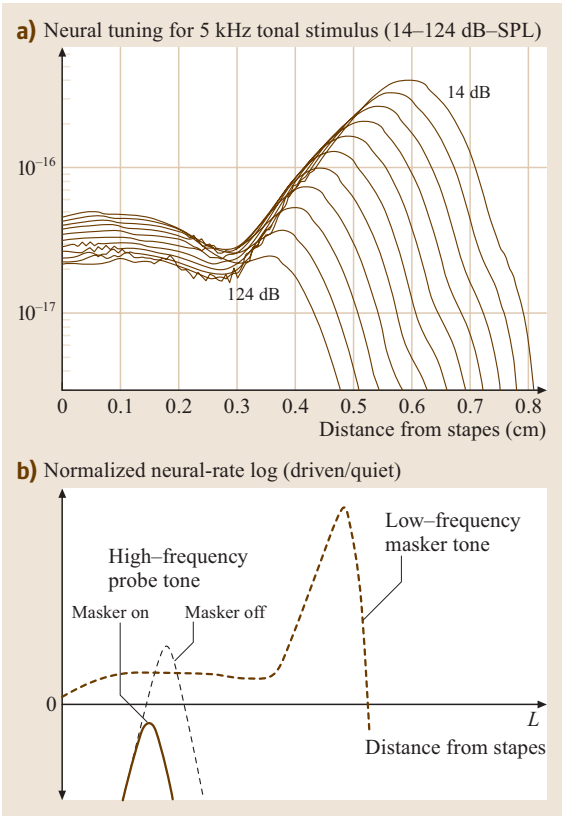
$$\begin{pmatrix} P \\ V \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} f \\ v \end{pmatrix} , \qquad (3.15)$$

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

Passive BM Models

The most successful *passive* model of cochlear tuning is the resonant tectorial membrane (RTM) model [3.9, 104]. 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. 3.4b. This sharpening is accounted for by a reflection in the tectorial membrane, introducing an antiresonance (*spectral zero*) at frequency F_z (Fig. 3.4b), which is about half an octave below the resonant frequency F_{cf} of the basilar membrane. As described by *Allen* and *Neely* [3.9], the detailed A, B, C, D elements of (3.15) are given by *Allen* [3.104], *Allen* and *Neely* [3.9].

As described in *Allen* [3.105], the response ratio of IHC cilia bundle displacement to basilar membrane displacement is defined as $H_{ihc}(x, s)$. The parameters of the RTM model may be chosen such that model results fit the experimental neural threshold tuning curves closely, as shown in Fig. 3.9a.



The Nonlinear RTM Model. The resonant tectorial membrane (RTM) model is made NL by control of the BM stiffness via OHC's stiffness is based on Fig. 3.1b. The OHC soma stiffness has been shown to be voltage dependent by *Dallos* et al. [3.106] and dependent on prestin in the membrane wall [3.107]. 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 [3.51, 103, 108, 109]. Such models are actively under consideration [3.102].

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 (3.6). As shown in Fig. 3.3, 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. 3.10. This migration of the excitation pattern, combined with the assumption that the TM has a high-pass characteristic, means that the cilia excitation gain at CF is nonlinearly compressed

Fig. 3.10a,b In **(a)** results of model calculations by *Sen* and *Allen* [3.103] 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. 3.3a that, if the BM stiffness is reduced, the location of the maximum will shift to the base, as is seen in real data. **(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 $0 \le X \le 0.3$ 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., $K_p(I_s)$)

as the intensity increases. This compression effect is shown in a cartoon format in Fig. 3.10b, while Fig. 3.10a shows the actual calculated model results. Note how the bandwidth $\Delta_f(X)$) remains approximately constant as a function of input intensity.

Sewell [3.110] 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, we show next that the NL RTM model can explain:

1. the basal-ward half-octave traveling-wave migration as a function of increasing intensity [3.76],
2. the upward spread of masking (USM) [3.20, 21], two-tone suppression (2TS) (see Sect. 3.2.1),
3. distortion product generation [3.49,55,56,111–113],
4. normal and recruiting loudness growth, and
5. hypersensitive tails [3.45].

From the steep 2.5 dB/dB slope of the USM and 2TS (Fig. 3.6a) 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. 3.10b.

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 [3.93, 114–117].

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 [3.94]. 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 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 [3.118, 119]. A general discussion of this model is presented in *Geisler* [3.90], and in *Allen* and *Fahey* [3.91].

It is presumed that this OHC action amplifies the BM signal energy on a cycle-by-cycle basis, increasing the sensitivity [3.69, 92]. 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 [3.120]. Nonlinear compression is introduced by assuming that the resistance is signal level dependent. This NL resistance model was first described by *Hall* [3.84] 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. In mathematical physics, NL damping resonators are described by *van der Pol equations*, while NL stiffness resonators are described by *Duffing equations* [3.121].

Allen and *Fahey* [3.91] 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 order 40–50 dB this is difficult to understand. A nice summary of this situation has been recently published in *Shera* and *Guinan* [3.120]. 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.

Discussion and Summary

Discussion. Both active and passive BM models are reasonably successful at simulating the neural thresh-

old response tuning curves. Thus we may 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 [3.122].

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. 3.4b) generally being much smaller than that of neural responses [3.97]. 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 [3.123], 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. 3.4b rather than traditional ad hoc bandwidth measures, is a useful tool in this regard. 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* [3.10].

Summary. This section has reviewed what we know about the cochlea. The *Basics* 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 Sect. 3.1.2 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. 3.2b. We presented the basic tools of cochlear modeling, *impedance*, and introduced the *transmission matrix* method (two-port analysis). We describe how these models work in intuitive terms, including how the basilar membrane may be treated as having a frequency 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 review 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 Sect. 3.2.3 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.

3.3 Neural Masking

When modeling human psychophysics one must carefully distinguish the external *physical* variables, which we call Φ variables, from the internal *psychophysical* variables, or Ψ variables. 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 [3.124]. 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* [3.125] in 1860, who raised the question of the quantitative transformation between the physical and psychophysical intensity. For a recent review of this problem, and a brief summary of its long history, see *Schlauch* et al. [3.126]. This section is based on an earlier report by *Allen* [3.79], and *Allen* and *Neely* [3.127].

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 remain unclear [3.128–130].

The contribution of *Allen* and *Neely* [3.127] and *Allen* [3.79] 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 [3.10, 131] with auditory signal detection theory [3.132].

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 wide-band 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*, $\mathrm{SNR_L}$.

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* [3.133]. Tonal data follows the *near miss to Weber's law* (thus does not obey Weber's law), while the wide-band noise data does obey Weber's law. 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 $\mathrm{SNR_L}$ for the tone case [3.127]. This work has applications 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* [3.75] 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 Sect. 3.3.1 and a review of historical models (e.g., Weber and Fechner), in Sect. 3.3.2, we explore issues surrounding the relation between the intensity JND and loudness, for the special cases 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* [3.134], which provided the inspiration for the present work. Next, we empirically estimate the loudness SNR, defined as the mean loudness over the loudness variance, and proportional to $L/\Delta L$, as a function of both intensity and loudness, using the tonal JND data of *Riesz* [3.75] and the loudness growth function of *Fletcher* and *Munson* [3.41]. We then repeat this calculation for Miller's wide-band 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.

3.3.1 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 [3.133]. 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,

$$I_s(t) \equiv \frac{1}{\varrho c T} \int_{t-T}^{t} s^2(t)\,\mathrm{d}t\,, \tag{3.16}$$

where T is the integration time 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 $|s|/s_{\mathrm{ref}}$, where the reference intensity is I_{ref} or $10^{-10}\,\mu\mathrm{W/cm^2}$ and the reference pressure $s_{\mathrm{ref}} = 20\,\mu\mathrm{Pa}$. These two reference levels are equivalent at only one temperature, but both seem to be in use. Equivalence of the pressure and intensity references requires that $\varrho c = 40\,\mathrm{cgs}$ Rayls. At standard atmospheric pressure, this is only true when the temperature is about 39 °C. Such levels are typically expressed in dB units.

Intensity of Masker plus 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 nonnegative 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

$$I_{\mathrm{m+p}}(t,\alpha) \equiv \frac{1}{\varrho c T} \int_{t-T}^{t} [m(t) + \alpha p(t)]^2\,\mathrm{d}t\,. \tag{3.17}$$

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 $I_{\mathrm{m}}(t)$ be the intensity of the masker with no probe ($\alpha = 0$), and $I_{\mathrm{p}}(t,\alpha) = \alpha^2 I$ be the intensity of the scaled probe signal with no masker. Thus

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

Because of small fluctuations in I_{m} and I_{p} due to the finite integration time T, this equality cannot be exactly true. We are specifically ignoring these small rapid fluctuations – when these rapid fluctuations are important, our conclusions and model results must be reformulated.

Beats. Rapid fluctuations having frequency components outside the bandwidth of the period T_{second} rectangular integration window are very small and will be ignored (T is assumed to be large). Accordingly we drop the time dependence in terms I_{m} and I_{p}. The beats between $m(t)$ and $p(t)$ of these signals are within a common critical band. 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 $I_{\mathrm{m+p}}(t,\alpha)$ and other slow–beating time dependent terms. In the Φ domain these beats are accounted for as a probe–masker correlation function $\rho_{\mathrm{mt}}(t)$ [3.132, p. 213].

Intensity Increment $\delta I(t,\alpha)$. Expanding (3.17) and solving for the *intensity increment* δI we find

$$\delta I(t,\alpha) \equiv I_{\mathrm{m+p}}(t,\alpha) - I = \left[2\alpha\rho_{\mathrm{mp}}(t) + \alpha^2\right] I\,, \tag{3.18}$$

where

$$\rho_{\mathrm{mp}}(t) = \frac{1}{\varrho c T I} \int_{t-T}^{t} m(t)\,p(t)\,\mathrm{d}t \tag{3.19}$$

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 increased from zero, the probe can eventually be detected. We specify the probe *detection threshold* as α_*, where the asterisk indicates the threshold value of α where a subject can discriminate intensity $I_{\mathrm{m+p}}(t,\alpha_*)$ from intensity $I_{\mathrm{m+p}}(t,0)$ 50% of the time, corrected for chance (i. e., obtain a 75% correct score in a direct comparison of the two signals [3.132, p. 129]). The quantity $\alpha_*(t,I)$ is the probe to masker root-mean-square (RMS) pressure ratio at the detection threshold. It is a function of the masker intensity I and, depending on the experimental setup, time. α_* summarizes the experimental measurements.

Masked Threshold Intensity. When $\rho_{mp} = 0$, 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 [3.135]. As a second example, when using the method of beats [3.75], the just–detectable modulation depends on the beat frequency. With *modern* 2AFC methods, the signals are usually gated on and off (100% modulation) [3.136]. According to *Stevens* and *Davis* [3.137, p. 142]

> *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, and depends on ρ_{mp}, and therefore T.

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 (3.16), rather the integration time is determined in the Ψ domain. This important Ψ-domain model parameter is called *loudness temporal integration* [3.138]. It was first explicitly modeled by *Munson* in 1947 [3.139].

The Φ-domain temporal resolution (T) is critical to the definition of the JND in Riesz's experiment 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 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* [3.75], *Miller* [3.133], and *Littler* [3.16], 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

$$\Delta I(I) \equiv \delta(t, \alpha_*) \,, \tag{3.20}$$

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 (3.18) with α set to threshold α_* and $\rho_{mp}(t) = 1$

$$\Delta I(I) = (2\alpha_* + \alpha_*^2) I \,. \tag{3.21}$$

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 $\Delta I(\alpha)$].

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* [3.75]. 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, $\rho_{mp}(t) = \sin(2\pi f_b t)$. Thus

$$\Delta I(t, I) = \left[2\alpha_* \sin(2\pi f_b t) + \alpha_*^2\right] I \,. \tag{3.22}$$

If the beat period is less than the Ψ temporal resolution window, the beats are *filtered* out by the auditory brain (the effective ρ_{mn} is small) and we do not hear the beats. In this case $\Delta I(I) = \alpha_*^2 I$. This model needs to be tested [3.139].

Internal Noise. It is widely accepted that the pure-tone intensity JND is determined by the *internal noise* of the auditory system [3.140, 141], 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 the experimental design; σ_I is the intensity standard deviation of the Φ-domain intensity due to Ψ-domain auditory noise [3.15, 17, 127].

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.

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 wide-band 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 [$L(I_{ref}, f_{ref}, T_{ref}) = 1$ defines a loudness scale]. For the sone scale, the reference signal is a $I_{ref} = 40\,\text{dB} - \text{SPL}$ tone at $f_{ref} = 1\,\text{kHz}$ with duration $T_{ref} = 1\,\text{s}$. For Fletcher's LU scale the reference intensity is the hearing threshold, which means that 1 sone = 975 LU [3.42] for a normal hearing person. Fletcher's LU loudness scale seems a more-natural scale than the sone scale used in the American National Standards Institute (ANSI) and International Organization for Standardization (ISO) standards.

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 space [3.132, Chap. 5]. For early discussions of this point see *Montgomery* [3.142] and p. 144 of *Stevens* and *Davis* [3.137]. To clearly indicate the distinction between random and nonrandom variables, a tilde ($\sim$) is used to indicate a random variable. As a mnemonic, we can think of the $\sim$ as a *wiggle* associated with randomness.

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 $\widetilde{L}$

$$L(I) \equiv \mathcal{E}\widetilde{L}(I)\,. \tag{3.23}$$

The second moment of the single-trial loudness

$$\sigma_L^2 \equiv \mathcal{E}(\widetilde{L} - L)^2 \tag{3.24}$$

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

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 [3.138, 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

$$M \equiv \frac{I_p^*(I_m)}{I_p^*(0)}\,.$$

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

$$J(I) \equiv \frac{\Delta I(I)}{I}\,. \tag{3.25}$$

From the signal detection theory premise that $\Delta I = d'\sigma_I$ [3.17], J is just the reciprocal of an effective signal-to-noise ratio defined as

$$\text{SNR}_I(I) \equiv \frac{I}{\sigma_I(I)} \tag{3.26}$$

since

$$J = d'\frac{\sigma_I}{I} = \frac{d'}{\text{SNR}_I}\,. \tag{3.27}$$

One conceptual difficulty with the Weber fraction J is that it is an *effective* signal-to-noise ratio, expressed in the Φ (physical) domain, but determined by a Ψ (psychophysical) domain mechanism (internal noise), as may be seen from Fig. 3.11.

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 (Fig. 3.11), giving

$$L(I + \Delta I) = L(I) + \Delta I \left.\frac{dL}{dI}\right|_I + \text{HOT}\,,$$

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

$$\Delta L \equiv L(I + \Delta I) - L(I) \tag{3.28}$$

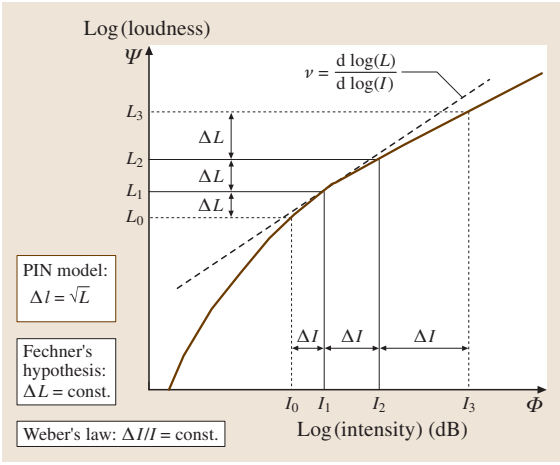


Fig. 3.11 Summary of all historical ideas about psychophysics and the relations between the Φ and Ψ variables. Along the abscissa we have the physical variable, intensity, and along the ordinate, the psychological variable loudness. The *curve* represents the loudness, on a log-intensity log-loudness 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 (3.32). These relations are verified in Fig. 3.12, as discussed in detail in *Allen* and *Neely* [3.127]

we find

$$\Delta L = \Delta I \left.\frac{\mathrm{d}L}{\mathrm{d}I}\right|_{\mathrm{I}} . \tag{3.29}$$

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

$$\Delta L = d' \sigma_{\mathrm{L}} . \tag{3.30}$$

A more explicit way of expressing this assumption is

$$\frac{\Delta L}{\Delta I} = \frac{\sigma_{\mathrm{L}}}{\sigma_{\mathrm{I}}} , \tag{3.31}$$

where d' in both the Φ and Ψ domains is the same and thus cancels.

Loudness SNR. In a manner analogous to the Φ-domain $\mathrm{SNR_I}$, we define the Ψ-domain loudness SNR as $\mathrm{SNR_L}(L) \equiv L/\sigma_{\mathrm{L}}(L)$. Given (3.30), it follows that

$$\mathrm{SNR_I} = \nu \mathrm{SNR_L} , \tag{3.32}$$

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 $\mathrm{d}y/\mathrm{d}x \approx \Delta y/\Delta x \approx \nu$. Since $\mathrm{d}\log(y) = \mathrm{d}y/y$ we get

$$\frac{\Delta L}{L} = \nu \frac{\Delta I}{I} . \tag{3.33}$$

Equation (3.32) is important because

1. it tells us how to relate the SNRs between the Φ and Ψ domains,
2. every term is dimensionless,
3. the equation is simple, since $\nu \approx 1/3$ is approximately constant above 40 dB – SL (i. e., Stevens' law), and because
4. 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* [3.143], *Fletcher* [3.21], *Wegel* and *Lane* [3.20], *Riesz* [3.75], *Fletcher* [3.144], and *Miller* [3.133].

To derive the JND counting formula, (3.29) is rewritten as

$$\frac{\mathrm{d}I}{\Delta I} = \frac{\mathrm{d}L}{\Delta L} . \tag{3.34}$$

Integrating over an interval gives the total number of intensity JNDs

$$N_{12} \equiv \int_{I_1}^{I_2} \frac{\mathrm{d}I}{\Delta I} = \int_{L_1}^{L_2} \frac{\mathrm{d}L}{\Delta L} , \tag{3.35}$$

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 [3.75,144]. The number of JNDs must be the same regardless of the domain (i. e., the abscissa variable), Φ or Ψ.

3.3.2 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 (3.21) and (3.25)

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

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 [3.145], has been successfully applied to many human perceptions. We refer the reader to the helpful and detailed review of these questions by *Viemeister* [3.129], *Johnson* et al. [3.146], and *Moore* [3.147].

Somewhat frustrating is the empirical observation that $J(I)$ is not constant for the most elementary case of a pure tone [3.75, 136]. This observation is referred to as *the near miss to Weber's law* [3.148].

Weber's law does make one simple prediction that is potentially important. From (3.35) along with Weber's law $J_0 \equiv J(\not{I})$ we see that the formula for the number of JNDs is

$$N_{12} = \int_{I_1}^{I_2} \frac{\mathrm{d}I}{J_0 I} = \frac{1}{J_0} \ln\left(\frac{I_2}{I_1}\right) . \tag{3.36}$$

It remains unexplained why Weber's law holds as well as it does [3.149, 150, 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 greatly clarifies the situation.

Fechner's Postulate

In 1860 Fechner postulated that the loudness JND $\Delta L(I)$ is a constant [3.125, 130, 151, 152]. We are only considering the auditory case of Fechner's more general theory. 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* [3.153], we shall show that Fechner's postulate is not generally true.

The Weber–Fechner Law

It is frequently stated [3.152] 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 any two intensities I_1 and I_2 is proportional to the logarithm of the ratio of the two intensities, namely

$$\frac{L(I_2) - L(I_1)}{\Delta L} = \frac{1}{J_0} \log\left(\frac{I_2}{I_1}\right) . \tag{3.37}$$

This is easily seen by eliminating N_{12} from (3.36) and 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 faulty assumptions, Weber's law and Fechner's postulate.

3.3.3 Models of the JND

Starting in 1923, *Fletcher* and *Steinberg* studied loudness coding of pure tones, noise, and speech [3.21, 154–156], and proposed that loudness was related to neural spike count [3.41], and even provided detailed estimates of the relation between the number of spikes and the loudness in sones [3.42, p. 271]. In 1943 *De Vries* first introduced a photon-counting Poisson process model as a theoretical basis for the threshold of vision [3.157]. *Siebert* [3.140] proposed that Poisson point-process noise, resulting from the neural rate code, acts as the internal noise that limits the frequency JND [3.136, 150]. A few years later [3.158], and independently [3.159] *McGill* and *Goldberg* [3.160] 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* [3.134] 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

$$\Delta L(L) \propto \sqrt{L} \,. \tag{3.38}$$

This may also be rewritten as $\sigma_L^2 \propto L$. We would expect this to hold if the assumptions of *McGill* [3.148] (i. e., the PIN model) are valid.

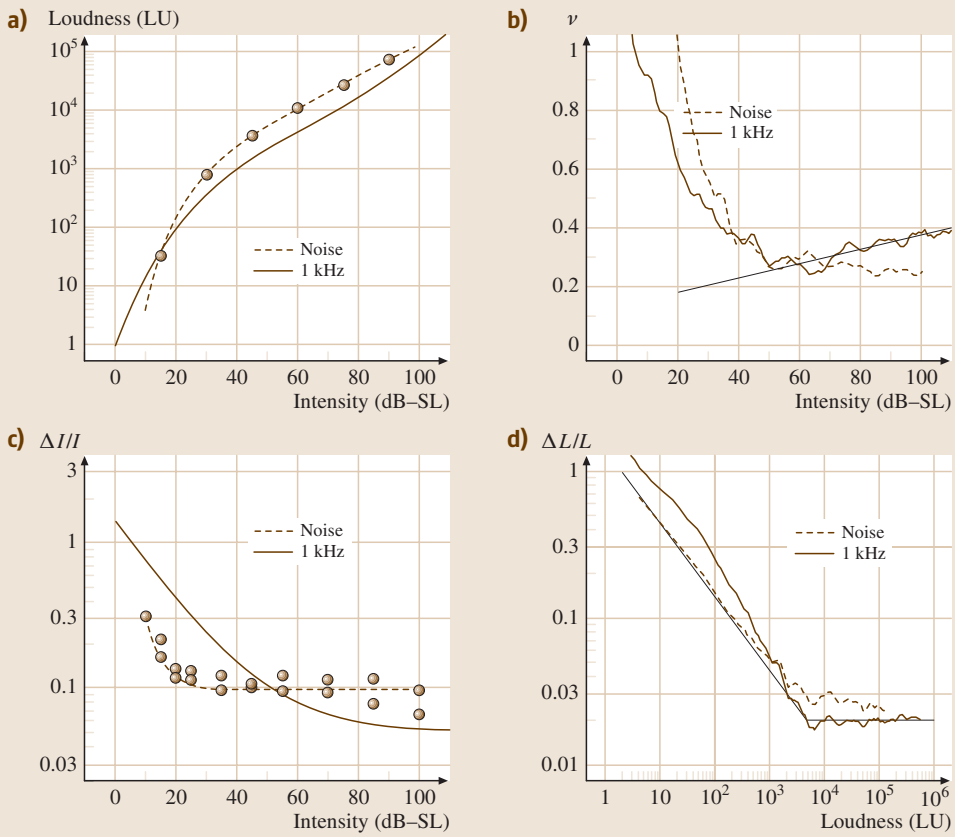


Fig. 3.12a–d In 1947 Miller measured the JND_I and the loudness level for two subjects using wide-band modulated noise (0.15–7 kHz) for levels between 3 and 100 dB – SL. The noise (*dashed line*) and pure tone (*solid line*) loudness are shown in **(a)**. The similarity between $\Delta L/L$ derived from 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. The direct derivation of ΔL based on pure tone JND and loudness data from *Miller* [3.133], *Riesz* [3.75], *Fletcher* and *Munson* [3.41].

In the following we directly compare the loudness–growth function of Fletcher and Munson to the number of JNDs N_{12} from Riesz [3.75, 127] to estimate $\Delta L/L$.

3.3.4 A Direct Estimate of the Loudness JND

Given its importance, it is important to estimate ΔL directly from its definition (3.28), 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 JND data to loudness using *Fletcher* and *Munson* [3.41] reference curve (i. e., Fig. 3.12a).

Loudness Growth, Recruitment, and the OHC

In 1924 *Fletcher* and *Steinberg* published an important paper on the measurement of the loudness of speech signals [3.155]. In this paper, when describing the growth of loudness, the authors state

> *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 [3.21].

In 1930 *Fletcher* [3.27] postulated that there was a monotonic relationship between central nerve firings rates and loudness. Given a tonal stimulus at the ear

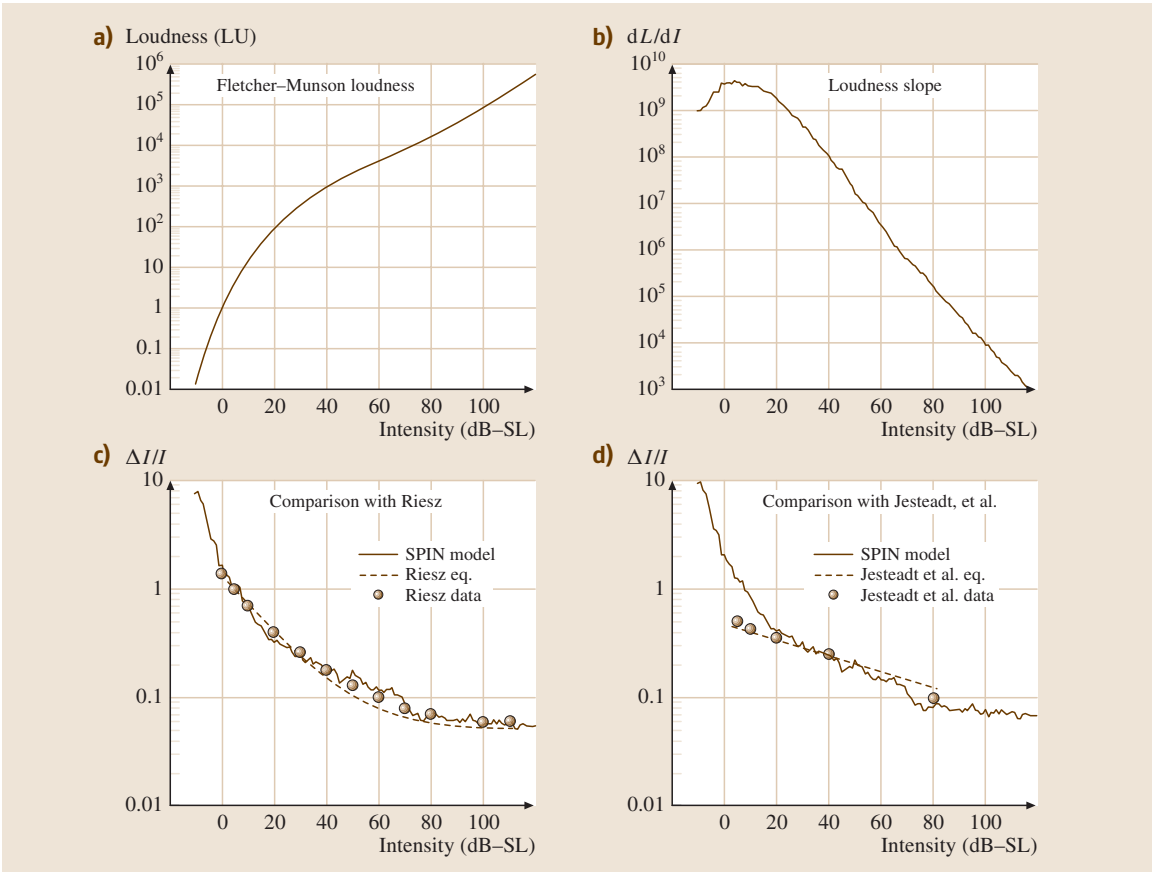


Fig. 3.13a–d Test of the SPIN model against the classic results of *Riesz* [3.75], *Jesteadt* et al. [3.136]. Test of the model derived on the *left* based on a comparison between loudness data and intensity JND data at 1 kHz, using the SPIN model

drum, Stevens' law says that the loudness is given by

$$L \equiv L(f, x, I) \propto I^{\nu} , \qquad (3.39)$$

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* [3.41] 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* [3.15].

Recruitment. What is the source of Fletcher's cube-root loudness growth (i. e., Stevens' law)? Today we know that cochlear outer hair cells 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 [3.3, 96]. *Steinberg* and *Gardner* 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) [3.3, 161]. 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 Nó* theorized that recruitment is due to hair cell damage [3.14]. *Steinberg* and *Gardner* clearly understood recruitment, as is indicated in the following quote [3.3, p. 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* [3.41], 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* [3.162] and *Steinberg* and *Gardner* [3.3], and work on modeling hearing loss and recruitment [3.122] support this view.

In summary, many studies conclude that the cube-root loudness growth starts with the NL compression of basilar membrane motion due to stimulus-dependent voltage changes within the OHC.

3.3.5 Determination of the Loudness SNR

In Fig. 3.12 we show a summary of $L(I)$, $\nu(I)$, $J(I)$ and $\Delta L/L = d'/\mathrm{SNR_L}$ for the tone and noise data.

The pure-tone and wide-band noise JND results may be summarized in terms of the loudness $\mathrm{SNR_L}(L)$ data shown in Fig. 3.12d where we show $\Delta L/L = d'/\mathrm{SNR_L}$, as a function of loudness.

For noise below $55\,\mathrm{dB} - \mathrm{SL}$ ($L < 5000\,\mathrm{LU}$) the loudness signal-to-noise ratio $\mathrm{SNR_L} \equiv L/\sigma_L$ decreases as the square root of the loudness. For a loudness greater than $5000\,\mathrm{LU}$ ($N \approx 5$ sones), $\Delta L/L \approx 0.025$ fn both tones and noise (Fig. 3.12d)

In the lower-right panel (Fig. 3.12d) we provide a functional summary of $\Delta L/L$ for both tones and noise with the light solid line described by

$$\frac{\Delta L(L)}{L} = h\,[\min(L, L_0)]^{-1/2}\ , \tag{3.40}$$

where $h = \sqrt{2}$ and $L_0 = 5000\,\mathrm{LU}$ (≈ 5 sone). We call this relation the *saturated Poisson internal noise* (SPIN) model. With these parameter values, (3.40) appears to be a lower bound on the relative loudness $\mathrm{JND_L}$ for both tones and noise. From (3.33) $\Delta L/L = \nu(I)J(I)$. Note how the product of $\nu(I)$ and $J(I)$ is close to a constant for tones above 5000 LU.

In Fig. 3.12b the second top panel shows the exponent $\nu(I)$ for both Fletcher and Munson's and Miller's loudness growth function. In the lower-left panel (Fig. 3.12c) we see $\Delta I/I$ versus I for Miller's subjects, Miller's equation, and Riesz's JND equation.

Near miss to Stevens' Law

For tones the intensity exponent $\nu(I)$ varies systematically between 0.3 and 0.4 above $50\,\mathrm{dB} - \mathrm{SL}$, as shown by the solid line in the upper-right panel of Fig. 3.12b. 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.

Figure 3.13a shows the Fletcher–Munson loudness data from Table III in [3.41]. The upper-right panel (Fig. 3.13b) is the slope of the loudness with respect to intensity ($\mathrm{LU\,cm^2/W}$).In the lower–right (Fig. 3.13d) we compare the SPIN model relative JND (3.43) (with $h = 3.0$), and the relative JND computed from the *Jesteadt* et al. [3.136] 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. 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 $80\,\mathrm{dB} - \mathrm{SL}$ 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.

3.3.6 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*

$$\Delta L = \Delta I L'(I)\,, \tag{3.41}$$

where $L'(I) \equiv \mathrm{d}L/\mathrm{d}I$. If we solve this equation for ΔI and divide by I we find

$$J(I) \equiv \frac{\Delta I}{I} = \frac{\Delta L}{IL'(I)}\ . \tag{3.42}$$

Finally we substitute the SPIN model (3.40)

$$J(I) = \frac{hL(I)}{IL'(I)}\,[\min(L(I), L_0)]^{-1/2}\ . \tag{3.43}$$

This formula is the same as that derived by *Hellman* and *Hellman* [3.134], when $L \le L_0$. In Fig. 3.13c we plot (3.43) labeled *SPIN-model* with $h = 2.4$ and $L_0 = 10\,000\,\mathrm{LU}$. For levels between 0 and $100\,\mathrm{dB} - \mathrm{SL}$, 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 [3.127]. The excellent fit gives us further confidence in the basic assumptions of the model.

3.4 Discussion and Summary

Inspired by the Poisson internal noise (PIN)-based theory of *Hellman* and *Hellman* [3.134], 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 $\mathrm{SNR_L}$ as the ratio of loudness (the signal) to standard deviation (a measure of the noise).

3.4.1 Model Validation

To evaluate the model we have compared the loudness data of *Fletcher* and *Munson* [3.41] with the intensity JND data of *Riesz* [3.75], for tones. A similar comparison was made for noise using loudness and intensity JND data from *Miller* [3.133]. We were able to unify the tone and noise data by two equivalent methods in Fig. 3.12d. 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 which we call the SPIN model. All the data are in excellent agreement with the SPIN model, providing support for the validity of this theory.

The above discussion has

- drawn out the fundamental nature of the JND,
- shown that the PIN loudness model holds below 5 sone (5000 LU) (the solid line in the lower right panel of Fig. 3.11 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) (one sone is 975 LU [3.127, p. 3631], thus 5000 LU = 5.13 LU. From the loudness scale this corresponds to a 1 kHz pure tone at 60 dB − SL),
- shown that above 5 sone the PIN model fails and the loudness SNR remains constant.

3.4.2 The Noise Model

The SPIN Model

Equation (3.40) summarizes our results on the relative loudness JND for both tones and noise. Using this formula along with (3.32), 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* [3.153].

The PIN Model

The success of the PIN model is consistent with the idea that the pure-tone loudness code 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 (Fig. 3.12d) with a constant loudness SNR between 30 and 50 for both the tone and noise conditions, as summarized by Ekman's law [3.163]. We conclude that the Hellman and Hellman theory must be modified to work above 5 sones.

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 $\mathrm{SNR_L}$ explains Weber's law for wideband signals (since ν and $\mathrm{SNR_L}$ 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. 3.12a).

Generalization to Other Data

If $\sigma_\mathrm{L}(L, I)$ 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 (3.40). It will be interesting to see how $\mathrm{SNR_L}$ 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 [3.126].

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* [3.164, pp. 141-143], JND data are quite sensitive to the modulation conditions. The *Riesz* [3.75] and *Munson* [3.165] 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. In 1928 Wegel, Riesz, and Munson were all members of Fletcher's department. *Riesz* [3.75] states that he used the same methods as *Wegel* and *Lane* [3.20], and it is likely that *Munson* [3.165] did as well.

Differences in the signal conditions are the most likely explanation for the differences observed in the intensity JND measurements of Riesz and Jesteadt shown in Fig. 3.13d. One difference between the data of *Riesz* [3.75] and *Jesteadt* et al. [3.136] 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 [3.164] 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 [3.136], one would need loudness data taken with identical (or at least similar) modulations. We are not aware of such data.

References

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

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

3.3 J. Steinberg, M. Gardner: Dependence of hearing impairment on sound intensity, J. Acoust. Soc. Am. **9**, 11–23 (1937)

3.4 J.O. Pickles: *An Introduction to the Physiology of Hearing* (Academic, London 1982)

3.5 P. Dallos: Cochlear neurobiology. In: *The Cochlea*, ed. by P. Dallos, A. Popper, R. Fay (Springer, New York 1996) pp. 186–257

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

3.7 S. Hecht: Vision II, The nature of the photoreceptor process. In: *Handbook of General Experimental Psychology*, ed. by C. Murchison (Clark Univ. Press, Worcester 1934)

3.8 G. Gescheider: *Psychophysics: The Fundamentals*, 3rd edn. (Lawrence Erlbaum, Mahwah 1997)

3.9 J.B. Allen, S. Neely: Micromechanical models of the cochlea, Phys. Today **45**(7), 40–47 (1992)

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

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

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

3.13 J. Steinberg: Stereophonic sound-film system-pre- and post-equalization of compandor systems, J. Acoust. Soc. Am. **13**, 107–114 (1941), B-1327

3.14 R.L. de Nó: The diagnosis of diseases of the neural mechanism of hearing by the aid of sounds well above threshold, T. Am. Otol. Soc. **27**, 219–220 (1937looseness1)

3.15 S.T. Neely, J.B. Allen: Relation between the rate of growth of loudness and the intensity DL. In: *Modeling Sensorineural Hearing Loss*, ed. by W. Jesteadt (Lawrence Erlbaum, Mahwah 1997) pp. 213–222

3.16 T.S. Littler: *The Physics of the Ear* (Pergamon, Oxford 1965)

3.17 W.M. Hartmann: *Signals, Sound, and Sensation* (AIP Press, Woodbury 1997)

3.18 H.L.F. Helmholtz: *On the Sensations of Tone* (Dover, New York 1954), 1863

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

3.20 R.L. Wegel, C. Lane: The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear, Phys. Rev. **23**, 266–285 (1924)

3.21 H. Fletcher: Physical measurements of audition and their bearing on the theory of hearing, J. Franklin Inst. **196**(3), 289–326 (1923)

3.22 G. Campbell: On loaded lines in telephonic transmission, Phil. Mag. **5**, 313–331 (1903)

3.23 G. Campbell: Physical theory of the electric wave filter, Bell System Tech. J. **1**(1), 1–32 (1922)
3.24 G.A. Campbell: Telephonic intelligibility, Phil. Mag. **19**(6), 152–159 (1910)
3.25 H. Fletcher: The nature of speech and its interpretation, J. Franklin Inst. **193**(6), 729–747 (1922)
3.26 H. Fletcher, J. Steinberg: Articulation testing methods, J. Acoust. Soc. Am. **1**(2.2), 17–113 (1930), Intelligibility Lists pp. 65–113)
3.27 H. Fletcher: A space-time pattern theory of hearing, J. Acoust. Soc. Am. **1**(1), 311–343 (1930)
3.28 J. Zwislocki: Theorie der Schneckenmechanik, Acta Otolaryngol. **72**, 1–112 (1948)
3.29 J. Zwislocki: Theory of the acoustical action of the cochlea, J. Acoust. Soc. Am. **22**, 779–784 (1950)
3.30 O. Ranke: Theory operation of the cochlea: A contribution to the hydrodynamics of the cochlea, J. Acoust. Soc. Am. **22**, 772–777 (1950)
3.31 L.C. Peterson, B.P. Bogert: A dynamical theory of the cochlea, J. Acoust. Soc. Am. **22**, 369–381 (1950)
3.32 H. Fletcher: On the dynamics of the cochlea, J. Acoust. Soc. Am. **23**, 637–645 (1951)
3.33 H. Fletcher: Acoustics, Phys. Today **4**, 12–18 (1951)
3.34 S. Puria, J.B. Allen: Measurements and model of the cat middle ear: Evidence for tympanic membrane acoustic delay, J. Acoust. Soc. Am. **104**(6), 3463–3481 (1998)
3.35 J.B. Allen, P.S. Jeng, H. Levitt: Evaluating human middle ear function via an acoustic power assessment, J. Rehabil. Res. Dev. **42**(4), 63–78 (2005)
3.36 H. Fletcher, W. Munson: Relation between loudness and masking, J. Acoust. Soc. Am. **9**, 1–10 (1937)
3.37 H. Fletcher: Loudness, masking and their relation to the hearing process and the problem of noise measurement, J. Acoust. Soc. Am. **9**, 275–293 (1938)
3.38 H. Fletcher: Auditory patterns, Rev. Mod. Phys. **12**(1), 47–65 (1940)
3.39 J. Steinberg: Positions of stimulation in the cochlea by pure tones, J. Acoust. Soc. Am. **8**, 176–180 (1937), cochlear map estimate, Monograph B-973
3.40 D.D. Greenwood: Auditory masking and the critical band, J. Acoust. Soc. Am. **33**(4), 484–502 (1961)
3.41 H. Fletcher, W. Munson: Loudness, its definition, measurement, and calculation, J. Acoust. Soc. Am. **5**, 82–108 (1933)
3.42 H. Fletcher: *Speech and Hearing in Communication* (Krieger, Huntington 1953)
3.43 D.D. Greenwood: Critical bandwidth and the frequency coordinates of the basilar membrane, J. Acoust. Soc. Am. **33**, 1344–1356 (1961)
3.44 M. Liberman: The cochlear frequency map for the cat: Labeling auditory-nerve fibers of known characteristic frequency, J. Acoust. Soc. Am. **72**(5), 1441–1449 (1982)
3.45 M. Liberman, L. Dodds: Single neuron labeling and chronic cochlear pathology III: Stereocilia damage and alterations of threshold tuning curves, Hearing Res. **16**, 55–74 (1984)
3.46 W. Rhode: Some observations on cochlear mechanics, J. Acoust. Soc. Am. **64**, 158–176 (1978)
3.47 W. Rhode: Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique, J. Acoust. Soc. Am. **49**, 1218–1231 (1971)
3.48 P. Sellick, I. Russell: 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*, ed. by F. Naunton, C. Fernandez (Academic, New York 1978) pp. 113–140
3.49 D. Kemp: Stimulated acoustic emissions from within the human auditory system, J. Acoust. Soc. Am. **64**, 1386–1391 (1978)
3.50 W. Brownell, C. Bader, D. Bertran, Y. de Rabaupierre: Evoked mechanical responses of isolated cochlear outer hair cells, Science **227**, 194–196 (1985)
3.51 J.B. Allen: OHCs shift the excitation pattern via BM tension. In: *Diversity in Auditory Mechanics*, ed. by E. Lewis, G. Long, R. Lyon, P. Narins, C. Steele, E. Hecht-Poinar (World Scientific, Singapore 1997) pp. 167–175
3.52 J.L. Goldstein, N. Kiang: Neural correlates of the aural combination tone 2f1-f2, Proc. IEEE **56**, 981–992 (1968)
3.53 G. Smoorenburg: Combination tones and their origin, J. Acoust. Soc. Am. **52**(2), 615–632 (1972)
3.54 D. Kemp: Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea, Arch. Oto-Rhino-Laryngol. **224**, 37–45 (1979)
3.55 D. Kim, J. Siegel, C. Molnar: Cochlear nonlinear phenomena in two-tone responses, Scand. Audiol. **9**, 63–82 (1979)
3.56 P.F. Fahey, J.B. Allen: Nonlinear phenomena as observed in the ear canal and at the auditory nerve, J. Acoust. Soc. Am. **77**(2), 599–612 (1985)
3.57 M.B. Sachs, Y.S. Kiang: Two-tone inhibition in auditory-nerve fibers, J. Acoust. Soc. Am. **43**, 1120–1128 (1968)
3.58 R. Arthur, R. Pfeiffer, N. Suga: Properties of "two-tone inhibition" in primary auditory neurons, J. Physiol. (London) **212**, 593–609 (1971)
3.59 N.-S. Kiang, E. Moxon: Tails of tuning curves of auditory-nerve fibers, J. Acoust. Soc. Am. **55**, 620–630 (1974)
3.60 P. Abbas, M. Sachs: Two-tone suppression in auditory–nerve fibers: Extension of a stimulus–response relationship, J. Acoust. Soc. Am. **59**(1), 112–122 (1976)
3.61 X. Pang, J. Guinan: 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 (1997), beautiful data showing the slope of suppression for low frequency suppressors

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

3.63 H. Duifhuis: Consequences of peripheral frequency selectivity for nonsimultaneous masking, J. Acoust. Soc. Am. **54**(6), 1471–1488 (1973)

3.64 D. Kemp: The evoked cochlear mechanical response and the auditory microstructure – evidence for a new element in cochlear mechanics, Scand. Audiol. **9**, 35–47 (1979)

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

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

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

3.68 E. Fowler: A method for the early detection of otosclerosis, Archiv. Otolaryngol. **24**(6), 731–741 (1936)

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

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

3.71 D. He, P. Dallos: Properties of voltage-dependent somatic stiffness of cochlear outer hair cells, J. Assoc. Res. Otolaryngol. **1**(1), 64–81 (2000)

3.72 A.M. Mayer: Research in acoustics, Phil. Mag. **2**, 500–507 (1876), in Benchmark Papers in Acoustics, Vol. 13, edited by E. D. Schubert

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

3.74 J.B. Allen: A short history of telephone psychophysics, J. Audio Eng. Soc. Reprint **4636**, 1–37 (1997)

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

3.76 D. McFadden: 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*, ed. by R. Salvi, D. Henderson, R. Hamernik, V. Coletti (Plenum, New York 1986) pp. 295–312

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

3.78 B. Strope, A. Alwan: A model of dynamic auditory perception and its application to robust word recognition, IEEE Trans. Acoust. Speech Signal Proc. **5**(5), 451–464 (1997)

3.79 J.B. Allen: Psychoacoustics. In: *Wiley Encyclopedia of Electrical and Electronics Engineering*, Vol. 17, ed. by J. Webster (Wiley, New York 1999) pp. 422–437

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

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

3.82 B. Delgutte: Physiological models for basic auditory percepts. In: *Auditory Computation*, ed. by H. Hawkins, T. McMullen (Springer, New York 1995)

3.83 L. Kanis, E. de Boer: Two-tone suppression in a locally active nonlinear model of the cochlea, J. Acoust. Soc. Am. **96**(4), 2156–2165 (1994)

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

3.85 C.D. Geisler, A.L. Nuttall: 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 (1997)

3.86 M. Régnier, J.B. Allen: *The Importance of Across-Frequency Timing Coincidences in the Perception of Some English Consonants in Noise* (ARO, Denver 2007), Abstract

3.87 M. Régnier, J.B. Allen: *Perceptual Cues of Some CV Sounds Studied in Noise* (AAS, Scottsdale 2007), Abstract

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

3.89 E. deBoer: Mechanics of the cochlea: Modeling efforts. In: *The Cochlea*, ed. by P. Dallos, A. Popper, R. Fay (Springer, New York 1996) pp. 258–317

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

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

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

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

3.94 D. Kim, S. Neely, C. Molnar, J. Matthews: 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*, ed. by G. van den Brink, F. Bilsen (Delft Univ. Press, Delft 1980) pp. 7–14

3.95 J.B. Allen: DeRecruitment by multiband compression in hearing aids. In: *Psychoacoustics, Speech, and Hearing aids*, ed. by B. Kollmeier (World Scientific, Singapore 1996) pp. 141–152

3.96 W.F. Carver: Loudness balance procedures. In: *Handbook of Clinical Audiology*, 2nd edn., ed. by J. Katz (Williams Wilkins, Baltimore 1978) pp. 164–178, Chap. 15

3.97 J.B. Allen: Nonlinear cochlear signal processing. In: *Physiology of the Ear*, 2nd edn., ed. by A. Jahn, J. Santos-Sacchi (Singular, San Diego 2001) pp. 393–442, Chap. 19

3.98 S. Neely: A model of cochlear mechanics with outer hair cell motility, J. Acoust. Soc. Am. **94**, 137–146 (1992)

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

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

3.101 K. Iwasa, R. Chadwick: Elasticity and active force generation of cochlear outer hair cells, J. Acoust. Soc. Am. **92**(6), 3169–3173 (1992)

3.102 I. Russell, P. Legan, V. Lukashkina, A. Lukashkin, R. Goodyear, G. Richardson: Sharpened cochlear tuning in a mouse with a genetically modified tectorial membrane, Nat. Neurosci. **10**, 215–223 (2007)

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

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

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

3.106 P. Dallos, D.Z. He, X. Lin, I. Sziklai, S. Mehta, B.N. Evans: Acetylcholine, outer hair cell electromotility, and the cochlear amplifier, J. Neurosci. **17**(6), 2212–2226 (1997)

3.107 P. Dallos: Prestin and the electromechanical responses of outer hair cells, ARO-2002 **25**, 189 (2002)

3.108 J.B. Allen: Derecruitment by multiband compression in hearing aids. In: *The Efferent Auditory System*, ed. by C. Berlin (Singular, San Diego 1999) pp. 73–86, Chap. 4 (includes a CDROM video talk by J. B. Allen in MP3 format)

3.109 J.B. Allen, D. Sen: Is tectorial membrane filtering required to explain two tone suppression and the upward spread of masking?. In: *Recent Developments in Auditory Mechanics*, ed. by H. Wada, T. Takasaka, K. Kieda, K. Ohyama, T. Koike (World Scientific, Singapore 1999) pp. 137–143

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

3.111 J.B. Allen, P.F. Fahey: Nonlinear behavior at threshold determined in the auditory canal on the auditory nerve. In: *Hearing – Physiological Bases and Psychophysics*, ed. by R. Klinke, R. Hartmann (Springer, Berlin, Heidelberg 1983) pp. 128–134

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

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

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

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

3.116 E. deBoer, A. Nuttall: 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 (1999)

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

3.118 G. Zweig: Finding the impedance of the organ of Corti, J. Acoust. Soc. Am. **89**(3), 1276–1298 (1991)

3.119 E. deBoer, A. Nuttall: The "inverse problem" solved for a three-dimensional model of the cochlea. II Application to experimental data sets, J. Acoust. Soc. Am. **98**(2), 904–910 (1995)

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

3.121 L.A. Pipes: *Applied Mathematics for Engineers and Physicists* (McGraw-Hill, New York 1958)

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

3.123 I.J. Russell, G.P. Richardson, A.R. Cody: Mechanosensitivity of mammalian auditory hair cells in vitro, Nature **321**(29), 517–519 (1986)

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

3.125 G. Fechner: Translation of "Elemente der Psychophysik". In: *Elements of Psychophysics*, Vol. I, ed. by H. Adler (Holt Rinehart Winston, New York 1966)

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

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

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

3.129 N.F. Viemeister: Psychophysical aspects of auditory intensity coding. In: *Auditory Function*, ed. by G. Edelman, W. Gall, W. Cowan (Wiley, New York 1988) pp. 213–241, Chap. 7

3.130 C. Plack, R. Carlyon: Loudness perception and intensity coding. In: *Hearing, Handbook of Perception and Cognition*, ed. by B. Moore (Academic, San Diego 1995) pp. 123–160, Chap. 4

3.131 J.B. Allen: Harvey Fletcher 1884–1981. In: *The ASA edition of Speech, Hearing in Communication*, ed. by J.B. Allen (Acoust. Soc. Am., Woodbury 1995) pp. 1–34

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

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

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

3.135 J. Egan, H. Hake: On the masking pattern of a simple auditory stimulus, J. Acoust. Soc. Am. **22**, 622–630 (1950)

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

3.137 S. Stevens, H. Davis: *Hearing, Its Psychology and Physiology* (Acoust. Soc. Am., Woodbury 1983)

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

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

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

3.141 D. Raab, I. Goldberg: Auditory intensity discrimination with bursts of reproducible noise, J. Acoust. Soc. Am. **57**(2), 437–447 (1975)

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

3.143 P.G. Nutting: The complete form of Fechner's law, Bull. Bureau Standards **3**(1), 59–64 (1907)

3.144 H. Fletcher: *Speech and Hearing* (Van Nostrand, New York 1929)

3.145 E.H. Weber: Der Tastsinn und das Gemainfül. In: *Handwörterbuch der Physiologie*, Vol. 3, ed. by R. Wagner (Vieweg, Braunschweig 1988) pp. 481–588, Chap. 7

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

3.147 B.C.J. Moore: *An Introduction to the Psychology of Hearing*, 2nd edn. (Academic, London, New York 1982)

3.148 W.J. McGill, J.P. Goldberg: A study of the near–miss involving Weber's law and pure tone intensity discrimination, Percept. Psychophys. **4**, 105–109 (1968)

3.149 D. Green: Audition: Psychophysics and perception. In: *Stevens' Handbook of Experimental Psychology*, ed. by R. Atkinson, R. Herrnstein, G. Lindzey, R. Luce (Wiley, New York 1988) pp. 327–376, Chap. 6

3.150 D. Green: Application of detection theory in psychophysics, Proc. IEEE **58**(5), 713–723 (1970)

3.151 S. Stevens: Mathematics, measurement, and psychophysics. In: *Handbook of Experimental Psychology*, ed. by S. Stevens (Wiley, New York 1951) pp. 1–49, Chap. 1

3.152 R. Luce: *Sound and Hearing* (Lawrence Erlbaum, Hilldale 1993)

3.153 S. Stevens: To honor Fechner and repeal his law, Science **133**(3446), 80–86 (1961)

3.154 H. Fletcher: Physical measurements of audition and their bearing on the theory of hearing, Bell System Tech. J. **ii**(4), 145–180 (1923)

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

3.156 J. Steinberg: The loudness of a sound and its physical stimulus, Phys. Rev. **26**, 507 (1925)

3.157 H. De Vries: 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 (1943)

3.158 W. Siebert: Stimulus transformations in the peripheral auditory system. In: *Recognizing Patterns*, ed. by P. Kolers, M. Eden (MIT Press, Cambridge 1968) pp. 104–133, Chap. 4

3.159 W. Siebert: *Personal communication* (1989)

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

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

3.162 H. Fletcher: A method of calculating hearing loss for speech from an audiogram, J. Acoust. Soc. Am. **22**, 1–5 (1950)

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

3.164 S. Stevens, H. Davis: *Hearing, Its Psychology and Physiology* (Acoust. Soc. Am., Woodbury 1938)

3.165 W.A. Munson: An experimental determination of the equivalent loudness of pure tones, J. Acoust. Soc. Am. **4**, 7 (1932), Abstract