Auditory Distortions: Origins and Functions

Paul Avan, Béla Büki, and Christine Petit

Laboratory of Neurosensory Biophysics, University of Auvergne, School of Medicine, Clermont-Ferrand, France; Institut National de la Santé et de la Recherche Médicale (INSERM), UMR 1107, Clermont-Ferrand, France; Centre Jean Perrin, Clermont-Ferrand, France; Department of Otolaryngology, County Hospital, Krems an der Donau, Austria; Laboratory of Genetics and Physiology of Hearing, Department of Neuroscience, Institut Pasteur, Paris, France; Collège de France, Genetics and Cell Physiology, Paris, France; INSERM UMR 1120, Paris, France; and Université Pierre et Marie Curie, Paris, France

Avan P, Büki B, Petit C. Auditory Distortions: Origins and Functions. Physiol Rev 93: 1563–1619, 2013; doi:10.1152/physrev.00029.2012.—To enhance weak sounds while compressing the dynamic intensity range, auditory sensory cells amplify sound-induced vibrations in a nonlinear, intensity-dependent manner. In the course of this process, instantaneous waveform distortion is produced, with two conspicuous kinds of interwoven consequences, the introduction of new sound frequencies absent from the original stimuli, which are audible and detectable in the ear canal as otoacoustic emissions, and the possibility for an interfering sound to suppress the response to a probe tone, thereby enhancing contrast among frequency components. We review how the diverse manifestations of auditory nonlinearity originate in the gating principle of their mechanoelectrical transduction channels; how they depend on the coordinated opening of these ion channels ensured by connecting elements; and their links to the dynamic behavior of auditory sensory cells. This paper also reviews how the complex properties of waves traveling through the cochlea shape the manifestations of auditory nonlinearity. Examination methods based on the detection of distortions open noninvasive windows on the modes of activity of mechanosensitive structures in auditory sensory cells and on the distribution of sites of nonlinearity along the cochlear tonotopic axis, helpful for deciphering cochlear molecular physiology in hearing-impaired animal models. Otoacoustic emissions enable fast tests of peripheral sound processing in patients. The study of auditory distortions also contributes to the understanding of the perception of complex sounds.

I. INTRODUCTION

A. Scope of the Review

Hearing has evolved to cover a huge range of sound intensities. The mammalian ear is sensitive enough to detect weak sounds [20 µPa, or 0 dB sound pressure level (SPL)] when the vibrations of molecules in the transmitting medium barely exceed displacements due to thermal, Brownian motion (19). This impressive sensitivity is achieved by a subtle sound-amplifying mechanism operated by the sensory cells themselves. This amplifier in the mammalian auditory organ, the cochlea, has been reviewed previously (228). The acoustic powers processed by the auditory system extend over a thousand billion times above detection threshold (20 million µPa or 120 dB SPL). Furthermore, in the presence of noise, the cochlear ability to extract the characteristics of sound remains robust, thereby allowing efficient central processing (“cocktail-party effect”; Ref. 39). If amplification were linear, the displacements of sound-detecting cochlear receptors, of the order of a nanometer around threshold, would reach the potentially damaging micrometer limit around 60 dB SPL, the normal level of conversational speech. Moreover, the resulting responses could not fit in the restricted dynamic range of auditory neurons innervating the sensory cells, as their action potential rates can only vary little, between a few per second, the spontaneous rate of their activity, and a few hundred per second at saturation. Actually, amplification gradually diminishes at increasing stimulus intensities, thereby producing the required compression. So, the displacements of liv-
ing auditory detectors are a highly nonlinear version of the vibrations of the external transmitting medium.

Nonlinear systems share a generic definition: the response to a combination of stimuli is not a simple algebraic sum of individual responses. However, depending on the design of a nonlinear system, the mathematical rules describing its behavior can be very different. For example, compressive nonlinearity may be achieved by instantaneous clipping of the stimulus when it exceeds some limit. This strategy has a drawback. It deforms the stimulus strongly and causes waveform distortion (WD). An alternative method of compression, the sluggish automatic gain control (AGC) acting over several stimulus periods, popular in the engineering and hearing-aid worlds, generates, in contrast, minimal WD. Finally, critical oscillators poised on the verge of self-oscillation (34, 60) could also combine reasonably fast-reacting compressive amplification and limited WD. The scope of this review is to unravel the origin and design of auditory nonlinearity. It will lead us to examine how its manifestations, compression and WD, relate to the mechanical-electrical transduction (MET) process. Its operation inescapably produces WD, and the mechanical coupling of auditory sensory cells to their surroundings transforms nonlinear products into cochlear vibrations. However, evolution seems to have managed to reach the best balance between compressive (nonlinear) amplification, faithful WD-free processing of sound, and the potential beneficial effects of WD. Indeed, although nonlinearity results in conspicuous distortions in the frequency spectrum of complex sounds, with the suppression of existing components and the creation of novel ones not present in the stimulus, perception puts up with these distortions. As we will discover, WDs also have a practical implication. Being easy to detect noninvasively, they may be used to probe the peripheral auditory function.

B. Anatomical and Physiological Framework

In mammals, the cochlear duct is divided longitudinally, from base to apex, by the cochlear partition made of the basilar membrane (BM) and the overlying auditory sensory epithelium, the organ of Corti. This partition separates two of the three internal compartments of the cochlea, scala tympani (ST) and scala media (SM). Reissner’s membrane separates scala media from the third cochlear compartment, scala vestibuli (SV). Both ST and SV are filled with perilymph (FIGURE 1A, light yellow area) while scala media contains endolymph (FIGURE 1A, light-blue area), a fluid that differs from perilymph and other extracellular fluids by its high K⁺ concentration and a positive electrical potential of about +100 mV. Sound pressure waves are transmitted to cochlear fluids via the outer and middle ears to SV by an ossicle, the stapes, which, by vibrating in the oval window of the cochlea, produces a pressure gradient through the BM.

The organ of Corti contains the auditory sensory cells, the inner hair cells (IHCs) and outer hair cells (OHCs). The transduction of sound is achieved by a unique structural feature of hair cells, protruding microvilli, namely F-actin filled processes called stereovilli or commonly, though improperly, stereocilia, which are organized in bundles (FIGURE 1B AND C). When mechanical vibrations produced by sound propagate along the BM (see sect. IVA), the resulting shearing motion between the BM and the tectorial membrane overlying the hair cells displaces stereocilia bundles. In these bundles, stereocilia organized in three rows are connected by multiple fibrous links, among which the so-called tip-links extend from the tip of a stereocilium to the side of the adjacent taller one (FIGURE 1C). When tensed by vibrational movements, tip-links produce a conformational change of MET channels located at the tip of stereocilia of the small and medium rows (18), thereby modulating their conductance (gating). In consequence, driven by their positive electrochemical gradient, K⁺ ions, which abound in the endolymph, enter the cells, together with Ca²⁺ ions. The ion currents lead to cell depolarization and to interactions between Ca²⁺ ions and various molecules inside the hair cell (35, 71).

It has been suggested in the 1940s (77), then confirmed in the 1980s that amplification of acoustic stimuli in mammals stems from a positive feedback mechanism (45) resulting from bidirectional transduction in OHCs (287), namely, their ability to respond mechanically to membrane-potential changes produced by MET (30; for a review, see Ref. 8). Electromotility is due to the presence of a motor protein, prestin, in their lateral plasma membrane (307). The resulting forces influence the mechanical impedance of the cochlear partition by feeding energy back into its movements. When applied with the proper timing, this feedback partly compensates for the loss of energy due to friction with cochlear fluids, thereby achieving amplification in an intensity-dependent, nonlinear, compressive manner (228). The sharpened BM resonance elicited by the positive feedback process converts the originally passive and broadly tuned frequency filtering of the cochlear partition into a highly tuned frequency representation, in which the high and low frequencies are processed at the base and apex of the cochlea, respectively, forming a tonotopic map along the longitudinal cochlear axis (200). Whereas OHCs are essentially confined to a micromechanical activity, IHCs aligned along the longitudinal axis in a single row of ~3,500 cells (in humans) and innervated by afferent neurons of the auditory nerve (FIGURE 1A) form the actual sensory cells of the cochlea.

Hair cells fulfilling comparable functions are found in a broad spectrum of vertebrate orders, from frogs through lizards and birds, both in hearing and vestibular organs (104, 158). Compared with mammals, the ways sound-induced vibrations reach the hair cells and set their stereo-
Cilia bundles into motion are different, and auditory hair cells are less specialized than OHCs and IHCs. Frequency tuning can stem from the intrinsic resonance of hair bundles or from electrical properties of basolateral membranes (158). Compressive amplification is also ensured by a positive feedback principle even though the somatic, prestin-driven electromotility is substituted by “active” hair-bundle motility (104). The topic of this review pertains to hair

Figure 1. A: schematic cross-section of a mammalian cochlea, showing the three fluid-filled compartments scala vestibuli (SV), scala media (SM) and scala tympani (ST), the basilar membrane (BM), and Reissner’s membrane (RM). The organ of Corti contains one row of inner hair cells (IHCs, black arrow) and three rows of outer hair cells (OHCs, red arrow), overlaid by the tectorial membrane (TM). Afferent auditory neurons forming synapses at the base of IHCs make the auditory nerve (AN). B: scanning electron micrograph of the organ of Corti of a mouse. The tectorial membrane (TM) is lifted, revealing the V-shaped stereocilia bundles protruding from the tops of OHCs organized in three spiraling rows. Beneath the reticular lamina (RL), the cylindrical bodies of OHCs can be seen behind the thin, winding elongations (phalangeal processes, PP) of supporting cells (the Deiters cells). The IHCs are not visible; they form a spiraling longitudinal row hidden beneath the tectorial membrane. On the lower side of the tectorial membrane, the imprints of the tallest OHC stereocilia can be recognized (one imprint is delineated by arrows). C: scanning electron micrograph of the top of an OHC, with its stereocilia bundle organized in three rows. The tips of small and intermediate stereocilia are connected to the sides of their taller neighbors by tip-links (arrows), essential for opening the MET channels during auditory stimulation. At this early stage (postnatal day 7), the tops of stereocilia are not yet connected by transverse links (see Figure 5). Scale bars: 4 μm (B) and 1 μm (C). [Courtesy of Aziz El Amraoui (A) and Vincent Michel (B and C).]
cell-driven nonlinearities and not to the exact feedback mechanism affording amplification and frequency tuning. Examples will be borrowed from nonmammalian auditory sensory organs when needed.

The gating principle of a mechanosensitive channel necessarily produces cycle-by-cycle WD (101) and even, clipping at saturating levels. Because of the feedback process, the sound wave of the stimulus itself gets contaminated by WD. Flattening and clipping add harmonics $2f$, $3f$, etc., to the sinusoidal wave of a tone at frequency $f$. If the stimulus contains a combination of tones, as the processing of auditory stimulation on the BM is spatially distributed, waves at neighboring and even remote frequencies travel together and can interact. The response to one tone of a complex stimulus may result in smaller displacements of the BM than it would have done if the tone had been presented in isolation. The stimuli undergoing this so-called suppression become less audible. Even though suppression diminishes the audibility of weak probe tones in the presence of noise, it has also been suggested that by increasing contrasts, it has overall positive influences on sound analysis (262). Another result of WD is that new sinusoids at arithmetic combinations of the original frequencies emerge in the response. This cochlear nonlinearity is found even at the lowest measurable stimulus levels, which led to call it “essential” (79) in contrast to the nonlinearities only observed at saturation. Although WD deforms the frequency spectrum of incoming sounds, it does not dramatically disturb perception as the cochlear compressive and filtering activity, applied to WD, ensures that it remains small in most circumstances.

All manifestations of auditory nonlinearities, compression, WD, suppression, even spontaneous activity, have been proposed to emerge from the same processes and coexist or vanish together. Thus their study should highlight basic cochlear mechanisms of sound processing including the analysis of complex sounds. Otoacoustic emissions (OAEs), sounds emitted by a nonlinear cochlea (121), deserve a particular emphasis. Although usually mixed with the stimulus that evokes them, they are easily singled out by their characteristic nonlinear properties and long latencies. They have become unique tools for noninvasive evaluation of peripheral hearing, whether for human audiology or in animal models with cochlear deficits.

II. HISTORICAL PERSPECTIVE: AUDITORY STIMULUS PROCESSING AND THE NONLINEARITY PRINCIPLE

The existence of WD in a system supposed to faithfully convert sound messages into information for the brain is counterintuitive, e.g., for high-fidelity specialists who obsessively fight any form of distortion. Thus before 1978, almost all reports of objective measurements of WD, apart from a few exceptions (193, 289), ascribed them to nonlinearities outside the cochlea, e.g., either in the middle ear (95) or in the earphones. Actually, the middle ear does not distort at physiological SPLs (4), and earphone nonlinearities can be controlled (260). However, a few early psycho-physical studies, which will be reviewed first, noticed unique features of WD that identified the cochlea as its source. With regard to objective measurements, it took two sets of seemingly unconnected observations to make clear that an intact cochlea works nonlinearly and is not a passive resonator (FIGURE 2). The first set of data (228) concerned the compressive growth of BM motion with the level of the stimulus at the cochlear places tuned to the frequency of the probe tone (225). The second set of data described instantaneous WD in normal cochleas, the central topic of this review.

A. Perceptive Evidence

It has been acknowledged for a long time that when two tones at frequencies $f_1$ and $f_2$ are mixed, intermodulation...
AUDITORY DISTORTIONS

WD results in the production of audible tonal components at arithmetic combinations \((f_2 f_1, 2f_1 f_2, 2f_2 f_1, \text{etc.})\), absent in the stimulus, called distortion products (DPs) or combination tones. Their discovery is attributed to the musician Tartini (267), and the first physically based investigation, to Helmholtz, although doubts can be raised as to the exact physiological significance of these early observations (see sect. XI). The very unusual dependence of DPs on stimulus level and frequency spacing were first disclosed by psychoacoustical methods (79, 259) ascertaining their cochlear origin.

Goldstein (79) evaluated the most prominent endogenous DP, i.e., at 2\(f_1 - f_2\), by cancelling it by an adjustable external tone at exactly 2\(f_1 - f_2\). At a precise combination of level and phase of the external tone, trained subjects, in whose ear the experiment was performed, no longer heard the 2\(f_1 - f_2\) component, as it was at the same level and exactly out of phase with the cancelling tone. The concept of “essential nonlinearity” coined by Goldstein stemmed from his key observation that when the amplitude of stimuli decreased, that of the response at 2\(f_1 - f_2\) decreased in similar proportions (1 dB per dB decrease of the primary tones at \(f_1\) and \(f_2\)) so that this nonlinear response was always approximately at the same dB level relative to the stimuli. In contrast, nonlinearities due to saturation in a passive system do not behave this way. It can be shown (see sect. VF) that the amplitude of the 2\(f_1 - f_2\) component depends on the square of the amplitude at \(f_1\) times the amplitude at \(f_2\) so that when both increase by 1 dB at the input of a passive, saturating nonlinear system, the 2\(f_1 - f_2\) component should increase by 3 dB.

The importance of Goldstein’s work was not immediately grasped. Today, its groundbreaking consequences are clear. Goldstein’s results indicate that there must exist a compressive amplifying device in the cochlea. More precisely, one has to assume that the amplitude of BM displacement increases by only 0.33 dB/dB increase of the stimulus, to explain why the 2\(f_1 - f_2\) response increases by the observed 1 dB/dB, i.e., 3 times 0.33 dB/dB. Rhode was the first to detect compression by direct mechanical measurements of BM vibrations (225), but the significance of his results was also overlooked for years. Another finding of Goldstein pointing to the cochlear origin of the DP was its strong dependence on frequency ratio \(f_2/f_1\). Here was the first explicit statement that WD, cochlear sensitivity, compression, and frequency selectivity may be associated (79). Last, the audibility of the 2\(f_1 - f_2\) DP for stimuli down to 20 dB SPL (79) called for another explanation of this distortion than saturation.

B. Objective Evidence: Otoacoustic Emissions

Goldstein’s discovery relied on perceptive reports from two subjects, including himself. Objective evidence of intermodulation from the cochlea was provided by the pioneer studies of David Kemp (120, 121), revealing the existence of OAEs from the ear. These low-intensity sounds emitted by healthy ears, either spontaneously or in response to sound stimuli, could be collected by a microphone sealed in the ear canal. Three different stimulus paradigms, using transient sounds (clicks, with a broad frequency spectrum), a single pure tone or a pair of pure tones, evoke three main categories of sound-evoked OAEs. The first observations were reported by Kemp (121) as transient-evoked OAEs (TEOAEs) initially called “Kemp echoes” (FIGURE 3). This finding was as groundbreaking a discovery as that of Goldstein because the existence of OAEs showed that the cochlea responds to sound by creating vibrations. Kemp echoes could not be regular linear echoes as they grew nonlinearly, not in proportion to click intensity. The second category of OAEs, evoked by a pure-tone stimulus, was found at the same frequency as the stimulus, stimulus-frequency OAEs (SFOAEs), which made them technically more difficult to separate from the stimulus. Last, intermodulation OAEs showed up in the spectrum of sound in the ear canal, in response to pairs of pure-tone stimuli with close enough frequencies \(f_1\) and \(f_2\), less than half an octave apart (79). In 1977, Kemp (120) had found that the DP heard at 2\(f_1-f_2\) was also emitted in the ear canal as an OAE, the so-called DPOAE, yet its weak level in human ears led its discoverer to set it aside. Henceforth, a distinction will be made between DPs, which encompass intracochlear vibrations at combination frequencies, their neuronal correlates and their perception, and DPOAEs, the acoustic correlates of DPs detected in the spectrum of sound in the ear canal after backward propagation of DP vibrations through the middle ear.

![FIGURE 3. Transient-evoked otoacoustic emission (TEOAE) evoked by a click and detected by a miniature microphone in the sealed ear canal of a normal subject. Stimulus ringing (in red, left-hand pressure scale) does not last beyond 3 ms. After this time, the averaging procedure adds the responses to clicks at opposing polarities and different amplitudes, chosen in such a way that the overall average of clicks is zero. Because the linearly growing responses are cancelled, a derived TEOAE remains, demonstrating the existence of a compressively growing response.](http://physrev.physiology.org/Downloadedfrom)
Strong doubts were initially raised whether otoacoustic emissions might be an experimental artifact. The original experiment was carefully replicated (295, 298), and the only alternative explanation for the compressive growth of OAEs with stimulus intensity, middle-ear muscle function, was excluded. Clear evidence of a close relationship between the presence of OAEs and cochlear sensitivity (7) ruled out the possibility that emissions were produced elsewhere than in the cochlea.

C. Cochlear Origin of Nonlinearities

Because of their large levels above the noise floor in laboratory animals (127, 128), DPOAEs have been extensively used in pathophysiological experiments conducted between 1980 and 1996, which unequivocally demonstrated their cochlear origin. In non-mammals and despite the differences in the physiology of their auditory organs, WD and OAEs were also found (15, 160). Evidence of the physiological links between DPOAEs and OHCs was collected from a variety of experiments enumerated in Table 1, the principles of which were either to contrast the effects on DPOAEs of damage to OHCs versus other cells, or to activate known modulators of OHC function. Nonetheless, when explored more thoroughly, the correlations between DPOAEs and the status of OHCs at places tuned to stimulus frequencies were less straightforward than initially reported, which required the issue of the places and mechanisms of DP generation to be revisited.

Objections to the idea that OHCs in the cochlea could drive BM motion strongly enough to produce detectable sound in the ear canal were lifted by studies revealing large changes in the stiffness of the cochlear partition upon tip-link disruption. These changes were large enough to show that OHC stereocilia can indeed influence the impedance of the cochlear partition (36). In turn, the cochlear partition can drive sound pressure in a sealed ear canal, and measurements of the reverse middle-ear transfer function in guinea pig (157) and gerbil (52) have shown that intracochlear DPs come out as DPOAEs with a level approximately reflecting the middle-ear transformer effect, that is, ~35 dB regardless of frequency.

It was soon suggested that the links between noninvasively measured OAEs, regardless of their category, and OHC functions could be exploited for inferring useful information regarding cochlear frequency tuning. In the frequency domain, so-called suppression tuning curves of OAEs were built. These represent for an interfering tone, how the level required to influence an OAE varies with interfering frequency. They reveal that interference happens only in a narrow frequency interval with a width similar to that of cochlear resonators (27). In the time domain, the delay of OAEs relative to their stimuli betrays the interplay of underlying resonators, such that the sharper their tuning, the longer they ring, as do OHC responses (200, 228). Non-physiological, equipment-related causes of distortion show only a delay corresponding to the time it takes for sound to travel a few centimeters in the air and through the external or middle ear (in humans, a fraction of a ms). Conversely, TEOAEs ring several tens of milliseconds after the click evoking them has faded away (121) [FIGURE 3]. Measurements of the latency of the other two types of OAEs are more delicate as they are usually produced by continuous tones with no definite onset. Either special methods utilizing

<table>
<thead>
<tr>
<th>Experiment</th>
<th>OHC Activity</th>
<th>OAEs</th>
<th>Reference Nos.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acoustic overstimulation</td>
<td>Abolished</td>
<td>Abolished</td>
<td>7, 196, 309, 285</td>
</tr>
<tr>
<td></td>
<td>Temporarily affected</td>
<td>Decrease then recovery</td>
<td>127, 241, 252</td>
</tr>
<tr>
<td>(Before the mid 1980s, the role of OHCs being</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>unknown, the effects of noise on OAEs could</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>be attributed to cochlear, but not OHC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>impairment)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aminoglycoside administration</td>
<td>Abolished</td>
<td>Abolished</td>
<td>28</td>
</tr>
<tr>
<td>Medial olivocochlear efferent neural</td>
<td>Slightly inhibited</td>
<td>Slightly Decreased</td>
<td>40, 128, 211</td>
</tr>
<tr>
<td>bundle stimulation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electrical stimulation of OHCs</td>
<td>Increased</td>
<td>Increased</td>
<td>188</td>
</tr>
<tr>
<td>Assayed losses of OHCs by platinum salts</td>
<td>Abolished</td>
<td>Abolished</td>
<td>269</td>
</tr>
<tr>
<td>Assayed losses of IHCs by platinum salts</td>
<td>Unaffected</td>
<td>Unaffected</td>
<td>269</td>
</tr>
<tr>
<td>DP grams in human hearing losses due to</td>
<td>Abolished</td>
<td>Abolished</td>
<td>89, 166</td>
</tr>
<tr>
<td>sensory-cell disease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DP grams in human hearing losses due to neural</td>
<td>Unaffected</td>
<td>Unaffected</td>
<td>196, 261</td>
</tr>
<tr>
<td>dysfunction</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OAE, otoacoustic emission; OHC, outer hair cell; IHC, inner hair cell; DP, distortion product.
interrupted stimuli allow DPOAE onset latency to be visualized directly (293), or the frequency dependence of the phase of DPOAEs reveals their so-called group delays (see sect. IVF) (130, 185, 186, 221, 231).

D. Essential Links Between Nonlinearity and Cochlear Activity

Around 2000, several publications independently emphasized that the Hopf-bifurcation formalism provides a parsimonious, generic model of an active ear that unites all its aforementioned properties. It describes the behavior of critical oscillators, poised (e.g., by negative feedback) for operating near their dynamic instability (34, 60, 105). With the assumption that the damping term of an oscillator contains some adjustable parameter, a bifurcation occurs when this parameter sets damping to zero. On one side of the bifurcation, damping is negative and the (unstable) system oscillates spontaneously, while on the other side, damping is positive and the oscillator is stable, yet neither sensitive nor finely tuned. It is when poised near the bifurcation that the critical oscillator displays the key features of an active cochlea, i.e., high sensitivity, compressive amplification, and level-dependent frequency tuning, combined with the above-described nonlinearities. Stability is ensured by a distortion-generating damping term increasing with the third power of oscillation amplitude (see sect. VIF). Several specific models of the cochlea present a Hopf bifurcation, even some of them not explicitly built for this purpose (265).

The confirmations of a shared origin for cochlear activity and nonlinearity have changed the status of OAEs since the 1980s. Initially, OAEs were considered as anecdotal by-products, worth being studied only because they might have a clinical utility, such as universal screening of congenital hearing impairment in neonates (22, 108, 290). From the 1990s on, their potential as tools for assessing cochlear nonlinearities has been understood. OAEs open a unique window on the intimacy of auditory stimulus processing in a hearing organ.

III. MECHANO-ELECTRICAL TRANSDUCTION AND NONLINEARITY

Mammalian OHCs act within a feedback loop, which processes the auditory stimulation. For modeling purposes, four functional stages of this feedback loop can be separated, all of them being theoretically able to distort to some extent. Locally, transverse sound-induced BM vibrations propagated from the basal cochlea produce a radial shearing motion between the reticular lamina, formed by the apical cell surface of hair cells and pillar cells supporting them, and the tectorial membrane, which deflects the stereocilia bundles. The tallest stereocilia embedded in the tectorial membrane are directly deflected in the excitatory direction, away from the modiolus, when the BM moves upward toward scala media (stage 1). The displacement of OHC bundles away from the modiolus opens MET channels by pulling on the tip-links (see sect. IB; FIGURE 1C). The endocochlear potential, about +100 mV, combined with the negative polarization of OHCs, near −40 mV (109), strongly drives endolymphatic K+ ions into the cells through open MET channels (stage 2). The sound-modulated current through an OHC produces a change in its membrane potential. As OHCs are endowed with electromotility (see sect. I), forces are generated (stage 3) which, by acting with the appropriate timing, produce positive feedback and a negative-damping effect that enhances BM vibration (stage 4) (FIGURE 4). In the hair cells of nonmammalian species, as mentioned previously, no somatic electromotility has been reported. The active feedback loop, instead, involves the stereocilia bundles (104). It has been proposed that hair-bundle mechanics contributes to the active process also in the cochlea (194, 205), which, in the four-stage model, would imply active force generation at stages 2 and 3, that is, not only stage 3.

FIGURE 4. Left: stages of OHCs activity organized in a feedback loop at the origin of amplification (see text for explanation). Right: at stage 2 (mechano-electrical transduction; MET), the current versus stereocilia bundle deflection transfer function is a sigmoid-shaped Boltzmann function (plot labeled P0). MET-channel functioning is also at the origin of the nonlinear stiffness of hair bundles, which reaches a minimum when the opening probability of MET channels is 0.50 (plot labeled “stiffness”). As a result, the force exerted on a stereocilia bundle against its deflection (solid line labeled “force”) departs from a straight line (dotted line labeled “linear”). [From van Netten (276). Copyright 1997, with permission from Elsevier.]
A. Which Stage Is Nonlinear?

At stage 1, the shearing motion between BM and tectorial membrane contains nonlinear terms due to the complex trigonometry of their relative displacements but only at large deflections (>10°), beyond the physiological scale of a few hundred nanometers (deflection <3°). Stage 3 is linear even though in vitro experiments pinpoint two potential causes of nonlinearity. The somatic current versus voltage characteristic of OHCs (97) is indeed complex, due to the interplay of ionic fluxes through membrane channels of the cell soma, and the OHC cell-body length relates to membrane potential in a slightly nonlinear manner when its changes exceed tens of millivolts (8). However, for stimulations above 1 kHz, the somatic current-voltage relationship becomes linear, and in the physiological range of changes, the voltage-dependent electromechanical activity of OHCs distorts negligibly (240).

Therefore, relevant contributions of OHCs to significant mechanical nonlinearity and WD should be searched for at stage 2. Here, two possible mechanisms can be found. First, the directly measured stiffness of hair bundles is asymmetrical, larger in the excitatory than inhibitory direction (68). Second, when zooming in on the OHC MET channels, the plot relating K⁺ ion current versus hair-bundle displacement, the so-called transducer transfer function, is sigmoid shaped. Experimental plots can be fitted by a first-order Boltzmann transfer function describing the statistical process by which MET channels in a hair bundle shuffle from an open to a closed state (288). These two nonlinear contributions, deflection-dependent change in hair-bundle stiffness and MET function, actually are tightly interconnected, as shown by experiments that used antibiotics to block MET function and also abolished the asymmetry of hair-bundle stiffness (102, 235). Therefore, gating of MET channels and activity-dependent hair-bundle stiffness might be two sides of the same coin, which suggested the term gating stiffness (or its inverse, “compliance”) to describe this nonlinearity (see sect. IIIB).

Nowadays, second-order asymmetrical Boltzmann transfer functions assuming one open and two closed states are more widely utilized for describing the activity of OHC MET channels (153). Their point of inflection lies closer to the lower saturation level than in a first-order function. The current varies from zero, when all channels are closed, to a saturated value corresponding to all channels open, and the operating point (OP) describes the percentage of open channels at rest, thought to be near 50% in OHCs (109, 234). The slope of the transfer function around the OP is \( S = \Delta I / \Delta \alpha \), where \( \Delta I / \Delta \alpha \) is the current through the cell produced by a deflection \( \Delta \alpha \) of the hair bundle. The larger \( S \), the larger the change in membrane potential in response to a given \( \Delta \alpha \), thus the larger the strength of feedback onto the BM and the gain of the OHC-based cochlear amplifier. From the maximum-gain position at the inflection point, if the OP shifts toward a saturating part of the transfer function, its curvature and ability to generate WD change too. The sinusoidal input tends to get clipped more on one side, as opposed to symmetrical clipping when the OP coincides with the center of symmetry of the function.

The exact shape of the transfer function at the OP, where the hair bundle operates, is of importance because, as will be shown in section VF, the magnitudes of DPs depend on the coefficients of its expansion in a Taylor power series. The lowest-order DP at \( f_2 - f_1 \) is called quadratic, and the next terms at \( 2f_1 - f_2 \) and \( 2f_2 - f_1 \), cubic, as they depend on the terms of the series to the powers 2 and 3, respectively.

B. Mechanisms of Nonlinearity on a Subcellular Level

The mechanosensitive channels of auditory sensory cells, whose gating is directly controlled by hair-bundle deflection and not by enzymatic reactions as in nonmechanical sensory modalities, can respond to frequencies up to ~100 kHz. The stereocilia of a mature OHC hair bundle, arrayed in three rows of increasing height, are coupled together by tip-links (207), made of cadherin-23 and protocadherin-15 (119) and likely anchored to the MET channel. The deflection of a hair bundle by the acoustic stimulus in the excitatory direction increases the tension in tip-links and related gating springs, putative elastic elements which pull on MET-channel gates and increase their open probability \( (P_o) \).
At present, even though attractive candidates have been proposed for the MET channel (118, 129), the exact molecular architecture of the MET machinery remains unknown. However, many basic characteristics of its gating have been inferred from ex vivo measurements of ionic current through the channel and of stiffness measurements, against hair-bundle deflection, using calibrated flexible fibers to deflect stereocilia tips (e.g., Refs. 68, 102). To understand the origin of the deflection dependence of stereocilia bundle stiffness (68), the first requirement is the identification of its components. Apart from the tip-link molecular complex and the stereociliary rootlets around which stereocilia rotate as rigid rods (23, 132), the stereocilia of OHCs are expected to be stiffened by zipperlike clustered fibrous links, the horizontal top-connectors that join the upper parts of adjacent stereocilia within and between rows (81, 280, 282) (FIGURE 5). All contributions are thought to be independent of deflection, which was confirmed experimentally for the first two ones (44, 100). Let $K_s$ the sum of these contributions, be the deflection-independent hair-bundle stiffness.

The gating process also introduces a deflection-dependent stiffness term $K_g$ due to the fact that when a MET channel opens, the associated conformational change leads to a significant shortening of the associated gating spring by a few nanometers, which relaxes it partially. As a consequence of this decrease in stiffness, the gating springs of the remaining channels bear additional tension, a gating force $F_g$ that increases $P_o$ (101). As shown by the current versus deflection mechanoelectrical transfer function, $P_o$ obeys a sigmoid Boltzmann statistical law (FIGURE 4B)

$$P_o = 1/\left(1 + \exp \left[ -\frac{F_g(X - X_0)}{k_BT} \right] \right)$$

in which $F_g$ is the gating force, $X - X_0$ is the displacement of the tip of the hair bundle, $k_B$ is the Boltzmann constant, and $T$ is the absolute temperature. If one considers more than one open and one closed states for the MET channel, higher-order Boltzmann statistics replace the first-order one in Equation 1. Thermodynamics then predicts that the stiffness of a hair bundle carrying $N$ transduction units is (101)

$$K_{HB} = K_o + K_g = K_o - (N F_g^2/k_BT) P_o (1 - P_o)$$

When all channels are either closed or open, $K_{HB} = K_o$. When a MET channel is pulled open by a stimulus so that $P_o$ increases, $K_g$, the nonlinear $P_o$-dependent term in Equation 2, describes a decrease in stiffness, i.e., an increase in compliance, with momentous consequences. Indeed, as the gating force $F_g$ adds up to the stimulus (FIGURE 6), it becomes easier to deflect the hair bundle and thus to open more MET channels. The single-channel gating force can be expressed as a function of the transducer current $I(X)$ and its derivative $dI/dX$ (277)

$$F_g(X) = k_B T / I(X) dI(X)/dX$$

Thus both the nonlinear gating stiffness of the hair bundle and the gating force that modulates the stimulus directly relate to the sigmoid Boltzmann transfer function and to MET-channel operation. As a result, the larger the size of $K_g$ relative to $K_o$ in Equation 2, the more conspicuous the overall hair-bundle nonlinearity. In hair-cell preparations from the frog saccule, a vestibular organ, $K_g$ is so large that the overall hair-bundle stiffness becomes negative within the range of MET channel operation, which leads to bundle instability and spontaneous oscillations (172). In mammalian hair cells, the linear stiffness $K_o$ dominates so that gating forces elicit much less mechanical nonlinearity ($K_o$ and the $K_o/K_g$ ratio are 10 times larger than in frogs) (277).
C. Horizontal Top Connectors: An Unexpectedly Large Part

Measurements on mutant mice that lack stereocilin (Strc<sup>−/−</sup> mice), a model for the hereditary recessive form of human hearing impairment DFNB16 (281), raised a surprising issue regarding the part played by MET channels in the generation of cochlear nonlinearity (282). Stereocilin is a protein associated with horizontal top connectors joining adjacent stereocilia within and between rows, and with the links that attach the tallest stereocilia to the tectorial membrane (280). In the absence of stereocilin, in Strc<sup>−/−</sup> mice, horizontal top connectors do not develop, and the tips of OHC stereocilia are less clearly aligned than in control mice (FIGURE 7A). Their tip-links persist until postnatal day 15, P15 (FIGURE 7B), and MET currents are found within normal range, ensuring near-normal auditory thresholds during a few days after the onset of hearing. Yet, DPOAEs and WDs in the electrical responses of OHCs (FIGURE 7, C, right panel, and D, dashed line) are abolished in response to stimuli below 90 dB SPL, whereas in wild-type mice they are evident in the 20- to 90-dB SPL interval (FIGURE 7, C, left panel, and D, solid line). A strong decrease of masking, in its suppressive component, is also observed (FIGURE 7E). As discussed, suppressive masking is viewed as enhancing contrasts among simultaneously present spectral components (262, 282). Whether decreased suppressive masking is detrimental to the intelligibility of speech in noisy environments remains to be tested in patients affected by DFNB16. Between P15 and P60, Strc<sup>−/−</sup> mice gradually entirely lose OHC function, which is attributed to the irreversible loss of the tip-links. Of note, in a recently published cohort of affected patients (69), mutations in the stereocilin-encoding gene, STRC, emerged as a major contributor to pediatric bilateral sensorineural hearing impairment. Yet it frequently came with only mild increase in hearing thresholds, which thus indicates that OHC hair bundles without the stereocilin-composed horizontal top-connectors can still function in humans.

The intriguing issue raised by the phenotype of Strc<sup>−/−</sup> mice is that it apparently contradicts the tenet that distortion is an exclusive property of MET channels in relation to their ability to open and close normally. By functioning enough in young Strc<sup>−/−</sup> mice to ensure the detection of low-level sounds, MET channels must distort. Yet no sign of WD was detected between 20 and 90 dB SPL. Of note, this divorce between normal responses at SPLs near the auditory threshold, and the absence of WD and defects in masking (FIGURE 7) may only be apparent, and be compatible with a normal functioning of MET channels when hair-bundle deflection is minimum ~40 dB SPL for threshold measurements, but could become abnormal at larger hair-bundle deflections ≥70 dB SPL for WD, DPOAE, and masking measurements.

Several hypotheses may be brought forward to explain these results. The most radical line of thought would be to challenge the current view by positing that the recorded distortion stems, not from MET channel nonlinearity (K<sub>g</sub> in Equation 2 would be too small), but from some passive nonlinear mechanism of the hair-bundle, so far overlooked, that involves the horizontal top connectors. This implies that the molecular assembly containing stereocilin has specific biophysical properties able to confer nonlinear stiffness to a hair bundle. Notably, some models of hair bundles predict that nonlinear stiffness can stem from changes in bundle geometry due to deflection when a cohesive hair bundle is formed (190). An even simpler view would assume the horizontal top connectors to be tight when a hair bundle is deflected in one direction, and relaxed in the other direc-

FIGURE 7. A 20- to 90-dB SPL interval (uli below 90 dB SPL, whereas in wild-type mice they are evident in the 20- to 90-dB SPL interval (FIGURE 7, C, left panel, and D, solid line). A strong decrease of masking, in its suppressive component, is also observed (FIGURE 7E). As discussed, suppressive masking is viewed as enhancing contrasts among simultaneously present spectral components (262, 282). Whether decreased suppressive masking is detrimental to the intelligibility of speech in noisy environments remains to be tested in patients affected by DFNB16. Between P15 and P60, Strc<sup>−/−</sup> mice gradually entirely lose OHC function, which is attributed to the irreversible loss of the tip-links. Of note, in a recently published cohort of affected patients (69), mutations in the stereocilin-encoding gene, STRC, emerged as a major contributor to pediatric bilateral sensorineural hearing impairment. Yet it frequently came with only mild increase in hearing thresholds, which thus indicates that OHC hair bundles without the stereocilin-composed horizontal top-connectors can still function in humans.

The intriguing issue raised by the phenotype of Strc<sup>−/−</sup> mice is that it apparently contradicts the tenet that distortion is an exclusive property of MET channels in relation to their ability to open and close normally. By functioning enough in young Strc<sup>−/−</sup> mice to ensure the detection of low-level sounds, MET channels must distort. Yet no sign of WD was detected between 20 and 90 dB SPL. Of note, this divorce between normal responses at SPLs near the auditory threshold, and the absence of WD and defects in masking (FIGURE 7) may only be apparent, and be compatible with a normal functioning of MET channels when hair-bundle deflection is minimum ~40 dB SPL for threshold measurements, but could become abnormal at larger hair-bundle deflections ≥70 dB SPL for WD, DPOAE, and masking measurements.

Several hypotheses may be brought forward to explain these results. The most radical line of thought would be to challenge the current view by positing that the recorded distortion stems, not from MET channel nonlinearity (K<sub>g</sub> in Equation 2 would be too small), but from some passive nonlinear mechanism of the hair-bundle, so far overlooked, that involves the horizontal top connectors. This implies that the molecular assembly containing stereocilin has specific biophysical properties able to confer nonlinear stiffness to a hair bundle. Notably, some models of hair bundles predict that nonlinear stiffness can stem from changes in bundle geometry due to deflection when a cohesive hair bundle is formed (190). An even simpler view would assume the horizontal top connectors to be tight when a hair bundle is deflected in one direction, and relaxed in the other direc-
tion, without having any specific nonlinear mechanical property per se. Equation 2 would be replaced by \( K = K_{\text{HB}} / K_{\text{d}} \), with \( K_{\text{d}} \) depending on the state of tension or relaxation of horizontal top connectors, i.e., on the direction of hair-bundle deflection.

Alternative explanations integrate the role of MET channel nonlinearity in the generation of WD. The coordinated opening of MET channels implicating the horizontal top connectors might combine weakly nonlinear transfer functions of individual MET channels into a markedly nonlinear mechanoelectrical transfer function of the hair bundle. The interest of a coordinated motion of bundled stereocilia has been recently emphasized in the context of transduction sensitivity rather than nonlinearity, but the same arguments may be relevant for both topics. One important issue is whether MET channels work in series or in parallel. In a mode of action in series of the MET channels, only the stereocilia in a radial relationship mediated by tip-links are involved. The opening of one MET channel is expected to reduce the force exerted on the other MET channels within the same radial column, which results in negative channel cooperativity (117). A possible exception was observed in a model of hair bundle that examined the series engagement of two transducer channels, with one of them assumed to be sensitive at small displacements and the second one at larger ones. In this case, the series mode of activation allowed complementary activation with improved dynamic range and intact sensitivity (278). It is noteworthy that hair-cell activation curves in this two-channel series model showed greater asymmetry and nonlinearity and much more compression at higher levels, relative to the single engagement case. Cooperativity between only two channels was sufficient for deeply influencing the nonlinearity of hair cell transfer functions (278). This strong impact of radial mechanical coupling between MET channels of one column of a stereocilia bundle suggests that a possible contribution of the inter-row horizontal top connectors to channel cooperativity and the ensuing nonlinearity might have to be considered in addition to the tip-links invoked (278).

Actually, several investigations of hair-bundle preparations whereby stereocilia splaying is prohibited suggested that the transduction channels work mechanically in parallel, rather than in series. Parallel coupling results in positive cooperativity in response to a force stimulus, i.e., the more MET channels open, the larger the force transmitted to the closed ones (116, 136). In hair cells from the bullfrog sacculus, the constraints on hair-bundle geometry were found to be strong enough for spontaneous stereocilia movements at opposite edges of the
bundle to show high coherence and synchrony (136). In the same type of hair cell, fast video analysis of hair-bundle motion (117) revealed the sliding displacement of adjacent stereociliary tips, indicative of parallel gating. In these saccular hair cells, links other than the horizontal top connectors could be severed without influencing bundle motion, thereby suggesting that sliding motion, at least in these cells, rests on horizontal top connectors. Mathematical models of three-dimensional motion of a whole hair bundle (190) support the idea that the asymmetry of the MET current versus hair-bundle deflection relationship increases with the number of MET channels working together. Altogether, these studies suggest that strong lateral coupling among MET channels might impinge the \( P_o \) Boltzmann pattern, and thereby may increase distortion enough to produce the major WD that the \( \text{Strc}^{-/-} \) mice fail to produce.

Could the W shape of the OHC hair-bundles have a special impact on the proposed stereocilin-mediated positive MET channel cooperativity? The fact that the basal insertion points, around which neighbor stereocilia pivot, lay at different distances from the bilateral symmetry axis of the hair bundle generates shearing motion between stereocilia of the same row when the hair bundle is deflected. Parallel coupling via intrarow horizontal top connectors, possibly reinforced by the anchoring of the tallest stereocilia in the overlying tectorial membrane, should allow, when a number of MET channels open, the tension release of their gating springs to affect the force acting on other MET channels of the hair bundle. In the absence of stereocilin in a W-shaped hair bundle, only the three stereocilia (FIGURE 1C) of the same radial column can interact via their tip-links, because the lateral constraint between stereocilia of neighboring radial columns has vanished. [Notably, in hair bundles organized in a straight line, as in IHCs, whether connected or not, stereocilia of neighboring radial columns are expected not to slide relative to each other, or to slide to a far more modest extent.] At hearing threshold, cooperative opening of channels should be irrelevant so that one hardly expects the presence of top connectors to make any difference on hearing sensitivity. But at large enough sound pressure levels for hair-bundle deflection to engage several MET channels, in the 20–90 dB SPL range where WD is explored and was not found in the \( \text{Strc}^{-/-} \) mice, the large difference in nonlinearity between the normal and \( \text{Strc}^{-/-} \) mice might be related to presence versus absence of positive cooperativity resulting in parallel coupling. At present, horizontal top connectors can be concluded as essential to the generation of WD, but which of the proposed mechanisms underlies their contribution remains to be clarified.

D. Other Consequences of Coordinated Stereocilia Motion Possibly Contributing to Hair-Bundle Nonlinearity

The effects of coordinated stereocilia motion have been mostly studied in terms of interaction between stereocilia and surrounding fluid. One possibly detrimental consequence of this interaction is damping by friction. The complex hydrodynamics of stereocilia have been studied either by finite-element modeling or interferometric measurements. Their outcomes concur in showing that in a cohesive hair bundle such that the fluid trapped between stereocilia is immobilized, the viscous drag is reduced compared with a vibration mode allowing relative squeezing among stereocilia shafts (135). An additional advantage of trapping the viscous fluid between stereocilia might consist of a more concerted activation of MET channels (135).

Brownian motion in the fluid bathing stereocilia is another phenomenon that might hamper MET efficiency by introducing noise in MET channel openings. It has been suggested that parallel coupling of transduction elements should reduce the negative impact of noise by increasing the synchrony of MET channels (50, 117, 191).

Numerical simulations of hair bundle motion, in the simpler case of the straight rows of IHC hair bundles (258), also emphasized the importance, for MET-channel dynamics, of controlling the degree of splay among stereocilia. The authors stressed the role of tectorial-membrane positioning in this dynamics, and furthermore, predicted the formation of nano-vortices between rows of stereocilia. Of note, nano-vortices vanish when stereocilia lose cohesion (Chadwick, personal communication). Vortices being a known source of WD, this may provide one more link between hair-bundle cohesion and nonlinearity.

E. OHC Bundles: Their Environment and Operating Point

If the processes allowing appropriate gating of MET channels form the source of WD and are an essential stage of nonlinear amplification, this does not preclude that linear elements in the environment of hair bundles modulate WD or contribute to transferring it into a detectable form. The electromotility of OHCs, for example, transfers the nonlinearity of the membrane potential due to MET-channel operation into a mechanical nonlinearity (see sect. IIIA). Despite the absence of DPOAE at low stimulus levels in prestin-null mice mutants lacking somatic motility and amplification (143), the observation of physiologically vulnerable DPOAEs around 70 dB SPL indicates, however, that prestin is not necessary to distortion generation when a high stimulus level overcomes the lack of amplification (144).

The OP defined by the percentage of open MET channels at rest controls the gain of the cochlear amplifier and influences the amount and symmetry of WD. What known processes modulate OP position? The percentage of open channels at rest tends to be stabilized by the process of adaptation (101). Conversely, the OP can be shifted by biasing the
BM by infrasound (21, 239); endolymphatic hydrops, i.e., swelling of the scala media (11, 268, 270); and modulation of OHC function by medial olivocochlear efferent neurons activated by the ipsi- or contralateral presentation of sound (3, 83, 189). It has even been suggested that the resting position of the BM and the OP of OHCs are continually adjusted by a control mechanism that modulates cochlear sensitivity and tuning (140). Displacement-sensitive experimental setups indeed reveal long-lasting shifts of the BM from its resting position (31) to which regular, velocity-sensitive measuring devices are insensitive. Fluid pressure in scala media may provide a suitable control mechanism, with hydrops representing an extreme situation (142).

The tectorial membrane, which embeds the tallest stereocilia of OHCs and incidentally, possibly contributes to the cohesion of OHC stereocilia bundles when stereocilin is present, might influence sound-evoked OAEs. Nonetheless, OAEs do exist in ears of lower vertebrates without tectorial membrane (17). In Tecta (deltaENT/deltaENT) mice with mutations in the Tecta gene encoding α-tectorin (139), and a tectorial membrane detached from the cochlear epithelium, DPOAEs are still produced at stimulus levels above 65 dB SPL (151). Their extant level can be explained by an abnormal mode of OHC stimulation by viscous coupling to the endolymph bathing the hair bundles, resulting in a 35-dB increase in hearing thresholds. The source of nonlinearity thus requires higher stimulus levels to compensate the decreased cochlear amplification. It produces DPOAEs that decrease in amplitude with increasing frequency, whereas DPOAE amplitude depends little on frequency in control mice, with an effect of the frequency ratio $f_2/f_1$ more irregular than in control mice (151). This supports the view of a tectorial membrane acting as a mechanical filter shaping the properties of sound-evoked OAEs, without influencing their generating mechanism, which depends on OHC hair bundles with normal MET channels and horizontal top connectors, as they are in Tecta mutants (280).

In summary, overwhelming evidence points to the nonlinear processes that ensure the gating of MET channels in sensory-cell hair bundles, i.e., their gating compliance and the cohesive motion of stereocilia, as the source of WD, one of the manifestations of auditory nonlinearity. In OHCs, MET processes also are the first stage of a feedback loop that dynamically controls BM motion and decreases cochlear amplification at increasing input levels. In this way, MET contributes to the other manifestation of auditory nonlinearity, the compressive (±0.33 dB/db) growth of responses with stimulus level at frequencies near resonance. Off resonance, compression no longer exists. This will lead us, in the next sections, to distinguish between WDs coming from hair bundles at resonance (nonlinearly growing, “essential” WD) versus off-resonance (more linearly growing, associated with saturation of MET currents on either side of the Boltzmann transfer function, i.e., saturation WD), to examine the spatial distribution of WD sources, and to review how WDs from different locations propagate and recombine to form noninvasively recorded OAEs. Although OAEs conveniently reveal the existence of WD and reflect many of its important properties, it does not mean that their generation mechanisms are purely nonlinear, and these mechanisms will have to be reviewed according to OAE type.

**IV. NONLINEAR RESPONSE OF THE COCHLEA TO A SINGLE TONE**

This section examines the simplest case of a pure tone at frequency $f$. It briefly reviews how acoustic vibrations travel along the cochlear BM and peak according to the rules of cochlear tonotopy (200); what nonlinearities the vibrating OHCs generate; from where and how the resulting OAEs can propagate backward to the external auditory canal.

**A. Forward Travel**

The thickness and width of the BM, separating ST from scala media and on which rests the organ of Corti (FIGURE 1A), gradually vary from base to apex, and so does the local resonance frequency, which decreases from base to apex. The Reissner’s membrane, rather uniformly compliant, is usually thought to play little mechanical part. Incoming sounds enter SV at the oval window, where the pistonlike vibration of the stapes produces a compression wave traveling at the speed of sound in water, ~1,500 m/s (206); however, it likely causes little significant motion of sensory cells. As a general rule, the impedance of a resonant system driven below its resonance frequency is dominated by its stiffness. Thus with respect to audible frequencies except the highest ones, the basal-most cochlea behaves as a stiff partition. Thus the amplitude of the pressure wave in ST is small, and there is a pressure difference across the BM. Energy transfer occurs as the stiff BM responds to this pressure difference by rapid vibrations. As shown by Békésy (14a), coupled interactions between the BM and cochlear fluids of ST and SV initiate a transverse, forward-moving wave traveling along the stiff BM surface at a few hundred meters per second, which increases in amplitude and gradually slows down to a few meters per second while approaching the site of resonance, called CF (for characteristic frequency) location (218, 227). This traveling wave excites auditory sensory cells by efficiently moving their stereocilia bundle. Toward CF location, the wavelength (ratio of speed to frequency) decreases and a large phase delay accumulates. Theory indicates that the place of maximum vibration along the longitudinal cochlear axis is just basal to CF location. Exactly at CF location, at resonance, the stiffness and mass terms of the cochlear impedance cancel each other, and the purely resistive impedance absorbs the energy of the tone, by so-called critical-layer absorption (147),
so that the amplitude of vibration is less, as the envelope of BM vibrations at $f$ peak very near CF location. For the sake of simplicity, no distinction will be made henceforth between CF location and this peak. More apically, the pressure difference between SV and ST vanishes, and BM vibrations decline sharply. If by some means a vibration is generated at frequency $f$ at a location tuned to a frequency lower than $f$ (it will be shown later on that WD may generate such frequency components), this vibration finds too slack a BM to propagate in either direction. Frequency is an important variable controlling the shape of BM vibrations. Yet, an even more relevant variable is the scaled ratio $f/CF$ between the stimulus frequency $f$ and the characteristic frequency of the measurement location. From accumulating observations of BM motion, Zweig (310) was the first to pinpoint that the amplitude and phase patterns of BM motion are determined solely by $f/CF$, a property which he called cochlear scaling symmetry [except in the most apical part of the cochlea where scaling symmetry apparently breaks (49, 168)].

In sensitive cochleas, the traveling wave is amplified by OHCs along its way. Their feedback, by compensating damping, boosts the BM response to soft sounds provided feedback occurs at an appropriate phase. This happens only between 1 and 2 mm basal to the CF location (in the chinchilla, at CFs around 10 kHz) (228). The enhancement of cochlear resonance and sensitivity comes with finer tuning between 0.2 and 0.5 dB increase in SPL), but again only along the short section on the BM where OHCs efficiently boost the traveling wave. Where OHCs are far from resonance (passive) or damaged, feedback exerts no amplification along the broad tail, OHC MET activity and BM vibrations at $f$ grow linearly 1 dB/dB (FIGURE 2). BM displacement thus has two components, near CF location, a high and sharp peak strongly depending on the “activity” of the OHC-driven cochlear amplifier, and between stapes and CF location, a broad “passive” tail region, hardly sensitive to OHC status. Around the peak, amplification ensures a large input to the nonlinearly working MET channels of active OHCs, accounting for essential distortion at low SPLs. Along the broad tail, OHC MET channels still open if stimulated high enough and produce “passive” nonlinear phenomena including WD. Their existence will be discussed in the next sections. Compared with essential distortion from around CF, their growth is expected to be steeper (see sect. IIA), and their time delays shorter as their generation place is nearer the oval window.

B. Always Forward?

Section IVA may give the false impression that cochlear waves can only propagate from stapes to CF location. Actually, from a vibrating site on the BM, mechanics prescribe that its elementary contribution to the wave, a “wavelet,” starts traveling both ways. The physicist Huygens introduced the principle that every point of a wavefront acts as the source of wavelets that, by spreading out in all directions, not only that of the incoming wave, combine together to continue the propagation. Fresnel (70a) later completed this principle by stating that the amplitude of a wave at any given measurement point equals the vector sum of the amplitudes of all the wavelets reaching that point. The result of this summation depends on whether wavelet phases are similar or widely different. Fast phase variations among wavelets lead to small contributions even though individual amplitudes may be large (251). A combination of large amplitudes and slowly varying phase (phase coherence; Ref. 247) leads to a large outcome. Thus the way that spatially distributed wavelets can or cannot efficiently combine introduces emergent directionality, critical for accepting the concept of OAEs. In a linear medium, the wavelets have the same frequency as the stimulus, but in the nonlinear cochlea, the incoming wave also generates wavelets at frequencies different from its own (see sect. V).

For a regular pure tone coming via the stapes or the bony wall, from the nondirectional principle of Huygens, forward directionality emerges. Wavelets from basal places, launched when the incoming stimulus reaches their place, incur an increasing phase lag while going forward. In this direction, they meet more apically produced wavelets, in phase with them because they have been launched with an increasing phase lag, as the stimulus reached their places later. The phase lags accumulate in parallel, which favors wavelet recombination in the forward direction. Wavelets starting basal-ward, conversely, experience strong cancellation as those from apical places, launched with already a large phase lag, incur an additional lag while traveling basally and interfere with wavelets produced more basally and presenting much less phase lags (launched earlier and having less to travel). The case of OAEs generated within the cochlea is different and will be examined in sections IVF and V.

C. Vibrations at f Generate Harmonic Distortion

In the cochlea stimulated by a single tone, WD is revealed by the production of acoustic harmonic distortion at integer
multiples of the stimulus frequency, the fundamental frequency \( f \), i.e., \( 2f \), \( 3f \), etc. (FIGURE 8). Through a tiny opening, a laser beam is focused on the BM (41, 125, 232), or a displacement-sensitive probe is placed near it (197), and the stimulus frequency is usually set near the CF of the tested site, \( \text{CF}_{\text{site}} \). Few reports exist on basal harmonic distortion, possibly because it seems small and labile, absent in Reference 232, \(-30\) dB relative to stimulus level in Reference 41, and \(-25\) dB in Reference 197. Apical harmonic distortion is conspicuously larger, \(-7\) dB relative to stimulus level (125), perhaps in relation to the different geometry of OHCs in the cochlear apex addressed in section VII. In Reference 197, the sensitivity of the preparations was not optimal, and the origin of all detected harmonic pressure components, at \( 2f \) and \( 3f \), was tracked to the place tuned to \( f \) (because the delays of the harmonics and of the fundamental frequency \( f \), inferred from their phase accumulation against frequency, were the same). Cooper also attempted to shift the stimulus frequency to \( f = \frac{\text{CF}_{\text{site}}}{2} \) (41). A response was measured at \( \text{CF}_{\text{site}} \), due to the harmonic \( 2 \times (\frac{\text{CF}_{\text{site}}}{2}) \). Harmonics, wherever produced, are vibrations and propagate accordingly. Where had this vibration been generated? It could hardly be the location tuned to the stimulus frequency, as it was when \( \text{CF}_{\text{site}} \) tones were used, because the second harmonic cannot travel along the BM at places tuned to lower frequencies. The only alternative hypothesis is that it was produced at basal places and traveled, forward, to the detector. While reaching there, its CF location, it was amplified and became detectable, hence its name “amplified distortion” (41). The existence in this experiment of harmonics generated at their frequency place on the BM and not at the CF location highlights the importance of considering basal contributions to WD and not only CF ones.

Two noteworthy facts were reported at basal sites (41). First, and this will be encountered again in nonlinearities produced by spectrally richer sounds, the observed harmonic distortion levels are low. The energy carried by harmonics never exceeds \( 4\% \) of that of the stimulus. There is likely an AGC in the cochlea that tends to lower the cochlear gain, when the sound level increases, enough for WD to be kept to a minimum (315). Harmonic distortions can be heard only at large stimulus levels (286). Their contributions produced at places tuned to \( f \) are filtered out; furthermore, they cannot propagate to their CF locations. As for “amplified distortions,” they likely suffer from the upward spread of masking by the louder, lower-frequency tone at \( f \) (see sect. IXA). Second, at variance with the outcome of two-tone stimulus experiments (see sect. VF) favoring odd-order WD, the \( 2f \) (even-order) harmonic is the largest, not the \( 3f \), \( 5f \), \( \ldots \) harmonics. The balance between odd- and even-order WD is very sensitive to the position of the OP of the nonlinear source (see sects. VF and VIII), which may differ from one experiment to another depending on how the invasive measurements affect cochlear homeostasis.

### D. Waveform Distortions at Off-Resonance Places

Section IVA suggested the existence of WD even at off-resonance basal places. This can be checked easily. An electrode at the round window detects several electric correlates of the vibrations of basal hair cells, an oscillating change in electric potential at the stimulus frequency called cochlear microphonic (CM) potential, and a steady change following the envelope of the stimulus, called summating potential (SP). They are thought to relate to the oscillating and steady components, respectively, of the receptor potentials of hair cells. The electrode collects spatially averaged responses in which, at stimulus frequencies below a few kHz, basal hair-cell contributions to the CM prevail. Being off-resonance,
they are small yet combine in unison, due to the long wavelength of the traveling wave. In contrast, the larger responses from hair cells near the CF location undergo fast phase rotations leading to spatial cancellation. The CM is largely dominated by contributions from OHCs, as its amplitude changes little after selective destruction of IHCs (57). The SP originates from the asymmetry of the transfer function of the hair-cell transduction process at the OP, which generates a sustained depolarizing component. The main sources of SP are IHCs, but a sizeable component from OHCs can be identified all along the cochlear spiral (57).

The CM waveforms display considerable distortion with large harmonics in response to a pure tone (193, 289). The CM WD is attributed to OHCs (38, 139). Its presence shows that basal OHCs do distort, even though excited by off-resonance tones. The sinusoidal deflection of stereocilia produces a saturating, nonsinusoidal current through the MET channels, which reveals the underlying Boltzmann-shaped deflection/current transfer function and the OP of basal OHCs (FIGURE 7). Mechanically induced position changes of the organ of Corti, e.g., by gel injections (238) or application of a biasing force on the bony cochlea (308), come with shifted OP and modified CM harmonics, but also an increased SP. This pinpoints the practical interest of monitoring any of these compound signals from off-resonance places, as they are sensitive to subtle changes in hair-cell transduction.

E. Emissions Produced by a Single Tone: Stimulus-Frequency Otoacoustic Emissions

Reemission of additional sound from within the cochlea is illustrated by the measurements of SFOAEs emitted at the same frequency as the stimulus. Although their measurement is straightforward, their properties and relation to cochlear nonlinearity are complex. It is worthwhile to study them carefully, however, because they reveal many properties of an active cochlea in a noninvasive manner. It had been reported by Elliott (61) that auditory threshold plots against frequency from normal ears display quasi-periodic ripples of a few decibels. Following up on psychoacoustic investigations of this auditory threshold microstructure, Kemp (120) slowly varied the frequency of a low-level pure tone while recording the SPL in the sealed ear canal. He detected narrow ripples, superimposed on broad changes in SPL due to earphone calibration. The ripples signaling periodic changes in the ear’s impedance correlated with auditory threshold microstructure. The most straightforward interpretation was that of an interference mechanism between the stimulus itself and a pressure wave emitted from within the ear and likely partially reflected at the middle ear boundary. Fast phase rotation of the acoustic emission with frequency would create a regular pattern with periodic

crests when emission and stimulus add together in phase, and troughs when they are out of phase and cancel each other. This OAE, elicited by a single continuous tone at the stimulus frequency, was called SFOAE.

How can one SFOAE be separated from its stimulus as they share the same frequency? Their different growths at increasing stimulus level provide a first solution. The relative size of ripples decreases and eventually vanishes around 70 dB SPL as the SFOAE saturates. The SFOAEs can thus be extracted by subtraction of the scaled, high-level pattern from the low-level one. The suppressive action of a tone superimposed on the stimulus at a slightly lower frequency affords a convenient alternative (120). Suppression removes the stimulus-related wave traveling inside the cochlea, so the SFOAE may be extracted by subtraction of the suppressed from the unsuppressed plot (FIGURE 9, A for the amplitude and B for the phase).

F. Backward Propagation of SFOAEs

Kemp and Chum (123) suggested that the SFOAE came from the peak of the incoming traveling wave, where nonlinear reflection occurred on some impedance discontinuity. The current dominant view is that this reflection mechanism is linear. Its description is necessary, however, as its interplay accounts for many properties of all types of OAEs and shapes their nonlinear characteristics. How can wavelets of an SFOAE combine in a regular manner to form a significantly large backward wave (see sect. IVB)? Bragg scattering offered a possibility (264). Strong positive interference happens in the backward direction when an X-ray wave falls upon a distributed array of periodic scattering centers, with a spatial period of half its wavelength. The difficulty is that there seems to be no regular scattering array of OHCs in the cochlea (150, 304), unless regularity would arise from periodic disruptions induced by the curvature of the coiled cochlea (159), but there are SFOAEs from uncoiled auditory organs.

A breakthrough occurred when a class of models of an active cochlea was shown (311) to generate a periodic structure of coherent reflecting sites out of spatially irregular distributions of scattering centers. The theoretical analysis showed that Bragg-like scattering could occur from a disorderly array satisfying several assumptions. First, if the peak of the mechanical response to the stimulus is tall, scattering is localized from the place where the response is largest. At this place, if the peak of the response is broad enough to encompass many scattering centers, some will present a spatial frequency matching the wavelength of the incoming sound. These will be singled out, by a sort of bandpass filter-like mechanism extracting a few coherent components from noise. The requirements of a tall and broad peak may seem conflicting as, to produce a tall peak, the cochlear amplifier generates frequency tuning which sharpens the
peak. Actually, the wavelength is around 0.2 mm at a 16 kHz CF location, in the gerbil (218), and the 1- to 2-mm-wide region over which amplification generates a tall peak (228) covers several wavelengths and many scattering centers.

The backward contributions from scattering centers in the peak region will add in phase and generate a large backward-emitted SFOAE. Bragg scattering is fundamentally linear and additive, contrary to what was initially suggested (123). Nonetheless, cochlear activity requiring the nonlinear action of OHCs is essential for the generation of a tall and broad peak allowing spatial filtering and, ultimately, efficient Bragg-like emission of SFOAEs. The nonlinear characteristics that SFOAEs carry with them are their saturation at increasing stimulus level and their suppressibility.

This view of purely linear coherent reflection has been challenged by recent observations based on time-delay considerations. If an SFOAE results from backward scattering on irregularities near the peak of the BM response to the stimulus, theory predicts that the time it takes for the SFOAE to reach the ear canal is the sum of two delays of similar size, one for the forward journey to the characteristic place on the BM and one for the backward journey (249). Because an SFOAE is a steady response to a steady tone, its timing cannot be directly evaluated. However, a classic way to derive temporal information from sinusoidal responses is to study how their phase relative to the stimulus varies with frequency. Let us consider a system through which the travel time is $t$, for example, to some characteristic place and back. For sound at frequency $f$, this travel time, or delay, translates into a phase shift $\Delta \varphi = 2\pi ft$ (every period, i.e., every $1/f$, the phase of the sound rotates by one cycle). At frequency $f + \delta f$, the travel time leads to a phase shift $\Delta \varphi' = 2\pi(f + \delta f)t$, assuming that $t$ varies little with $f$. It follows that $t = 1/2\pi d\varphi/df$; thus the travel time can be deduced from how much the phase of the response to a pure tone varies with frequency (FIGURE 9, C AND D). The delay measured by this technique is called group delay in signal processing, as the envelope of a group of waves at different frequencies (or wave-packet) moving at the group velocity $v_g$ undergoes a group delay $t = \Delta x/v_g$ when moving by $\Delta x$. Average SFOAE delays are $\sim 2$ ms in cats and guinea pigs at 1 kHz (2 cycles), and even longer in humans, 10 ms at 1 kHz (10 cycles) (247).

By comparing BM group delays to SFOAE ones in a broad frequency interval in the chinchilla cochlea, Siegel et al. (251) found that SFOAEs had too short a delay relative to a doubling of BM group delays, especially at frequencies $> 4$ kHz. They ranged from 0.7 to 2 ms, well below the esti-

---

**FIGURE 9.** A: amplitude of the residual sound signal corresponding to an SFOAE, plotted against frequency, in response to a swept tone at 30 dB SPL in a chinchilla ear. Open circles, initial measurement; solid line, repeated measurement after -1 h. B: phase versus frequency function. C: group delay against frequency, extracted, at each frequency, from the slope of the phase versus frequency function. Arrows marking the positions of amplitude notches, $\pi$ phase reversals and extreme group-delay values, respectively, happen to coincide [From Siegel et al. (251). Copyright 2005, with permission from Acoustical Society of America.] D and E: principle of group-delay measurements. The phase rotation of a signal from a local site, when its evoking stimulus is swept in frequency, accumulates in proportion to the time it takes for the signal to propagate from its generation place to the measurement point (e.g., the stapes or external ear canal), and therefore, to their distance. The red and green sine waves start from the same site at different frequencies, thus undergoing different phase rotations in a given time interval. D: short propagation delay, small phase gradient. E: long propagation delay, large phase gradient, for the same change in frequency.
mated round-trip travel between stapes and CF location. The authors proposed two possible explanations. One is that the backward travel of the SFOAE is much shorter than the forward travel of the stimulus along the BM, which suggests a different propagation mechanism, perhaps by a compression wave through cochlear fluids (see sect. V.H). Another possibility is that contributions to the SFOAE are widely distributed, not only near the peak of the traveling wave, but also basally with shorter delays. These basal sources, although being smaller in size than the CF-location ones, would undergo smooth spatial summation in long-wavelength cochlear regions, while cancellation in the short-wavelength region would offset the advantage of larger levels, finally emphasizing short-latency contributions (251). Since then, the criticism of Siegel has been addressed (248) by qualifying the previous model and incorporating more general considerations no longer based on “SFOAE delay = twice the BM delay.” New evidence has been produced that SFOAEs (in cats) are as sensitive to low-frequency interference as cochlear responses from the place tuned to \( f \), which is thus likely their dominant site of production (145). Other evidence, in the guinea pig, accommodates combined mechanisms of SFOAE generation, by showing that nonlinear mechanisms (independent of the presence of coherent reflection sites, but dependent on the high-level wave at \( f \)) can strongly influence the emission at high stimulus levels, whereas linear coherent reflection dominates at moderate and low levels (80). Subtle mechanisms such as position-dependent reflectance along the cochlea, multiple intracochlear reflections, or generation of “intermodulation” SFOAEs (difference tones occurring between harmonics, e.g., \( 2f - f \) or \( 3f - 2f \); see sect. VB and Ref. 64) might also have to be taken into account.

G. SFOAEs Without Traveling Wave

The previous explanation refers to the BM and its traveling wave, but the existence of scattering and of SFOAEs does not require them. For instance, it is possible to measure SFOAEs in lizard ears, which lack a BM-supported traveling wave (16), resonance being incorporated in hair-cell bundles (161, 204). Lizard SFOAEs, including their latency, are similar to those measured in mammalian ears. Obviously, in this case, latency cannot be attributed to the propagation time of a traveling wave that slows down while approaching its CF location. Delays arise not only from travel times, but also filter responses. The more finely tuned a filter, the later the beginning of its response and the longer its duration. It has been shown with the help of a physical model (16) that the only requirement for long-delay SFOAEs is the presence of a slightly disarranged battery of active filters. Similar qualitative conclusions have been reached regarding the presence of long delays in TEOAEs evoked by short impulse stimuli. Their mathematical principles, simple and generic, are described in Appendix I. They account for how MET nonlinearity produces detectable OAEs in all tetrapod ears, however diverse their anatomy. The basic requirement is the presence of an active filtering principle, regardless of how it is achieved.

In summary, this section shows how the properties of OAEs evoked by a single tone can be understood by considering several key functions of an auditory organ; how waves propagate and interact with sensory cells; how fast they travel; what amplification and compression they undergo and whether a tall and broad peak is produced at CF location; and how finely waves are filtered. By measuring SFOAE properties, it is possible to probe several aspects of auditory stimulus processing and perception without disrupting their fragile anatomical support. The next section elaborates on these methods by examining a spectrally richer situation.

V. INTERACTION OF TWO TONES IN THE COCHLEA

When two tones at frequencies \( f_1 \) and \( f_2 \), called the primaries (\( f_2 > f_1 \)), are presented simultaneously, the envelopes of BM motion, that peak at two different places, overlap over a broad interval. The amplitudes of the peaks at \( f_1 \) and \( f_2 \) on the BM still grow with stimulus level in a compressive manner and harmonics are still produced at \( 2f_1, 2f_2, 3f_1, \) etc. However, as BM and OHC vibrations are nonlinear, two other important nonlinear manifestations show up from where the vibrations at \( f_1 \) and \( f_2 \) interact, i.e., all over the basal part of the cochlea and particularly near the place tuned to \( f_2 \), the higher frequency. The presence of vibrations at one frequency tends to suppress the vibrations at the other frequency, and conversely, intermodulation due to the combined presence of vibrations at two frequencies generates vibrations at new frequencies not present in the stimulus.

A. Two-Tone Suppression

Two-tone suppression (2TS) occurs when the presentation of a (suppressor) tone at frequency \( f_s \) decreases the mechanical response to a probe tone at frequency \( f_p \) usually set near the resonance frequency of the measured spot (228). Suppression is one of the mechanisms of masking whereby a weak probe tone is no longer audible in the presence of a louder masker sound (see sect. IX.A). An objective correlate is two-tone rate suppression in auditory-nerve fibers (107). As the firing rate of neurons that code for a probe tone is reduced by the presentation of a suppressor tone at a different frequency, so is the audibility of the probe tone. Early single-fiber recordings of auditory neurons revealed two domains of the frequency versus level map of neuronal activity, on both sides of the excitatory region centered on the probe frequency, such that a suppressor tone decreased the...
activity due to the probe even when too weak to elicit an activity of its own (236). As IHCs and neurons were not excited by the suppressor tone, suppression had to be a micromechanical mechanism, rather than a synaptic or neural one. Direct evidence of 2TS in BM vibrations (226, 233) confirmed this mechanical interpretation.

The characteristics of 2TS on the BM have been dissected in the hope that they would shed light on its underlying nonlinear mechanism. The growth of BM response to a probe tone presented alone is compressive at intermediate intensities, at frequencies near the CF of the measured place. Extensive experimental studies of 2TS have evaluated how the BM vibration at the probe frequency changes when a second suppressor tone is added, at a frequency either higher or lower than CF (233). Although suppressors presented alone produce responses smaller than the probe, they induce a clear suppression, monotonically increasing with suppressor intensity, and decreasing with probe intensity. In addition to decreasing the sensitivity to the probe at CF, 2TS decreases the compressive nonlinearity of BM growth. Suppression is also a tuned phenomenon, so that for a probe at CF, suppressor frequencies close to CF are most efficient (42, 43). When the probe frequency departs from CF, low- or high-frequency suppressors lose their suppressive influence (233). The phase changes of BM vibrations at CF in the presence of a suppressor tone are reminiscent of those produced by stimulus-intensity changes (42, 228). Last, 2TS is labile and requires healthy OHCs (202, 233).

In summary, experimental facts concur in suggesting a close correlation of 2TS with the interaction of OHCs and BM and with its results, enhanced sensitivity, compressive amplification, and frequency tuning. To give these facts a common theoretical framework requires an analysis of the complex behavior of a nonlinear oscillator driven by a mixture of input signals at two different frequencies and different combinations of levels. Of note, the simplest experimental model eliciting realistic 2TS is a single hair cell from the bullfrog’s sacculus placed in a two-compartment preparation allowing cell activity and resonance (14), which indicates that 2TS is not specific of the complex architecture of the mammalian cochlea, but can be ascribed to the active resonance of a nonlinear oscillator. This suggests the use of a generic model of a compressive, self-tuned resonator with mass, stiffness and damping, such as the one provided by the mathematical formalism of a critical oscillator near a Hopf bifurcation (105), solvable using reasonable approximations (110). At low stimulus levels, a controlled negative resistance almost cancelling the viscous drag poises a critical oscillator near self-oscillation and allows cochlear amplification. The resistive term also contains a stabilizing amplitude-dependent term $|z|^2z$ (where $z$ denotes the amplitude) that generates compression near resonance by decreasing amplification at increasing input levels (34, 60).

Most characteristics of 2TS can thus be predicted in terms of a competition between probe and suppressor to drive a critical oscillator (FIGURE 10) (64, 110). In the presence of a single tone, the oscillator tuned to it behaves as a sharp filter with narrow bandpass. Thus, if a second, interfering tone is added at a frequency far enough from resonance, it is filtered out and bears no influence. When the interfering tone gets nearer to the resonance of the filter, all the more when it is louder, the oscillator starts responding to it, which increases $z$ and the damping, $|z|^2z$. Increased damping decreases the response to the probe tone, which gets suppressed. A critical oscillator becomes linear if it is stimulated at frequencies differing from its resonance frequency, which explains why the suppressive effect of a low- or high-frequency masker decreases when the probe tone frequencies moves away from CF (233). On FIGURE 10 and in
Physiol Rev  VOL 93  OCTOBER 2013  www.prv.org

Reference 110, it can be seen how the general critical-oscillator framework accounts for the outcomes of most experimental situations of 2TS mentioned above. The model also reveals how suppression acts the other way round as the probe tone itself, by also influencing the damping term, competes with the responses to simultaneously presented tones and tends to suppress them. A general rule that holds regardless of the exact mathematical model of nonlinearity is that for suppression to be significant, the suppressor level at the input of the nonlinearity must get within 10 dB of the probe stimulus (64), which can be achieved either when the suppressor frequency moves closer to CF or, at fixed frequency of the suppressor, when its intensity increases. Thus the fact that a realistic 2TS stems from the parsimonious formalism of Hopf oscillators does not refute other models of cochlear nonlinearity (see sect. VIII A).

Whereas, for a single nonlinear oscillator, suppressors below and above the resonance frequency play symmetrical parts, it is not the case for the cochlea. The pattern of Békésy’s traveling wave accounts for the so-called upward spread of masking (see sect. IX A), whereby a low-frequency suppressor influences higher frequency probes, whenever the tail of the suppressing envelope presents a significant amplitude at the place tuned to the probe (see sect. IVA). In contrast, high-frequency interfering tones, unable to reach beyond their characteristic place, can suppress only if they peak near enough to the probe to influence its traveling wave along the short section where its amplification takes place (203). The extent of the active mechanism along the BM can be evaluated in this manner. Patuzzi (203) concluded that activity spans a relatively broad distribution of OHCs, e.g., 650 μm basal to the characteristic place of an 18-kHz probe tone in the guinea pig (203).

It might seem that suppression means “loss of information.” However, by ensuring that when a given filter receives several inputs, the larger one (likely the one at resonance) will win the competition, 2TS also improves spectral analysis. The spectral components most easily suppressed at one site of the cochlea are off-resonance ones. Actually, they will propagate to their own CF locations, where the competition will favor them at the expense of off-frequency competitors. In this respect, 2TS acts as a contrast enhancer (14, 262). Nonetheless, very loud low-frequency suppressors will irretrievably suppress all competitors.

B. Distortion-Product Otoacoustic Emissions Conserved Throughout Evolution

Two-tone suppression can lead to the complete masking of a soft tone by a loud one, but in the two-tone paradigm discussed in this section, with similar moderate levels at $f_1$ and $f_2$, 2TS is weak. The second nonlinearity typical of this two-tone paradigm is an enriched frequency spectrum of BM motion, with new spectral components at integer combinations of frequencies $f_1$ and $f_2$. As already seen in section II, these audible DPs objectively detected in the ear canal (120) differ from instrumental nonlinearities by their existence even at low stimulus levels (essential nonlinearity, Ref. 79) and their spectral content with a predominance of odd-order DPs (see sect. XI; Ref. 267).

The presence of DPs has been ascertained at many spots along the mammalian cochlea, and in the first place, on the BM, with the help of optic interferometric measurements at basal locations (FIGURE 11) (229, 230) and near the apex, with slightly different characteristics (43, 86, 125). DPs are found in the pressure wave in SV throughout the cochlea (13), and near the BM in ST (54). Last, the CM reflecting transduction currents through basal OHCs contains DPs (289).

Because the common source of DPs is the nonlinear mechanics inherent to the gating of mechanosensitive channels, it is not surprising to observe DPs and DPOAEs in lower vertebrates, as ubiquitous as SF-OAEs (15). Even in insects, DP equivalents in vibrating structures have been described (82, 285). The idea to take advantage of this evolutionary-rooted similarity has led to the development of simple ex vivo systems in which DPs and attendant characteristics, activity, frequency tuning, and compressive gain can be studied together in the absence of complex propagation issues specific of the cochlea, such as BM resonance, propagation of vibrations, interaction of DPs from many neighboring cells, and recombination from discrete DP sites of generation. For example, hair cells from the neuroepithelium of bullfrogs, in settings containing two-fluid compartments preserving their natural ionic environment, can be excited individually either by fluid jets or calibrated glass micropipettes acting on their stereocilia bundle. In this non-mammalian, nonauditory hair cell, the MET channel gating compliance is qualitatively similar to that of OHCs, and amplification activity exists even though it involves only the hair bundle.

DPs have been observed in the current through hair cells, in their membrane potentials (106), and in hair-bundle displacements (14). The latter case offers a neat illustration of how 2TS and DP properties emerge together from the active or passive properties of hair-bundle motion (14). The non-linearities in the mechanical response of the frog saccular hair bundle were explored. FIGURE 12A, which is modified from Barral and Martin (14), depicts the compressive increase of the hair bundle’s displacement observed for a stimulus at the characteristic frequency of spontaneous oscillations of the hair bundle (its resonance frequency). An interfering tone of about the same frequency produces strong 2TS and decreases the compression (FIGURE 12A). Off-resonance, the mechanical response of the hair bundle to a single stimulus, is decreased at low stimulus levels but grows faster with increasing level than near resonance, as...
compression does not exist (FIGURE 12B). Likewise, 2TS no longer acts (FIGURE 12B). The narrow range of efficiency of 2TS was as predicted by the model of Jülicher et al. (110) (FIGURE 10). As regards intermodulation, with stimuli near resonance, the DPs differed from the off-resonance case by the shallow slope of their decrease when stimulus levels decreased [near 1 dB/dB, as in Goldstein’s experiment (79) vs. 2.7 dB/dB off-resonance], that ensured their persistence at low stimulus levels, and the fact that fewer DP frequencies stood out, essentially the \( n f_1 - (n-1) f_2 \) and \( n f_2 - (n-1) f_1 \) odd-order series (with \( n = 2, 3, 4 \ldots \)) flanking \( f_1 \) and \( f_2 \) (FIGURE 12, C VERSUS D) (14). As we will see in sections V, C–F, and VIII A, many of these characteristics of single-hair cell DPs also show up in mammalian DPOAEs.

C. DPOAEs, a (Not so Transparent) Window on the Inner Ear

Direct access to cochlear micromechanics and OHC operation to assess the enhanced sensitivity, compression, and filtering requires the opening of the bony cochlea, with a high risk of damaging the vulnerable active mechanisms. Even minimal damage would affect the validity of measurements. Furthermore, only a few cochlear spots are accessible near the round window and at the apex. Conversely, DPOAE measurements are totally harmless and straightforward over a large range of frequencies, and DPOAEs are easier to extract than SFOAEs as their frequencies differ from those of the stimuli. Yet, their relation to cochlear amplification is not straightforward, and this requires the issues of sites of generation and of propagation mechanisms [absent from the single hair-cell model of Barral and Martin (14)] to be teased apart.

A prerequisite is to understand the logics behind the conspicuous changes in DP amplitudes and phases when stimulus characteristics are modified. Consider for example the \( 2f_1-f_2 \) DP, usually the largest one. For a given \( f_2 \), its amplitude varies considerably when the \( f_2/f_1 \) ratio varies, from 1.60 to near 1.00 (so-called \( f_1 \)-sweep). This dependence clashes with the intuitive assumption that the dominant contributions to a DP come from where maximum interference occurs between \( f_1 \) and \( f_2 \), near the peak response of the BM to \( f_2 \). As the DPs of the \( f_1 \)-sweep come from the same OHCs, should not their amplitude be constant? Experimentally, a “bandpass filter”-like characteristics is observed, such that the DPOAE at \( 2f_1-f_2 \) reaches a maximum near \( f_2/f_1 = 1.20–1.25 \) (FIGURE 13) (5, 88). The heard DP is not constant either, but gets louder at decreasing ratios (79).

A tentative explanation for the “bandpass filter”-like outcome of an \( f_1 \)-sweep is the interplay of a mechanical filter near the generation site, tuned to a lower frequency. At optimal ratios, \( 2f_1-f_2 \approx f_2/1.4 \), so that a suitable filter would have to be tuned at half an octave below the main filter. The
tectorial membrane does present such a characteristic radial resonance (84, 314), which suggested it as a possible candidate (5, 26). Another, nonexclusive possibility, would be that at $f_2/f_1$ ratios near 1, the $2f_1-f_2$ and other lower band DPs, by falling near the primary frequencies, undergo 2T$S$ (113) strongly decreasing their amplitudes. In these frameworks, the noninvasively accessible DPOAEs would uncover the influence of subtle cochlear micromechanics. Yet, the most conclusive interpretation so far relies on the phase behavior and directionality of interfering traveling waves from distributed DP generation sites (65, 244), developed in the following section.

D. Distributed Sources and Propagation(s?) of DPOAEs

The propagation of DPs from their generation places follows the same principle as pure tones and SFOAEs (see sect. IV, B and E). As intermodulation creates new frequencies and wavelengths, the prediction of phase relationships between the wavelets emitted from different generation sites at the frequency of each DP is significantly more complex. Its mathematics (244, 266) are outlined accessibly in APPENDIX II. The particular case of $2f_1-f_2$, the largest DP, is examined there without loss of generality.

Potential DP generation sites are distributed over a broad interval where the envelopes at $f_1$ and $2f_1-f_2$ overlap and OHCs respond to both frequencies, particularly, but not only near $f_2$. The lowest stimulus levels in most publications range from 40 – 60 dB SPL, and at these levels, the envelope of BM motion definitely has a broad tail (228). If we consider a DP-generating region along the BM, stimulation evokes DP-wavelets, each at a particular location of this region, from which it travels in two directions, apically to the CF location and basally. At the stapes or at their CF location, all wavelets are vectorially added and, depending on their respective phases, cancellation or enhancement occurs. The DP level at the stapes determines the DPOAE level, and the DP level at CF determines the loudness of the heard combination tone.

Due to scaling symmetry, the vibration patterns of $f_2, f_1,$ and $2f_1-f_2$ are just translated versions of each other along a logarithmic representation of the cochlea. The phase at $f_2$ rotates fast when this primary nears its CF location, while at the same spot, $f_1$ and $2f_1-f_2$ accumulate less and less phase with increasing $f_2/f_1$ ratio, as the peaks at $f_1$ and $2f_1-f_2$ move farther apically (see APPENDIX II). A DP wavelet emitted at some place along the region of generation, starts with a phase lag due to the phase lag of the stimuli on their way to

FIGURE 12. A: growth function of the motion of the stereocilia bundle of a hair cell from a frog sacculus in response to a tone at frequency $f_1$ ($f_1 = 13$ Hz, near the characteristic frequency of spontaneous oscillations, 11 Hz), in the presence of a suppressor tone at frequency $f_2 = 12$ Hz, added to the test tone at increasing amplitudes (symbols: right panel). B: same hair bundle as in A, tested off-frequency, with $f_1 = 130$ Hz and $f_2 = 120$ Hz. C: frequency spectrum of hair-bundle motion in response to a two-tone stimulation at $f_1 = 20$ Hz and $f_2 = 22$ Hz, near resonance, with rather high stimulus forces (red labels: main intermodulation frequencies). D: same cell as in C, with off-resonance two-tone stimulation at $f_1 = 200$ Hz and $f_2 = 220$ Hz. At lower stimulus forces [FIG. 3, A AND C, in Ref. 14], the two-tone stimulus of C still elicited visible DPs, whereas that of D did not. [From Barral and Martin (14). Copyright 2012, with permission from National Academy of Sciences, USA.]
the generation site, then accumulates another phase lag while traveling from its generation site to the measurement place, either the stapes or the CF. Here the DP results from vector addition of wavelets from all generation sites. The dependence of the DP amplitude on the $f_2/f_1$ ratio is explained by the computation, for each ratio, of the difference in overall phase among wavelets. A large DP stems from wavelets with the same phase accumulation, phase coherence (244). In the absence of phase coherence, the DP is small. Inspection of the vibration patterns of $f_2$, $f_1$, and $2f_1-f_2$ leads to predict an $f_2/f_1$ ratio-dependent directionality, properly accounting for a large DPOAE amplitude (corresponding to backward propagation of wavelets) at intermediate ratios (near 1.25), and for a loud heard DP cubic difference tone at small ratios (near 1).

The fact that a theory robustly explains an experimental fact is encouraging yet constitutes no definite proof. This theory makes use of the powerful mathematical wavelet description of Huygens, the main difficulties of which are to make sure that all possible wave sources are included, and that all propagation modes are properly described. It cannot be excluded that new sources be uncovered and that the reverse propagation of DP waves eventually turns out to use another mode than the BM traveling wave (see sect. VIH, although we will see there that it is unlikely). Then the theoretical basis of the ratio dependence of DPOAEs and DPs would have to incorporate the new sources and propagation paths.

**E. One Nonlinearity, Several Paths**

According to the two-mechanism model of Shera (246), the forward-propagating wavelets at $2f_1-f_2$, when their phases are coherent enough to let them reach their CF location, are back-scattered on cochlear irregularities there, by linear coherent reflection (see sect. IVF). The DPs reflected by this mechanism, primarily arose from the basal nonlinear interference between $f_1$ and $f_2$. At the stapes, the DP from the coherent reflection site combines with the DP directly emitted backward from the places of nonlinear interaction. The two mechanisms and propagation modes identified by this model for the measured DPOAEs share the same initial nonlinear stage. As we will see, experiments reveal two distinct phase behaviors of the DPOAEs: one with a uniform phase independent of primary frequencies, and a second one with a phase that rapidly rotates with primary frequencies. This accords quite well with a two-path model, either the one developed by Shera (246) or any alternative one predicting the two phase behaviors. The first path is straightforward and its existence unchallenged. Part of the DP directly comes backward from the place of nonlinear interaction, and its uniform phase relates to scaling symmetry (see sect. VD).

As for the second path, the model of coherent reflection on local BM irregularities (246) implies a suitable frequency-dependent phase. One alternative to it has been proposed recently (217), in which the Reissner’s membrane (FIGURE 1A) is assumed to support a bidirectional propagation mode, the backward component of which would contribute to the overall DP at the stapes. Reichenbach’s model postulates that distortion on the BM evokes a traveling wave along the Reissner’s membrane which, below 1 kHz but not at high frequencies, can significantly interact with the BM. It also predicts that DPs, both at $2f_1-f_2$ and $2f_2-f_1$, travel along the Reissner’s membrane, forward and backward, by a short-wavelength mechanism similar to capillary waves at the surface of a water tank. Interestingly, along the nonresonant Reissner’s membrane, these waves do not incur the limitation that forbids propagation across a resonant position. For the BM-propagated $2f_2-f_1$ DP, backward travel from the place where maximum nonlinear interaction between primary tones occurs is impossible, as it is apical to the $2f_2-f_1$ CF location. Upper-band DPs experience a shift of their generation place toward their own CF location (163, 165), where nonlinearity is less. The Reissner mode, by letting $2f_2-f_1$ DPs travel from the place tuned to $f_2$ along the whole cochlea, might produce a larger outcome. Experimental validation, by interferometric measurements in rodents (217), was restricted to showing the presence of a 1-kHz $2f_1-f_2$ DP on the Reissner’s membrane, traveling forward beyond its CF location with a fast rotating phase, over places where the BM could not vibrate at $2f_1-f_2$. Important unsettled issues are whether large DPs could show up on the Reissner’s membrane, particularly at $2f_2-f_1$, and not only travel backward, but also influence the BM-propagated mo-
tion at $2f_2-f_1$, even at higher frequencies where the motion of Reissner’s membrane is supposed to influence BM motion very little.

The extent of generation sites and mixture of propagation paths are essential issues for the outcome of DPOAE measurements from a locally damaged cochlea to be interpreted usefully. An unfortunate combination of stimulus parameters, favoring the coexistence of several contributions with similar levels and different phases in a normal cochlea, might generate intractable complexity offsetting the advantage of DPOAEs as local probes of active cochlear mechanics (51). Local damage removing only one contribution might even fail to be noticed on the overall map of DPOAE levels. This is why intense research has been devoted to identifying choices of stimulus parameters that emphasize one mechanism and path at the expense of the other ones. It is helpful in practice that these choices for the $2f_1-f_2$ DP extend to all other lower-band DPs ($3f_1-2f_2$, $4f_1-3f_2$, etc.). Phase criteria have been developed for separating the sources when they coexist (134).

As seen above, the lower-band DPs traveling straight back from the nonlinear place dominate at intermediate $f_2/f_1$ ratios (>1.10 for $2f_1-f_2$). Their principle of generation and propagation essentially depends on how waves at $f_1$ and $f_2$ overlap. When $f_2$ and $f_1$ vary at constant $f_2/f_1$ ratio, scaling symmetry applies, vibration patterns are conserved, and the phase at any point moving with the traveling wave envelopes changes little, thus the generated DPs are “wave-fixed” (134). In two-dimensional plots of phase against DPOAE frequency at varying $f_2/f_1$ ratios (133, 168) (FIGURE 14), the independence of DPOAE phase on DPOAE frequency at constant $f_2/f_1$ ratio translates into a horizontal banding pattern (FIGURE 14B, top). Conversely, the DP components traveling forward from the generation place then backscattered on local irregularities near the DP CF place (dominant at small $f_2/f_1$ ratios) are “place-fixed” as they strongly depend on the irregular places. As the stimuli are swept in frequency, the excitation patterns move along the BM and different reflecting sites are encountered, with different phase rotations as a result. The reflection-mechanism DPs are thus characterized by a vertical banding pattern.

![FIGURE 14. Maps of DPOAE level (A and C) and phase (B and D) against DPOAE frequency (horizontal axis) and $f_2/f_1$ frequency ratio (vertical axis) in the ear of a normally hearing rabbit. The horizontal dashed white line (corresponding to $f_2/f_1 = 1.00$) splits the plots into an upper and a lower part representing the characteristics at $2f_1-f_2$ and $2f_2-f_1$, respectively. Different colors are used to represent DPOAEs with amplitudes or phases falling in different dB or 45°-wide intervals. Horizontal banding in the phase map means that DPOAE phase at fixed $f_2/f_1$ ratio is independent of DPOAE frequency: it is determined by $f_2/f_1$ and obeys frequency scaling. Vertical banding means that DPOAE phase, regardless of the $f_2/f_1$ ratio, is determined by some coherently reflecting microstructure, e.g., at the (fixed) place tuned to the DPOAE. Yet, compared with A and B, C and D are plotted in the presence of an interfering tonal stimulus near $2f_1-f_2$, meant to jam the coherent-reflection process at the place tuned to $2f_1-f_2$. Its failure to influence the banding pattern suggests that vertical bands may be due to coherent reflection on a more basal microstructure. [From Martin et al. (168). Copyright 2009, with permission from Acoustical Society of America.]](attachment:figure14.png)
pattern (FIGURE 14B, bottom) contrasting with the horizontal pattern of wave-fixed DPs. Within one vertical band, the measured phase is independent of $f_2/f_1$ as, for a chosen DPs frequency, the scattering places remain fixed (since coherent reflection is supposed to occur at the place tuned to the DP frequency).

A caveat is that, even though the lower-band DPs arising from coherent-reflection at the place tuned to their frequency are expected to generate a vertical-banding pattern, evidence has been produced recently that some vertically banding DP components originate from another place, basal to $f_2$, using methods more extensively described in section VIA and APPENDIX III. For example, FIGURE 14D shows a persistent vertical banding pattern despite the presence of an interference tone, which by being placed near the DP frequency, suppresses coherent reflection occurring at the DP CF place (hence the decrease in overall DPOAE level seen in FIGURE 14C). In other words, although consistent with coherently reflected DP components, vertical banding is not a signature of apical coherent reflection (167). This issue is also raised by upper-band DPs. In their case, only place-fixed components have been observed, with fast phase rotation and vertical banding at all frequency ratios (see FIGURE 14B, bottom). It is a common observation in mammals that their levels are smaller than their lower-band cousins of the same order, but not in chicken, lizards, or frogs (15). A first explanation considers that, as the place near $f_2$ with maximum nonlinear interaction has a resonance frequency less than DP frequencies, the BM is too slack for launching these DPs directly toward the stapes. Upper-band DPs would be more efficiently produced more basally (163, 169), with a region of place-fixed reflections occurring somewhere between the DP generation region and the DP CF place (133).

With regard to both upper- and lower-band DPs, it is noteworthy that at high primary levels, the tails of the envelopes of BM motion at $f_1$ and $f_2$ extend far basally, where the absence of compression allows a steeper growth than at resonance. At $>70$ dB SPL, they are almost as large as they would be more apically at the CF location. Conceivably, the interactions in OHCs responding at both frequencies remain nonlinear enough, and phases along the interaction interval vary smoothly enough, that significant DPs could come from basal places. Some authors found no evidence of it (302), while others did (170). The complexity of wave combinations is such that DPs from basal sites may be easily overlooked because their removal (by basal lesions or by application of a third interacting tone well above $f_2$; see APPENDIX III) hardly influences the overall DP levels. Vector subtraction has to be used for separating the basal contributions from the standard ones. Computations of vector differences between $2f_1 - f_2$ DPOAEs without versus with an interfering tone (that turns off the contribution from the place where it peaks) recently led to the observations that in small rodents, there is no significant contribution from the CF place of the DPOAE; and that a large DPOAE component sensitive to high-frequency interfering tones, one-third octave above $f_2$, thus likely generated basally, displays a vertical banding pattern (167). The proposed explanation still rests on the idea that vertically banding components come from a relatively fixed location when primary frequencies are swept, but assumes that what determines this fixed basal location is some constraint to which the DPOAEs have to conform. For example, for the $2f_2 - f_1$ DPOAE, contributions apical to its CF location would be absorbed when traveling over this place, and for the $2f_1 - f_2$ DPOAE, suppression by high-level primary stimuli (113) would spare only the contributions from sites well basal to stimulus CF locations (167). Without contradicting the existence of coherent reflection, and without the need to invoke the Reissner’s membrane as an alternative propagation medium (217), such evidence strongly supports a complementary place-fixed mechanism involving basal generation sites. At increasing stimulus intensities (70 dB SPL or more), other combinations of interactions have been suggested. For example, $2f_2 - f_1$ could be produced in two stages, generation of the harmonic $2f_2$, which in turn could interact near its own basal peak with $f_1$, to generate a quadratic product $[2f_2 - f_1]$ (64). However, these mechanisms can likely be neglected as long as stimulus levels are kept $\leq 65$ dB SPL, as was the case for rodent ears measured in Reference 167.

A currently active issue is whether this complex behavior of DPOAEs in a cochlea can be found in other vertebrates without tonotopically organized BM, electromotile hair cells, or tectorial membrane (15). In birds in which the auditory sensory-cell dichotomy is less marked than in mammals, the dependence of DPOAE phase on primary ratio is similar to mammals, with clear place- and wave-fixed components. In geckos with no BM traveling wave, more similarities with the DPOAEs of mammals have been found above than below 1 kHz, as for the latter frequencies, no evidence exists for the distortion versus reflection classification. In frogs with no flexible BM, no wave-fixed behavior has been found (15), and the question is open whether a traveling wave could still exist along another flexible structure. Solving these issues would help to establish what minimum architecture leads to a physiological complexity comparable to mammalian DPOAEs, which in the previous sections had been explained mainly in terms of bidirectional wave propagation from distributed places, with little reference to anatomy.

F. Spectral Characteristics and Growth

From an engineering standpoint, the spectral pattern of distortions coming out of any nonlinear black box provides specific insights into its distorting mechanism, nonlinear transfer function, and OP. A caveat for applying this ap-
proach to the cochlea is the requirement that DPs come directly from one emission place, which we now know is at best an approximation. Nonetheless, it reveals several relevant features.

At small enough input amplitudes, a nonlinear transfer function can be developed as a Taylor power series for which the coefficient of the $n$th power term is the $n$th derivative of the transfer function evaluated at the OP and divided by $n!$ (For a linear system, only the first derivative differs from 0 as the transfer function is a straight line.) Input signals considered throughout this section are of the form $A_1 \sin 2\pi f_1 t + A_2 \sin 2\pi f_2 t$. In the power series, terms such as $A_1^2 A_2 \sin^2 2 \pi f_1 t \sin 2 \pi f_2 t$ show up and, with the help of some elementary trigonometric formulas, can be converted into sums of terms corresponding to combination tones, among which $\sin 2\pi (2f_1 - f_2) t$, the $2f_1-f_2$ DP (its cubic nature mentioned in sect. IIIA) reveals itself in its coefficient $A_1^2 A_2^2$, proportional to $A^3$ in the commonest situation with $A_1 = A_2 = A$. The resulting expression of the amplitude at $2f_1-f_2$, for example, is far from simple, except its lowest-order term (62, 66)

$$A_{DP} = A_1^2 A_2^2 \left[ \frac{3}{4} a_3 + \frac{5}{8} a_5 (2A_1^2 + 3A_2^2) + \frac{105}{64} a_7 (A_1^4 + 4A_1^2 A_2^2 + 2A_2^4) + \ldots \right]$$

(4)

A similar treatment of all terms of the Taylor series predicts the distortion spectrum as a series of equidistant peaks at intermodulation frequencies.

Indeed, this is observed in physiological responses from auditory sensory organs (230), the largest DPs being odd-order ones at $(n + 1)f_1 - nf_2$ (lower band) and $(n + 1)f_2 - nf_1$ (upper band) $(n = 1, 2, \ldots)$ surrounding the main two peaks at primary frequencies. The higher the order (defined by $p = 2n+1$), the smaller the DP as the terms in the Taylor series (assuming small amplitudes) get increasingly smaller with increasing exponent. At increasing primary levels, the observed tentlike pattern (FIGURE 11) changes little as all DP peaks grow at approximately the same rate of decibel per decibel increase of stimuli. In linear physical units, the amplitude of the $p$th-order $DP$ grows as the amplitude of the input to the nonlinear system to the power $p$ (thus at $p$ dB per dB increase in input). Other features are not generic but specific of the distorting system, and here, of its physiology, as already seen for the slow average growth of DPs with SPL, which betrays the compressive amplification (see sect. II A; Ref. 79).

Another specific property of physiological DPs is that even-order components, such as the quadratic DP at $f_2-f_1$, are smaller and more labile in many species (37). In contrast to odd-order DPs that depend on asymmetric components of the nonlinear transfer function of the MET channels ($a_3, a_5, a_7, \ldots$, in Equation 4), even-order ones rely on symmetric components (70). The local symmetry of the MET transfer function and the exact content in odd- and even-order terms of its Taylor expansion, depend on the OP position. It is thus expected that its displacements influence the whole series of DPs in a coordinated manner. With the assumption that the OP moves along a second-order Boltzmann function, it was modeled that when the OP is at the inflection point where the steepest slope ensures maximum cochlear amplification (at low sound levels), the even-order $2f_1-f_1$ DP is minimum whereas the $2f_1+f_2$ DP is maximum (70). This provides one explanation of the predominance of odd-order components in usual DPOAE spectra (FIGURE 2). However, as soon as the OP moves away from the inflection point, the predicted DP pattern dramatically changes with a sharp increase of the $2f_1-f_1$ DP, together with the $2f_1$ and $2f_2$ (even-order) harmonics, whereas the $2f_1+f_2$ DP sharply decreases in amplitude. Larger OP shifts lead to reciprocal changes of the even- versus odd-order DPs (70).

Experimentally, low-frequency biasing of the cochlear partition has been used as a tool for displacing the OP along the cochlear transducer function and reconstruct its shape (21, 29). Other experiments have exploited the changes in the balance between odd- versus even-order DPs to monitor endolymph volume disturbances produced by direct fluid injection in scala media (253). The outcome of these manipulations in terms of complex combinations of $2f_1-f_2$ and $2f_2-f_1$ DP changes was similar to those of former experiments attempting to modulate the cochlear amplifier by decreasing the endocochlear potential using furosemide, a loop diuretic (180, 187, 253), which suggests that volume-induced OP shifts occurred in such circumstances. Even-order DPOAEs may also provide a convenient tool for monitoring the medial olivocochlear efferents, a neuronal pathway that innervates the OHCs via cholinergic synapses. Physiological activation of these efferents by the presentation of sound in either ear produces a decrease in the gain of the cochlear amplifier, likely beneficial in the presence of background noise (83). By affecting OHC function, efferent activity influences DPOAEs which therefore can serve for assessing the efferent control (40, 131, 187, 211). The $2f_2-f_1$ DPOAE responds to contralateral noise by a disappointingly small decrease in amplitude, but the initially weak $2f_2-f_1$ DPOAE component can almost reach the level of its 10-dB-larger $2f_1-f_2$ counterpart when medial efferents are activated by low noise levels (30 dB SPL) (3). As, in this experiment, efferent activation also strongly modified the phase of changes exerted by a biasing infrasound, a common contribution of OP shift in both paradigms was suggested (3). The changes in even- versus odd-order DPOAEs after continuous low-level sound stimulation, noticed long ago, had already been attributed to efferent activity (25).
ously nonmonotonic DPOAE growth functions with deep minima (notches) that separate two portions with different slopes (66, 152, 173, 180, 292). Such notches are consistently observed in small mammals for primary levels around 60–70 dB SPL, at many combinations of primary frequencies (179, 292), but they are less frequent in human ears (291). A single measurement at the notch could lead a naive experimenter to conclude that absent DPOAEs pinpoint damaged OHCs, whereas a slight arbitrary shift of stimulus level would have produced different results. The theoretical background of these notches, therefore, has to be explained; otherwise, one would have to deny the practical validity of DP measurements.

A first generic explanation of a notch has been brought forward (179), drawing on the characteristic 180° phase change observed whenever a nulling happens (152, 180, 182, 292) (FIGURE 15). This pattern suggests interference between two DPs of opposite phases coming from two different sites or mechanisms. When they happen to have equal amplitudes, cancellation happens. On one side of the notch, one mechanism dominates while on the other side, the other mechanism takes over, and their physiological meanings may be totally different. The observation of different slopes on both sides of a notch has suggested that the DPs from active compressive mechanisms, the ones discussed so far, dominate at lower levels while at high levels, passive DP sources with a steep 3 dB/dB growth would take over. Their lack of vulnerability to cochlear damage whereas at the same time, low-level, slow growing DPs have vanished, might point to their being valueless. Actually, there is mounting evidence that the nonlinear sources of slow- and fast-growing DPs likely are the same MET channels of the same OHCs. When impaired, these OHCs no longer exert compression so that the DPs that they produce are only the fast-growing ones (9, 183). The reported lack of vulnerability depends on what means were used to “silence” OHCs. Loop diuretics or transient hypoxia, for example, by decreasing the endocochlear potential without affecting OHC structures, kill low-level DPs without affecting high-level ones (181, 292). Conversely, factors destroying OHCs and their MET channels [aminoglycosides (28, 292); genetic defects (33, 96); excessive noise (9)] readily remove any trace of the “passive” DP component, right where OHCs have been damaged, while in the same cochlea, places where OHCs survive in an inactive state still produce DPs at high stimulus levels (9). In hair-bundle preparations from the bullfrog’s sacculus (14), the DPs produced when a hair bundle is passive, either when stimulated far from its intrinsic resonance or at high level, keep providing specific information on the function of MET channels, even though the hallmarks of activity have been lost. In the more complex situation of the cochlea in vivo, the possibility that significant DP contributions come from basal OHC-stimulated off-resonance is now strongly substantiated (170).

Nevertheless, an alternative generic explanation of growth-function notches has taken shape, implying a single compressive nonlinearity and changing level (66, 153). Several mathematical models of nonlinear transfer functions have been tested for their influential parameters, in search of signatures of a single-site mechanism producing a notched pattern. The origin of nulls has been interpreted in terms of the cancellation of terms in the Taylor series describing the nonlinear transfer function, due to the mixing of an odd-order nonlinear term with the next highest odd-order nonlinear term (66). In this framework, nulling turned out to be a commonplace result of the nonlinear nature of a transfer.

**FIGURE 15.** Dependence of the 2f1-f2 DPOAE level (A) and phase (B) on primary levels in the ear of a guinea pig before (thick line) and at varying times (as indicated in the inset) after injection of furosemide, a loop diuretic which, by decreasing the endocochlear potential, affects cochlear amplification. A deep notch splits the initial growth function into a shallower low-level segment and a steeper high-level segment (A), with a half-cycle difference in phase between the two segments (B). Furosemide essentially affects the low-level segment, as if two different sources of DPOAEs were combined, an active one at low levels, vulnerable to furosemide, and a passive one at higher levels and with a steeper growth, not physiologically vulnerable (180). Their combination would generate a notch when their levels are similar because they happen to be out of phase. An alternative explanation assumes a single nonlinear transfer function with changing gain and operating point. This function would generate both segments and even predict the observed shift (ΔL) in notch position with decreasing amplifier gain. [From Khashkin et al. (152). Copyright 2002, with permission from Acoustical Society of America.]
function. Several properties of this function (even symmetry, degree of asymmetry, severity of clipping) could equally easily produce nulling, while a mere shift of the OP could profoundly alter the ability to produce it (TABLE 2).

Small changes in the frequencies of stimuli, not only their levels, can also lead to large changes in DPOAE levels. A so-called fine structure shows up in the DP grams (DP vs. frequency plots) of human ears, consisting of up to 20-dB-deep troughs (92) (FIGURE 16) that modulate the overall level-versus-frequency pattern in a quasi-periodic manner. Notches are more densely packed in human ears than in other species, about 1/10th of an octave apart (92). It appears that their density is inversely proportional to DP time delay, of the order of 10 ms around 1 kHz in humans. The fine structure persists in an interval of primary levels over 20 dB wide (FIGURE 16, A versus B). Its acknowledged explanation is an already familiar one (see sect. V E), building on the existence of the two discrete DP contributions from the nonlinear and coherent-reflection sites (111, 133, 263). As their relative phases rotate with frequency, an interference pattern of regularly spaced nulls is created in the DP gram. The lack of fine-structure in small mammals comes with the lack of reflection component from the $2f_1-f_2$ place, inferred in Reference 167 from the failure of an interfering third tone (see APPENDIX III) to exert any change in the DP pattern when set near $2f_1-f_2$, where it was meant to suppress the reflection mechanism.

Human ears contrast with rodent ears in that no notch in growth function is observed around 60–70 dB SPL primary levels; nonetheless, another type of notch in the growth function has been reported in close relation to the commonly observed deep fine structure (92). These notches correspond to the fact that the fine structure tends to shift in frequency when the primary level changes so that the DP grams at different primary levels are not parallel. At fixed frequency, this generates irregular growth functions with dips around 50 dB SPL, with little difference between the slopes below and above a dip (92). For human DPOAEs, growth notches and fine structure thus rest on a common background, the distortion versus coherent reflection dualism. In rodents, the explanation of growth notches is more contentious. To decide whether the notch in a DPOAE growth function is due to the interference between two contributions from different sites, not the coherent-reflection one but other possibilities have been invoked (170), or

| Table 2. Possible (not mutually exclusive) explanations for the presence of notches in DPOAE ($2f_1-f_2$) growth functions and of microstructure in DPOAE versus frequency plots |
|-----------------|-----------------|
| Explanation | Reference Nos. |
| Two emission principles, one active and vulnerable, one passive and robust | 180, 292 |
| Two emission sites, one near $f_2$ or basal to it (distortion), one at the place tuned to $2f_1-f_2$ (coherent reflection) and/or two propagation modes with different phase rotations | 133, 170, 246 |
| One generation site, two propagation modes at different speeds along BM and Reissner’s membrane | 217 |
| One generation site, one propagation mechanism, notch due to a change in saturation on the transfer function at increasing level | 152 |
| One generation site, one propagation mechanism, shift in OP | 70 |
| Retarded action of an AGC on each primary, generating two spectral series, and interference between them. | 271 |
| Single hair bundle in its passive (off-resonance) regime, with a gating compliance strong enough to elicit negative stiffness | 14 |

DPOAE, distortion product otoacoustic emission; BM, basilar membrane; AGC, automatic gain control.

FIGURE 16. The high-resolution plot of DPOAE amplitude against frequency at fixed stimulus level and frequency ratio commonly displays pseudo-periodic notches, 5–10 dB deep (example in one human ear; A: high-level stimuli, $f_1 = 70$ dB SPL and $f_2 = 60$ dB SPL; B: low levels, $f_1 = 55$ dB SPL and $f_2 = 40$ dB SPL). Salicylate intake, which decreases cochlear amplification by affecting the electromotility of OHCs, affects the depth of notches and shifts their frequency (open circles: initial measurement; dashed and dotted black lines: measurements at various times during salicylate treatment; closed circles: post-control showing the stability of DPOAE fine structure). (From Rao and Long [213]. Copyright 2011, with permission from Acoustical Society of America.)
to the dynamic behavior of a single nonlinear source, a suitable experimental test would be to perform well-defined local damage, e.g., to OHCs in the basal cochlea, and examine whether it affects the notches produced by primary frequencies with more apical CF locations. Alternatively, DPs produced by the same stimuli could be measured at different places to check whether a notch is stable at all locations, in which case it would be an inherent part of a single nonlinearity. The two-site model predicts that the notch exists because of a particular 180° phase difference which would not hold if the measurement site moved closer to one site and further from the other one. The latter test was performed in a few guinea pigs as a complement of the protocol of Reference 13, the $2f_1 - f_2$ signal being measured in the ear canal as a DPOAE and in the perilymph of SV1 as a pressure wave, and it revealed that the notch in the DP growth function was present at both measurement sites (P. Avan, P. Magnan, and A. Dancer, unpublished data). In contrast, in the cat, the stable dip visible in the DPOAE fine structure at one $f_2/f_1$ ratio vanished when the intracochlear DP was noninvasively assayed using a third tone as a calibration probe (245), in support of a two-mechanism model. This discrepancy, yet to be explored, suggests that indeed, notches in DPOAE plots may arise from different mechanisms, among those theoretically available (TABLE 2), thus bear different significances.

H. DPOAEs: Their Timing and Backward Propagation

Evaluation of cochlear tuning is an important challenge in audiology, as decreased tuning may account for impaired detection of sounds in noisy environments. Direct measurements of tuning on the BM or in single-neuron responses are highly invasive, thus prone to damage the system. Behavioral measurements of tuning (psychophysical tuning curves or notched-noise masking experiments; Ref. 184) are time-consuming even in experienced subjects. Indirect objective measurements are suggested by the rule that the more narrowly tuned a filter, the longer the delay of its response. The cochlea can be considered as a filterbank, which led Shera (248) to analyze how cochlear tuning, cochlear filter delay, and OAE delay covary. Thus the topic of this section, i.e., how to measure the timing of OAEs, might provide reliable information about the other two.

A thorny issue is that filter delay is only one part of OAE delay beside the time traveling waves need to reach the nonlinear place(s) and the backward propagation time(s) of OAEs from the generation place(s) to the recording site in the ear canal (FIGURE 17). This is why so much energy had to be devoted to the first task of identifying the sites where OAEs are produced and their propagation pathways to the detection system. Strong evidence is now available concerning the existence of distributed sites of DPOAE production, all along the sites where primary envelopes overlap and not only where they peak. One direct backward pathway from the nonlinear generation sites and one forward then backward pathway ascribed to coherent reflection have been described, both influenced by the $f_2/f_1$ ratio (see sect. VE).

Noninvasive measurements of OAEs only provide access to their round-trip travel time, and this indirectly, except when viewed in the time domain using special averaging techniques (293). The phase plot of DPOAEs evoked by a fixed tone and a swept-frequency tone is approximately linear and from its negative slope, the DPOAE group delay in the ear canal can be inferred (see sect. IVF). The filter delay is the DPOAE group delay minus the travel times of the stimulus to the generation place and of OAEs along their propagation path(s). The forward travel time is given by the onset delay of BM response (231). The mainstream hypothesis is that DPs reach the stapes via retrograde BM traveling waves by the same mechanism and at the same speed as the forward traveling wave, in which case, DPOAE delay = filter delay + onset delays. In a thorough analysis of available data, Ruggiero (231) listed all possible alternatives to this view, among which that whereby OAEs propagate via very fast acoustic compression waves in the cochlear fluids, implying a different prediction: DPOAE delay = filter delay + onset delay. For DPs to move the stapes and produce DPOAEs, the important parameter is the pressure difference between stapes footplate and round window, not BM motion as was the case for sensory cells, and a compression wave might achieve pressure transmission efficiently (296). The stapes footplate of mammals is far from the basal BM so that to reach the oval window, DPs have to travel through the bulk of SV fluid at least for the last stage of their journey (219).

A first observation is that the formula giving the filter delay combines several components, and cumulative errors can lead to largely erroneous estimations. The most recent experimental attempt at improving the accuracy of OAE delay measurements (176) used a modified emission technique designed for emphasizing the contribution of the return journey to the emission delay. Partial suppression of the SFOAE at frequency $f_2$ was produced by a complex of closely spaced low-frequency tones around frequency $f_1$, inspired from Reference 192. The resulting modulation of the $f_2$ SFOAE is nothing but intermodulation, which thus translates into DPOAEs consisting of a series of spectral components, at $f_2$ plus the frequency separation between the components of the $f_1$ complex. Frequency $f_1$ itself bears no influence in this series thus can be shifted at will, and for instance can be set very low. Then, the forward journey of an $f_1$ tone to the place tuned to $f_2$, where OAEs emerge, is shortened to a minimum, and the OAE phase behavior is dominated by the reverse travel of the OAE complex. The authors found a reverse travel time of $\sim 400 \mu s$ or 2.5 cycles...
As indirect measurements are inconclusive even when confounding factors are kept to a minimum, invasive experiments are required to definitely establish by which mechanism DPOAEs reach the stapes to exit the cochlea: slow backward traveling waves along the BM, fast compression waves through SV fluid, or alternative paths such as that recently suggested along the Reissner’s membrane (217). The common goal of the scarcely published experiments was to assess the timing of DPs along their way, from comparisons of phase data collected at different spots along the cochlear partition. Three different experimental protocols were used. In the first one (13), the authors inserted a sensitive piezoresistive hydrophone in the SVs and STs of the first and second turns of the guinea pig cochlea. A second group (219) performed interferometric measurements of gerbil BM motion at different places extending for ~1 mm in the longitudinal direction (93) and of stapes motion (94). Last, spatial variations in intracochlear pressure were measured with the help of a fiber-optic pressure detector at different places close to the BM in ST in the basal turn of gerbil cochleas (53). All investigators measured DPs in response to intermediate level primaries (40–70 dB SPL) in cochleas with preserved sensitivity, with the thresholds of neural responses after the DP measurements being within a few decibels of their initial values.

What sparked the controversy were the consistent findings by Ren and co-workers (94), with continually improving experimental setups, of DPs reaching the stapes earlier than any other BM measurement spot in the cochlea (FIGURE 18A). Measurements at different longitudinal places of the BM, thought to be well basal to the place where the DP was allegedly generated (the $f_2$ CF place), should have revealed a backward traveling wave with an increasing phase lag toward the stapes. Instead of that, systematically increasing phase lags were found at increasing distances from the stapes, suggesting that on the BM, DP waves traveled in the forward direction and not the reverse one. This led Ren to propose that the DPs reach the stapes and round window almost instantaneously via a compression wave through the fluid, and that from there, the asymmetric motion of cochlear windows due to their difference in impedance generates a forward traveling wave along the BM.

FIGURE 17. Main contributions to the delays of cochlear responses from different locations and mechanisms, when the cochlea is excited by a pair of pure tones at frequencies $f_1$ and $f_2$ (vibration patterns in blue and red, respectively; delays, likely similar at $f_1$ and $f_2$, in violet). The 2$f_1$-f$_2$ DPOAE propagates both forward and backward via different putative mechanisms associated with different speeds and delays (vibration patterns and delays in green). Upward-pointing arrows mark measurement sites between which propagation delays can be evaluated. The basal site from which the propagation delay to the stapes is $\tau_{basal}$, actually corresponds to multiple locations, likely broadly distributed between the stapes and the generation site tuned to $f_2$. Straight horizontal lines indicate propagation via a fast compression wave, and undulating ones, via a slow transverse traveling wave.
In the other experiments using a fiber-optic pressure sensor (53), the technique, although requiring stimulus levels of 70 dB SPL, allowed physiologically vulnerable DPs to be mapped near the BM at one cochlear location approximately tuned to 20 kHz. The complexity of DP patterns confirmed the combination of basally produced DPs traveling forward and apically produced DPs going backward. The fact that the DP level fell off when the tip of the pressure sensor was moved away from the BM by a few hundred microns (FIGURE 18B) revealed the existence of a slow DP traveling wave along the BM, without precluding a fast DP pressure wave near the noise floor of the system, i.e., 55–60 dB SPL. Previous data with the same setup (54) had shown evidence of a fast, spatially invariant compression wave dominating at 100 m/s from the BM, whereas the slow backward propagating wave took over at closer distances. Most of Ren’s findings were thus confirmed including, in a number of configurations, DPOAEs leading in phase their corresponding BM DPs. The interpretation clashed with Ren’s in terms of which DP wave was most relevant, as the larger slow wave near the BM was emphasized. The phase lead of DPOAEs was explained, not in terms of the stapes vibrating at the DP frequency before the BM, but by assuming the combination of a forward traveling DP produced basal to the measurement point and propagating toward its CF location, and of a flat-phased DPOAE obeying scaling symmetry, i.e., the wave-fixed component of Kemp.

The fiber-optic work, though showing DP waves traveling on the BM in the reverse direction, could not assess which, these or the fast compression waves away from the BM, dominate in the external ear canal. Another experiment (13) addressed this issue, by measuring the properties of the pressure wave far from the BM, near the bony wall of the cochlea, both in SV and ST in turns 1 and 2 of the guinea pig cochlea (sites SV1 and SV2 tuned to 9 and 3 kHz, respectively). The single pressure detector could be moved back and forth without perturbing cochlear sensitivity and ear-canal DPOAEs, and its sensitivity allowed stimuli at 50 dB and 60 dB SPL to be used for generating DPs at 2f1-f2 well above the noise of the detector. Phases were compared between the two measurement sites for stimulus frequencies between 18 and 1.5 kHz (FIGURE 18C). The DPs, larger in SV than ST by 10 dB, were comparable in magnitude (55–60 dB SPL in SV) in the first and second turns, within a few decibels on average (FIGURE 18D). Their phase differ-

**FIGURE 18.** A: delay between stapes and BM responses to a DP at one basal cochlear site computed from phase measurements, suggesting that the BM response lags, as if the stapes were reached almost instantaneously by the DP via a compression wave, before the DP started traveling forward as a transverse wave. [From He et al. (94). Copyright 2010, with permission from Elsevier.] B: DP pressure against DP frequency in the ST of one preparation, at two distances from the BM [thick line, 37 μm; thin line, 117 μm]. The DP amplitude drops off by >5 dB with increasing distance from the BM, suggesting a dominant traveling-wave mode close to the BM, as a compression mode should fill up the ST. [From Dong and Olson (53). Copyright 2008, with permission from Acoustical Society of America.] C: phase difference between pressures at the DP frequency in SV of the second and first turn of guinea pig cochleas (SV2 vs. SV1), against DP frequency. For f2 = 5 kHz, the average difference is 0 (dashed vertical line) as the DPOAE site of generation falls midway between the two measurement points. A propagation speed of 40 m/s from apex to base is inferred, in line with published interferometric BM measurements at similar frequencies, i.e., between 20 and 100 m/s (227). D: the DP level in SV displays little spatial variations and matches the DPOAE level in the ear canal when taking into account the reverse middle-ear transfer function. [From Avan et al. (13). Copyright 1998, with permission from John Wiley & Sons.]
ence between SV2 and SV1 was about +90° for \( f_2 = 3 \) kHz, i.e., with the main DP generation site near SV2, thus the DP pressure wave took \( \sim 160 \mu s \) to travel the 6 mm from its generation site toward SV1, at an average speed of \( \sim 36 \text{ m/s} \). For comparison, the pressure difference between SV and ST at stimulus frequencies traveled from SV1 to SV2 at \( \sim 40 \text{ m/s} \). As for levels, as the reverse middle-ear transfer function applies a flat 35-dB loss (52, 157), the 55- to 60-dB SPL of the DP wave filling up the SV, correctly matched the DPOAE level in the ear canal, i.e., 22 dB.

This experimental setup reveals a clearly shorter backward travel time than the one derived from OAE studies, including recent ones (160 vs. 400 \( \mu s \); Ref. 176). It conclusively indicates that the measured pressure throughout SV is the relevant signal at the origin of DPOAEs beyond the stapes footplate, thus addressing a question left unanswered in Reference 53. Nonetheless, it contradicts the compression wave theory in several respects. In the first place, 40 m/s is 40 times slower than the speed of sound in water, and actually comparable to the average speed of transverse traveling waves at the cochlear base (227). Second, a compression wave should spread through the whole cochlea and produce, almost instantly, similar pressures in SV and ST, whereas the finding of a >10-dB difference suggests that the BM, acting as a high-impedance device, was coupled to the SV pressure wave, as would a membrane serving as support for a backward traveling wave. The relationships between the space-filling pressure wave in SV measured in Reference 13 near the bony wall and the pressure measured in Reference 53 near the BM on its ST side, at two places where the wavelengths were clearly very different, remain to be clarified, i.e., how do sound waves propagate between these two places, and how fast?

Going back to data modeling, the most thought-provoking finding of Ren is undisputed, the negative slope of phase data. As the initial explanation, i.e., that the DPs reached the oval window ahead of the BM responses, has lost ground, this negative slope calls for an alternative explanation. A recently proposed one (250) rests on the outcome of the so-called Allen-Fahey paradigm (6), invoked by Ren in support of the compression-wave theory (222). This consists in keeping constant, not only the DP CF site, but also the DP level at this site by monitoring the responses of auditory neurons tuned to the DP frequency. The ensuing measurements of DPOAEs under different combinations of stimulus frequencies and ratio \( f_2/f_1 \) reveal a strong influence of the latter parameter. At \( f_2/f_1 \approx 1 \), wavelets from distributed sources recombine much more strongly toward the DP CF site than toward the stapes. This can be explained only if wavelet phases undergo the large rotations that come with slow wave retro-propagation (see sect. VD). A hybrid model might solve the conundrum by assuming DP source coupling into both slow and fast waves, with the DPOAE arising predominantly by fast-wave coupling, and the DP at its CF location by slow-wave coupling. The simplest class of hybrid models fails to improve the accuracy of predictions of the Allen-Fahey experiment. More complex models, that remain to be designed, are criticized on the grounds of suspiciously ad hoc requirements such as the need to radiate similar amounts (250). Nevertheless, both fast and slow waves have been detected and their interferences produced notches (54), which suggests that they do radiate similar amounts. Another recent theoretical approach (254) examines several conventional models of sound propagation in the cochlea, which consistently predict that slow reverse-traveling transverse waves can produce negative phase slopes if two conditions are met: a broad enough distribution of DP sites of generation and a reasonably large stapes reflectivity. Even more recently, application of a two-dimensional nonlinear hydrodynamic cochlea model (283) to the analysis of Ren’s experiments suggested that the DP generation sites were well basal to the measurement sites, thus reconciling the negative phase slopes with the possibility of slow-wave backward propagation. The distribution of generation sites basal to the \( f_2 \) CF location was estimated to 0.5 mm for \( f_2 = 12 \) kHz, which emphasizes, once again, two important caveats. What relates DPOAE levels to local cochlear status is not a simple proportionality relationship, as contributions undergoing significant phase rotations within the generation sites interfere in a complex manner; and DPOAE versus lesion maps may not coincide as DPs basal to local lesions still contribute (170).

At present, evidence that the dominant mode of OAE backward propagation is by slow transverse waves along the BM has become overwhelming, supported as it is both by theoretical models and experiments, even though hints of a fast compression mode and perhaps, of a second slow transverse mode along the Reissner’s membrane have been reported. However, this propagation controversy had the positive consequences of fostering innovative experimental setups allowing fine measurements of longitudinal BM responses and intracochlear sound pressures, and of emphasizing the inescapable complexity of DPs. For a model to realistically predict even the simplest outcomes (\( f_2/f_1 \), Allen-Fahey paradigm, suppression patterns), it must acknowledge that DP sites of production span a broad cochlear interval, much broader than the place tuned to \( f_2 \), and that DPs travel and recombine along intricate paths. Thus the naive idea that a DP gram provides an “undistorted” map of cochlear function no longer safely supports audiological applications. On the other hand, DP grams provide much more useful insights into the intimacy of cochlear micromechanics, for experimenters willing to spend some time designing proper paradigms and deciphering their outcomes.
VI. FROM COMPLEX STIMULI TO NO STIMULUS

A. Three or More Tones

Suppression induced by cochlear nonlinearity provides a simple means for probing DP generation sites. As suppression arises from compressive amplification, which is tuned, a suppressor tone is most efficient near the place tuned to its frequency. Thus when a signal from an unknown site undergoes strong suppression, this site is likely to be located near the characteristic place of the suppressor. The addition to two DPOAE-generating primaries of a third tone swept in frequency and level allows suppression tuning curves to be built, depicting the suppressor level required to induce a given amount of DPOAE level change, against suppressor frequency (FIGURE 19). These curves display a sharp tip around \( f_2 \), indicating that DPOAE suppressibility is tuned. It confirms the sensitivity of the \( 2f_1-f_2 \) DPOAE to cochlear status around \( f_2 \), the place of maximum interaction between primaries (27, 163, 165). The suppression patterns of DPOAEs also display a fine structure, notably confirming the existence of the secondary DP mechanism (the coherent-reflection one) around the place tuned to \( 2f_1-f_2 \) (72).

The suppression paradigm has been refined to explore all possible sources of DPs and all possible nonlinear interactions in the cochlea, and the maps now include the effect of suppression on DPOAE phases (64, 169). The addition of a third tone \( f_3 \) to a two-tone stimulus \( (f_1, f_2) \) at the input of a nonlinear system also creates a huge increase in complexity by multiplying the possible arrangements of combination tones. Their full analysis is outlined in APPENDIX III. It has revealed the importance of basal generation of DPs, more than one octave above \( f_2 \) even at moderately loud stimulus levels (103, 168, 169) (FIGURE 19). These results stress the need to take into account the broad distribution of nonlinear sources to correctly explain initially unexpected properties of DPs (see sect. VH for the paradoxical phase lead of DPs at the stapes, for example).

Novel combinatorial possibilities that lead to the production of additional tones have been discovered or inferred from three-tone studies, extending the two-mechanism taxonomy of nonlinear versus coherent-reflection contributions (246). The need for faster paradigms, e.g., for collecting phase-gradient data to infer from them cochlear travel times and tuning information, has also inspired attempts at replacing \( f_1 \), in the \( (f_1, f_2) \) pair of stimuli, by a group of closely spaced tones with carefully chosen spacings (176, 272). A whole new set of DPs has been disclosed in this manner, thus opening a new field of investigations (see APPENDIX III).

B. Impulses

The briefer a sound impulse (a click), the broader its spectral range. Thus a click spreads over most of the BM. An advantage is that the entire cochlea can be probed for its ability to emit OAEs simultaneously by collecting the TEOAEs generated by this sudden excitation (121). The tradeoff is that, while the response of a linear system to a click is a mere superposition of the responses to each spectral component of the click, the click-evoked response of the nonlinear cochlea bears complex relationships to the responses to tones. These already require a careful analysis of complex issues regarding the processing and propagation of sound-induced vibrations in the cochlea (see sects. IV and V), and TEOAEs add two degrees of complexity in relation to nonlinearity and traveling-wave dispersion along the BM, i.e., the fact that the lower frequency OAE components being processed more apically will increasingly lag the higher-frequency ones and spread out in time. The wide popularity that TEOAEs have gained as an audiological tool, for giving an idea of OHC function over a broad range of frequencies in a single recording, justifies a brief overview of their prop-

![FIGURE 19. Interference response areas in a rabbit ear, obtained as color maps of the changes in amplitude (top) and phase (bottom) of a \( 2f_1-f_2 \) DPOAE (primary tones near 3 kHz, vertical arrows, at 80 dB SPL), in the presence of a third interfering tone swept in frequency (horizontal axis) and level (vertical axis). In addition to a first suppression lobe in the \( (f_1, f_2) \) region, a second prominent lobe peaks about an octave higher. The phase change in the high-frequency suppression lobe differs from that in the primary region (dark red vs. blue-green), suggesting the presence of two different DPOAE sources with different phases, each in turn being removed by the interference tone at different places. [From Martin et al. (169). Copyright 1999, with permission from Elsevier.]
properties in relation to cochlear nonlinearity and other types of OAEs.

Recordings of TEOAEs in human ears display a complex pattern of oscillations (FIGURE 3) generally lasting 10–20 ms after click onset, with the lowest frequency components tending to fade out the latest. The onset of TEOAEs is generally hidden by the fact that the recording equipment removes the first millisecond so as to ensure that the processed response cannot be contaminated by stimulus ringing in the external ear canal. This procedure may remove a sizeable part of the short-lived, high-frequency TEOAE components (>6 kHz, typically). The spectrum of TEOAEs thus extends from 6 kHz down to ~0.5–0.7 kHz, below which middle-ear transfer is thought to be unfavorable. Characteristically, the growth of TEOAEs is compressively nonlinear, much less than 1 dB per dB increase in click level, which inspires the design of the so-called nonlinear-artifact rejection protocol. In the original equipment designed by Kemp (122), one block of regularly spaced stimuli is made of three clicks at level L and positive polarity, followed by one click at level L+10 dB (i.e., three times larger) and with negative polarity. In the averaging process, linearly growing contributions, among which the echo of the stimulus on passive structures, get cancelled and only the nonlinearly growing part of the TEOAE remains.

Another typical feature of TEOAEs, as conspicuous in the time domain as in the spectral one, is an idiosyncratic distribution of beats among higher and lower frequency oscillations, and of rather irregular peaks and troughs in the spectrum (210). Such irregularities suggest that TEOAEs arise from place-fixed perturbations along the BM, as their exact distribution, differing from cochlea to cochlea, would strongly influence how backward-scattered wavelets would recombine to form the final TEOAE. Indeed, simple hardware or computational models of banks of filters (300, 313) excited by impulse stimuli produce outputs closely resembling natural TEOAEs once some degree of disarray has been introduced in their frequency-to-place map. Such models also called into question the idea that narrow-band frequency components extracted from a TEOAE could be literally interpreted in terms of round-trip travel time of the corresponding frequency component along the BM (which does not mean that it does not relate to it). First, some models of the cochlea as a travel-less filterbank accurately predict the observed spectral structure and temporal delays of TEOAEs (300), and second, the temporal properties of oscillations generated by filter disarray reflect the way oscillations from neighboring filters, by interfering with one another, produce waxing and waning envelopes, the peak delays of which depend on the frequency differences among filters and not on their central frequencies (313).

One anticipates that TEOAEs may relate to SFOAEs, in response to every single frequency component of the click, or to DPOAEs, responses to any pair of components of the click \((f, f')\) such that \(n f_j - m f_j = f\), with \(n\) and \(m\) integers, or to both, which raises the issue of which relationship is closest. As SFOAEs mainly stem from a coherent-reflection mechanism (part IV) whereas DPOAEs result from a mixture of contributions from a nonlinear, wave-fixed mechanism and from (linear) coherent reflection (part V), which mechanism accounts best for TEOAE properties must be examined. The mechanism issue is also important for interpreting the pattern of changes in TEOAEs produced by OHC impairment over some frequency interval along the BM. It has been seen with DPOAEs that the nonlinear mechanism of generation can spread over a broad basal interval where OHC damage may impinge on OAE levels at lower frequencies.

The dominant view is that TEOAEs arise mainly from (linear) coherent reflection (112, 246). If TEOAEs are principally made of a mixture of SFOAEs, a close correspondence is expected between frequencies such that OHCs tuned to them are healthy (efficient emitters giving rise to coherent reflection) and frequency bands represented in the TEOAE, and indeed often reported (210). The structure of TEOAE spectral peaks, highly stable in a given ear, supports the idea that place-fixed irregularities serve as primers for coherent reflection. Furthermore, TEOAE and SFOAE growth functions at fixed frequency, and TEOAE and SFOAE spectra at fixed stimulus intensity tend to match one another in a given ear (112). The rapid phase variation of TEOAE phase with frequency and that already reported with SFOAEs closely resemble each other and differ from the nearly frequency-independent phase of DPOAEs (244).

Yet, a few discrepancies suggest a more complex picture for the mechanism of TEOAEs. For example, the correlation between the map of healthy OHCs and the TEOAE spectrum is only approximate. After basal OHCs had been selectively exposed to acute damage, TEOAE responses were found to be decreased in lower-frequency intervals, although OHCs tuned to these frequencies had been unaffected (10) (FIGURE 20). Cogent evidence of a nonlinear mechanism other than coherent reflection has been brought about by the finding in TEOAE responses of frequency components not present in the stimulus and, thus, produced by a distortion kind of mechanism (305). Such a mechanism predicts a not-so-straightforward mapping of healthy versus damaged cochlear intervals, as indeed observed (10). Insight into the coherent reflection versus distortion contributions (place- versus wave-fixed), was provided by reports that in guinea pigs and humans, with an open-canal recording technique allowing the onset of TEOAEs to be reliably analyzed (301, 303), TEOAE generation appeared to shift from a wave-fixed to a place-fixed mechanism over the time course of the response, with the phase slope shifting with poststimulus onset time from a shallow to a steep pattern.
This framework helps understand the differences between TEOAEs in humans or non-human primates, e.g., in Reference 164, and small laboratory rodents (TABLE 3). In the latter, TEOAEs are smaller in amplitude and much shorter in duration (e.g., Refs. 10, 12, 295). Conversely, DPOAEs usually have much larger levels in small laboratory animals (30–40 dB SPL are common) than in humans (~10 dB SPL). These observations, summarized in TABLE 3, fit the two-mechanism model of Shera (246) if one assumes that the coherent-reflection mechanism is larger in primates than nonprimate mammals. Indeed, in four species of small rodents, the presence of an interacting tone at the frequency of the DP, meant to suppress the coherent-reflection component, actually does not affect any of the DP properties (167), which suggests a weak or nonexistent coherent-reflection mechanism. In addition to accounting for large rodent DPOAEs only made of the nonlinear wave-fixed component, combined with the conclusion of Reference 301, it explains why the TEOAEs of rodents are small and short-lasting, as only their early, wave-fixed component exists. The longer-latency TEOAEs relying on coherent reflection are strong only in primates. The short latency of TEOAEs in small rodents finds its counterpart in the large-frequency spacing of the periodic crests found in SFOAE spectra, as it is inversely proportional to the emission latency.

Taking advantage of the sharp onset of the TEOAE-evoking stimulus, the timing of suppression of active mechanisms has been demonstrated directly, which was impossible with other types of OAEs. A click suppressor presented at varying times before the click stimulus (114) produces a TEOAE change depending on postsuppressor delay (90, 279). This change discloses some typical features of an AGC, a system able to adjust its gain according to the input level integrated over the immediate past and such that the larger the integrated level, the smaller the updated setting of the gain. Experiments revealed the presence of a cochlear AGC acting with a time constant of about two periods, with a controlling slope of 0.5 dB/dB (90, 279). Its general significance will be examined more closely in section VIII.B.

### C. Spontaneous Vibrations

Even without stimulation, the ear can spontaneously emit continuous tones, SOAEs (210). They were objectively detected and identified for the first time by Kemp (120). It was initially argued that SOAEs might be mere filtered noise; however, the statistical properties of the temporal distribution of SOAE amplitudes strongly suggest the existence of a

---

**Table 3. OAE mechanisms in humans versus rodent ears, in TEOAEs versus DPOAEs**

<table>
<thead>
<tr>
<th></th>
<th>Human Ears</th>
<th>Rodent Ears</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Distortion mechanism</strong></td>
<td>Present</td>
<td>Present</td>
</tr>
<tr>
<td></td>
<td>Coherent reflection mechanism</td>
<td>Present, yet not strong: −3 to −15 dB relative to the distortion component (1); present in only 3 of 10 ears (167)</td>
</tr>
<tr>
<td><strong>DPOAE amplitude</strong></td>
<td>Small (120, 167); ceiling 15 dB SPL</td>
<td>Large (127, 167); ceilings 35–40 dB SPL</td>
</tr>
<tr>
<td><strong>DPOAE growth notches</strong></td>
<td>Few, around 50 dB SPL (92)</td>
<td>Many, around 60–70 dB SPL (291)</td>
</tr>
<tr>
<td><strong>DPOAE fine structure</strong></td>
<td>Conspicuous (91)</td>
<td>Weak or absent (167)</td>
</tr>
<tr>
<td><strong>TEOAE duration</strong></td>
<td>Long (121)</td>
<td>Short (10)</td>
</tr>
<tr>
<td><strong>TEOAE amplitude</strong></td>
<td>Large (121)</td>
<td>Small (if integrated over the same 20-ms window as for human ears) (10)</td>
</tr>
<tr>
<td><strong>TEOAE structure</strong></td>
<td>Two parts, the early one attributed to distortion, the later one to coherent reflection (303)</td>
<td>Only the early (distortion-related) part (10)</td>
</tr>
</tbody>
</table>

TEOAE, transient evoked otoacoustic emission; DPOAE, distortion product otoacoustic emission; SPL, sound pressure level. Reference numbers are given in parentheses.
genuine intracochlear source of energy (19). Many species from all four tetrapod classes have SOAEs so that the presence of a basilar or tectorial membrane is not critical. These SOAEs are usually made of one or a few narrow spectral peaks, 0.15–40 Hz wide in humans (19, 274), with highly stable frequencies and levels. Apart from rare exceptions (76), a normal cochlear sensitivity is required for SOAEs to be present in humans. However, a sizeable percentage of normally hearing individuals have no SOAE, from 35% in human neonates to ~70% in adults (210). As expected, SOAEs have correlates on the BM (195). In his pioneer work, Gold (77) predicted that spontaneous oscillations would occur in his local regenerative oscillator model, close to the region of instability. However, he wrongly assumed that SOAEs would be objective correlates of tinnitus. His efforts at recording them in tinnitus-complaining subjects failed because tinnitus is a phantom neural percept in relation to peripheral deafferentation due to sensory-cell failure that precludes OAEs.

Despite his conceptual error, Gold had made an important point. In the engineering domain, families of limit-cycle oscillators (e.g., the van der Pol oscillator) poised on the verge of auto-oscillation can, in the absence of input other than the inherent noise, generate outputs presenting many highly nonlinear characteristics of SOAEs (104), even though isolated oscillators cannot emulate all SOAE properties (274). A pure tone in the ear canal can enhance or suppress an SOAE at a neighboring frequency, and entrain SOAE frequencies (148). Repeated click stimuli synchronize the existing SOAEs that show up among the averaged SOAE frequencies (148). Repeated click stimuli synchronize the existing SOAEs that show up among the averaged TEOAEs as stable, inordinately long-lasting oscillations. Changes in cochlear status, whether general [aspirin- or temperature-induced (148, 149)] or targeted to aspecific areas, affect SOAEs (47). Similarly, stress or temperature-induced (148, 149]) or targeted to a specific frequency interval (exposure to high-level sound) affect the SOAEs by decreasing their level and shifting their frequency, before recovery brings them back to their initial state. Minor changes in the boundary conditions at the stapes, produced by a change in ear-canal, middle-ear, or intracochlear pressure (47), similarly affect SOAEs (Figure 21). When the cochlear partition is biased by the application of a loud infrasound, SOAEs exhibit dynamic behaviors consistent with the shifting of the OP of OHCs (20). Privileged frequency spacings between adjacent SOAEs have been reported (24), similar to the average frequency spacings between the crests in SFOAE and TEOAE spectra (313). All these properties illustrate typical characteristics of cochlear micromechanics, i.e., subtle impedance adaptation for optimal wave propagation; strong nonlinearities adjusting the amplification such that near threshold, when the cochlea amplifies optimally, cochlear oscillators are close to instability; as a result of the former two properties, a tendency for coupled oscillating elements to lock to a common frequency linear coupling (297, 299), in sharp contrast, always pulls resonance frequencies apart.

Two theories of SOAE generation have been proposed. The first one is inspired by the behavior of van der Pol oscillators, and supported by experiments demonstrating that isolated hair bundles can spontaneously oscillate. This striking phenomenon has been described in sensory cells from bullfrog saccules surviving in controlled ionic environments (171). Oscillations relate to the existence of an unstable interval of negative stiffness, calcium-dependent and due to the influence of the gating compliance. Every time a deflected bundle moves toward its resting position, it enters the unstable region and flips back and forth without stopping. If deflection-independent stiffness terms exceed the negative stiffness associated with gating compliance, which is likely the case in mammalian cochleas (277), spontaneous bundle oscillations may no longer occur. An alternative mechanism for SOAE generation is that of cochlear standing-wave resonances. In this view, SOAEs result from mul-
Multiple internal reflections of forward- and backward-traveling waves between the stapes boundary and a place with disarrayed reflection sites giving rise to coherent reflection. In its passive version, the energy comes either from sounds from the environment or physiological noise, while in the active, prevailing view, the cochlea acts as a biological “laser cavity” maintaining stable standing waves by coherent wave amplification (243).

VII. THE PARTICULAR STATUS OF THE APICAL COCHLEA

Experimental data from the cochlear apex are scarce, mostly from interferometry (43, 85, 125). Little amplification can be observed even in a sensitive cochlea, and the growth function is said to be “linear,” a misnomer only referring to the absence of compression. Broad and symmetrical tuning curves are observed, in stark contrast with the high-frequency pattern of a sharp tip with a steep cutoff on the high-frequency side and a U-shaped low-frequency tail. Nonetheless, WD is definitely present and physiologically vulnerable, and large harmonic components have been recorded from the motion of the apical organ of Corti in healthy guinea pig cochleas (125). A revealing pattern of changes was observed after death, at different spots in the organ of Corti where the probing laser beam was focused. Whereas the motion of reticular lamina connecting the tops of OHC bodies decreased, as intuition would have suggested, BM motion increased almost a hundredfold, with a sharp increase in tuning. It suggests that at variance with the rest of the cochlea, the apex is subject to a different feedback mechanism with a different purpose. Active negative feedback (instead of the positive basal one) exerted by OHCs would tend to decrease BM motion. Fundamental principles of negative feedback are to reduce noise, to stabilize the amplification, and to make the growth of BM motion “linear” even if it was compressive without feedback (124, 125). Feedback manipulations did not affect the size of harmonics, which the authors saw as an indication that their hypothesized feedback loop coupling fluid velocity acting on the BM to the Hensen support cells on which they focused their interferometer, did not include the WD-generating OHC stereocilia bundles (125).

The apical DPs, not only harmonics but also intermodulation products at \( f_2 - f_1 \) and \( 2f_1 - f_2 \), can be heard, have correlates in the responses of neurons in the auditory midbrain (2), and usefuly underpin psychophysical tasks (99, 174, 209). Below some corner frequency, measured at 1.3 kHz in gerbils (2), DPOAEs are smaller than would be expected from the large intracochlear DPs. This might be due to a decrease in middle-ear gain, as happens in human temporal bones (212). Yet, large interspecies differences in the frequency dependence of the middle-ear gain have been reported, perhaps in relation to differences in ear-canal volumes (157, 212), making it difficult to find a general explanation for the issue of low-frequency DPOAE levels.

Another model exploring the peculiarities of apical responses has been proposed (216), hypothesizing a ratchet mechanism that strives to explain how frequency selectivity may be achieved in low-frequency auditory-nerve fibers, despite the fact that the resonance frequency of the apical BM is likely higher than the frequencies processed apically. Indeed, predictions from realistic values of mass and stiffness indicate a discrepancy between apical BM resonance and the actual limit of low-frequency hearing. The ratchet model also examines why the steep high-frequency cutoff that characterizes wave propagation along the basal BM is not visible apically. The ratchet mechanism (216) is such that every second half-cycle, OHC electromotility uncouples hair-bundle forces from the BM. The resulting unidirectional coupling is unfavorable to amplification of BM motion, hence its observed lack of significant compressive nonlinearity. The main perceptive interest would be that contrary to BM motion, hair-bundle motion would be tuned, with open and symmetrical tuning curves, as observed in auditory neurons, extending the hearing sensitivity to frequencies well below the low-frequency limit of BM resonance. This ratchet model may well be anatomically grounded. The orientation of OHCs in the apical cochlear turn strikingly differs from basal ones in that, instead of standing perpendicular to the BM, their axis is tilted by \( \sim 67^\circ \) relative to the reticular lamina joining the apexes of hair cells (125), which may favor periodical uncoupling of hair bundles from the tectorial membrane.

The specificity of low frequencies affects not only the tuning mechanism, but also wave propagation. A break of scaling symmetry has been pinned down at the cochlear apex (49). Although suppression experiments show that human DPOAEs below 1.5 kHz still come directly backward from the nonlinear site of generation, and not from secondary coherent reflection, they lose their characteristic phase invariance at constant \( f_2/f_1 \), the one underpinning the horizontal-banding pattern of Figure 14B. A possible explanation is suggested by the recent model of cochlear mechanics positing two sound propagation modes, the classical one along the BM and another one along Reissner’s membrane (217). Although the latter wave may be of uncertain relevance at high frequencies, the model derives from impedance and wavelength considerations that, at low frequencies, it can interact with BM motion and disturb BM resonance and DP propagation.

VIII. MATHEMATICAL MODELS OF NONLINEARITY

The ultimate sources of peripheral auditory nonlinearities are clearly identified as the MET channels of hair cells. Their nonlinear behavior can thus be incorporated in models of cochlear mechanics that attempt to predict realistic characteristics. Many of these models correctly describe the nonlinear compressive amplification and frequency tuning.
Auditory WDs are more difficult to accurately predict, as they present distinctive features such as a rate of growth that depends on the passive or active modality at the site of the tone, and the preeminence of odd-order distortions. These idiosyncrasies were ignored at the time of Helmholtz when he tried to propose the first mathematical model of auditory nonlinearity (95), which led him to dismiss some of Tartini’s observations that contradicted the outcome of his model (see sect. XI: Historical Insert). In the field of cochlear mechanics, modeling efforts endeavor to address a few questions, e.g., what is the most parsimonious model of active resonators in which nonlinear compressive amplification, frequency tuning and distortions are coupled and present realistic features? Which parameters of this model are critical for accurate predictions? If the cochlea gives sluggish responses to brief excitations, is it due to mere filter delays that come with any battery of highly tuned oscillators, or does it betray the presence of stimulus envelope-extracting devices from which an AGC can be built, with the advantage that its compression is almost WD-free? Last, if a model predicts a correct outcome, does it vindicate the architecture of the model or can its correct outcome stem from generic considerations that any other model of a similar class would also satisfy?

A. Generic or Specific?

The last decade has witnessed thriving literature using the Hopf-bifurcation formalism to describe the motion of auditory hair cells or of their stereocilia bundles as critical oscillators poised on the verge of self-oscillation (34, 60, 105). The advantage of this model lies in its ability to describe many features of auditory stimulus processing by hair cells, compressive amplification and nonlinearity with limited risk of instability, just by introducing a control parameter in the behavior of a mass-stiffness resonator with damping. This control term allows viscous damping to be offset at very low oscillation amplitudes. Damping contains an additional nonlinear term varying as the third power of displacement. Its sharp increase at higher amplitudes reduces the gain of the oscillator and avoids instability. The dynamics of a critical oscillator described by the complex variable $Z$, in the presence of a force $f(t)$, is described by the differential equation

$$\frac{dZ}{dt} = -(r + i\omega_0)Z + B|Z|^2Z + e^{i\theta}/\Lambda f(t) \quad (5)$$

where $\omega_0$ is the resonance radian frequency, the real parameter $r$ controls the proximity to the bifurcation (which happens when $r = 0$). The multiplicative factor acting on the external force expresses the existence of friction ($\Lambda$) and of a phase shift $\theta$ at which $f$ acts on $Z$. The amplitude-dependent $|Z|^2$. Z cubic damping term predicts compressive amplification at a rate of 0.3 dB/dB, consistent with experimental data. By slowing down the decrease of DPs with decreasing stimulus level, this rate also underpins the prediction of essential nonlinearity with the characteristics that Goldstein described long ago. Two-tone suppression, DPOAEs, and many of their characteristics can also be correctly predicted, at least within the approximation of small displacements as a Taylor expansion. Refinements such as the incorporation of a traveling wave for exciting the resonators have been successfully implemented (56, 110).

Other systems of time-dependent differential equations can be built from attempts to model a realistic auditory organ with its BM and OHCs. For example, they can include MET transfer functions as sole source of nonlinearity, longitudinal coupling between OHCs for introducing some degree of delayed feed-forward, and a combination of hair-bundle dynamics and OHC-body electromotility reconciling the two alternative models of an amplifying mechanism (265). A Hopf bifurcation may even emerge from such systems, without the need for individual oscillators to be critically poised near a bifurcation (265). Modeling the active nonlinear cochlea is still a very open field in which many details remain to be sorted out. To choose the most suitable set of models, the ability to simulate realistic nonlinearities may be a good criterion. It is thus important to establish which characteristics are generic and which are specific to a particular model. For example, does the success of Hopf-bifurcation views of auditory stimulus processing in hair cells, for predicting realistic cochlear nonlinearities and not only the compressive growth, definitely vindicate the underlying model?

Actually, theories published in the 1930–50s in electrical engineering (55, 64) have shown that several properties of a (time independent) nonlinear system mapped by an input/output sigmoid transfer function automatically derive from a minimum set of hypotheses. Any transfer function can be modeled as a combination of odd and even-order terms [an odd function is such that $f(x) = -f(-x)$, while $f(x) = f(-x)$ for an even one] that underlie odd and even DPs, respectively. Experimentally derived voltage/displacement hair-cell transfer functions share with any arbitrary nonlinear function displaying large second and third derivatives, many nonspecific properties of their combination tones. For example, the condition allowing two-tone interference to generate maximum DPs is always that the amplitudes of both tones be equal at the input of the nonlinear system, that is, at the interference place (137). The growth of a DP with the amplitude of one primary tone when the amplitude of the second primary is kept constant is also stereotyped. With $A_1$ (amplitude at $f_1$) held constant, the amplitude of the $2f_1 - f_2$ DP increases linearly when $A_2$ (at $f_2$) increases from a small value. With $A_2$ constant, the amplitude of the $2f_1 - f_2$ DP increases in proportion to $A_1^2$, reaches a maximum, then decreases. As for suppression, it always requires that the suppressor drives the nonlinear function into its saturation region, and that if the suppressed signal is, itself, into saturation, the suppressor must be no less than 10 dB below it. Not surprisingly, Equation 5 predicts these fea-
tured equally well. Another intriguing, yet generic property is the occurrence of sharp nulls in the output of the system (Table 2). Although sometimes interpreted as proof of interference between sources that would happen to be equal in amplitude and out of phase, they can actually mark transitions between saturations from one to both sides of a nonlinear transfer function (64, 154).

However, modeling issues are far from straightforward because in addition to local properties of nonlinear oscillators, several other aspects of auditory processing in the cochlea shape the outcome of nonlinearities. Wave mixing from distributed cochlear sources of DPs is one compounding factor. A second one is that the input to the nonlinear transfer function producing cochlear nonlinearities is not the sound in the external ear canal, but what reaches the DP-generating system after propagation, amplification, and compression have occurred. Likewise, a measured DP is what eventually reaches the measurement spot, and after leaving its site of generation, the DP has undergone several modulating interactions, such as filtering at the generation place (tuned to another frequency), absorption at its own CF location, coherent reflection, reflection at the impedance discontinuity at the cochlea/middle ear boundary.

B. Instantaneous or Sluggish?

Another intriguing question has been little explored, namely, whether cochlear nonlinearities are static or dynamic (59, 155, 315). Above, this review has largely rested on descriptions of the nonlinearity associated with hair bundle motion as instantaneous, with immediate distortion of the current through the bundle. A dynamic nonlinearity would introduce some degree of sluggishness, due to its dependence on the past history of the system. The most typical example is provided by an AGC, a design that engineers use to obtain compressive amplification, similar to that in the cochlea, while minimizing waveform distortion. It has already been mentioned that the moderate levels of distortion in the ear are striking, in view of the strong nonlinearity of the transfer function of MET channels. In an AGC, the gain is adjusted as a function of the average level of the output from the system. This is extracted with the help of some low-pass filter, providing an integrated view of the stimulus over a finite interval of recent history. Different integration times define a whole continuum of AGC systems, and even a Hopf-bifurcation system implicitly assumes sluggish gain control (271). However, neither sluggishness nor the moderate distortion level suffices to imply the existence of an AGC. Regarding the time issue, any tuned system stores energy and responds to a stimulus with a delay that increases with the degree of tuning. As for the pattern of distortions, their frequencies, particularly those of harmonics and even-order intermodulation, are far enough from the CF of their generation site to be filtered out more than the odd-order intermodulation components, again in proportion to the narrowness of tuning. To probe the existence of an AGC experimentally, more distinctive features must be sought.

Several experimental observations strongly support the presence of a noninstantaneous control of cochlear gain. For example, in the auditory nerve, the period histogram of neuronal discharges in response to a pure tone, up to a few kilohertz, remains a half-wave sinusoid over a broad range of levels (the sinusoid is rectified because there cannot be any action potential when hair bundles move in the inhibitory direction). Yet the growth of the synchronized (i.e., phase-locked) rate of action potentials with stimulus level is highly compressive (74). With a system displaying an instantaneous nonlinearity, even though its tuning made it sluggish, it would be paradoxical that such a compression result in a sinusoidal, thus undistorted period histogram, but not with an AGC system.

Initially, the proposed AGC was a neural mechanism, either a reservoir of neurotransmitter or a possible slow modulation of cochlear gain by an efferent control of OHCs via the medial olivocochlear neural pathway (155). Later, indirect evidence of a cochlear AGC has been produced (315). Electric responses to tone bursts from supporting cells, thought to be quite similar to OHC responses yet easier to record directly, displayed several signatures of an AGC in addition to small WD despite high compression, i.e., asymmetry of the on- and off-transients; a frequency increase with time during the off transient, within a modulated envelope. In an AGC, the gain is turned off during signal onset so that the on-response builds up faster, while the increase in gain after signal termination generates a complex shape in the off-transient.

Direct evidence of the size of the AGC-like delay has emerged from several paradigms. The first one (see sect. VII B) consisted in measuring the timing of TEOAE changes when suppressed by a click emitted a few milliseconds before the regular TEOAE-evoking stimulus (90, 115). A time constant of two periods was inferred. Another protocol (273) studied the timing of 2TS in auditory neurons tuned to a broad interval of frequencies, by evaluating the group delays of their responses. The delay of suppression was larger than the delay of excitation by several hundred microseconds, which suggested an AGC with, again, an integration time of about two periods at CF location. Last, responses of the BM of guinea pig cochleas to white noise were analyzed by cross-correlating the input noise to the BM velocity response (189), which yields a so-called “first-order Wiener kernel” (see sect. VIIC, Appendix V). In a linear system, this is a standard procedure providing its impulse response. Comparison of the BM responses to noise and clicks, therefore, evaluated by how much the cochlea departed from linearity. It was found to depart very little so that cochlear filtering could be said to be “quasi-linear.”
This is as surprising as the fact that cochlear responses to tones exhibit so little harmonic distortion (43, 224, 230), despite the high measured compression rate. If cochlear processing worked as a band-pass linear filter combined with an instantaneous nonlinearity, the wave shapes of responses to noise could not be accurately predicted by a linear method. Thus it is not only convenient but also necessary to invoke an AGC in the cochlea.

The existence of an AGC drastically reduces harmonic distortion, but interestingly, without necessarily precluding intermodulation. Assume an AGC with a response time extending over several periods, and an amplitude-modulated signal. Its envelope, extracted by the AGC filter, adjusts the gain. A sine wave is transformed, after a few periods, into a sine wave with smaller amplitude. The amplitude-modulated signal, conversely, gets distorted as its envelope is flattened by the decrease in gain triggered by the growth of this envelope. This generates DPs, and it can even be shown that these DPs, even when generated at a single location, present a rather realistic microstructure suggestive of interference (this time among terms with different mathematical origins; Ref. 271). Overall, a variety of experimental clues point to the existence of a functional AGC in the cochlea, even though an open issue is the actual identity of the low-pass filter mechanism extracting the average level of the output from the system, to provide an integrated view of the stimulus over about two periods.

C. From Particular Stimuli to a General Characterization

Most published investigations of cochlear function have used only a very restricted set of stimuli, i.e., either pure tones, pairs of pure tones for the specific topic of DPOAEs, or brief impulses. What is straightforward for a linear system, that is, to predict the response of the system to a spatially complex stimulus from the responses to each spectral line of this stimulus, becomes only approximately true in the case of a nonlinear cochlea. One has to consider that the prominent nonlinearities of the cochlea are small enough for being mere perturbations in an almost linear cochlea. When using a pair of tones, for example, 2TS is conveniently evaluated by measuring how the amplitude of the response to the probe tone changes in the presence of a second, suppressing tone. Likewise, it is possible to study the interaction of two simultaneous tones at similar levels by detecting the combination tones that their interference generates, and perturbation theory is valid as combination tone levels are generally found at more than 30 dB below primary levels, i.e., they are more than 30 times smaller.

The Wiener-Volterra kernel formalism provides a powerfully general method in strongly nonlinear systems. Briefly presented in Appendix IV, it describes the responses of a system as a sum of functionals unfolding the nonlinearities at increasing orders, extracted from cross correlations between the output of the system, and its input stimulus at varying time delays between input and output. Gaussian white noise is the privileged input, as its mathematical interest is to provide Wiener kernels that are independent of each other and can be numerically computed. Its ecological interest is that, better than tones or clicks, it allows both simultaneous and prolonged stimulation of the whole cochlea. First-order Wiener kernels from a linear system would give its impulse response, but in a nonlinear one, it contains the impulse response plus corrective terms. Second-order Wiener kernels give a measure of how much the presence of a first impulse stimulus influences the response to a second one (162). They have sometimes been successfully applied to characterizing the temporal coding of auditory nerve fibers even when the carrier frequency of the stimulus is too high for phase-locking of the action potentials to occur. In this case, it is from the envelope modulations also produced by the stimulus that, indirectly but efficiently, temporal coding in these neurons can be inferred. The extended use of Wiener kernels for studying cochlear mechanics, seldom attempted (156, 275), usually requires their outcome to be compared with that of a model, in search of an optimal match ensured by adjusting the model parameters. Thus the advantages of these general methods, unconstrained by the limitations of low-level approximations to explore the nonlinearities of a system, are somewhat offset by the need to start with a few hypotheses regarding the block diagram of the system under study.

IX. AUDITORY MASKING AND THE COCHLEA

Auditory masking is one typical nonlinear phenomenon. It is revealed by the psychophysical task of detecting a probe tone, first alone and at a level slightly above detection threshold, enough to be easily detected, and then, at the same level but mixed with a second sound with varying frequency and level, called masked. The masker characteristics are then adjusted in order that in the masking configuration, the probe tone is no longer heard, or the auditory potential evoked by the probe is no longer detected. Differences between masking and nonmasking configurations probe the nonlinearity of the system. Mixtures of sounds at different frequencies and levels are frequently encountered in everyday life, either when the vibrations from different sources are mixed in the transmitting medium before reaching the ears, or when a single source produces a combination of components at different frequencies (e.g., a musical instrument or a voice). Nonlinear interactions among frequency components of the mixture allow some components to mask others, according to well-acknowledged empirical rules (58, 286). The detection of speech alone or in the presence of background noise can become difficult if crucial components get masked. The study of masking should help quantify the perceptive consequences of cochlear nonlinear-
ity. However, whether measured by perceptive tests in subjects or by neuronal recordings, masking actually results from a blend of three mechanisms, only one of which, the suppressive masking, is of cochlear origin while the other two mechanisms are neuronal ones. This section examines which are the distinctive features of the suppressive masking, from which information on cochlear nonlinearity can be derived.

A. Basic Characteristics of Auditory Masking

For a tone, serving as a probe, presented with a narrow-band masker (it is not recommended to use a pure tone as masker as the amplitude modulation of the mixture of probe and masker might provide unwanted perceptive cues) the threshold of masking increases roughly in proportion to masker level for probe frequencies around the masker frequency region. In other words, starting from a probe tone just detected in the presence of a masker, when the level of this masker increases by 10 dB, the probe tone has to be set 10 dB higher to be heard again. In contrast, thresholds increase more rapidly for probe frequencies well above the masker frequency region (58, 286). This so-called “upward spread of masking” expresses the detrimental effect of high-level, lower-frequency maskers on the detection of higher frequency probe tones. As the probe must increase by over 1 dB increase of masker level, to remain audible, psychophysical references often call this masking “nonlinear,” as opposed to a “linear,” 1 dB/dB race for audibility between probe and masker levels. With reference to the meaning of linear and nonlinear throughout this review, these terms are misnomers, and as such, when necessary, will be mentioned with quotes, as even the “linear” masking is an essentially nonlinear effect.

B. Line-Busy Masking

The earliest systematic investigations of the mechanisms of masking relied upon recordings of auditory nerve fibers (48, 107, 256), and contributed to the identification of the three masking mechanisms. The earliest to be invoked (67) is purely neuronal, the line-busy mechanism. Let us consider a probe tone presented alone, above the threshold of a recorded auditory nerve fiber. This induces an increase in discharge rate of this fiber. If, when the same probe tone is added to a continuous masker, no change happens to the discharge pattern produced by the masker alone, a simple interpretation is that the neuron is busy responding to the masker so that the probe tone can no longer modify its activity; thus it can no longer make itself detectable. [A caveat is that even though the neuronal activity in response to the probe tone is no longer detectable in the group of auditory fibers tuned to the probe, this does not mean that the probe can no longer be detected at all. It may still activate other neurons enough to modify their activity. This detection is called off-frequency listening as it is mediated by neurons not tuned to the probe. The perceptive properties of the detected signal may of course differ from what they should be if detected through the neurons tuned to it, in terms of pitch for example, but since the psychoacoustic tasks are usually based on mere binary detection, the fact that tested subjects exploit off-frequency listening to achieve a task may be difficult to guess even by the subjects themselves.]

The line-busy explanation is a sketchy one, positng that, as neurons have a limited rate of action potentials due to their refractory period of a few milliseconds, if the masker drives them at the maximum rate, the occurrence of the probe tone, which when presented alone increases the discharge rate, can no longer influence it as it is already saturated. This explanation does not hold when the masker is weak enough not to saturate the neuron, and yet masking still occurs. The term swamping is more conservative, suggesting that masking results from the fact that the probe tone increases the discharge rate, but not enough for being detected because the probe-induced activity is swamped by the masker. Swamping happens for two reasons. The first one is that the rate-level functions have a compressive slope so that the increment in discharge rate is smaller when the tone is mixed up with the masker than when alone. The second reason is the increase in variance of the discharge rate of a nerve fiber when its mean rate increases, which makes it more difficult to detect the slight increase in rate due to the probe tone (48). The swamping masking mechanism requires that the masker itself excite the neurons responding to the probe tone (excitatory masking). It is not cochlear, and not even specific of hearing. Yet, whenever masking experiments investigate the nonlinear cochlear mechanisms, their results will most often be contaminated by this nonlinear behavior of neurons.

C. Adaptation

The second mechanism by which masking happens in auditory neurons is called adaptation, again an excitatory, specifically neuronal mechanism happening when a masker has been presented long enough for adapting the neurons responding to it, thus decreasing the response of these neurons when the probe tone is presented (87, 257). Adaptation still goes on for a few milliseconds after the masker has been turned off. This phenomenon is called nonsimultaneous masking, here, forward masking.

D. Suppressive Masking

The third mechanism, suppressive masking, is one inherent property of nonlinear cochlear mechanics, whereby the masker suppresses the mechanical response to the probe
tone along the basilar membrane, by jamming the active mechanism by which the probe tone is processed. The simplest protocol revealing this masking mechanism is 2TS, described in section VA. In contrast to line-busy and adaptation mechanisms, suppressive masking is not caused by an excitatory neuronal mechanism. Even before direct measurements on the BM and hair cells revealed the mechanical nature of the suppressive nonlinearity (201, 233), experiments using microelectrode recordings of neuronal activity showed that a masker can suppress the response of a neuron to a probe tone despite not exciting this neuron when presented alone (236). More complex situations than the 2TS paradigm have been studied, i.e., noise (126) and speech sounds (237), for which evidence of suppression in the discharge rate of auditory neurons has also been obtained.

Suppression is strong enough that, for a probe tone at the CF of a neuron, the level required to produce a given average discharge rate may shift upward by 40–50 dB in the presence of a suppressor (48, 73, 75). Despite this large effect and more generally, the impact of nonlinearity on the cochlear processing of sound, traditional psychophysical models rest on an excitatory picture of masking explaining the influence of a masker by the way its pattern of excitation spreads along the cochlea to the place of the probe tone (67, 286), and swamps neuronal activity in response to the probe. Accordingly, many models used for interpreting masking, loudness, and frequency and intensity discrimination do not take nonlinear interactions into account. Nowadays, the accuracy of these linear masking models is increasingly called into question (48). Recent reports of strong changes in masking, in the particular case of Strc<sup>−/−</sup> mutant mice with no suppressive interaction despite normal cochlear sensitivity and frequency tuning, raise the same issue (282) (see sect. IIII). In an extensive study at the level of the auditory nerve, Delgutte (48) therefore sought to differentiate between excitatory and suppressive mechanisms over ranges of masker levels and of frequency spacings between probe and masker. Comparisons were made of masking thresholds obtained in simultaneous and forward masking. Suppression may contribute to the former, but in no way to the latter one. Thus the differences in masker levels required to mask a fixed signal at the CF of the recorded fiber, could be attributed to suppression. For masker frequencies well below the probe frequency, the amount of suppressive masking was found to be large, increasing with masker level more rapidly than did excitatory masking. The results of Delgutte, obtained with probes <40 dB SPL, thus indicate that the upward spread of masking is largely due to the growth of suppression rather than to that of excitation. This view, supported by other physiological studies, e.g., Reference 199, suggests that psychoacoustic models should be revisited to incorporate suppression. Accordingly, physiologically motivated models that describe cochlear compression with the help of nonlinear, velocity-dependent feedback characteristics similar to those discussed in section VIIA, now achieve realistic predictions of OAE properties, of 2TS and of the outcome of some psychoacophysical tasks (63).

The psychoacoustic finding by Oxenham and Plack (198) of a strong “nonlinear” (i.e., >1 dB/dB) upward spread of masking in nonsimultaneous masking apparently contradicts the physiological results of Delgutte (48), who found no suppression in the nonsimultaneous masking condition, no strong upward spread of masking, and a merely “linear” (1 dB/dB) growth of masking. A likely cause of this discrepancy is that Oxenham’s probes fell between 40 and 90 dB SPL. It is in this range of levels that the compressive behavior of the BM is most intense, whereas at <40 dB SPL, compression decreases and probably vanishes below 20 dB SPL (228). In the region tuned to the probe tone, BM vibrations in response to maskers between 40 and 90 dB SPL and near the probe frequency (“on-frequency”) grow slowly as they undergo strong compression, whereas the responses to off-frequency maskers, masker frequency well below that of the signal, grow at a rate of 1 dB/dB increase of the masker level. This differential growth of BM responses to the probe versus the masker suffices to explain a large growth of masking for probes with a frequency above the masker frequency even in the nonsimultaneous, i.e., nonsuppressive masking conditions observed by Oxenham and Plack. In addition, and in agreement with physiological studies, the same authors did also find a major role for suppression in determining thresholds at high masker levels, for an on-frequency masker (198). Playing with simultaneous versus nonsimultaneous-masking conditions and with on- versus off-frequency maskers turns out to be suitable for separating the various causes of masking and study their distinctive properties.

When more than two tones are presented simultaneously to the cochlea, a more complex situation than 2TS pertaining to the intelligibility of speech sounds, unmasking phenomena may occur and reinforce auditory-contrast sharpening. A two-tone stimulus can be less effective to mask a probe tone than one of its components (98, 242). In a nonsimultaneous masking experiment, the masker, made up of a mixture of two sounds, is presented first, then a gap, then the probe tone. This probe tone can be better heard if muscular suppression happens between the two simultaneously present components of the forward masker, as this suppression decreases the efficiency of the overall masker. Conversely, simultaneous presentation of all sounds would lead to no clear result because the signal, being present simultaneously, would be suppressed by the masker, and not only the weak component of the masker by the stronger one. This explains an apparent paradox. It is widely accepted in the psychoacoustical literature that effects of suppression are not revealed in simultaneous masking (98), which might sound weird since suppression requires simultaneous presence of suppressor and probe. The par-
X. CONCLUSIONS AND PERSPECTIVES: NONLINEARITIES AT THE CORE OF AUDITORY MECHANISMS

While the physiological and psychophysical relevance of compressive amplification in the nonlinear cochlea is well acknowledged, WD, the other manifestation of auditory nonlinearity, for the time being, seems to draw its major interest as a tool for probing cochlear mechanisms. A coherent picture has gradually emerged from three decades of investigations of the active mechanics in auditory organs and of OAEs, relating amplification and frequency tuning to all manifestations of nonlinearity. An amplitude-dependent positive damping term ensures mechanical stability and produces compression. Nonlinearity itself is a mandatory consequence of MET channel functioning, and WD and its consequences afford a privileged insight not only on the processes that ensure an appropriate gating of MET channels, but also on the integrated mechanism of feedback which, within auditory sensory cells, actively processes auditory stimuli.

While the basic functional principles of auditory nonlinearity are increasingly well understood from mammals to lower vertebrates, the molecular assembly surrounding its core, the MET channel, remains incompletely established. The solution to this protracted molecular hunt would not radically change any physical principle of the gating process leading to the generation of WD, anchored as they are in thermodynamic laws. However, it would make it easier to dissect the interaction of MET-machinery subparts, e.g., the pore, gating spring, tip-link, and the mechanisms that reset the resting position of the MET channel. Our knowledge of how interstereociliary links work also remains imperfect. An improved understanding of hair-bundle architecture might help to explain why DPs become undetectable in the absence of top connectors even though MET channels still operate, and more generally, how the shape of the MET transfer function and the characteristics of cochlear nonlinearities can depend on a coordinated motion of the stereocilia bundle.

The most readily detected WDs are the DPOAEs, yet their propagation from stereocilia bundles to the stapes is a complex topic so that, in practice, it may be useful to choose eliciting stimuli such that one generation site or one propagation mode dominates the DPOAE pattern. The controversy regarding backward propagation of sound by fast compression waves versus slow traveling waves has led to considerable refinements of measuring devices and theoretical tools. The part played by slow traveling waves seems dominant; however, intermediate structures and paths ensuring DP propagation between hair bundles and the stapes remain a matter of active research. DPs present a strikingly complex microstructure that challenges simplistic interpretations, e.g., of a change in DP amplitude. Some DP features, deep notches for example, are reminiscent of interference patterns among different sources, yet some possibly are mere products of the complexity of the nonlinear transfer function of a single source, with its saturations, asymmetries, and mobile OP. No less convincingly, other DP features have been shown to stem from genuine interference among sources or paths. Is it possible that still other paths and mechanisms exist that have been missed? Propagation is only one aspect of the broader issue of DPs as noninvasive windows on cochlear frequency tuning. Their use heavily draws on the possibility to extract reliable timing data which, with steady tones as stimuli, remains a nagging issue. In particular, despite evidence that the cochlea operates as a (minimally distorting) AGC exploiting a noninstantaneous nonlinearity, surprisingly little hard data can be found regarding the timing and hard-wiring of what controls the noninstantaneous extraction of stimulus level to adjust the gain. Several reports point to an interaction time of a few periods (90, 273), while thought-provoking, extreme models incorporate the whole history of cochlear activity over hours (142), by assuming that the scala media may act as a pressure accumulator, with pressure and the resulting shift in resting position (140) as a control parameter of cochlear operation. In contrast, Hopf bifurcation-based models predict noninstantaneous distortion, but its mechanisms are not explicit, and at present, a tangible way to control critical oscillators, e.g., by Ca$^{2+}$ ions entering the MET channels, is suggested only in models of nonmammalian ears.

The perceptive benefits of WD remain an open issue even though the coherence of all cochlear nonlinear properties grants it, at least, the status of a sensitive probe. Several mechanisms tend to maintain WD at low levels, i.e., compression by controlling their growth; filtering (14) as WDs are produced at places not tuned to them; and an AGC mode of functioning. One likely advantage of WD is that the resulting suppressive interactions may improve the analysis of complex sounds by enhancing local contrasts and favoring resonant components (262). Suppressive interactions within the spectral components of complex masks might decrease their masking effect on a probe tone (98, 242). Last, DPs may usefully contribute to the processing of complex sounds as the activity of auditory centers contains a robust representation of the $f_2-f_1$ and $2f_1-f_2$ components (2). Pitch extraction is likely assisted by neuronal responses to these DPs, as suggested by masking experiments that decrease pitch salience and neuronal representations of DPs in parallel (209, 255). [The missing-fundamental perceptive phenomenon is different (see sect. XI, Historical Insert), as
Regardless of its influence on perception, WD still deserves the efforts required to disentangle its intricate origin and propagation mechanisms, as its detection moved from the laboratory to everyday use in audiological setups decades ago. Several examples deserve to be mentioned owing to their everyday usefulness. With sound-evoked OAEs, universal screening of sensorineural hearing impairment has been implemented in neonates, which most often includes OHC dysfunction (22, 122). Being intracochlear sources of sound, OAEs act as a virtual calibrated earphone on the internal side of the stapes, probing the influence of acoustic impedance mismatches along sound-propagation pathways. Impedance changes due to increased pressure in labyrinthine fluids creating endolymphatic hydrops, or even to causes outside the field of otology such as increased cerebrospinal fluid pressure, can be monitored and applied to clinical and physiological investigations (32). Longitudinal studies in large human cohorts have suggested that sound-evoked OAEs rapidly sieve the auditory function, sound-evoked OAEs rapidly sieve the auditory function and help unveil the associated impaired subcellular structures. Last, central modulations of cochlear function via the olivocochlear medial efferent nerve bundle, possibly affording protection against noise overstimulation and attention-driven improvement of signal-to-noise ratio, can also be assayed by OAEs (83).

**XI. HISTORICAL INSERT: TARTINI’S TONES**

The composer Giuseppe Tartini (1692–1770) reported the perception of an additional tone (“terzo suono”), when two tones at frequencies $f_1$ and $f_2$ played in a loud and sustained manner by two violins or oboes were listened to by an experimenter, himself, placed midway, near enough for the sounds to be loud and with similar levels, thus 5–6 steps away. He published a detailed account of his experiments in *Trattato di Musica secondo la vera scienza dell’armonia* printed in 1754 in Padua, Italy. This captivating book (available online as a free e-book) classically leads his name to be used in the alternative designation as “Tartini tones” of intermodulation WD components at arithmetic combinations ($2f_1-f_2$, $3f_1-2f_2$, $f_2-f_1$, etc.). A few years before, the organist Sorge had published similar findings in his book *Vorgemach der musikalischen Komposition* (Lobenstein, 1745–1747).

However, doubts arise whether Tartini and, especially, Tartini’s readers were actually interested in combination tones. From Tartini’s own words, “one has discovered a new harmonic phenomenon . . . if two sounds of just intonation be sounded clearly and loudly together, there will result a third sound lower in pitch than the other two, and which will be the fundamental sound of the harmonic series of which the first two sounds form an integral part . . . ” The suspicion increases at the reason invoked by Tartini in support of his claim to have discovered combination tones well before anybody else, in another of his books published in 1767, *De’ principii dell’armonia musicale contenuta nel diatonico genere*. There he tells that many people could testify that when aged 22, at Ancona, he had discovered the third sound, later (1728) used in Padua to teach pupils how to obtain pure intonation on the violin. When playing double stops (by pressing two strings simultaneously), the pupils’ ability to hear the “terzo suono” served as a guide to correct intonation, since its occurrence indicated that the intonation was pure. As a music teacher, Leopold Mozart had in mind a similar scope when he wrote that Tartini tones are a device that “a violinist can use in playing double-stopping, and which will help him to play with good tone, strongly, and in tune.” Tartini also reported that any two consecutive sounds of the series of harmonics produced by a vibrating body invariably produce the same “terzo suono,” and often used the term “basso fondamentale” throughout his works.

Pitch, the key element invoked in all these reports, is the quality that distinguishes two different notes played on the same instrument with equal force, and that allows notes to be organized on a scale from low to high. While the pitch of a pure tone and its frequency are isomorphic, the pitch of a mixture of tones relates to physical cues in a less straightforward manner. The so-called “periodicity pitch” (146) derives from the detection in central auditory nuclei of periodic patterns of discharges in afferent neurons at the period $1/f$ of the fundamental component, when several harmonics of the same fundamental $f$ (mf, nf, etc., with $m$ and $n$ integers) coexist. The pitch of a sound complex (mf, nf) matches that of $f$ presented alone, as the main periodicity of the discharge patterns is the same regardless of the presence of energy at $f$. When it is missing, the so-called “missing fundamental” (306) shows up as a phantom perception from the neural auditory system. This phenomenon is universally experienced by telephone users. Most telephone lines cut low frequencies <300 Hz so that all male voices are deprived of their fundamental, but the perception of their pitch is not in the least affected. The identification of the fundamental pitch can even persist, although in a weaker form, if each harmonic of a pair is presented in a dichotic manner, showing that the reconstruction of pitch from the two ears is made in the central auditory system. Both the missing fundamental and the Tartini tones are spectacular perceptive effects involving additive mechanisms in relation to the occurrence of a series of regularly spaced spectral lines. Yet they are fundamentally different in that a Tartini’s “terzo suono” is physically produced in the cochlea irrespective of any particular relationship between $f_1$ and $f_2$, while a missing fundamental can be perceived, centrally, only if the primaries are consonant.
The insistence of Tartini and especially his followers to focus on pitch may suggest that if they fell on combination tones, as they were actually interested in harmonic series and for Tartini, in the use of double-stops in the pieces he composed, their interest was largely caught by the missing-fundamental phenomenon. An incidental paragraph on page 17 of Tartini’s Trattato, however, shows that Tartini, as the keen observer he was known to be, did hear genuine combination tones: “perche il terzo suono si ha non solo dagl’ intervalli composti da quantita razionale, ma si ha ancora dagl’ intervalli composti da quantita irrazionale” (“because the third sound stems not only from intervals composed of rational quantities, but also from intervals composed of irrational quantities,” in which case no missing fundamental can exist).

A century later, Hermann von Helmholtz was the first to propose a mathematical theory of Tartini’s tones, based on a Taylor expansion of the mechanical response of a nonlinear system to two primary tones, limited to the quadratic term assumed to generate the largest contribution (it does, unless the nonlinear system displays a peculiar kind of symmetry). Regardless of any harmonic or inharmonic relationship between \( f_1 \) and \( f_2 \), components at \( f_2-f_1 \) and \( f_2+f_1 \) emerge (Die Lehre von den Tonempfindungen, APPENDIX XVI). Helmholtz thus ascribed Tartini’s observations to the \( f_2-f_1 \) quadratic combination tone. Intriguingly, however, he noticed discrepancies between his work and Tartini’s, subtle yet most revealing. According to Helmholtz, Tartini had been an unreliable observer... In his chapter IV, Helmholtz wrote that “Tartini notated all secondary pitches an octave too high.” In hindsight, now that the properties of cochlear distortion are understood, this rebuke provides a forceful vindication of Tartini’s outstanding skills as an experimenter, and of his indisputable discovery of Tartini’s tones as a characteristic cochlear phenomenon: not only did Tartini note his pitches correctly, but this remark shows that Helmholtz could not have detected Tartini’s tones in his studies!

Among the ratios tested by Tartini were 4:3:(2), 5:4:(2), 6:5:(2), 8:5:(2) and 5:3:(2), the first two numbers of each series referring to the primary sounds played by each violin, deliberately chosen as integer multiples of the same fundamental frequency, and in parentheses, the terzo suono identified by Tartini (177). With the difference tone at \( f_2-f_1 \) in mind, Helmholtz predicted that the difference tones reported by Tartini should have been 4:3:(1), 5:4:(1), 6:5:(1), 8:5:(3) and 5:3:(2), thus in many cases, secondary pitches different from Tartini’s. Helmholtz thought, wrongly, as we now know, that higher-order combination tones did not exist as such, but were first-order combination tones of particular harmonics of the primaries (208). For reasons explained in section V, it is now acknowledged that physiological odd-order DPs are much louder than even-order ones, the \( 2f_1-f_2 \) tone reaching \(-15 \text{ dB} \) below the primary tones while the \( f_2-f_1 \) is not audible at stimulus levels below 50 dB SL (79). It is therefore likely that Tartini detected the \((n+1)f_1-nf_2\) combination tones, with the right pitch and not an octave too high, and to give another example, 8:5:(2 = 2 x 5 - 8) and not at all the 8:5:(3 = 8 - 5) of Helmholtz. Tartini’s combinations 5:4:(2) and 6:5:(2) correspond to higher odd-order combinations than \( 2f_1-f_2 \) [that would have been 5:4:(3) and 6:5:(4), respectively], the 3f_3-2f_2 and 4f_1-3f_2 DPOAEs that can be quite large and fully audible. Moreover, Tartini also understood that the terzo suono was best heard when the primary tones were of similar level so that the musicians playing each primary could not hear it as clearly as himself standing midway. This is now attributed to the need for the two primaries to have envelopes of similar size on the BM to maximize their nonlinear interaction and the resulting DPs.

Tartini thus deserves full credit for the discovery of physiological combination tones. His power of analysis led him to disclose several of their unusual properties, now known to be signatures of cochlear nonlinearity, i.e., odd-order predominance and envelope-overlap requirement. But this raises a new shocking question: could an outstanding scientist like Helmholtz have so deeply misunderstood the issue of Tartini’s tones? The (obvious) answer is that he did not. Helmholtz actually never claimed to have specifically studied physiological combination tones; on the contrary, he explains thoroughly, in chapters VII and XI of his book, that many of his sound sources distorted sound, especially when the primary sounds producing combination tones were played on the same instrument or by mechanically coupled sound sources. They also produced higher harmonics. Helmholtz sometimes used purer tones from tuning forks, but mostly < 200 Hz (177). Helmholtz also mentioned his misgivings regarding his, and others’ ability to identify the pitches of combination tones, all the more when buried in a complex mixture of spectral components, as seen above, hence his lack of confidence in Tartini’s reports. In many cases, his measurements were done with the help of one of his resonators, a cumbersome, yet remarkable ancestor of digital spectral analyzers. This could be done only for sounds generated, outside the ear, by the aforementioned nonlinear sources.

Helmholtz’s only mistake was to have assumed that the characteristics of these sources, and of their carefully studied exogenous combination tones, also applied to the case of combination tones from the ear. This would have been justified if the distorting element had been a passive structure, such as the eardrum or ossicles as Helmholtz proposed. Actually, being produced by the MET channels of OHCs, the endogenous combination tones bear the signatures of a slow rate of increase with stimulus level and of odd-order symmetry. All the quantitative characteristics reported by Helmholtz, at variance with these signatures, point to exogenous combination tones, i.e., that they are...
produced by even-order, quadratic distortion (his appendix XVI); that their growth function is steeper than that of primary tones (in the second page of his chapter VII; we know that it should be 2 dB/dB in the absence of compression); that the dominant spectral lines in the case of quadratic distortion are at \( f_2 - f_1 \) and \( f_2 + f_1 \), the summation tone, that Helmholtz most likely heard, not as an endogenous DP as this one is inaudible (208), but as the product of a distorting music instrument. The reasons why the properties of cochlear DPs are drastically different from instrumental ones stem from the dynamic properties of the hair bundle, discussed above, and that have only been recently unveiled.

XII. APPENDIX I: DTOACOUSTIC EMISSIONS WITHOUT TRAVELING WAVE

The emission of SFOAEs with long delays by the auditory organs of all tested tetrapods, i.e., mammals, birds, lizards, and frogs, regardless of the existence of a BM-supported traveling wave, raises the issue of the minimum system able to produce responses with long delays. Delays can be seen as the sum of traveling and filtering contributions. The current view is that the only requirement for producing realistic SFOAEs is the presence of a slightly disarrayed battery of tuned resonators (16). This paper starts from a detailed realistic model of a nonmammalian inner ear, to thoroughly derive an analytic expression of the SFOAE delay. Other authors have reached similar qualitative conclusions regarding the presence of long delays in another category of OAEs evoked by short impulse stimuli [clicks; these transiently evoked emissions are thought to stem, at least partially, from a backscattering mechanism similar to that of SFOAEs (see sect. IVE)].

The core of their argument rests on minimal assumptions. Assume a bank of overlapping bandpass filters connected in parallel, regularly spaced, with gradually increasing center frequencies. Let the bank be excited by an impulse. Every frequency component of the impulse preferentially goes through the filter centered on its frequency. The outputs of this filter and of all other filters recombine to generate an output signal. As a whole, the battery passes all frequencies equally; thus, it responds to the impulse by a similar impulse. However, the output of each individual filter is a slowly damped oscillation at its center frequency. As the oscillations from neighboring filters gradually shift in frequency, when combined they oscillate out of phase with one another and cancel each other, except at the very beginning when they oscillate in unison, hence the impulse as overall output. If the filters get slightly disarrayed, then the cancellation between neighboring components becomes imperfect so that remnants of individual responses show up, with time delays corresponding to several periods of the center frequencies of the disarrayed area. Zwicker (312), in his hardware model of the cochlea simulated as a battery of electrically tuned filters, observed this phenomenon and aptly remarked that although the disarrayed system seemed to emit transient-evoked signals with delays that could suggest the propagation of traveling waves, these delays, in this case, merely corresponded to the time it took for the oscillations of the disarrayed filters to build up.

In a workshop held in 1989 in honor of Thomas Gold and his pioneer idea of an active cochlea, Gold (78) brought forward the analogy of the sound of a concert piano when clapping one’s hands nearby. Every string significantly vibrates for several seconds and ringing is heard, but it would not if the accuracy of frequency mapping was improved by putting 20 times the number of strings. Conversely, ringing would get louder if strings were mistuned. Wit et al. (300) quantitatively modeled the behavior of a disarrayed bank of gammatone filters and showed that not only do its “emissions” resemble TEOAEs in the time domain, but they also present, in the spectral domain, a realistic quasi-periodic pattern of peaks and troughs with the periodicity emerging, not from the structure of the disarray (there was none as it was random) but from the tuning properties of individual filters, just as shown for SFOAEs (16). This happened with no intervention whatsoever of traveling wave or specific cochlear micromechanics, just a battery of tall and broad resonators with neighboring resonance frequencies and some degree of disarray.

XIII. APPENDIX II: BACKWARD VERSUS FORWARD PROPAGATION OF DISTORTION PRODUCTS

The principles allowing DPs to travel backward from their generation place to give rise to DPOAEs should remind the reader of those discussed for SFOAEs, even though here more mechanisms and more nonlinearities come into play. Assume that potential DP sites of generation are distributed along the places where the envelopes at \( f_1 \) and \( f_2 \) overlap and OHCs respond to both frequencies. Consider the case of the \( 2f_1 - f_2 \) DP. Likely, its dominant contributions come from where the displacement is largest at primary frequencies, that is, near the peak response of the BM to \( f_2 \). This place covers a broad interval. Let \( B \) and \( A \) be its limits on the basal and apical sides (FIGURE 22). Since the generation segment is long and not punctual, during stimulation DP-wavelets are evoked along its entire length, every one of them at a particular location \( X \) between \( A \) and \( B \) (FIGURE 22). From \( X \) they start to travel in two directions, apically to their CF location \( D \), and basally to the stapes \( S \). At the stapes or at their CF location, all of them are vectorially added and, depending on their respective phases, cancellation or enhancement occurs. The DP level at the stapes determines the DPOAE level, and the DP level at CF location determines the loudness of the heard combination tone. As seen in section IVA, the phase rotation of a tone increases during propagation, by one cycle every time the wave travels by one wavelength, and as the
wavelength gets shorter when the incoming wave gets closer to its CF place, the phase accumulation speeds up. Scaling symmetry tells that the vibration patterns of $f_2$, $f_1$, and $2f_1-f_2$ are just translated versions of each other along a logarithmic representation of the cochlea. At the apex, where the scaling symmetry is broken, this representation is probably not valid and its predictions fail. It is important for what follows that the phase at $f_2$ rotates fast when this primary nears its CF location, while at the same spot, $f_1$ and $2f_1-f_2$ accumulate less and less phase.

FIGURE 22. Sketch of the interactions between waveforms at $f_1$ and $f_2$ along the BM (envelopes, solid lines; instantaneous vibrations, dotted lines), that lead to the generation of DPs and DPOAEs. Three different frequency ratios $f_2/f_1$ are represented from top to bottom, with $f_2$ fixed. S, stapes, from where DPOAEs are collected; D, place tuned to the $2f_1-f_2$ DP frequency; locus X, within the BA segment, hypothetical site of nonlinear interactions between waves at $f_1$ and $f_2$, along which wavelets at $2f_1-f_2$ are launched forward (toward D) and backward (toward S). Boxplots on the right and left sides depict, for each $f_2/f_1$ ratio, the phase accumulation of $2f_1-f_2$ wavelets launched from X along the paths SXD (on the right) and SXS (on the left). Phase coherence is ensured when phase accumulation no longer depends on X between B and A, and allows a large DP to be measured where all wavelets recombine.
with increasing $f_2/f_1$ ratio, as the peaks at $f_1$ and $2f_1-f_2$ move farther apically.

The DP amplitude dependence on the $f_2/f_1$ ratio can be explained according to the following line of thought [FIGURE 22; from the sketches of the phase behavior of primary tones and DP, diagrams have been built displaying the phase accumulation of wavelets as a function of their generation place $X$, when they travel either forward (right-hand side) or backward (left-hand side); 3 different $f_2/f_1$ frequency ratios are illustrated from top to bottom].

**A. Forward Propagating Wavelets**

In the case of wavelets traveling in the forward (apical) direction, to the place tuned to the DP, the total phase accumulation can be shown to be $2\varphi_1 - \varphi_2 - \varphi_{DP}$ (244, 266). The journey splits into two segments, SX and XD. Along SX, primary tones $f_2$ and $f_1$ accumulate phase rotations $\varphi_2$ and $\varphi_1$ while reaching places of nonlinear interaction $X$, somewhere between $A$ and $B$ near the $f_2$ peak. A DP wavelet starts from $X$ with a $2\varphi_1 - \varphi_2$ initial phase, and travels the next segment of the journey, XD, at a new frequency $2f_1/f_2$. The $\varphi_{DP}$ term expresses the phase accumulation of the DP wavelet along its forward path XD. The nearer $X$ is to $B$, the smaller the initial DP-wavelet lag, because both $f_2$ and $f_1$ accumulated less phase, on their way to their resonance place. Yet the nearer $X$ to $B$, the longer XD, so the larger the phase lag that the wavelet accumulates on its own. Because of the sign “=” in $(2\varphi_1 - \varphi_2) - \varphi_{DP}$, wherever $X$ lays along the BA segment, the stimulus- and wavelet-associated phase lags at the CF location, $D$, act in opposite directions.

At small $f_2/f_1$ ratios (close to 1.0), the phase accumulations $\varphi_2$, $\varphi_1$, and $\varphi_{DP}$ tend to equalize, despite the necessary phase rotation along the $A-B$ interval. For more basal $X$s, wavelets experience this phase rotation during their journey forward; in the case of more apical $X$s, their stimuli accumulated the same phase change instead. That wavelets experience similar phase accumulations regardless of where they were generated between $B$ and $A$ (bottom right diagram on FIGURE 22) is favorable to their adding up coherently to produce a high-amplitude vibration on the BM at CF location (point $D$) so that the DP is clearly audible. At large $f_2/f_1$ ratios however, the forward phase accumulation of DP wavelets, $2\varphi_1 - \varphi_2 - \varphi_{DP}$, is dominated by $-\varphi_2$. It can be seen on FIGURE 22 (top sketch) that between $B$ and $A$, as the places tuned to $f_1$ and $2f_1-f_2$ are much more apical that that tuned to $f_2$, at large $f_2/f_1$ ratios, $\varphi_1$ and $\varphi_{DP}$ remain small as most of the phase accumulations happen apically to $A$. Conversely, $\varphi_2$ rotates fast from $B$ to $A$, and the term $-\varphi_2$ dominates the phase difference between DP wavelets launched from different $X$s. The diagram depicting how $2\varphi_1 - \varphi_2 - \varphi_{DP}$ varies against $X$ for wavelets traveling forward to $D$ (FIGURE 22, top right) shows a spreading spectrum of phases, unfavorable to the generation of a DP at its CF location.

**B. Backward Propagating Wavelets**

Backward propagating wavelets are responsible for the DPOAE in the ear canal. On FIGURE 22, the paths along which phase rotations are to be compared are SXS (diagrams on the left side of FIGURE 22, from top to bottom according to the $f_2/f_1$ ratio), with the same patterns of vibrations at $f_2$, $f_1$, and $2f_1-f_2$ as for the forward-propagation case supporting the evaluation of phase accumulation. Its cumulated value at the stapes S is now $2\varphi_1 - \varphi_2 + \varphi_{DP}$, with $2\varphi_1 - \varphi_2$, as before, representing stimulus phase accumulations along SX, their forward path to the interference places $X$, and $\varphi_{DP}$ representing the phase accumulation of the DP wavelet at $2f_1/f_2$ from $X$ to $S$. The sign flip of $\varphi_{DP}$, compared with the previous case, expresses backward propagation of the DP.

Once again, the more basal $X$, the smaller the initial phase lag with which the DP wavelet is launched. However, this time, the more basal $X$, the less the DP wavelet has to travel to the stapes [actually, the quantitative evaluation of $\varphi_{DP}$ depends on what mechanism is accepted for the reverse propagation of DPs. Whether it is a slow reverse traveling wave along the BM according to the same mechanism as forward propagation, or a fast compression wave in scala vestibuli (see sect. VH) is a topic of intense controversy. The slow traveling-wave mechanism predicts the same $\varphi_{DP}$ forward and backward while the fast compression-wave mechanism predicts a smaller $\varphi_{DP}$ backward.] At large $f_2/f_1$ ratios, the phase accumulation of DP wavelets at $S$, $2\varphi_1 - \varphi_2 + \varphi_{DP}$, is the same as at D, dominated by $-\varphi_2$, as the resonance at $2f_1/f_2$ occurs well beyond $A$ in the apical direction so that $\varphi_{DP}$ is negligible. This term $-\varphi_2$ is again unfavorable to the generation of a large overall DPOAE. For $f_2/f_1 = 1$, $\varphi_1 \approx \varphi_2 \approx \varphi_{DP}$, as for the forward-propagation case, but because of the sign flip of $\varphi_{DP}$ in $2\varphi_1 - \varphi_2 + \varphi_{DP}$, the phase difference among backward-propagating wavelets varies approximately like $+2\varphi_2$. This is too fast a phase rotation to build a large DPOAE. At decreasing $f_2/f_1$ ratio, the phase lag $(2\varphi_1 - \varphi_2 + \varphi_{DP})$ varies monotonically, and this, from $-\varphi_2$ at high ratios to $+2\varphi_2$ at low ratios. The change in sign implies that between these extremes an intermediate $f_2/f_1$ ratio can be found where the result is 0. Near this ratio, the phases of all wavelets remain coherent (FIGURE 22, left side, intermediate diagram) and the resulting DPOAE is maximum. According to the models and the results of human measurements, this happens around $f_2/f_1 = 1.20–1.25$. 
XIV. APPENDIX III: COCHLEAR DISTORTION IN THE CASE OF THREE OR MORE TONAL STIMULI

The presence of at least three tones in a nonlinear cochlea, $f_1, f_2$ (the primary tones at the origin of a DPOAE), and $f_3$ (a suppressor tone probing the two-tone interaction) creates a large complexity as the number of possible combinations, notably if sound levels are high enough, increases considerably compared with the two-tone situation. The problem faced by experimenters is to avoid technical issues such that different combinations lead to the same spectral line, and more generally to disentangle the possible combinations to identify the sources of observed combination tones.

A first issue is order aliasing, happening when two different combinations of $f_1, f_2,$ and $f_3$ lead to the same combination tone, for example, if there are integers $L, M, N$ and frequencies $f_3$ such that $L f_1 + M f_2 + N f_3 = 2 f_1 f_2$. The resulting spectral line at $2 f_1 f_2$ is a complex vector addition of two contributions, one of which comes from places and mechanisms that were not acting in the absence of $f_3$. The choice of $f_3$ can be made so as to avoid obvious aliasing problems, and data can be collected using phase rotations of the suppressing tone. For example, if the $f_3$ suppressor is presented four times with a phase set at $0, \pi/2, \pi,$ and $3\pi/2$ in succession, most combination tones containing $f_3$ are cancelled without the directly produced $2 f_1 f_2$ being affected. To characterize the influence of an interfering tone at $f_3$, interference response areas are built (Figure 19) representing how this $f_3$ tone influences the amplitude and phase of a given DPOAE, against interfering level (the term interference is more appropriate than suppression, as the DPOAE amplitude can sometimes increase in the presence of $f_3$). A striking feature of these areas is the presence, at moderately loud primary levels, of sharp high-frequency interference lobes when $f_3$ is set at frequencies that can be more than one octave above $f_2$ (Figure 19) (66, 163, 169). A simple explanation is that there is a secondary site of DP generation in the high-frequency region in addition to those already described; that the DP generated in this region is comparable in level to the other sources from the $f_2$ and $2 f_1 f_2$ sites; that enhancement or suppression is observed depending on the phase difference among the mixed DP contributions so that when the $f_3$ tone removes the basal contribution, the overall level can increase if this contribution is out of phase with the other ones. Therefore, if the plotting procedure takes into account only level changes and not phase shifts, the interference effect can well be missed (66, 302).

The mechanism of basal generation of two-tone DPOAEs, revealed by a third tone, calls for an explanation. A first possibility is a harmonic mechanism (64), for example, a sound at $2 f_1$ could be produced around the peak response to $f_1$, and induce a response at its characteristic $2 f_1$ place. Although small because having to travel in the wrong direction over a too loose BM, in certain circumstances, the $2 f_1$ harmonic might remain large enough to interact nonlinearly with the lower frequency $f_2$ stimulus, thereby generating a $2 f_1 f_2$ difference tone, from a quadratic mechanism akin to that producing $f_2 - f_1$ at the place tuned to $f_2$. Direct measurements at the cochlear apex confirm the existence of a very large mechanical component at $2 f_1$ (85), and the high-frequency cutoff of the BM may be less sharp at higher levels than it is usually. Another mechanism termed catalyst by Fahey et al. (64) assumes the existence of richer nonlinear interactions in the presence of a third tone, that could generate $2 f_1 - f_2$ and $2 f_2 - f_1$ responses via intermediate stages involving $f_1, f_3$ and $f_2, f_3$ combinations. Contrary to the harmonic mechanism for which the third tone serves only for unveiling an existing source, the catalyst mechanism requires the $f_3$ component. As an example of catalytic pathway, the combination tones $f_3 - f_1$ and $f_3 - 2 f_2$ could combine at the $f_3 - f_1$ place and generate a difference tone at $2 f_2 - f_1$. Likewise, mixes ($f_3 - f_2$) - ($f_3 - f_1$) + $f_1$ or ($f_3 - f_2$ + $f_1$) - $f_3 + f_1$ could generate a $2 f_1 f_2$ component adding up to the cubic $2 f_1 f_2$ produced near $f_2$. Last, cogent evidence has been produced (170), with the help of accurately mapped lesions of OHCs, of basal interactions between the tails of $f_1$ and $f_2$ able to produce sizeable components. Surprisingly, phase characteristics previously attributed to a coherent reflection mechanism that was thought to occur at the $2 f_1 f_2$ CF place were observed (vertical banding), in humans and rabbits, which suggests that the two-mechanism taxonomy may have to be completed. Whether the possible contribution of a traveling mode along the Reissner’s membrane would better account for this vertical banding pattern remains to be investigated (217).

As for the simpler case of OAEs produced by a single tonal stimulus (SFOAEs), a similar application of a catalyst principle may account for the outcome of suppression experiments when a suppressor tone at $f_3$ is added to the SFOAE-evoking probe stimulus at frequency $f_p$ (251). Nonlinear combinations of $f_2$ and $f_p$ can produce, for example, $3 f_p - f_2, 2 f_p - f_2, f_3 + 2 f_p, f_3 + f_p,$ etc., distortion components, from which $(|3 f_p - f_2| - (2 f_p - f_3))$ or $(|f_3 - 2 f_p| + (3 f_p - f_2))$ combinations produce $f_p$, perhaps more efficiently than the $f_p$ produced by coherent reflection, the strength of which has been called into question in small rodents (167). The production of an SFOAE at $f_3$ by catalyst processes would occur at more basal sites than by coherent reflection, thus with shorter group latency (see the timing and backward-propagation controversy in sect. IVF).

Recent protocols designed combinations of even more than three tones. DP phase gradient paradigms (for example, an $f_1$-sweep at fixed $f_2$) afford a noninvasive means to evaluate cochlear travel times. With the goal of simultaneously acquiring data for avoiding interrecording variability and improving the computation of phase gradients, the lower of the two tones of the traditional $f_1, f_2$ stimulus can be replaced by a closely spaced group of tones ($g_n, g_n,$ etc.) (175).
The resulting DPs are made of several groups of simultaneous frequency components from which group delays can be derived from a single recording. Among the DPs, a novel category emerged, made of sidebands around the single tone primary \( f_2 \). This group has no two-tone equivalent, as it consists of a set of sidebands around the single stimulus component \( f_2 \) at frequencies \( f_2 + g_i - g_j \), with \( i \neq j \). With stimulus frequencies chosen such that all possible difference and sum frequencies were unique (a so-called “zwezis” complex) (272), each \( i \neq j \) combination resulted in a unique DPOAE frequency, unambiguously affected to a single pair \((g_i, g_j)\).

Another interest of this sideband complex of DPs is that it persists at wide frequency separation (several octaves) between \( f_2 \) and \( g_i \), whereas with a \((f_1, f_2)\) pair, the ratio \( f_2/f_1 \) can hardly exceed 1.5–1.6. Moreover, the DPOAE frequencies depend only on the frequency differences \( g_i - g_j \), and are unaffected by a collective frequency shift of the entire tone complex \( g \). It is therefore possible to efficiently use a quite low frequency for the average \( g \), such that the forward travel time of the \( g \) stimuli to the place of nonlinear interaction at the \( f_2 \) CF place is very brief. Thus the time properties of DPs should be associated principally with the reverse journey of DPs, lately a hot matter of controversy (231) (see sect. VH).

**XV. APPENDIX IV: CHARACTERIZATION OF NONLINEARITIES BY WIENER KERNELS**

In terms of the mathematics for describing a biological system such as the cochlea, it is customary to test it with the help of a restricted set of stimuli, for example, pure tones, then extrapolate the obtained results to predict the response of the system to any stimulation, however complex, for example speech. This is straightforward for a linear system, as it can be equally well described by a transfer function in the spectral domain (for example, the bell-shaped transfer function of a bandpass filter), or in the time domain, by its impulse response (the damped oscillation obtained when the bandpass filter is excited by a short click). A third method is to use Gaussian white noise as a stimulus; the first-order cross-correlation \( h_1(\tau) \) between this stimulus and the system’s response is called first-order Wiener kernel. In a linear system, \( h_1(\tau) \) is identical to the impulse response of the system. In a nonlinear system, the impulse response of the system also contains other components. The use of Gaussian white noise as input to the auditory system, with spikes in auditory nerve fibers as the output, was first introduced by Egbert de Boer (46) who computed the revcor (for “reverse correlation”) function as the average value of the stimulus \( x(\tau) \) at time \( \tau \) before the occurrence of a spike: normalization to the stimulus power spectral density provides the cross-correlation \( h_1(\tau) \).

In a linear system, from any of these equivalent descriptions, it is possible to correctly predict any response, for example, the response to a sum of sinusoids is the sum of the responses to each sinusoid presented alone. Nonlinear systems are more difficult to analyze as their response to a sum of sinusoids is not the sum of the responses to each sinusoid presented alone. The use of perturbation theory is one convenient solution for weakly nonlinear systems. For more strongly nonlinear systems, a first possibility is to use a model and derive from its analysis a set of differential equations describing the response of the system. The parameters of the model can then be adjusted by comparison between model predictions and experimental data. The topology of a strongly nonlinear system, however, is seldom well known, and a more general and systematic method relies on Volterra and Wiener formalisms (284, 294; reviewed in Ref. 59), describing the responses of a system as a sum of functions including correction terms unfolding the nonlinearities at increasing orders. With Gaussian white noise as input to the system, its Wiener kernels are independent of each other and can be computed from input-output cross correlations. So, the outcome of the revcor method of de Boer is proportional to the first-order Wiener kernel. It provides an idea of the impulse response of the system connected to the auditory nerve fiber under investigation, from which by Fourier transform, the bandpass filter characteristics (i.e., frequency tuning) can be derived. For nonlinear systems, the first-order Wiener kernel is only one component of the impulse response to which higher kernels also contribute.

In general, second-order kernels give a measure of the “cross-talk” between the responses to two impulses, i.e., of how much the presence of the first impulse influences the response to the second one (162). Another prominent interest of the second-order Wiener kernel, regarding auditory nerve fibers, is that, thanks to even-order cochlear nonlinearities encoded in the low-frequency response envelopes, it still provides information on temporal coding in auditory-nerve fibers innervating the entire length of the cochlea, whereas the revcor function vanishes as neuronal phase-locking decreases (215, 275). Studies of the nonlinearities themselves have been less rewarding so far because their translation in physiological terms often requires models that can be probed against experimental data (215). Two types of conclusions can then be reached. In Reference 275, the authors, who described the hearing organ and auditory neurons of frogs with the help of a sandwich model incorporating filters and nonlinearities in cascade, deduced the orders of the filters from a careful inspection of the second-order Wiener kernel. Yet, they also had to acknowledge the failure of their model to account for the two-tone interactions revealed by Wiener analysis. In Reference 156, Wiener analysis applied to noise-evoked OAEs demonstrated the consistence of the first-order kernel and of click-evoked OAEs. The compressive level dependence of cochlear responses was revealed in the nonlinear dependence of the first-order Wiener kernel on stimulus level, but the fact that no emission component
was found in the higher order kernels indicated that at each particular noise level the cochlea behaved nearly linearly, as would do a system with an AGC once set up after a time delay.

ACKNOWLEDGMENTS

We are particularly indebted to Jean-Pierre Hardelin at Institut Pasteur for his many apt suggestions for improving the manuscript and for his meticulous editing work. The recommendations of two anonymous reviewers were of great help for improving the consistency of arguments from section to section and better emphasizing the harmony of all manifestations of distortion reviewed in this text.

Address for reprint requests and other correspondence: P. Avan, Laboratory of Neurosensory Biophysics (INSERM UMR 1107), School of Medicine, 28 Place Henri Dunant, 63000 Clermont-Ferrand, France (e-mail: paul.avan@udamail.fr).

GRANTS

The authors’ research was supported by grants from ERC 2011-AdG nr 294570, Fondation Louis-Jeantet, Novalis-Taitbout, Réunica-Prévoyance, ANR MNP P007354 “Presbycusis,” ANR 08-ETEC-001-01 “Audiapic,” Groupe Entendre, Fondation de l’Avenir pour la Recherche Médicale (ET1-617).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

REFERENCES


