Mode conversion in the cochlea?

Robert S. MacKay*
Mathematics Institute and Centre for Complexity Science,
University of Warwick, Coventry CV4 7AL, U.K.

August 28, 2008

Abstract

It is suggested that the frequency selectivity of the ear may be based on the phenomenon of mode conversion rather than critical layer resonance. The distinction is explained and supporting evidence discussed.

PACS numbers: 43.64.Kc, 43.66.Ba
Keywords: Cochlea, travelling wave, mode conversion
Running title: Mode conversion in the cochlea

*Electronic mail: R.S.MacKay@warwick.ac.uk

1 Introduction

The cochlea is the part of the ear where mechanical vibrations (forced by sound waves in the air) are turned into neural signals. There is a huge literature about it, both experimental and mathematical (for summaries, see Robles & Ruggero, 2001, or Ch.23 of Keener & Sneyd, 1998). There are large differences between the cochleae of different species (especially between mammals, birds and reptiles); in this paper attention is restricted to mammalian and usually human cochleae.

Most current explanations for the function of the mammalian cochlea are based on "critical layer resonance", the build up of wave energy at a frequency-dependent place where the wavelength goes to zero, e.g. Lighthill (1981), Nobili, Mammano & Ashmore (1998), de Boer (1996), Hubbard & Mountain (1996), even if the words and mathematical theory are not always used and it is disputed by some (e.g. the discussion in section VII of Olson, 2001).

The goal of this paper is to suggest that critical layer resonance is not the right explanation: rather it might be "mode conversion", the propagation of a wave to a frequency-dependent place where it turns round and comes back out in another wave mode, the energy density forming a peak at the turning point. Critical layer resonance models fail to fit several experimental observations, whereas mode conversion produces consistent results.

A notable alternative to both critical layer and mode conversion models is given by de Boer & Nuttall, 2000, on which I'll comment at various points.

Critical layer resonance modelling is reviewed in section 2, with particular reference to the cochlea, first in the frequency domain and then in the time domain. In section 3 a range of inadequacies of critical layer resonance models for the cochlea is listed. Section 4 surveys the phenomenon of mode conversion and section 5 suggests that it could occur in the cochlea. Section 6 mentions some precedents. Section 7 gives some ideas about how mode conversion could arise physiologically. Section 8 discusses stability constraints.

Despite clear experimental evidence for nonlinear effects (e.g. summaries in Robles & Ruggero, 2001, and section 8 of de Boer, 1996), and the large amplitude that occurs near a mode conversion point, this paper is limited to treatment of the linear regime, as the first step in the analysis.

2 Critical layer resonance

First the phenomenon of critical layer resonance is recalled, together with how it arises in a simple model of the cochlea. I've given quite a long treatment of this, starting from elementary considerations, because it is important to appreciate how mode conversion differs from it, but readers who are already familiar with critical layer resonance can skip the section.

2.1 Frequency domain

The context for critical layer resonance is waves in a medium whose properties vary smoothly in space. If the medium can be treated as linear (i.e. any linear combination of two solutions is another solution) and time-independent, one can analyse its behaviour by considering solutions with a single frequency ω, i.e. time-dependence like (the real part of) $e^{\mathrm{i}\omega t}$ (ω is the "angular frequency"; some may prefer to reserve the word "frequency" for $f = \omega/2\pi$). Then one can attempt to write solutions as superpositions of waves $e^{\mathrm{i}(\omega t - k.x)}$, or better, WKB (after Wentzell, Kramers and Brillouin; also attributed to Liouville and Green and contributed to by Jeffreys) solutions $A(x,\omega)e^{\mathrm{i}(\omega t - \int^x k(y,\omega)dy)}$ with A determined by the evolution of wave energy density along the rays, where the wave vectors $k(x,\omega)$ are solutions of the local *dispersion relation* for the medium. For introductions to linear wave theory in fluid mechanical contexts, see Whitham (1974), Lighthill (1978). A *critical layer* is a place x (or line or surface, depending on the ambient dimension) where the local wave number $k(x,\omega)$ goes to infinity for some mode at the given frequency ω.

The phenomenon of critical layer resonance is as follows. Waves can propagate on only one side of the critical layer. Waves propagating towards it slow down and take infinite time to reach it. Their energy density increases in inverse proportion to their group speed until damping effects take over. Virtually all the energy is absorbed in a zone near the critical layer and almost nothing is reflected.

Fluid mechanics attribute the discovery and analysis of the critical layer phenomenon to Booker & Bretherton (1967), but it had already been done by Budden (1955) in plasma physics, and it had already been found numerically in a model of the cochlea by Peterson & Bogert (1950)!

The best way in the cochlear context to illustrate the critical layer phenomenon is provided by the passive 1D cochlear model of Peterson & Bogert (1950), so it is now recalled. Consider the cochlea to consist of a rigid tube separated into two by a flexible membrane along its length, called the "basi-

lar membrane", and the fluid motion to be incompressible (this is good as long as the time for acoustic waves in the fluid to equilibrate the pressure to that corresponding to the boundary conditions at the ends is small compared with the timescale on which the boundaries move, though there are authors who claim this fails above 7 kHz, referred to in Lighthill, 1981) and for the moment only longitudinal (this is inaccurate, cf. Fig. 7 of Olson, 2001, but the effects of 2D and 3D fluid flow are recalled shortly and make little qualitative difference (Lighthill, 1981)). So the volume fluxes in the two tubes are equal and opposite, denoted by $\pm j(x,t)$ at longitudinal position x and time t, and the pressures are uniform over each part of the cross-section. Denote the area displaced by the membrane from its equilibrium position in the two parts of the cross-section by $\pm a(x,t)$ and the pressure difference across the membrane by $p(x,t)$ (Lighthill, 1981, denotes it $2p$). If the membrane deforms in a mode shape $\zeta(y)$ with respect to lateral position y (which may depend on longitudinal position x), then a is related to the displacement z at a chosen y_0 by $a = \frac{z}{\zeta(y_0)} \int \zeta \; dy$, so some authors convert a to z. Then fluid mass conservation leads to

$$\frac{\partial a}{\partial t} = \frac{\partial j}{\partial x}, \tag{1}$$

and horizontal momentum conservation (ignoring viscous effects) to

$$\sigma \frac{\partial j}{\partial t} = -\frac{\partial p}{\partial x}, \tag{2}$$

where

$$\sigma(x) = \frac{\rho_1}{A_1} + \frac{\rho_2}{A_2},$$

with ρ_i the fluid density in the two tubes and $A_i(x)$ their equilibrium cross-sectional areas. It would be reasonable to take $\rho_1 = \rho_2$, but the analysis is no different when $\rho_1 \neq \rho_2$ (and actually the upper channel consists of two fluids of different ionic composition separated by a very flexible membrane). Combining equations (1) and (2),

$$\frac{\partial^2 a}{\partial t^2} = -\frac{\partial}{\partial x} \frac{1}{\sigma} \frac{\partial p}{\partial x}. \tag{3}$$

The vertical momentum equation is

$$m \frac{\partial^2 a}{\partial t^2} = -p - \lambda a, \tag{4}$$

where $\lambda(x)$ is the stiffness of the membrane (meaning the pressure difference to produce unit area displacement in the plane at constant x) and $m(x)$ an

effective "density" of the membrane. If the membrane deforms in mode ζ as above and has mass $\mu(x, y)$ per unit area in (x, y), then (Lighthill, 1981)

$$m(x) = (\int \mu\zeta^2 dy)/(\int \zeta \ dy)^2. \tag{5}$$

A more realistic evaluation of m requires incorporation of how the organ of Corti moves, which in the above treatment is regarded as moving rigidly with the basilar membrane. In live cochlea it is well established that there are additional forces generated by the outer hair cells (OHCs), e.g. Robles & Ruggero, 2001. Effects of active feedback can be added to (4), as could any other contributions to the transverse force, but for simplicity of illustration of the phenomenon we continue without them.

Combining (3) and (4), and looking for solutions with time dependence $e^{\mathrm{i}\omega t}$, yields

$$\left(\frac{\lambda}{\omega^2} - m\right) \frac{\partial}{\partial x} \frac{1}{\sigma} \frac{\partial p}{\partial x} + p = 0. \tag{6}$$

Thus if the parameters σ, λ, m are treated as locally independent of x one obtains $p \propto e^{\pm \mathrm{i}kx}$ with dispersion relation

$$(\lambda - m\omega^2)k^2 - \sigma\omega^2 = 0. \tag{7}$$

The term $\sigma\omega^2$ proportional to ω^2 is sometimes described by saying the response of the longitudinal fluid flow to transverse membrane motion acts like an "added mass" for the membrane, with density σ/k^2.

Lighthill (1981) surveyed ways to calculate 2D and 3D corrections to this added mass, which involve modifying (3) to give a as the convolution of p with a nonlocal version of the second derivative of a delta-function, so one no longer obtains a differential equation, but one can still perform WKB analysis, inserting

$$p = -\omega^2 I(k) a$$

into (4) with

$$I(k) \sim \begin{cases} \frac{\sigma}{k^2} \text{ for } kh \ll 1 \\ \frac{2\rho}{w|k|} \text{ for } kh \gg 1, \end{cases} \tag{8}$$

where w is an effective width of the membrane and h an effective height of the two channels.

Sticking to the 1D model for ease of exposition, (7) yields

$$k^2 = \frac{\sigma\omega^2}{\lambda - m\omega^2}.$$

It is usually stated that σ and m do not vary much along the length of the cochlea. On the other hand, Fig. 1 of Mammano & Nobili (1993) shows a significant variation of A_1 and A_2, hence of σ, and it seems difficult to me to estimate m; the simplest guess from (5) would be that m is inversely proportional to the width w of the membrane. Nevertheless, all agree that λ decreases by a factor of about 10^4 from base to apex (partly because the width of the membrane increases by a factor of about 4), so k^2 increases. Taking the direction of increasing x from the base to the apex, k^2 goes to $+\infty$ as x approaches a place where

$$\lambda(x) = m(x)\omega^2 \tag{9}$$

and is thereafter negative (i.e. k is imaginary, so solutions grow or decay exponentially with respect to x locally). Thus there is a critical layer at the solution $x(\omega)$ of (9). It moves from base to apex as the frequency ω decreases from an upper limit to a lower limit.

It is useful to consider the *affine impedance model* obtained from (6) by making the straight-line approximation

$$\frac{1}{\sigma(x)}\left(\frac{\lambda(x)}{\omega^2} - m(x)\right) \approx \beta(x(\omega) - x)$$

for x near $x(\omega)$, and supposing σ locally constant. In the scaled variable

$$X = \frac{x - x(\omega)}{\beta}, \tag{10}$$

this yields

$$Xp'' = p. \tag{11}$$

The solutions can be written in terms of Bessel functions of order 1 (de Boer & MacKay, 1980). Using (3) to convert p to a gives

$$a = \frac{\sigma p}{\omega^2 \beta^2 X},$$

and one obtains that the solutions on each side of the critical layer look like linear combinations of those in Fig. 1. In particular $a(X) \to \infty$ like $|X|^{-1}$ for most solutions as x approaches $x(\omega)$ from either side.

The differential equation (6) does not tell one how to connect the solutions on the two sides of the critical layer, because it is singular at $x = x(\omega)$ (the coefficient of the highest derivative goes through 0 there). Nevertheless, adding a little damping (e.g. a dissipative term $-\delta\frac{\partial a}{\partial t}$, $\delta > 0$, to (4)) allows

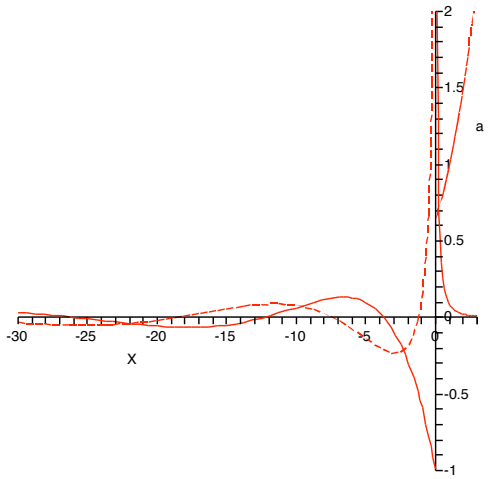

Figure 1: Graph of $a(x)$ for some solutions of (6), plotted against $X = \frac{x-x(\omega)}{\beta}$.

one to match the solutions on the two sides: for the affine approximation the answer is that the solution with no growing mode as $X \to +\infty$ (a physically reasonable requirement) has only in-going waves on the other side (Budden, 1955; Stix, 1962; Booker & Bretherton, 1967): at one phase of the oscillation the solution for a is the dashed curve of Fig. 1 on the left (zero on the right), and $\frac{\pi}{2}$ later it is the solid curve (both have infinite jump discontinuity). The dotted solution on the right is not excited. Then the amplitude of oscillation $|a|$ is proportional to a function $A(X)$, drawn in Fig. 2(a). Damping makes the maximum of $|a(x)|$ finite, albeit still large (inversely proportional to the damping strength) (de Boer & MacKay, 1980).

Similarly, one can plot the phase ϕ of the solution (Fig. 2(b)). Note the jump of $-\frac{\pi}{2}$ on crossing the critical layer, which is determined by considering the effect of small damping. Analysis of the Bessel functions entering the exact solution shows that the phase deviates from the WKB approximation $\phi = -\int k \, dX = 2\sqrt{-X} - \frac{\pi}{4}$ (choosing the constant to fit the asymptotics of the Hankel function) by losing an extra $\pi/4$ in the non-WKB regime, followed by this jump of $-\pi/2$.

Somewhere to the left of the critical layer one has to match the solution of the affine impedance model to the full model. This can be done with

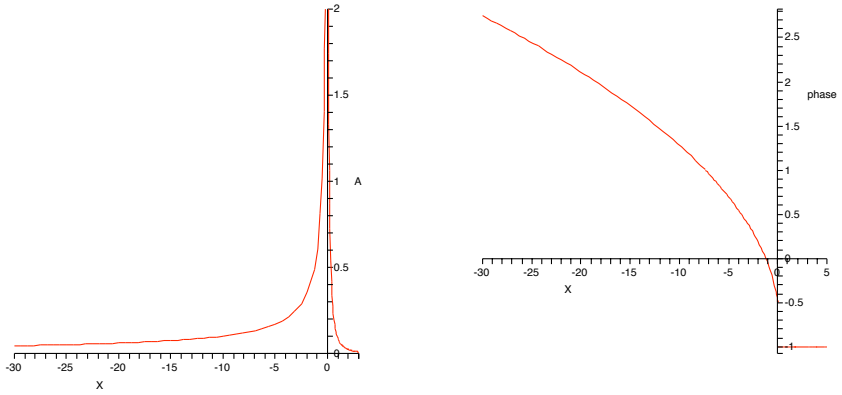

Figure 2: (a) Amplitude $A(X)$ and (b) scaled phase $\phi(X)/\pi$ for the affine impedance critical layer model.

ease if the affine approximation is good over a range $x(\omega) - C\beta < x \leq x(\omega)$ for some C significantly larger than 1, because then it can be matched to a WKB solution coming in from $x < x(\omega) - C\beta$. The condition for validity of the WKB approximation is $k' \ll k^2$ and we assume this holds up to a point where the affine impedance approximation (11) takes over (for the affine impedance model, $k'/k^2 = \frac{1}{2}|X|^{-1/2}$, so WKB is good for $|X| \gg \frac{1}{4}$).

When 2D or 3D effects for the motion of the fluid are taken into account, the local dispersion relation changes from (7) to

$$I(k) = \lambda/\omega^2 - m.$$

Then the criterion for the WKB approximation is just $\beta \ll \bar{A}/w$, where $\bar{A}$ is the harmonic mean of A_1 and A_2, which holds nearly everywhere in the cochlea (Lighthill, 1981) (if the membrane is assumed to fill the width of the cochlea then $\bar{A}/w \approx h$, so this shorthand will sometimes be used). It follows that the amplitude $|a|$ grows like $|X|^{-1}$ again near the critical layer, but the phase goes to $+\infty$ in proportion to $-\log|X|$ as X increases to 0.

2.2 Plausibility of critical layer resonance in the cochlea

Critical layer resonance sounds a perfect explanation for the cochlea. Indeed, Lighthill (1981) argued that any model of the cochlea must be based on critical layer resonance. Frequency ω is mapped to place $x(\omega)$, so a given inner hair cell (which transduces movement into neural signals to the brain)

is stimulated mainly by Fourier components with only a given frequency. One can work out the frequency selectivity, assuming some reference for forcing amplitude. Sound pressure level in the ear canal is a common choice in experiments; to translate this into models one could assume conservation of wave energy through the ossicles and ear drum; the sound energy flux in the ear canal is $\frac{A}{\rho c}|p|^2$, where $|p|$ is the amplitude of pressure fluctuation, A the cross-section of the ear canal, ρ the density of air and c the speed of sound in air; so constant sound pressure level corresponds to constant energy flux, independently of frequency. In the model, the energy flux for a right-going wave in the WKB regime is

$$\Phi = \langle \Re p \ \Re j \rangle = \frac{1}{2}\Re(p\bar{j}) = \frac{1}{2}\sigma \left(\frac{\omega}{k}\right)^3 |a|^2, \tag{12}$$

by (1) and (2), where $\langle\rangle$ denotes the average over a cycle. Thus the amplitude of oscillation as a function of incoming frequency ω for fixed incoming energy flux Φ in the WKB regime is (for the 1D model)

$$|a| = \sqrt{\frac{2\Phi}{\sigma}} \left(\frac{k}{\omega}\right)^{3/2} = \frac{\sqrt{2\Phi}\sigma^{1/4}}{(\lambda - m\omega^2)^{3/4}}.$$

Matching this to the energy flux of the solution of (11) gives

$$|a| = \pi\sqrt{\frac{\Phi}{2\sigma}}(\omega\beta)^{-3/2}A(X)$$

in the regime where (11) holds. Consequently, the frequency response at a given place looks like Fig. 3(a), computed for the *exponential model*: $\lambda(x) = Ce^{-\alpha x}$, σ and m constant (de Boer & MacKay, 1980).

The exponential model has a scaling symmetry so it suffices to study a single place or a single frequency and there is just one free parameter $\frac{\alpha^2 m}{\sigma}$, which is about $\frac{1}{40}$ in the cochlea (de Boer & MacKay, 1980); actually, the result in the WKB regime is independent of its value, so it is plotted for $\frac{\alpha^2 m}{\sigma} = \frac{1}{4}$, else the decay above the resonant frequency becomes too steep to see. Also the extension of the WKB curve to the left half of the affine region agreed to within one pixel with that for the affine model, so just the WKB curve and the right-hand half of the affine region are plotted.

A more common protocol in experiments is to measure the forcing amplitude required to produce a given amplitude of response at a given point, which also has the advantage for comparison with theory that the system is likely to remain in the linear regime. Also the measure of response amplitude is usually basilar membrane velocity $\dot{z}$, which on linear theory is

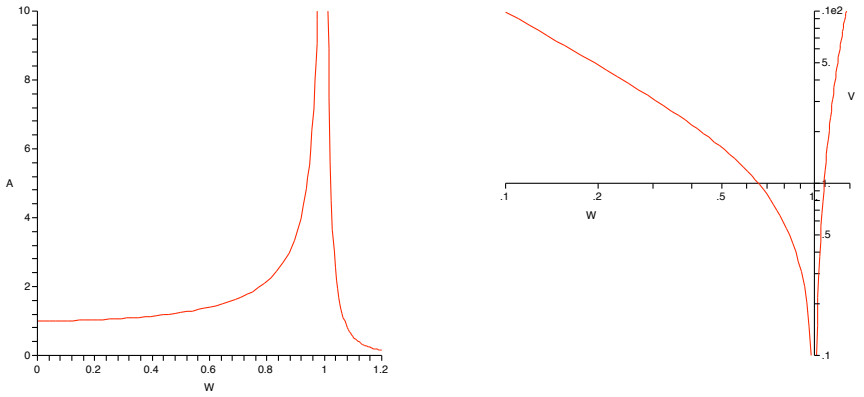

Figure 3: Exponential model with $\frac{\alpha^2 m}{\sigma} = 1/4$: (a) Amplitude $A = |a|$ of response at a given place x to forcing of a given energy flux, as a function of scaled frequency $W = \frac{\omega}{\omega_r(x)}$, with $\omega_r(x)$ being the resonant frequency at x; (b) a tuning curve (log-log graph of basilar membrane velocity amplitude $V = |\dot{a}|$ against scaled frequency W).

proportional to $\dot{a}$ at a given place. It is then plotted on a log-log scale and called a "tuning curve" (though sometimes a linear scale is used for the frequency). A tuning curve for the exponential model is shown in Fig. 3(b). If the effects of damping are incorporated, which rounds off the spike, the resulting tuning curves look roughly like observed ones, e.g. Fig. 6 of Robles & Ruggero, 2001. One can also compute the phase at a given position relative to that at the stapes, as a function of input frequency, and incorporate modifications for the effects of 2D and 3D fluid flow.

Good fits to the observed frequency response of the cochlea have been obtained by suitable dependence of parameters on x, including damping (both positive and negative) (de Boer & Nuttall, 2000), though interestingly their best fits have no critical layer. I will comment more on this is Section 7.

2.3 Time domain

On the view of the ear as a frequency analyser it is natural to work in the frequency domain, as this paper has done until now. Some issues are better discussed in the time-domain, however, such as response to an impulse ("click") (Robles et al, 1976; Recio et al, 1998), even though for a time-invariant linear system this is in principle deducible from the frequency

response. Almost any treatment of nonlinear effects requires formulation in the time-domain. Even the stability problem is best treated in the time-domain. So here the system, with or without 2D or 3D effects and OHC forces, is formulated as a time-evolution. This point of view was taken by Mammano & Nobili (1993), for example.

Given the deflection a of the membrane as a function of x along the membrane and boundary conditions at the base and apex, the pressure field is determined by the assumptions of incompressible and irrotational flow, in particular the pressure difference p across the membrane is determined as a function of x. The simplest case is the one-dimensional fluid flow approximation and no OHC force. Then the equation determining p is

$$m\frac{\partial}{\partial x}\frac{1}{\sigma}\frac{\partial p}{\partial x} - p = \lambda a. \tag{13}$$

This does not involve time. Physically what happens is that the pressure equilibrates rapidly to a solution of (13) via acoustic waves in the fluid. The situation is analogous to water waves (Whitham, 1974; Lighthill, 1978) where the velocity potential under the surface is determined instantaneously by a boundary condition at the surface, and it is another surface boundary condition that determines the evolution.

Equation (13) needs supplementing by boundary conditions at the base and apex, however, denoted $x = 0$ and x_h respectively. An appropriate boundary condition at the apex, where there is a hole in the membrane called the helicotrema which permits fluid to flow from one side to the other, is that $p = \frac{\rho}{d}\frac{\partial j}{\partial t}$ where d is the diameter of the hole (it has area about 0.06 mm^2) (Lighthill, 1981). Under the 1D flow approximation, (2) makes the boundary condition at $x = x_h$

$$p + \frac{\rho}{\sigma_h d}p' = 0, \tag{14}$$

where $\sigma_h = \sigma(x_h)$.

At the base, the channels end in flexible membranes called the oval and round windows. Actually these are on the sides of the channels near the base, but let us model as if at their ends. Both can flex in and out of the middle ear, and the oval window moves the stapes which moves the other ossicles and hence the ear drum. I'm not aware of a good treatment of this boundary condition, though much is written about the impedance of the stapes, so here is a proposed treatment. Denote by v the volume of fluid displaced by the oval window (positive if displaced into the scala vestibula). By the assumed compressibility of the fluid and rigidity of the bone around

the cochlea, an equal and opposite volume is displaced by the round window, and $j(0) = \dot{v}$. Using (2) again, this gives

$$p'(0) = -\sigma_0 \ddot{v}, \tag{15}$$

where $\sigma_0 = \sigma(0)$. Now suppose change of v and associated motion of the windows, ossicles and eardrum has an effective "mass" M (in g cm^{-2} because with respect to a volume-coordinate) and "elasticity" μ, so that

$$M\ddot{v} = -\mu v - Ap + F$$

where A is an effective mean area of the two windows and F is any external force due to interaction of the eardrum with sound waves in the outer or middle ear. Using (15), this gives boundary condition at $x = 0$

$$Ap - \frac{M}{\sigma_0} p' = -\mu v + F. \tag{16}$$

The ratio $\frac{M}{A\sigma_0}$ is an effective length of the stapes (if replaced by a cylinder of material of the density of the fluid with the same volume).

Equation (13), with boundary conditions (14) and (16), determines the function p in terms of the function a and the numbers v and F. Then a and v evolve in time according to

$$\begin{aligned} \frac{\partial^2 a}{\partial t^2} &= -\frac{\partial}{\partial x}\frac{1}{\sigma}\frac{\partial p}{\partial x} \\ \frac{\partial^2 v}{\partial t^2} &= -\frac{1}{\sigma_0} p'(0) = -\frac{1}{M}(\mu v + Ap(0) - F), \end{aligned} \tag{17}$$

the first being the result of volume and horizontal momentum conservation in the 1D approximation (as in (3)), and the second being (15). The analysis of Mammano and Nobili (1993) leads to a similar formulation.

In reality the evolution equations are more complicated. One has to allow 3D fluid flow (as Mammano & Nobili (1993) point out, that is in principle straightforward, though in practice requires numerical computation if one wants to include the real geometry of the cochlea), add damping effects and OHC forces, and one has to determine the radiation of sound by the eardrum and corresponding reaction force to include in F, but let us stick with system (13,14,16,17) for this section.

A first issue about the system is its stability. It is a question of the spectrum of the operator taking (a, v) to the right hand side of (17) being entirely negative. Thus, stability is decided not purely by the basilar membrane dynamics: interaction with the stapes plays a role. Since there is a

critical layer for every frequency of interest, we should expect the spectrum to consist mainly of a continuum (in contrast to acoustic waves in brass and woodwind instruments, for example, where the non-zero wave speed and finite length of the instrument makes the spectrum discrete), but there should also be at least one eigenvalue, corresponding to the discrete variable v. It would be interesting to estimate the corresponding frequency.

A second issue is its impulse response. This does not look easy to determine, but should presumably separate into a wave train with locally defined (in space and time) frequency and wave number, and energy flow given by the group velocity. In WKB approximation, a group of waves with given frequency ω moves with the group velocity $c_g = \frac{\partial \omega}{\partial k}$. Thus for the 1D critical layer model one obtains $c_g = \frac{(\lambda - m\omega^2)^{3/2}}{\lambda \sigma^{1/2}}$, which slows to zero at the resonant location. It slows even more when 2D or 3D effects are included and $kh \gg 1$, to $c_g = \frac{(\lambda - m\omega^2)^2 w}{4\rho\lambda\omega}$. This makes the time from the stapes for a group to reach a given x near the critical layer be $t = \frac{2\lambda}{\omega^3 \sigma \beta^2 \sqrt{-X}} + cst$ for the 1D model provided $X < -1$, and $\frac{4\lambda\rho}{\omega^3 \sigma \beta^3 w |X|} + cst$ with 2D or 3D corrections, provided $-(\frac{h}{\beta})^2 < X < 0$.

3 Inadequacies of critical layer models

Critical layer models of the cochlea, however, have several inadequacies.

Firstly, even after modification for 2D and 3D effects, adding damping processes like $-\delta \frac{\partial a}{\partial t}$ to (4) and $-\gamma j$ to (2), and many models for active processes (OHC feedback), it seems agreed that the peak in the response to periodic forcing does not come out the right shape if based on critical layer resonance (de Boer, 1996): "either the amplitude of the peak remains too low or the phase variations in the peak region are too fast".

Although this has been put right in a class of abstract models allowing a general basilar membrane impedance function (de Boer & Nuttall, 2000), it seems to me that realistic electro-mechanical assumptions would place strong constraints on the joint frequency and position dependence and it is not clear to me that they are satisfied by the basilar membrane impedance functions fit from data.

Returning to critical layer models, let us discuss particularly the phase variation. On a critical layer model, the wavelength should go to zero at the critical layer. Yet the wavelength (inferred as $2\pi/\phi_x$, where ϕ_x is the rate of change of phase with position) of the travelling wave along the basilar membrane is observed not to go to 0; indeed it seems to go to a minimum

of about 0.5–1mm at the resonance (Table 4 of Robles & Ruggero, 2001), with the shorter minimum for the higher frequencies. Now the 1D models can be allowed to escape this criticism because even though the slope of the phase does go to infinity, it does so in such a way that the phase change up to the critical layer is finite (Fig. 2), so any smoothing effects (damping, nonlinearity, other forces, experimental error) will saturate the slope. Note also that if instead one considers the wavelength to be twice the distance between nodes then the 1D case escapes even without adding extra effects because the WKB approximation is not valid for $|x(\omega)-x| \leq \beta$ so the image of slowly modulated waves is not correct there, and it can be inferred from the analysis behind Fig. 1 that there is a $c > 0$ (of order 1) such that there is at most one node between $x(\omega) - c\beta$ and $x(\omega)$. An order of magnitude estimate for β, however, for a model with exponential stiffness variation (de Boer & MacKay, 1980) is 0.075mm (independent of frequency), which is much shorter than the observed minimum wavelength.

Models with 2D or 3D fluid flow do not escape this criticism, because they predict that the phase goes to infinity on approaching the critical layer, which disagrees strongly with observations. One could argue that the observations are masked by an additional component after the resonant location ("cochlear fast wave"), due to the differences between the oval and round windows, so that the infinite phase change gets truncated to some large value. Indeed, the phase beyond the critical layer is always observed to adopt a specific relation modulo 2π with that of the stapes, so perhaps this is a valid escape. The observed sharpness of tuning at low amplitudes, however, suggests that damping is almost cancelled and then the phase should go to infinity.

Similarly, one can compare with experiments observing fixed place at varying frequency. The accumulated phase (usually measured as the phase difference between response at a fixed place and forcing at the stapes, on increasing the frequency ω from 0), does not go to infinity as ω approaches the critical frequency for the given place; instead it plateaus at around 3–4 cycles, e.g. Fig. 7 of Robles & Ruggero (2001) and Fig. 5 of Shera & Guinan (1999). This is consistent with a 1D critical layer but not a 2D or 3D one.

Secondly, a common objection to such critical layer models (e.g. Naidu & Mountain, 1998) is that the observed variation in stiffness (by a factor of 10^4, e.g. Olson & Mountain, 1991) from one end of the membrane to the other is inadequate to account for the observed range of frequencies that we can hear (10^3). According to the models, the natural frequency at a given place is $\sqrt{\frac{\lambda}{m}}$, which would change by a factor of only 10^2 from

one end to the other (as already mentioned, the effective membrane density m is usually assumed not to vary much). Some authors try to get round this by proposing alternative resonances below the minimum on the basilar membrane, to extend the range, for example using the helicotrema (Lighthill, 1981), but the effects look weak.

Thirdly, nearly all the models ignore longitudinal stiffness of the membrane, which, although apparently very small (Voldrich, 1978), can not logically be neglected in models which predict k^2 to go through infinity at some point and the basilar membrane displacement to suffer an infinite discontinuity there! Addition of longitudinal stiffness would add a term sk^6 to the dispersion relation, for some stiffness coefficient s, and instead of going through infinity as x increases, $k^2(x)$ would rise steeply and then plateau at $\sqrt{m\omega^2/s}$. Thus one would obtain an enhanced amplitude as the wave approaches what was the critical layer but the amplitude would remain high thereafter, unless damping was stronger for such short wavelength modes.

Critical layer models in any domain of science suffer from a generalisation of this problem: the phenomenon of critical layer absorption is not robust to addition of many other physical effects, however small (Stix, 1992; Swanson, 1998). To some extent this situation is mitigated by damping: then addition of further physical effects weaker than damping does not change the results much. In the (live and oxygenated) cochlea, however, damping can be considered to be very small because of the active feedback provided by outer hair cells, which among other things is likely to nearly cancel (or exceed) damping.

Fourthly, the impulse response is often observed to have a double-lobed structure, e.g. Fig. 9 of Robles & Ruggero (2001), and I don't think critical layer models can explain this. On a critical layer model, an impulse at the stapes should produce a response at a given location starting with low frequencies and increasing to the resonant frequency for that location and then ringing at that frequency for a number of cycles depending on the damping. The double-lobed response suggests that a second packet of waves arrives later and interferes with the first. From where does it come?

Fifthly, critical layer models have trouble explaining otoacoustic emissions (OAE). These are sounds that come out of ears. There are various sorts (Probst et al, 1991):

spontaneous : occurs in about one third of ears, at one or more well defined frequencies (e.g. Fig. 5.6 of de Boer (1991) and figures in McFadden & Plattsmier, 1984; Martin et al, 1988; a case with 23 has been reported (Probst et al, 1991). Note that although sometimes

correlated with tinnitus, a subjectively audible high frequency whistle, opinion seems to be that tinnitus is a neurological rather than mechanical phenomenon (Probst et al, 1991).

transiently evoked : response to a click or short tone-burst; it usually has a fairly well defined frequency and delay time (around 10 ms) before starting. The reflected energy can reach 100% of the incident energy at low stimulus amplitudes, e.g. Figs 5 and 8 of Wilson, 1983. This was the first form of OAE to be discovered (Kemp, 1978).

tone-evoked : response at the same frequency; interference with the incident waves causes modulation in the frequency response of the ear canal with a period of around 100 Hz, which approaches near complete cancellation at low amplitudes, e.g. Fig. 1 of Shera & Guinan, 1999.

evoked distortion-products : response at a combination frequency to input at two frequencies (as a clearly nonlinear effect, this will not be considered here).

Spontaneous OAEs (SOAEs) and the near full energy reflection of transiently and tone-evoked OAEs were the first evidence for an active component to the mechanics of the cochlea. Active processes are not a problem for critical layer models, but the first three classes of OAEs are incompatible with critical layer models in several ways. Critical layer models can produce one eigenvalue, but SOAEs in some subjects occur at more than one frequency. On linear theory, all those places with negative damping would produce SOAEs corresponding to their resonant frequency, and nonlinear effects could lock them to a discrete set of frequencies (or produce chaotic output), but it seems to me the set of frequencies would be highly sensitive to physiological conditions, in contrast to the observed stability (Martin et al, 1988). It is usual to try to explain transiently and tone-evoked OAEs in terms of errors in the WKB approximation, arising from non-smooth spatial variations of the medium, or nonlinear effects producing some equivalent. Shera & Zweig (1993) (also Zweig & Shera, 1995) made a particularly thorough attempt to make this approach work. Such explanations encounter a problem, however: the delay time for transiently evoked OAEs is around 10–14 periods, whereas the phase shift from stapes to resonance is of the order of 3, so if the echo were a reflection in the same mode before the critical layer it would have a time delay of at most 6 periods (de Boer, 1991).

Sixthly, if the active processes over-compensate for damping at the resonant location (as is proposed in a zone just before the resonant location),

then the connection formula for critical layer resonance would change, so that an incoming wave produces a response growing rapidly towards the apex and the solution decaying towards the apex matches to an outgoing wave from the critical layer. One might say this is SOAEs, but then the cochlea should be unable to match in-going waves at unstable frequencies to bounded displacement at the apex. In any case, using a model with such extreme sensitivity to the balance between damping and anti-damping feels risky. It is a good general principle that models should be robust to modifications unless there are strong constraints on their form.

In contrast, mode conversion is robust to modelling errors, and I will indicate that it can work with the observed range of stiffness, can explain multiple SOAEs, tone-evoked OAEs, and with some tweaking of models might give the right tuning curves.

4 Mode conversion

The context for mode conversion is a wave-bearing medium which varies smoothly in space in such a way that the square k^2 of the local wavenumber is a multivalued function of position x, with a fold point at a positive value of k^2 where two real values of k merge and split as two complex solutions (Fig. 4).

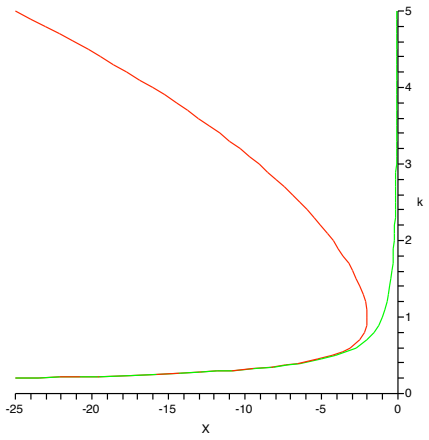

Figure 4: Wavenumber k as a function of position x for given frequency for a system exhibiting mode conversion, contrasted with a case of critical layer resonance.

It is crucial to distinguish this from the well known case where k^2 decreases through 0, which is usually called a "cutoff" (because waves can not propagate beyond it) or "turning point" (because it gives a fold in the graph of $k(x)$) and plays a fundamental role in semiclassical mechanics, radio propagation and many other domains of science. To avoid confusion, I will always refer to the place where the graph of $k^2(x)$ folds as a "fold point", although both "turning point" and "cutoff" would a priori have been equally good descriptions.

The phenomenon is that a wave entering in one mode slows down to zero group velocity at the fold point and then turns into the other mode and propagates back out the way it came. The intensity builds up in inverse proportion to the group speed. This phenomenon was found and analysed by Stix (1965) in a plasma waves context (warning: Stix called the fold a "critical layer", which again is reasonable terminology but was subsequently used by Booker & Bretherton (1967) for the case of $k^2(x)$ passing through 0 and the latter usage has dominated). In magnetised plasmas (ionised gases) it occurs near lower and upper hybrid resonances, the perpendicular ion-cyclotron resonance and the Buchsbaum two-ion resonance (Stix, 1992). I think it should also occur in many other contexts, e.g. ultrasound in elastic plates with thickness gradients, cf. Fig. 1 of Prada et al (2005).

Stix considered equations of the form

$$\eta p^{(4)} - Xp'' + p = 0, \tag{18}$$

representing the leading terms in a description of a lossless medium near a point ($X = 0$) where the coefficient of the second derivative changes sign. Actually, he used a parameter μ and spatial coordinate u, related to ours by $\eta = \mu^3$ and $X = -\mu u$, but the above form is more convenient for present purposes. The case of small positive η was studied earlier by Wasow (1950) in connection with Orr-Sommerfeld theory for linear stability of parallel shear flows (though application to that problem requires taking η imaginary).

Take $\eta > 0$ (the case $\eta < 0$ has no fold and corresponds qualitatively to the case with longitudinal stiffness). The local dispersion relation is

$$\eta k^4 + Xk^2 + 1 = 0,$$

so it has local dispersion curve

$$k^2 = \frac{-X \pm \sqrt{X^2 - 4\eta}}{2\eta},$$

equivalently

$$X = -(\eta k^2 + k^{-2}), \tag{19}$$

with a fold at $X = -2\eta^{1/2}$, $k = \eta^{-1/4}$. Stix considered waves in a warm magnetised plasma, but (18) can be viewed as a (singular) perturbation of the affine impedance cochlear model (11).

The general solution of (18) can be written as a linear combination of solutions of the form

$$p(X) = \int_\Gamma dz \exp\left(\frac{\eta}{3z^3} - \frac{X}{z} - z\right), \tag{20}$$

for various choices of contour Γ in the complex z-plane (some are shown in Stix (1992) in the plane of $u = \mu z$). Asymptotic analysis of these solutions showed that the solutions going to 0 on the far left do so like

$$\mathrm{i}\sqrt{\frac{2X}{\pi}}K_1(2\sqrt{X}),$$

and connect to linear combinations of solutions asymptotic to

$$-\sqrt{\frac{-\pi X}{2}}H_1^-(2\sqrt{-X}) + \frac{\eta^{3/4}}{\sqrt{2}(-X)^{5/4}}e^{\mathrm{i}\left(\frac{2}{3}\eta^{-1/2}(-X)^{3/2} - \frac{\pi}{4}\right)}$$

and its complex conjugate on the far right (a subdominant term on the left determines the linear combination). K_1 and H_1^- are Bessel functions: K_1 decays to 0 as its argument goes to infinity, H_1^- oscillates with clockwise-rotating phase.

The interpretation is that an incoming wave from large negative X in a mode where $\eta p^{(4)}$ is negligible (the term involving H_1^-; this is the usual mode approaching a critical layer) produces an evanescent wave for $X > 0$ (the term involving K_1; just as for a critical layer) and an outgoing wave for $X < 0$ in a mode where p is negligible (resulting from a balance between $\eta p^{(4)}$ and Xp''; this would be a new mode in the cochlear context).

Note that the problem depends on the parameter η which can not be removed by scaling, so unlike critical layer solutions, there is not a universal form for mode conversion solutions.

On the $k' \ll k^2$ criterion the WKB approximation is good for $|2\eta k^3 - \frac{2}{k}| \gg 1$. This corresponds to avoiding sufficiently a neighbourhood $[k_1, k_2]$ of the wavenumber at the fold, with $k_{1,2} \sim \eta^{-1/4} \mp \frac{1}{8}\eta^{-1/2}$ for $\eta \gg \frac{1}{4096}$, $k_1 \sim 2, k_2 \sim (2\eta)^{-1/3}$ for $\eta \ll \frac{1}{4096}$, equivalently, to X being to the left of X_1 on the lower branch, X_2 on the upper branch, with $X_{1,2} \sim -2\sqrt{\eta} - \frac{1}{16}$ for $\eta \gg \frac{1}{4096}$, $X_1 \sim -\frac{1}{4}, X_2 \sim -(\frac{\eta}{4})^{1/3}$ for $\eta \ll \frac{1}{4096}$. Since the crossover occurs at a small value of η, it is valid to use the large η formulae even for η of order 1.

5 Correspondence with the cochlea

Let us assume that there is some physical effect in the cochlea which produces a perturbation of the form $\eta p^{(4)}$ to (11), and analyse to what predictions it would lead for the cochlea.

Initially let us use (3) to convert p to a, though 2D and 3D effects are addressed later. Some solutions $a(X)$ were computed numerically by Kevin Painter for his undergraduate Applied Mathematics project with me in 1994. To select solutions which decay to the right one has to shoot from small final conditions on the right or treat it as a two-point boundary value problem giving at least two (small or zero) values at the right. Figure 5 shows one I produced recently: one can see two wave modes on the left – the underlying long one corresponds to the critical layer resonance and the superimposed short one is the new mode.

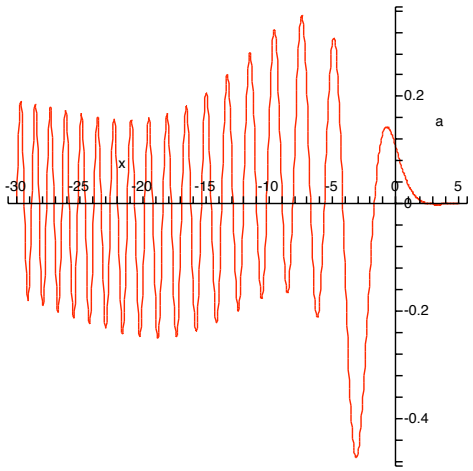

Figure 5: $a(X) \propto p''(X)$ for a solution of (18) for $\eta = 1$.

One might say that the large amplitude short wave mode is inconsistent with observations of the cochlea, but perhaps one would need greater than usual spatial resolution to see it; probably most current measurement procedures would average away the short wave leaving only the long one. Furthermore, I would suggest that except at low power the short mode is damped relatively fast, so decays to the left for a solution with input power

coming from the left in the long mode. Also, the solutions depend on the parameter η whose value I have not yet estimated (see Section 7 for a formula, but it requires knowledge of OHC responses), but numerics did not reveal a great dependence on η other than setting the wavelengths of the two modes.

To start some analysis of the shape of the solutions, note that (18) has a conserved quantity, "power" or "wave energy flux":

$$\Phi = \frac{1}{2\sigma\omega}\Im(\eta p''\bar{p}^{(3)} - p\bar{p}')$$

(the factor $\frac{1}{2\sigma\omega}$ is included to correspond with the physically derived (12) in the case $\eta = 0$). This can be checked by differentiation or derived as a consequence of the following Hamiltonian formulation for (18). Equation (18) defines a non-autonomous dynamical system on the space of $(p, p', p'', p^{(3)}) \in \mathbb{R}^4$. For a solution p let $H = \frac{1}{2}(\eta {p^{(3)}}^2 - Xp''^2 + 2pp'' - p'^2)$ and for two solutions p, q let $\omega(p, q) = \eta(p''q^{(3)} - q''p^{(3)}) - (pq' - qp')$. Then ω is a symplectic form on $\mathbb{R}^4$ and the Hamiltonian vector field of H with respect to ω (i.e. solution V of $\omega(V, \xi) = dH(\xi)\ \forall \xi \in \mathbb{R}^4$) is equivalent to (18). Conservation of ω is automatic; this is the Hamiltonian way to view conservation of Wronskians. The formulation can be extended to $\mathbb{C}^4$ by letting $H = \frac{1}{2}(\eta|p^{(3)}|^2 - X|p''|^2 + 2\Re(p\bar{p}'') - |p'|^2)$ and $\omega(p, q) = \Im(\eta p''\bar{q}^{(3)} - p\bar{q}')$. Then phase rotation invariance implies the associated Noether conserved quantity $P = \Im(\eta p''\bar{p}^{(3)} - p\bar{p}')$, solving $\omega(W, \xi) = dP(\xi)\ \forall \xi \in \mathbb{C}^4$, where W is the vector field for infinitesimal phase rotation. Up to multiplication by a physical factor, this P is the "power". It is not necessary for H to be conserved (nor even possible, because $\frac{dH(p)}{dX} = \frac{\partial H}{\partial X} \neq 0$). Note that all this remains true even if the coefficient of p'' in (18) is replaced by any real function of X. Indeed, such a Hamiltonian formulation is the mathematical expression for the physical concept of a "lossless medium" at the linear time-invariant level.

Now let us see what the WKB approximation tells us about the solutions, using conservation of power. As soon as $\eta \gg \frac{1}{4096}$ the domain of validity of the WKB approximation is all but an interval of length about $\frac{1}{16}$ in X to the left of the fold point (and even for smaller η it is all but an interval of about $\frac{1}{4}$ in X). For solutions bounded on the right, hence going to zero, $\Phi = 0$, so it follows that the incoming and mode-converted waves carry equal and opposite powers, which abusing notation I'll also write as Φ. Then in the WKB regime for each mode,

$$|p|^2 \quad = \quad \frac{2\sigma\omega\Phi}{(\eta k^4 + 1)k}, \tag{21}$$

$$|a|^2 \;=\; \frac{2k^3\Phi}{\sigma\omega^3(\eta k^4+1)} = -\frac{2k\Phi}{\sigma\omega^3 X},$$

using (3) to convert p to a (one could redo this to allow for 2D and 3D fluid flow effects, by taking $\eta k^2 = -X - 1/|k|$ for $kh \gg 1$). Using (19), note that the resulting amplitude as a function of position depends only on $X/\eta^{1/2}$ (up to scale in a which is arbitrary). Figure 6 shows the amplitude as a function of $X/\eta^{1/2}$ for the two waves (connected through the fold even though the WKB approximation is not formally valid there).

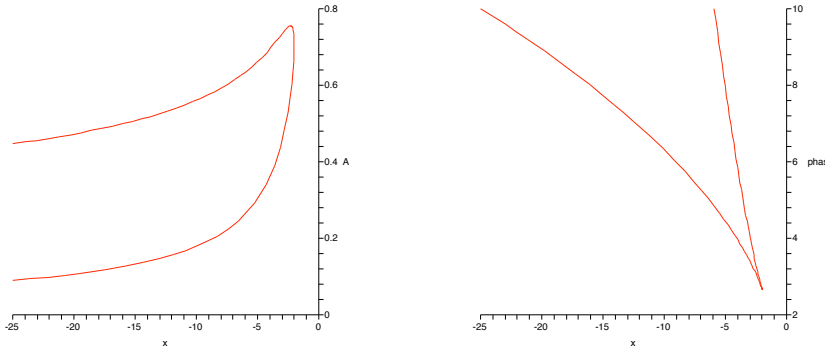

Figure 6: Amplitude $|a|$ and scaled phase $\phi/\eta^{1/4}$ of oscillation as functions of $x = X/\eta^{1/2}$ for the WKB approximation of (18). The long-wave mode has the lower amplitude and shallower phase curve.

The phase of the waves from WKB theory is $\phi = -\int k\, dX$. Using (19) and integration by parts one obtains $\phi = 2(\frac{\eta k^3}{3} + \frac{1}{k})$. Thus in terms of $\kappa = \eta^{1/4}k$ we have

$$X = -\eta^{1/2}(\kappa^2 + \kappa^{-2}),\;\; \phi = 2\eta^{1/4}(\frac{\kappa^3}{3} + \frac{1}{\kappa}).$$

This is also plotted in Fig. 6. For the phase of the true problem, there is a non-WKB correction and a non-trivial approach to a constant on the right, which I have not computed.

Since the maximum amplitude in mode conversion occurs near the fold rather than at the resonance, it is feasible for the range of mode-converting frequencies to be broader than that for critical layer resonances. In particular, the frequency range extends considerably downwards because for low frequencies one can have a fold on the basilar membrane at a frequency whose resonant location would be off the apical end.

With addition of suitable anti-damping to the incoming wave, damping to the outgoing one and 3D effects, I think it is likely that a fit with observed tuning curves could be obtained.

Mode conversion could explain transiently and tone-evoked OAEs, because it gives complete conversion of the incoming wave into an outgoing one, and the properties of the second mode can be widely different from the first, e.g. much slower group velocity, so it could fit the long time delay for the echo and the accumulated phase curves.

As a first step, I propose a specific form for the additional term, namely a force $\frac{\partial^2}{\partial t^2}\frac{\partial}{\partial x}\nu\frac{\partial a}{\partial x}$ with $\nu > 0$, to be justified physiologically in section 7. Let us compute the phase shift and group delays for a case of this with exponential stiffness variation but other properties constant. 3D fluid flow effects are allowed, but success or failure of this particular model should not be taken as definitive for the whole class of mode conversion models. Thus I take $\lambda(x) = Ce^{-\alpha x}$ and m, ν and the function I of (8) independent of x. So the dispersion relation is

$$\nu k^2 - (\frac{C}{\omega^2}e^{-\alpha x} - m) + I(k) = 0.$$

This inherits the simplifying feature from the exponential critical layer model that a change in ω is equivalent to a shift in x. It follows that

$$e^{-\alpha x} = \frac{\omega^2}{C}f(k) \tag{22}$$

where $f(k) = \nu k^2 + m + I(k)$.

First, compute the phase shift from the base to the fold in the long wave mode (using the WKB approximation):

$$\Delta\phi^f = -\int_0^{x_f(\omega)} k\ dx,$$

where the position $x_f(\omega)$ of the fold is given by $e^{-\alpha x_f(\omega)} = \frac{\omega^2 M}{C}$ with $M = \min_k f(k)$. Using integration by parts, and defining $k_-(\omega) < k_+(\omega)$ to be the positive roots of $f(k) = \frac{C}{\omega^2}$ for $\omega < \sqrt{C/M}$ (the maximum frequency for which the fold falls in the domain $x \geq 0$) and k_f to be the minimiser of f (note that this is independent of ω, whereas experiment shows the wavenumber at the characteristic place to increase weakly with frequency, but this could be fixed by allowing ν, m or I to vary with x, at the cost of more complicated calculations), one obtains

$$\Delta\phi^f = -k_f x_f(\omega) + \int_{k_-}^{k_f} x\ dk = -\frac{1}{\alpha}\left(k_- \log\frac{C}{M\omega^2} + \int_{k_-}^{k_f}\log\frac{f(k)}{M}\ dk\right).$$

This gives a phase delay of magnitude $1/\alpha$ times the area shown in Figure 7. As $\omega \to 0$ the area converges to a finite non-zero limit $L = \int_0^{k_f} \log \frac{f(k)}{M}\ dk$.

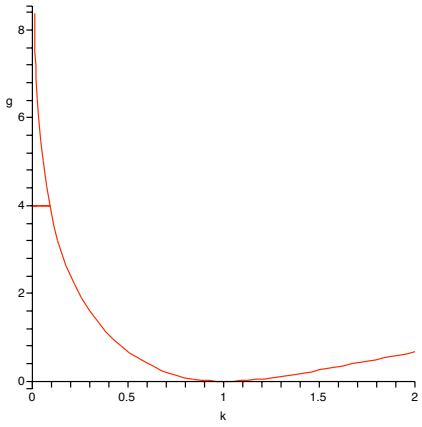

Figure 7: The area entering the calculation of the phase delay from stapes to fold for the exponential mode conversion model. The plotted function $g(k) = \log \frac{f(k)}{M}$, the horizontal cutoff is at height $\log \frac{C}{M\omega^2}$, and the integral extends to the minimum. The plot was made using $\nu = 1$ and $I(k) = \frac{1}{k \tanh k}$.

Thus for ω sufficiently below $\sqrt{C/M}$ the phase shift is close to $-L/\alpha$. This corresponds with observations of a phase delay from stapes to characteristic place (of about 23 radians) independent of frequency. This model predicts a reduction in the magnitude of the phase delay as frequency increases, at a rate $\tau_p^f = -\frac{\partial \Delta\phi^f}{\partial \omega} = \frac{2k_-(\omega)}{\alpha\omega} \approx \frac{2}{\alpha}\sqrt{\frac{\sigma}{C}}$ for low frequencies, with the phase delay going to 0 as ω approaches the maximum $\sqrt{C/M}$. It would be interesting to know whether such a change is observed.

Next, let us compute the group delay τ_g^f from stapes to fold in the long-wave mode. Disturbances near frequency ω propagate at the group velocity $c_g = \frac{\partial \omega}{\partial k}$. Thus the time delay for a disturbance from stapes to fold is

$$\tau_g^f = \int_0^{x_f(\omega)} \frac{dx}{c_g} = \int_{k_-}^{k_f} \frac{\partial x}{\partial k} \Big/ \frac{\partial \omega}{\partial k}\ dk,$$

the derivative in the numerator being performed at constant ω and that in the denominator at constant x. This evaluates to

$$\tau_g^f = \frac{2}{\alpha\omega}(k_f - k_-(\omega)) \sim \frac{2k_f}{\alpha\omega}$$

for ω sufficiently below $\sqrt{C/M}$. Thus the model predicts a group delay from stapes to characteristic place inversely proportional to frequency. I don't know if this has been measured.

Thirdly, let us compute the phase shift for a wave entering in the long wave mode and coming out in the short one (or equivalently, the other way round). The WKB phase shift (perhaps one should add something for the non-WKB region near the fold) is, integrating by parts,

$$\Delta\phi^0 = -\int k\ dx = \int_{k_-}^{k_+} x\ dk = -\frac{1}{\alpha}\int_{k_-}^{k_+} \log\frac{\omega^2 f(k)}{C}\ dk.$$

This is a phase advance, because the argument of the logarithm is less than 1 in the given range. I am not aware of a direct measurement of this phase change, but what has been measured is its derivative with respect to frequency, inferred from the 100 Hz modulation of the frequency response. This derivative is

$$\tau_p^0 = \frac{\partial\Delta\phi^0}{\partial\omega} = -\frac{2}{\alpha\omega}(k_+(\omega) - k_-(\omega)).$$

Note first that it is negative, so looks like a time delay. This agrees with the sign observed (Kemp, private communication). Secondly, for ω sufficiently below $\sqrt{C/M}$ we have $k_+(\omega) \sim \sqrt{\frac{C}{\nu}}\frac{1}{\omega}$ and k_- much smaller, so $\tau_p^0 \sim -\frac{2}{\alpha}\sqrt{\frac{C}{\nu}}\frac{1}{\omega^2}$; it decreases in magnitude as frequency increases but too fast compared with observations, which give $\omega\tau_p^0$ approximately constant at about $15 \times 2\pi$ (Wilson, 1980). So perhaps some modifications of the model are necessary, e.g. ν could vary exponentially with x too (it is thought to be larger at the base than the apex).

Fourthly, the group delay for the round trip from stapes to stapes is:

$$\tau_g^0 = \int \frac{dx}{c_g} = \int_{k_-}^{k_+} \frac{\partial x}{\partial k}\Big/\frac{\partial\omega}{\partial k}\ dk = \frac{2}{\alpha\omega}(k_+(\omega) - k_-(\omega)),$$

which, remarkably, is the same as $-\tau_p^0$. The equality of τ_g^0 and $-\tau_p^0$ agrees with observations (Wilson, 1980), but again the frequency dependence is incorrect.

In principle, one could work out a formulation in the time domain for a mode conversion system, analogous to that of section 2.3, but we can already deduce further qualitative features without this.

For example, consider generation of SOAEs. In contrast to critical layer models, mode conversion models have a discrete set of resonant frequencies, much as in a wind instrument but using reflection between the two modes

at the fold and a combination of reflections at the base. Of these, it could be that some are unstable or at least so close to unstable that they amplify noise and hence give rise to an array of SOAEs, cf. Fig. 1 of Martin et al (1988). Fig. 9 of Wilson (1983) suggests a situation close to a subcritical Hopf bifurcation for one of the modes.

Finally, consider the impulse response for a mode conversion model. An impulse at the stapes will generically produce waves in both modes, of which the short-wave mode is slower. Thus at a given location one should expect to see first a fast wave arrive, with frequency increasing to the value for the fold at that location and then decreasing as we get to see waves turning at more apical folds and coming back in the other mode, followed by a slow wave, with frequency also increasing to the value for the fold at that location and then decreasing. This could explain the two-lobed waveform of Recio et al (1998). If one waits longer one might also see the effects of waves reflecting at the stapes and coming round again.

To obtain realistic predictions of the shape of the tuning curves, impulse response and OAEs one would have to make realistic models, including realistic variation of the parameters along the length, active feedback and all the relevant transfer functions, and probably nonlinear effects too (which could have a drastic effect on peak shape!).

6 Precedents

The idea of mode conversion has already been proposed in cochlear mechanics, by Huxley (1969), albeit without the terminology or results. He pointed out that it was inconsistent to leave out the effects of longitudinal stiffness of the membrane and suggested that to obtain the observed frequency to place response additional effects should be included to make a path for waves from the incoming mode to an exponentially decaying one. He suggested this could be achieved by longitudinal compression of the membrane (with which I agree, though whether it is physical is another matter) or an effect of the spiral geometry of the cochlea (which I haven't managed to understand). Longitudinal stiffness would give rise to an irrelevant third mode with much shorter wavelength.

The idea of studying models with more than one mode has been proposed several other times (see section 7.1 of de Boer, 1996). In addition to the references given there, Kolston (1988) proposed an "OHCAP" model, which has two modes of deformation of the basilar membrane, de Boer (1990) proposed a two-membrane model (basilar membrane and reticular lamina),

and Hubbard (1993) proposed a two-mode model. Also Wilson and Bruns (1983) observed two modes of deformation of the basilar membrane in a bat. These and others are surveyed in section 4.3 of Hubbard & Mountain (1996). None of the authors appears to have commented on the possibility of mode conversion, however. Interestingly, de Boer's model does exhibit mode conversion, namely where $hZ_{BM} = 4h_3(Z_{OC} - Z_{HC})$, though this might not occur in the physiological range.

Note that the "second mode" of Watts (1993) is not a candidate, because it is just the evanescent wave on the apical side of a critical layer. Nor is the "fast wave" of Lighthill (1981) a candidate, because it has a much longer wavelength, this being an acoustic wave in the fluid supported by its slight compressibility.

7 Physiological origin of mode conversion?

What the mode conversion explanation requires before anything else is a plausible physical reason for a fold in the wavenumber as a function of position. I am not convinced by Huxley's suggestions. It is not clear to me that any of the two-mode models have mode conversion in the physiological range, and if they produce critical layers then they will fail for the same reasons as already discussed.

Here is one conjectured mechanism to create a fold. The outer hair cells (OHCs) exert a force (Brownell et al, 1985) on the membrane. What forces the OHCs produce is a matter of continuing debate (see section 7.2 of de Boer (1996) for a survey up to that date), but since they are contained within the organ of Corti they can not produce a net force nor a net longitudinal torque on the basilar membrane, so it seems to me than in the continuum approximation the force must be a second derivative with respect to x (or a change in transverse mode shape). This could be a natural result of evolution of the hair cells sharpening the response, since such a force acts as amplification of transverse displacement with inhibition. There are three rows of OHCs and they are slanted differently along the membrane so could easily produce a second derivative response (indeed, a first derivative response was already proposed by Kolston et al (1989), Steele et al (1993)). In any case, a second derivative in x of the given sign creates a fold in $k^2(x)$, analogous to Huxley's suggestion of longitudinal compression.

Let us suppose the OHC force has value

$$F = -\omega^2 \frac{\partial}{\partial x} \nu \frac{\partial a}{\partial x}, \tag{23}$$

for some positive function $\nu(x)$ of position, allowing for the observed variation in properties of the OHCs along the membrane. The factor ω^2 is required to keep the membrane stable (stability is analysed in the next section). It is also reasonable because Mammano & Nobili (1993) claim that despite the OHCs being low-pass filters, the "inertial reaction of the tectorial membrane makes the triggering mechanism of outer hair cells increase as the square of frequency over a wide range". This picture might also fit with those who see the OHCs as having a natural frequency, as for frogs and turtles (Duke & Julicher, 2003): balance between $\omega^2 \frac{\partial}{\partial x} \nu \frac{\partial a}{\partial x}$ and $(\lambda - m\omega^2)a$ could give oscillations at frequency $\omega_0 = \sqrt{\frac{\lambda}{\nu k_0^2 + m}}$ with k_0 corresponding to the minimum possible wavelength of two segments of the basilar membrane (20 μm), to achieve zero group speed.

Note that the proposed outer hair cell force (23) is reactive rather than resistive. The possibility of a reactive component was anticipated in Gold (1948). Of course there could be, and almost certainly is, an (anti-)resistive component too to cancel damping but it might be that the principal response of the outer hair cells is reactive (Kolston & Smoorenburg, 1990) (this would also be sensible to reduce the power requirement of the OHCs, which is a prohibitive concern for many models). The work of de Boer & Nuttall (2000) inferring the basilar membrane impedance by fitting data to a 3D critical layer model does not allow one to settle this question, because the separation of the imaginary part into active and passive parts is not possible. Also it seems strange to me that they obtain anti-damping only in the region just before the peak response at the tested frequency: I would have thought the anti-damping should be ready for any frequency of input and therefore distributed along the whole membrane; but see sections 6.1 and 8.3 of de Boer (1996) for a distinction between "undamping" and "local activity": I think the idea is that undamping is a force proportional to a velocity whereas local activity can depend on other time and space derivatives.

Assuming the OHC force law (23), one obtains dispersion relation (ignoring the now irrelevant longitudinal stiffness term $-sk^6$, but using 1D fluid flow for simplicity)

$$\nu\omega^2 k^4 - (\lambda - m\omega^2)k^2 + \sigma\omega^2 = 0.$$

If m, σ and ν are taken to be roughly constant and λ to decrease with x, then one obtains a fold as in Fig. 4 at the position x where $\lambda(x) = (m + 2\sqrt{\nu\sigma})\omega^2$, and the wavenumber at the fold is $k = (\frac{\sigma}{\nu})^{1/4}$. Thus k_f is roughly independent of ω, as observed, though there should be corrections from 3D fluid flow and an effect from variation of ν with position.

Taking the affine approximation (11), the parameter η of (18) is given by $\eta = \frac{\nu}{\sigma\beta^4}$, evaluated near the fold.

One feature of this model is that existence of the return mode depends on outer hair cell activity, so if it were inhibited (e.g. by oxygen deprivation or aspirin) then the fold would be replaced by the critical layer curve of Fig. 4 and there would be no transiently or tone-evoked OAEs, consistent with observations (Weir et al, 1988; Martin et al, 1988); but of course, oxygen deprivation would also reduce the cancellation of damping and it might just be that the OAEs become more damped so not noticeable.

One refinement that can be made is to replace the derivatives by finite differences, to reflect the fact that the basilar membrane is made up of segments of width about $10\mu m$ (Mammano & Nobili, 1993). This would saturate the wavelength of the new mode at two segment widths.

Another consideration is that in addition to active OHC length changes there is also active hair bundle motion (e.g. Kennedy et al, 2006), which might contribute to the function.

8 Local Stability

A fundamental requirement of a cochlear model is that the undisturbed state normally be stable, though spontaneous OAEs show it to be close to the threshold of instability. To obtain a first idea one can study "local stability" meaning that all properties like λ, σ, m are treated as locally constant along the membrane.

Longitudinal compression K (one of Huxley's suggestions) would lead to instability, because it would modify (4) to

$$m\frac{\partial^2 a}{\partial t^2} = -K\frac{\partial^2 a}{\partial x^2} - \lambda a - p,$$

which with (3) gives frequencies ω for real wave number k according to

$$\omega^2 = \frac{\lambda k^2 - Kk^4}{\sigma + mk^2}.$$

This is negative for $k^2 > \lambda/K$ so the membrane is unstable (to short wave buckling).

In contrast, the proposed outer hair cell force proportional to the second derivative in both time and space would give a stable result, because

$$m\frac{\partial^2 a}{\partial t^2} = \frac{\partial^2}{\partial t^2}\frac{\partial}{\partial x}\nu\frac{\partial a}{\partial x} - \lambda a - p \tag{24}$$

with (3) gives

$$\omega^2 = \frac{\lambda k^2}{\sigma + mk^2 + \nu k^4},$$

which is positive for all real k.

To treat the case of λ, σ, m, ν depending on position one ought to examine the spectrum of the generalised eigenvalue problem (eigenvalue ω^2)

$$(\frac{\lambda}{\omega^2} - m)\frac{\partial}{\partial x}\frac{1}{\sigma}\frac{\partial p}{\partial x} - \frac{\partial}{\partial x}\nu\frac{\partial^2}{\partial x^2}\frac{1}{\sigma}\frac{\partial p}{\partial x} + p = 0$$

with appropriate boundary conditions, for which the eigenfunctions can not be assumed to be of the form e^{ikx}, but the above treatment should suffice to gain a first impression. In fact, consideration of the correct boundary conditions leads to the requirement to include the volume displacement at the oval window (and its opposite at the round window) as an additional dynamical variable, as was already done for critical layer models in section 2.3.

9 Conclusion

The inadequacy of critical layer resonance models of the cochlea has been detailed and it is proposed that they be replaced by mode conversion models.

Mode conversion models can explain the double-lobed impulse response, the resonances in stimulated otoacoustic emissions, the minimum observed wavelength roughly independent of frequency, and other experimental observations.

The main prediction for experimentalists is that a short-wave mode should occur on the basilar membrane in addition to the "usual" one, depending on a balance between membrane stiffness and outer hair cell forces. This might be observable in transiently evoked oto-acoustic emission experiments, but the wavelength will be shorter than those observed so far, so it would require high spatial resolution to detect.

The most important extension for modellers to make to mode conversion models is to analyse the effects of nonlinearity.

Another issue that may need addressing is low frequencies ($\leq 200Hz$), for which it is possible that the mode conversion location would be off the apical end of the basilar membrane and hence different treatment required.

Acknowledgements

I dedicate this paper to the people who have inspired me about waves in inhomogeneous media or cochlea mechanics: Ken Budden, Michael McIntyre,

James Lighthill, Egbert de Boer, Tom Stix, Ted Evans, Andrew Huxley and Pat Wilson. I have been terribly slow about writing it (it could have been written in the early 1990s), and I'm afraid Budden, Lighthill and Stix have not lived to see this. I'm grateful to Jonathan Ashmore for suggesting it could still be useful, to David Kemp for important comments, and to IHES (France) for hospitality during the writing of a first version (MacKay, 2006).

References

[1] Boer E de, 1990, Wave-propagation modes and boundary conditions for the Ulfendahl-Flock-Khanna preparation, in Dallos et al, 1990, 333–9.

[2] Boer E de, 1991, Auditory physics. Physical principles in hearing theory III, Phys Rpts 203, 125–231.

[3] Boer E de, 1996, Mechanics of the cochlea: modeling efforts, in: The cochlea, eds Dallos P, Popper AN, Fay RR (Springer) 258–317.

[4] Boer E de, MacKay R, 1980, Reflections on reflections, J Acoust Soc Am 67, 882–890.

[5] Boer E de, Nuttall AL, 2000, The mechanical waveform of the basilar membrane II: from data to models and back, J Acoust Soc Am 107, 1487–96.

[6] Booker JR, Bretherton FP, 1967, The critical layer for internal gravity waves in a shear flow, J Fluid Mech 27, 513–539.

[7] Brownell WE, Bader CR, Bertrand D, de Ribaupierre Y, 1985, Evoked mechanical responses of isolated cochlear outer hair cells, Science 227, 194–6.

[8] Budden KG, 1955, The non-existence of a "fourth reflection condition" for radio waves in the ionosphere, in: Physics of the ionosphere (Physical Soc, London) p.320.

[9] Dallos P, Geisler CD, Matthews JW, Ruggero MA, Steele CR (eds), 1990, The mechanics and biophysics of hearing, Lect Notes Biomath 87 (Springer).

[10] Duke T, Jülicher F, 2003, Active traveling wave in the cochlea, Phys Rev Lett 90, 158101.

[11] Gold T, 1948, Hearing II: the physical basis of the action of the cochlea, Proc Roy Soc Lond B 135, 492–8.

[12] Hubbard A, 1993, A traveling-wave amplifier model of the cochlea, Science 259, 68–71.

[13] Hubbard AE, Mountain DC, 1996, Analysis and synthesis of cochlear mechanical function using models, in: Auditory Computation, eds Hawkins HL, McMullen TA, Popper AN, Fay RR (Springer) 62–120.

[14] Huxley AF, 1969, Is resonance possible in the cochlea after all? Nature 221, 935–940.

[15] Keener J, Sneyd J, 1998, Mathematical Physiology (Springer).

[16] Kemp DT, 1978, Stimulated acoustic emission from within the human auditory system, J Acoust Soc Am 64, 1386–91.

[17] Kennedy HJ, Evans MG, Crawford AC, Fettiplace R, 2006, Depolarization of cochlear outer hair cells evokes active hair bundle motion by two mechanisms, J Neurosci 26, 2757–66.

[18] Kolston PJ, 1988, Sharp mechanical tuning in a cochlear model without negative damping, J Acoust Soc Am 83, 1481–7.

[19] Kolston PJ, Smoorenburg GF, 1990, Does the cochlear amplifier produce reactive or resistive forces? in Dallos et al, 1990, 96–105.

[20] Kolston PJ, Viergever MA, Boer E de, Diependaal RJ, 1989, Realistic mechanical tuning in a micromechanical cochlear model, J Acoust Soc Am 86, 133–140.

[21] Lighthill J, 1978, Waves in Fluids (Camb U Press).

[22] Lighthill J, 1981, Energy flow in the cochlea, J Fluid Mech 106, 149–213.

[23] MacKay RS, 2006, Mode conversion in the cochlea? linear analysis, IHES preprint P/06/31,
http://www.ihes.fr/IHES/Scientifique/Preprint/prepub.php

[24] Mammano F, Nobili R, 1993, Biophysics of the cochlea: linear approximation, J Acoust Soc Am 93, 3320–32.

[25] Martin GK, Lonsbury-Martin BL, Probst R, Coats AC, 1988, Spontaneous otoacoustic emissions in a nonhuman primate I. Basic features and relations to other emissions, Hear Res 33, 49–68.

[26] McFadden D, Plattsmier HS, 1984, Aspirin abolishes spontaneous otoacoustic emissions, J Acoust Soc Am 76, 443–8.

[27] Naidu RC, Mountain DC, 1998, Measurements of the stiffness map challenge a basic tenet of cochlear theories, Hearing Res 124, 124–131.

[28] Nobili R, Mammano F, Ashmore J, 1998, How well do we understand the cochlea? Trends Neurosci 21, 159–167.

[29] Olson ES, 2001, Intracochlear pressure measurements related to cochlear tunings, J Acoust Soc Am 110, 349–367.

[30] Olson ES, Mountain DC, 1991, In vivo measurements of basilar membrane stiffness, J Acoust Soc Am 89, 1262–75.

[31] Peterson LC, Bogert BP, 1950, A dynamical theory of the cochlea, J Acoust Soc Am 22, 369–381.

[32] Prada C, Balogun O, Murray TW, 2005, Laser-based ultrasonic generation and detection of zero-group velocity Lamb waves in thin plates, Appl Phys Lett 87, 194109.

[33] Probst R, Lonsbury-Martin BL, Martin GK, 1991, A review of otoacoustic emissions, J Acoust Soc Am 89, 2027–67.

[34] Recio A, Rich NC, Narayan SS, Ruggero MA, 1998, Basilar-membrane responses to clicks at the base of the chinchilla cochlea, J Acoust Soc Am 103, 1972–89.

[35] Rhode WS, Robles L, 1974, Evidence from Mössbauer experiments for nonlinear vibration in the cochlea, J Acoust Soc Am 55, 588–596.

[36] Robles L, Rhode WS, Geisler CD, 1976, Transient response of the basilar membrane measured in squirrel monkeys using the Mössbauer effect, J Acoust Soc Am 59, 926–939.

[37] Robles L, Ruggero MA, 2001, Mechanics of the mammalian cochlea, Physiol Rev 81, 1305–52.

[38] Shera CA, Guinan JJ, 1999, Evoked otoacoustic emissions arise by two fundamentally different mechanisms: a taxonomy for mammalian OAEs, J Acoust Soc Am 105, 782–798.

[39] Shera CA, Zweig G, 1993, Noninvasive measurement of the cochlear traveling-wave ratio, J Acoust Soc Am 93, 3333–52.

[40] Steele CR, Baker G, Tolomeo J, Zetes D, 1993, Electro-mechanical models of the outer hair cell, in: Biophysics of hair-cell sensory systems, eds Duifhuis H, Horst JW, van Dijk P, van Netten SM (World Sci) 207–214.

[41] Stix TH, 1962, Theory of plasma waves (McGraw-Hill).

[42] Stix TH, 1965, Radiation and absorption via mode-conversion in an inhomogeneous collision-free plasma, Phys Rev Lett 15, 878–882.

[43] Stix TH, 1992, Waves in plasmas (Am Inst Phys).

[44] Swanson DG, 1998, Theory of mode conversion and tunneling in inhomogeneous plasmas (Wiley).

[45] Voldrich L, 1978, Mechanical properties of basilar membrane, Acta Oto-Laryngol 86, 331–5.

[46] Wasow W, 1950, A study of the solutions of the differential equation $y^{(4)} + \lambda^2(xy'' + y) = 0$ for large values of λ, Ann Math 52, 350–361.

[47] Watts L, 1993, Cochlear mechanics: analysis and analog VLSI, PhD thesis, CalTech.

[48] Whitham GB, 1974, Linear and nonlinear waves (Wiley).

[49] Wier CC, Pasanen EG, McFadden D, 1988, Partial dissociation of spontaneous otoacoustic emissions and distortion products during aspirin use in humans, J Acoust Soc Am 84, 230–7.

[50] Wilson JP, 1980, Evidence for a cochlear origin for acoustic re-emissions, threshold fine-structure and tonal tinnitus, Hear Res 2, 233–52.

[51] Wilson JP, 1983, Active processes in cochlear mechanics, in: Physiologie et Physiopathologie des recepteurs auditifs (Collège de France), 103–25.

[52] Wilson JP, Bruns V, 1983, Basilar membrane tuning properties in the specialised cochlea of the CF-bat, Rhinolophus ferrumequinum, Hearing Research 10, 15–35.

[53] Zweig G, Shera CA, 1995, The origin of periodicity in the spectrum of evoked otoacoustic emissions, J Acoust Soc Am 98, 2018–47.