Novel insights into cochlear information processing

Ruedi Stoop, Karlis Kanders, Leonardo Novelli, Florian Gomez [†] [‡]

[†] Institute of Neuroinformatics and [‡] Institute of Computational Science, University and ETH Zürich
Irchel Campus, Winterthurerstr. 190, 8057 Zürich, Switzerland
Email: ruedi@ini.phys.ethz.ch

Abstract—Already Helmholtz profoundly addressed the question how the nonlinearity of the human hearing sensor, the cochlea, might shape human sound perception. At his time, research was, however, obstructed by the lack of experimental data regarding the amplification properties of the inner ear. In the meantime, accurate measuring methods have permitted the comparison of models of the hearing sensor with empirical data, leading to a strong revival of the interest into Helmholtz's original research questions. In our paper, we describe some recent theoretical and modeling advances in the understanding of the nature of human pitch perception. We reveal a number of to date unexplained human auditory percept effects to be direct consequences of the nonlinear properties of the mammalian hearing sensor. Our insights also demonstrate, as a by-note, the limitations of the present reverse engineering approach towards cochlear implants.

1. Introduction

Audition has a long history, starting with the ancient Greeks who stipulated that inside the ear there should be air, as matter only communicates with matter of the same type. Only in the 1800s was it discovered by the physiologist Cotugno that in fact, the inner ear, the cochlea, is filled with perilymph comparable to that of plasma and cerebrospinal fluid, a salty fluid and endolymph that resembles more intracellular fluid [1]. A similar fate is related to the understanding of the information processing within the cochlea. While it was early realized by Helmholtz [2] that nonlinear processes play a major role, probably driven by von Bekesy's experiments and his profound understanding of the passive (i.e., essentially dead) cochlea, the picture prevailed that information processing in the cochlea would be very similar to a Fourier-analysis tool. Even after the active processes in the cochlea were postulated by Gould and finally corroborated by the detection of otoacoustic emissions by Kemp, this view basically remained, although it was widely known that nonlinear active amplification must be expected to generate strong deviations from a linear model. Combination tones ('CT') are a major effect of the nonlinearity of the physics underpinning amplification in the mammalian hearing process. They have a profound effect on different aspects of how we perceive sounds, one main manifestation is the perception of pitch. It was probably the consequence of the negation of the important role of nonlinearity that CT were also termed 'distortion products', with the erroneous assumption that CT would only have a noticeable role in hearing at very strong input amplitudes. In fact, their role is often more important at small amplitudes. Besides this more local effect (CT travel down the cochlear duct just as directly input-generated sounds do, and they are also locally amplified like the latter) a second, more global manifestation of the nonlinearity located in the cochlea's amplification properties, exists. This effect is the behavior of the mammalian hearing threshold that is known to be frequency-dependent. We will show that this is not because the amplifiers towards the ends of the accessible frequency interval are less efficient, but that this happens as the consequence of a more global behavior emerging from local nonlinearities and their couplings.

2. Combination-Tone Laws

For nonlinear systems, the superposition principle does not hold. This implies that, in general, the response to the sum of two inputs is not the sum of the individual responses to each input. In the cochlea, this manifests in a variety of effects, the most prominent one being the generation of CT within the cochlea. CT were already known to 18th century musicians Georg Sorge and Giuseppe Tartini, and can easily be heard also by non-trained listeners. CT are by-products that, however, are not filtered out on the way to the brain. Instead, they propagate along the cochlear duct, get amplified, and interact with other frequencies to create additional CT. They thus were detected at all stages of the auditory pathway, e.g., in the inner hair cell response, in the auditory nerve, or in the inferior colliculus. From this, it follows that the cochlear activation profiles often become highly nontrivial, despite the simple inputs used.

Single Hopf Oscillator: First, we will present the basic principles of combination tone generation using the example of a single Hopf oscillator. A detailed analysis however shows then that only the full Hopf cochlea model can exhibit combination tone responses comparable to the ones observed in biological and psychoacoustic experiments. The following analysis exhibits in detail how a single Hopf oscillator generates CT. Upon a stimulation with two frequencies, as a product of the nonlinearity in the Hopf equation, a well-defined set of CT is generated ('cubic CT'). Cubic CT show, as a function of their order, exponentially decaying amplitudes, a fact that has been mentioned in the

context of early psychoacoustic studies as well as more recent investigations of cochlear mechanics (see also [3]).

From the ω_{ch}-rescaled Hopf equation of a section [3, 4]

$$\frac{dz}{dt} = (\mu + i)\omega_{ch}z - \omega_{ch}|z|^2 z - \omega_{ch}F(t), \qquad (1)$$

where ω_{ch} is the characteristic frequency of the oscillator, $F(t)$ is the (complex) external forcing and μ is the bifurcation parameter, a harmonic two-tone forcing $F(t) = F_1 e^{i\omega_1 t} + F_2 e^{i\omega_2 t}$, with $\omega_1 = k\omega_0$ and $\omega_2 = (k + 1)\omega_0$ so that all CT are multiples of ω_0, generates a response that can be expanded in a Fourier series $z(t) = \sum_j a_j e^{ij\omega_0 t}$, with complex coefficients a_j (i.e. they include a phase). By inserting the expansion of $z(t)$ for a frequency $\omega_l = l\omega_0$ into Eq. (1), we obtain

$$(i(\omega_l - \omega_{ch}) - \mu\omega_{ch})a_l + c.i.t. = -\omega_{ch}F_l, \qquad (2)$$

where c.i.t. denote cubic interaction terms ($\propto \omega_{ch}a_{k'}a_{k''}a_{k'''}^*$ where $k' + k'' - k''' = l$).

We now may calculate the responses at ω_1, ω_2 and all combination tones, if we assume low to moderate sound levels and an ω_{ch} close to the forcing frequencies. Inserting the expansion of $z(t)$ into the Hopf equation, to lowest order in ω_1 and similarly for ω_2 we obtain

$$i\omega_1 a_1 = (\mu+i)\omega_{ch}a_1 - \omega_{ch}|a_1|^2 a_1 - 2\omega_{ch}|a_2|^2 a_1 - \omega_{ch}F_1, \quad (3)$$

$$i\omega_2 a_2 = (\mu+i)\omega_{ch}a_2 - \omega_{ch}|a_2|^2 a_2 - 2\omega_{ch}|a_1|^2 a_2 - \omega_{ch}F_2, \quad (4)$$

respectively. Solving Eqs. (3) and (4) simultaneously yields complex-valued expressions for a_1 and a_2. To lowest order (appropriate for low sound levels), for the first combination tone CT1 at $\omega_1 - \omega_0$, we obtain

$$i(\omega_1 - \omega_0)a_{CT1} = (\mu + i)\omega_{ch}a_{CT1} - \omega_{ch}a_1 a_1 a_2^*. \quad (5)$$

For higher sound levels, to the r.h.s. the cubic interaction terms

$$-2\omega_{ch}a_{CT1}|a_1|^2 - 2\omega_{ch}a_{CT1}|a_2|^2. \qquad (6)$$

would have to be added. This, however, retains the equation to be linear in a_{CT1}, since the $|a_{CT1}|^2 a_{CT1}$-term is of higher order. The next combination tone CT2 follows to lowest order ($\propto (a_1)^3(a_2^*)^2$) from

$$i(\omega_1 - 2\omega_0)a_{CT2} = (\mu + i)\omega_{ch}a_{CT2} - 2\omega_{ch}a_{CT1}a_1 a_2^*, \quad (7)$$

with $\omega_1 - 2\omega_0$, and is thus mainly a result of the interaction of the the first CT and both primary responses (and not a_{CT1} and a_1 alone, as seems to have been an assumption of Ref. [5]). The latter term and two other interaction terms, i.e.

$$-\omega_{ch}a_{CT1}^2 a_1^* - 2\omega_{ch}a_{CT2}|a_1|^2 - 2\omega_{ch}a_{CT2}|a_2|^2 \qquad (8)$$

need to be considered when including higher order contributions. Close to bifurcation ($\mu \approx 0$) and for $\omega_{ch} \approx \omega_1$, Eq. (7) yields

$$a_{CT2} \approx \frac{\omega_{ch}a_1 a_2^*}{i\omega_0}a_{CT1}. \qquad (9)$$

For the third combination tone CT3 at $\omega_1 - 3\omega_0$, we proceed correspondingly and get to lowest order

$$i(\omega_1-3\omega_0)a_{CT3} = (\mu+i)\omega_{ch}a_{CT3} - 2\omega_{ch}a_{CT2}a_1 a_2^* - \omega_{ch}a_{CT1}^2 a_2^*. \qquad (10)$$

Here, the last two terms are of the same order ($\propto (a_1)^4(a_2^*)^3$). Using Eqs. (5) and (9), for $\omega_{ch} \approx \omega_1$ and $\mu \approx 0$ we obtain $a_{CT3} \approx \frac{\omega_{ch}a_1 a_2^*}{i\omega_0}a_{CT2}$, which is exactly of the same form as Eq. (9). Handling the interaction terms of lowest order carefully shows that the same law holds for all subsequent combination tones CT4, CT5, ..., which leads to an approximate exponential decay of CT amplitudes as

$$a_{CTk} \approx \frac{\omega_{ch}a_1 a_2^*}{i\omega_0} a_{CT(k-1)} = \kappa\, a_{CT(k-1)}, \qquad (11)$$

with $\kappa := (\omega_{ch}a_1 a_2^*)/(i\omega_0)$. Approximate exponential decays emerge off-bifurcation as well (e.g. $\mu = -0.1$), and not exactly at resonance $\omega_{ch} = \omega_1$, see below.

Two examples confirm these results. For the first we choose a two-tone stimulus of $\omega_1 = 2\pi \cdot 2000$ rad/s and $\omega_2 = 2\pi \cdot 2200$ rad/s with amplitudes $F_1 = F_2 = 0.01$. In Fig. 1a). We then compare the numerical integration of Eq. (1) with our above-derived analytical calculations for $\omega_{ch} = \omega_1$ and $\mu = 0$: For the theoretical approach, a_1 and a_2 follow from solving Eqs. (3) and (4) simultaneously, which yields -14.1 and -22.3 dB respectively. a_{CT1} follows from Eq. (5) together with Eq. (6), yielding a response of $a_{CT1} = -33.06$ dB. Using this, we obtain (to lowest order) from Eq. (7) a_{CT2} and from Eq. (10) a_{CT3} (and correspondingly the other CT). For the numerical integration, we chose a sample rate $SR = 80$ kHz and a fourth-order Runge-Kutta scheme with integration step $h = SR^{-1}$, where we discarded the first 0.25s, whereupon we observe an excellent agreement. For the second example, we chose $\omega_2/\omega_1 = 1.05$ and $\omega_{ch} = 2\omega_1 - \omega_2$, which corresponds to the biological experiment of [6]. We use $\omega_1 = 2\pi \cdot 2000$ rad/s and $\omega_2 = 2\pi \cdot 2100$ rad/s for simplicity ($\Delta f := (\omega_2 - \omega_1)/2\pi = 100$ Hz), and amplitudes $|F| = 0.01$ (-40 dB, moderate to high sound level). Fig. 1 shows the obtained response b) for a biologically reasonable value of $\mu = -0.1$ and c) at criticality, $\mu = 0$. Both settings produce exponential CT amplitude decays, but with significantly too high decay exponents for corresponding sound pressure levels (Biology: 5-6 dB/Δf for the lower CT [6]). For the whole set of biological measurements from 30 to 80 dB SPL made in [6], a single Hopf oscillator underestimates CT levels substantially. This is, however, not the case for the compound Hopf cochlea, as will be exhibited below.

Full Hopf Cochlea: In biology, CT of frequencies lower than stimulus propagate down the cochlea until the waves are amplified and stopped where their frequency matches the characteristic frequency ω_{ch}. This situation that differs from the single Hopf element case, leads to an asymmetric (low-pass) and generally slower CT decay. We first exhibit the response corresponding to the first cubic combination
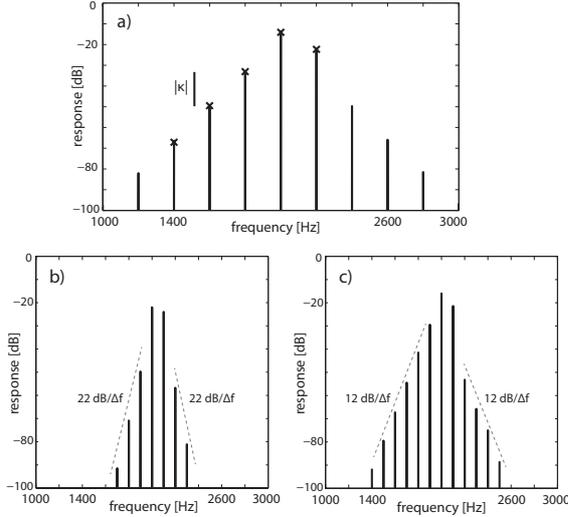
Figure 1: CT in a single Hopf oscillator. a) Fourier transform of the response of a single Hopf oscillator ($\omega_{ch} = 2\pi \cdot 2000$ rad/s, $\mu = 0$) when subject to two-tone forcing ($\omega_1 = 2\pi \cdot 2000$ rad/s and $\omega_2 = 2\pi \cdot 2200$ rad/s with amplitudes $F_1 = F_2 = 0.01$); spectrum: numerical integration, crosses: analytical results; exponential decay factor $|\kappa| = |\frac{\omega_{ch} a_1 a_2^*}{i\omega_0}|$ from Eq. (11). b), c) Same experiment as in a), with $\omega_{ch} = 2\pi \cdot 1900$ rad/s, $\omega_{1,2} = 2\pi \cdot (2000, 2100)$ rad/s and amplitudes $F_1 = F_2 = 0.01$, where b) $\mu = -0.1$ and c) $\mu = 0$ (oscillator at criticality).

tone 'CT1', i.e. the $2f_1 - f_2$-tone. For this, we focus on a location with $\omega_{ch} = \omega_{CT1}$, to see how strong a two-tone signal from two single pure tones of equal strength would be required to be to generate the same effect as a direct stimulation by a tone of frequency f_{ch} - for a single Hopf oscillator, and for a compound cochlea as well. We then compare these results to the biological measurements. The conventional quantification of this difference is the 'relative CT1-tone strength', that evaluates at a characteristic place x_{ch} how much stronger a two-tone input (the two inputs of equal strength) having a CT1 at ω_{ch} would be needed to generate the same response as a pure tone with ω_{ch} [6]. The horizontal distances between the black and the green lines in Fig. 2a, b) exhibit this measure. The obtained results show that CT1-amplitudes depend in a nontrivial manner on stimulation level, and also on whether we consider a Hopf amplifier alone or a section within the compound cochlea. Only the compound cochlea reproduces the effect correctly, and is also consistent with measurements from the apical part of the cochlea [7].

We now focus on the full set of cubic combination tones. These results (not shown) demonstrate that the CT are stronger for higher input levels; nevertheless, the first CT at frequency below the stimulation frequencies f_1 and f_2 (i.e. at frequency $2f_1 - f_2$) is clearly visible even for an input level as low as 30 dB SPL (-84 dB). CT are thus not only relevant at high sound levels. As we will see below,
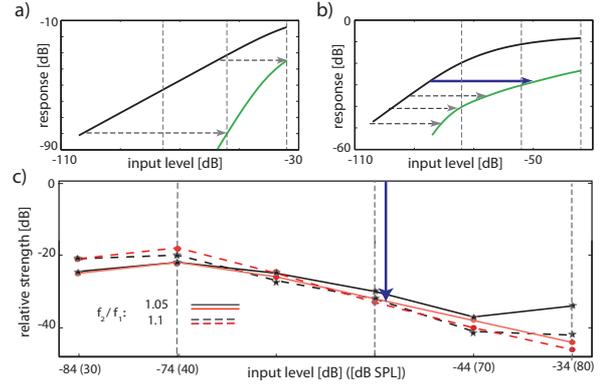
Figure 2: The behavior of the first cubic CT in a single Hopf element, in the compound cochlea and in biology. a,b) Response amplitude to a pure tone of frequency f_{ch} (black), response to a two-tone input (equal strength of components) with $f_{ch} = f_{CT1}$ (green). The difference (arrows) is the 'relative strength of CT1'. a) Single Hopf amplifier (no fluid comprised), and b) cochlea section 6 where $f_{ch} = f_{CT1}$. c) Relative strength of CT1 for two f_2/f_1-frequency ratios. Red: cochlea section 6, black: biological data [6] ($f_{ch} = 9000$ Hz). The blue arrows in b) and c) describe the same experimental result.

the CT below f_1, f_2 propagate further down the cochlea and become ever more dominant in the response. This has far-reaching consequences for pitch perception [8].

Comparing the results obtained for the full cochlea, we find that single Hopf oscillators are unsuited to reproduce biological CT responses. The feed-forward coupling of the oscillators and the low-pass filtering change CT responses in an essential way: CT amplitudes decay in general more mildly, and moreover asymmetrically. Last but not least, the results obtained for the full cochlea provide us with the opportunity to match our simulations to the sound level scales of biological measurements (dB SPL). For the experiment shown, an input level of -74 dB in the simulation is comparable to an input sound pressure level of 40 dB SPL. From this reference point, all other sound levels can be deduced (e.g. -64 dB $\cong$ 50 dB SPL). This matching has provided a detailed reproduction of the well-known pitch shift effects [8], but also seems to be relevant for the faithful reproduction of other salient biological data.

In order to characterize the full cochlear response to two-tone stimulation, we need to track the signal along its way through the cochlear sections. The first cochlea sections mildly amplify the two tones, without any further particular effect on the signals's waveform or spectrum. When the frequencies f_1, f_2 get close to their respective place of resonance (section 10), the first CT start to appear. Further on (section 12), the lower CT ($f < f_1, f_2$) get stronger, and new CT appear. At section 14, the waveform and the spectrum have changed completely: the higher frequencies

(including f_1, f_2) have been dissipated, and the remaining frequencies are grouped around the local characteristic frequency $f_{ch} = 1.31$ kHz. Towards the end of the cochlea, also the remaining lower CT vanish gradually.

3. Hearing Threshold and Generalized Hopf Cochlea

We then looked at the hearing threshold curve of the Hopf cochlea and compared it with the psychoacoustic hearing threshold curves of human subjects in age groups 10-21 and 56-65 years [11], where our Hopf cochlea comprised 23 sections ($\mu = -0.25$) with characteristic frequencies covering the frequency interval from 311 to 14.08 kHz. Input signals were pure tones (no higher harmonics). A second surprising manifestation of the, this time, inherently global, nonlinear character of the hearing sensor is that the hearing threshold curves that characterize the finest sound that elicits a neuronal response obtains a measurable response are very well reproduced by the Hopf cochlea. This observation is based on the conversion of the cochlea measurements to sound pressure units, where a cochlear amplification threshold of -50 dB corresponds to the (psychoacoustic) hearing threshold. Declaring a section excited, if the response reaches above this level, without any tuning an extremely good approximation of the u-formed psychoacoustical hearing threshold is obtained. Since this emerges despite this uniform tuning of the compound cochlea, this effect is of a global nonlinear nature of the hearing device, observed even for stimulations that are fully based on pure tones (no higher harmonics are present in the input signal).

Already v. Helmholtz [2] had speculated on the origins of the hearing nonlinearities; he saw it in the eardrum nonlinearities or the mechanical impedance-matching middle ear, an 'error' that was finally corrected by v. Bekesy's measurements [9] (a concise account of the time around v. Bekesy's discovery is found in Ref. [10]). Whereas Hopf systems with a cubic nonlinearity ($\gamma = 2$, below) are consistent with biological experiments, we may for completeness also study a generalization to values of γ different from two, by considering in Eq. (1) instead of the term $\omega_{ch}|z|^2z$ a term $\omega_{ch}|z|^\gamma z$, with variable exponent γ. While the decay of the combination tones is a property of the Hopf system, depending, in particular, on the exponent γ, systems with $\gamma \neq 2$ can generate the same effects, but show different compressive nonlinearities and different decay laws of the combination tones. Following the experiments by Smoorenburg [12], the pitch of a sound is extracted in the cochlea in a neighborhood where the lowest audible CT is found (the pitch itself then is the residue pitch of the spectrum obtained at this location [8]). A generalized Hopf system in particular leads to the response $| F |^{\frac{1}{1+\gamma}}$. This also alters the CT decay behavior and will hence also modify the pitch perception, as a function of γ.
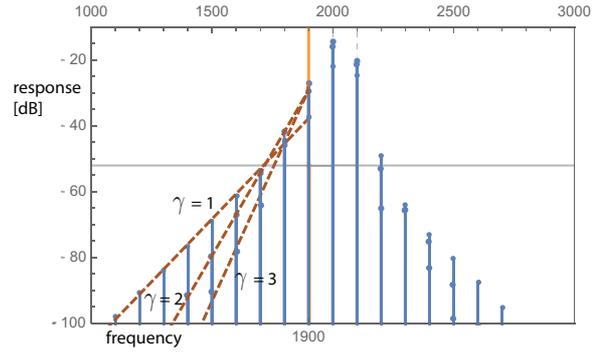
Figure 3: CT decay, generalized Hopf cochlea, $\gamma = 1, 2, 3$.

References

[1] D. Cotugno, "De aquaeductibus auris humane internae..", Typogr. St. Th. Aquinatis, Naples, Italy (1775).

[2] H. von Helmholtz, "Die Lehren von den Tonempfindungen", 6. Reprint, Braunschweig (1913).

[3] R. Stoop, A. Kern, "Two-Tone Suppression and Combination Tone Generation as Computations Performed by the Hopf Cochlea ", *Phys. Rev. Lett.* 93, 268103 (2004); R. Stoop, A. Kern, "Essential auditory contrast-sharpening is preneuronal ", *Proc. Natl. Acad. Sci. U.S.A.* 101, 9179 (2004).

[4] A. Kern, R. Stoop, "Essential Role of Couplings between Hearing Nonlinearities ", *Phys. Rev. Lett.* 91, 128101 (2003).

[5] F. Jülicher, D. Andor, T. Duke, "Physical basis of two-tone interference in hearing ", *Proc. Natl. Acad. Sci. U.S.A.* 98, 9080 (2001).

[6] L. Robles, M.A. Ruggero, N.C. Rich, "Two-Tone Distortion on the Basilar Membrane of the Chinchilla Cochlea ", *J. Neurophysiol.* 77, 2385 (1997).

[7] N.P. Cooper, W.S. Rhode, "Mechanical responses to two-tone distortion products in .. the mammalian cochlea ", *J. Neurophysiol.* 78, 261 (1997).

[8] F. Gomez, R. Stoop,"Mammalian pitch shaped by the cochlear fluid ", *Nat. Phys.* 10, 530 (2014).

[9] G. v. Bekesy, "Über die nichtlinearen Verzerrungen des Ohres ", *Ann. Phys.* 20, 809 (1934).

[10] F. Trendelenburg, "Klänge und Geräusche", Springer, Berlin (1935).

[11] Lee et al., " Behavioral Hearing Thresholds Between 0.125 and 20 kHz Using Depth-Compensated Ear Simulator Calibration ", *Ear. Hear.* 33, 315 (2012).

[12] G.F. Smoorenburg, "Pitch Perception of Two-Frequency Stimuli ", *J. Acoust.Soc.Am.* 48, 924 (1970).