IEICE Proceeding Series

The physical bases of the psychophysical pitch-shift effects

Florian Gomez, Ruedi Stoop

Vol. 2 pp. 429-432
Publication Date: 2014/03/18
Online ISSN: 2188-5079

The physical bases of the psychophysical pitch-shift effects

Florian Gomez[†] and Ruedi Stoop[†]

†Institute of Neuroinformatics, University and ETH Zurich
Winterthurerstrasse 190, 8057 Zurich, Switzerland
Email: fgomez@ini.phys.ethz.ch, ruedi@ini.phys.ethz.ch

Abstract—It has been known for a long time that the perceived pitch of a complex harmonic sound changes if the partials of the sound are shifted in frequency by a fixed amount. Based on simple nonlinear modeling, approximate rules for the shift of the pitch shift could be given (first pitch shift law). In psychoacoustic experiments, however, clear deviations from these predictions were observed (second pitch-shift effects). This raised the question of whether these deviations are due to the biophysics of the nonlinear hearing sensor, the cochlea, or an artifact generated higher up in the auditory pathway. In this article, we demonstrate that the second pitch-shift is generated in the cochlea, and that combination-tone generation, low-pass filtering and a feed-forward coupling are the key factors responsible for the phenomenon. In particular, we find that the scaling laws of Hopf cochlea combination tones explain the classical, to date poorly explained psychoacoustical data of G.F. Smoorenburg (1970).

1. Introduction

Pitch is a central and yet most intriguing trait of human hearing. For pure tones, pitch coincides with the physical frequency of the sound. This changes if a tone contains several partial sounds. Since the discovery of these so-called complex sounds, the mechanisms of the perception of their pitch has been under dispute and many aspects have remained unclear, although quite successful models of pitch perception have been developed ([1] for a review). Given a complex sound containing N *subsequent harmonics*

$$k\, f_0, (k+1)\, f_0, (k+2)\, f_0, ..., (k+N-1)\, f_0 \tag{1}$$

of some fundamental frequency f_0 (*i.e.* $k > 1$), for k not too high and if $N \geq 2$, the perceived pitch f_p is the fundamental f_0. In the case of $N = 2$ or $N = 3$, the residue frequency coincides with the *modulation frequency* of the signal. For a number of psychoacoustic experiments dealing with more complex sounds, this interpretation, surprisingly, falls short: If all higher partials ($k > 1$) are shifted by a fixed amount δf (keeping the modulation frequency f_0 fixed), a shift of the perceived pitch f_p is observed [2, 3, 4]. Simple models [2, 3] of the pitch propose for the shift the rule

$$f_p = f_0 + \frac{\delta f}{k'}, \tag{2}$$

where k' is the ’center’ of the set $\{k, k+1, k+2, \ldots\}$ (for $N = 3$: $k' = k+1$), or, for larger N, one of the lower frequencies present. A corresponding result has been evidenced when a neuronal threshold oscillator was stimulated by a signal $A(\sin f_1\, t + \sin f_2\, t + \ldots + \sin f_n\, t) + \xi(t)$, with frequency components chosen as in Eq. (1) and Gaussian white noise $\xi(t)$. The experiment yields interspike distributions centered at frequencies f_p as given by Eq. (2), with $k' = k + (N-1)/2$ [5]. The guiding idea behind this experiment was that the main resonance should be the dominant periodicity of the subsequent maxima in the stimulus waveform. This parallels the temporal pitch perception paradigm (c.f. [6]), in which f_p is given by the most prominent peak in the autocorrelation or in the auditory nerve interspike interval histograms [7]. For two-tone stimuli ($N = 2$), a pitch-shift of $\delta f/(k + 1/2)$ is predicted (this result also emerges from a pattern-matching perspective [6]). Subsequent psychoacoustic studies (most prominently Smoorenburg’s two-tone pitch-shift experiments) evidenced, however, that this rule [6] only holds for a restricted family of complex sounds. It has been proposed that combination tones (CT, also known as ’distortion products’) are at the origin of the problem [8, 6]. CT emergence is a well-known phenomenon from the cochlear nonlinearity. Given a sequence of harmonics

$$k f_0, \ ..., \ (k+N-1) f_0,$$

a cubic nonlinearity re-introduces all the missing partials

$$(k-1) f_0, \ (k-2) f_0, ... \text{ and } (k+N) f_0, \ (k+N+1) f_0, ...,$$

where the corresponding partials above the stimulus frequencies are generally not perceivable. The lower partials then shift the “center of gravity” of the stimulus towards lower frequencies, which may substantially increase the slope of the lines $f_p(\delta f)$ in Eq. (2).

Experiments with a nonlinear biophysical cochlear model have corroborated the role of CT as the origin of second pitch-shift effects [9]. However, the exact match of the deviations from de Boer’s rule based on a detailed cochlear model cooperating all the claimed generating principles and available biological (i.e. mostly psychoacoustic) data is still missing. In the present work, we generate CT from a biophysically realistic model of the cochlea and reproduce all salient findings reported in Ref. [10], to demonstrate that our model fully complies with presently accepted biophysical evidence. In the second part of our work, we demonstrate that under very mild cochlear read-out assumptions, this automatically leads to the second pitch-shift effect as reported by Smoorenburg [6].

2. CT - generation

An underlying problem with the application of Eq. (2) to complex sounds is that it is based on the inherent assumption of equal amplitudes of all partials. However, along the cochlear duct, CT of unequal amplitude decay rates are generated, due to the nonlinear processes present in the cochlea. CT were long thought to be relevant only in the context of high sound levels (hence their alternative naming 'distortion products'), assuming that the hearing system would be essentially linear at low to moderate sound levels. Although such a linearity at low sound levels is still sometimes incorporated in cochlear modeling (e.g. [11]), this view was challenged quite early by contrary psychoacoustic evidence [12, 6] demonstrating that CT are already perceived at relatively low sound levels. CT are thus not a high-level input artifact, but are ubiquitously present in the hearing system. A decade ago, it was proposed that a relaxation oscillator close to a bifurcation could account for all salient nonlinear properties of hearing [13, 14]. Based on these insights, we developed and realized a biophysically detailed Hopf cochlea in software and hardware [15, 16, 17], which reproduces the biological evidence extremely well (c.f., e.g., [9], Supplemental Material). In this model, the cochlea is discretized into sections, where each section hosts an amplification process that is the result of a stimulated Hopf process

$$\dot{\mathbf{z}} = (\mu + i)\omega_{ch}\mathbf{z} - \omega_{\mathbf{ch}}|\mathbf{z}|^2\mathbf{z} - \omega_{\mathbf{ch}}\mathbf{F(t)},\ \mathbf{z} \in \mathbb{C}, \tag{3}$$

where $\mathbf{F(t)}$ is the stimulation signal and μ measures the systems' distance to its Hopf bifurcation point.

In order to show that our model of the cochlea is firmly based on biophysical reality, we will first explicitly reproduce the biophysical findings related to CT as collected in Ref. [10]. To this end, we consider a signal composed of the harmonics of given angular frequencies $k\omega_0$, ..., $(k+N-1)\omega_0$ and amplitudes F_k, ..., F_{k+N-1} (some amplitudes possibly zero). Since all CT are multiples of ω_0, we expand the response of a single Hopf-oscillator in a Fourier series $z(t) = \sum_j a_j e^{\imath j\omega_0 t}$. For a frequency $\omega_l = l\omega_0$ we obtain

$$(i(\omega_l - \omega_{ch}) - \mu\omega_{ch})a_l + i.t. = -\omega_{ch}F_l, \tag{4}$$

where *i.t.* denotes cubic terms of three interacting modes

$$\propto \omega_{ch}\, a_{k'}\, a_{k''}\, a^*_{k'''},$$

with $k' + k'' - k''' = l$. The first term on the l.h.s. of Eq. (4) is linear in a_l and thus becomes dominant far away from resonance and bifurcation. For a single-frequency forcing F_k of low amplitude, at resonance and close to bifurcation, a single self-interaction term $|a_k|^2 a_k$ remains and the response $z \propto F^{1/3}$ emerges. In the presence of a second stimulus F_{k+1}, the response with respect to F_k is suppressed, due to an interaction term $2\omega_{ch}a_k|a_{k+1}|^2$ that is linear in a_k. A CT at frequency ω_{k-1} is then generated, via the interaction term $w_{ch}a_k^2 a^*_{k+1}$, the $2f_1 - f_2$-CT. Further cubic CT are generated at frequencies $\omega_l,\ l < k-1$, with amplitudes decreasing exponentially. (c.f. Fig. 1, region of exponential scaling).

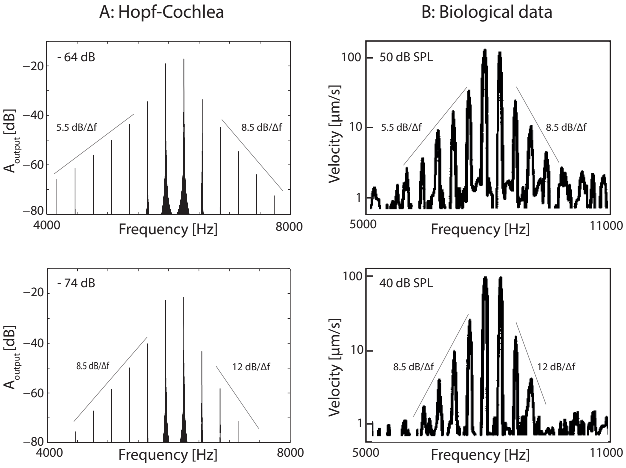

Figure 1: Basilar membrane response spectrograms for two-tone stimulation with different amplitudes (frequencies $f_2/f_1 = 1.05$ and $2f_2 - f_1 = f_{ch}$). Left: Hopf-cochlea model, 6th section ($f_{ch} = 5656$ Hz). Right: Biological data [10] ($f_{ch} = 7500$ Hz).

Exponential decays of CT levels were observed in psychoacoustic experiments some decades ago [12]. More recently, direct experimental observation of two-tone CT on biological inner ear basilar membranes became accessible by laser interferometry. In these experiments, for a situation where $f_2/f_1 = 1.05$ and $f_{ch} = 2f_1 - f_2$ and 30 to 80 dB SPL [10], exponentially decaying CT amplitudes were confirmed. From our device, we obtain an excellent agreement with the biological data (Fig. 1).

For high-level forcing (50-80 dB), CT amplitudes decay around $5 - 5.5$ dB/f_0, with higher slopes for lower sound levels. Here, we find a discrepancy between biology and the direct analytical calculations or numerical integration from Eq. (3) that yield stronger decays at all sound levels, even at bifurcation. Single Hopf elements (which is what we so far dealt with) substantially underestimate CT strength. This is because in biology, CT of frequencies lower than stimulus frequencies propagate down the cochlea until the corresponding waves are finally amplified and stopped where their frequency matches the characteristic frequency f_{ch} [12]. Cochlea models without a propagation medium do not account for this mechanism and therefore underestimate CT strength.

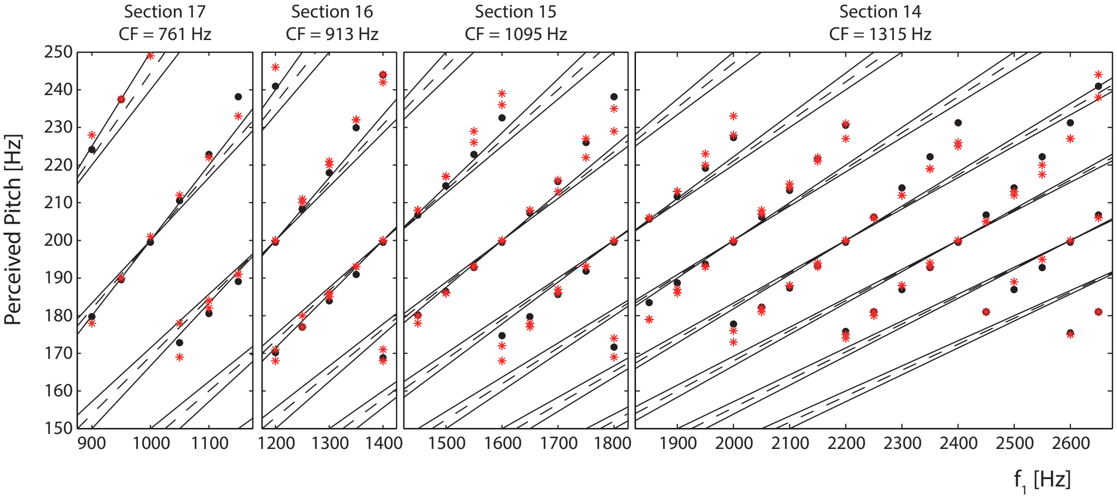

Figure 2: Smoorenburg's pitch-shift experiments: Two-frequency stimulation $f_2 = f_1 + 200$ Hz. Psychoacoustic data (partial sound levels each 40 dB SPL, red stars [6]) and simulation (Hopf cochlea, different sections, partial tones each -74 dB, black circles). Solid lines: Predictions by Eq. (2) for $k' = k$, $k' = k + 1/2$ and $k' = k + 1$, respectively.

3. Pitch fundaments

To corroborate the arguments made by model data, we use our Hopf cochlea that implements a chain of feed-forward coupled Hopf oscillators, including fluid internal friction by means of low-pass filtering. This model has been shown to reproduce biophysical evidence extremely well [15, 16, 17], and it will also do this in the present context. For all results presented, we used a software model representation composed of 20 sections with characteristic frequencies covering the range from 14.08 to 0.44 kHz (5 octaves), although equivalent hardware realizations exist. For the first five sections, the bifurcation parameters μ are set to -0.1, and decrease with -0.025/section afterwards. This leads to biologically plausible amplification and tuning curves at more apical sections. As a test, we ensure that the amplification of generated CT by the following sections essentially preserves the exponential scaling far from resonance, while the low-pass filtering now results in larger slopes of CT levels for frequencies above the stimulus frequencies. The results obtained indeed fully coincide with the biological observations (Fig. 1). The minor mismatch of the relative responses at f_1 and f_2 is just the result of the cochlea's discretization into sections.

4. Pitch extraction and second pitch-shift

Obviously, the perceived pitch of a general spectrum as in Fig. 1 cannot be predicted directly from Eq. (2), since unequal weights of the partial tones fail to exhibit a sharply defined k'. As a consequence, we have resorted to the autocorrelation picture, where pitch is defined in terms of the most prominent peak in the autocorrelation-function (ACF). Smoorenburg's pitch-shift data for two-tone experiments are compared in Fig. 2 to our Hopf-Cochlea data, obtained from an input of the form $F_1 e^{2\pi i f_1 t} + F_2 e^{2\pi i (f_1+200)t}$ and measuring f_p at a specific cochlea section by computing the ACF. Two key parameters determine the amount of pitch-shift observed: whereas the forcing amplitudes of the stimulus determine the exponent in the decay of CT via the responses a_1 (quadratic) and a_2 (linear), the choice of the read-off cochlea section governs the low-pass filtering and the overall-amplification of the lower CT. For Fig. 2, a sound level of both partials of -74 dB was chosen, which corresponds to the partial levels of 40 dB SPL used in Smoorenburg's experiments (see scale-correspondence in Fig. 1).

5. Conclusion

The quest for exactly how pitch should be defined and where in the cochlea pitch is exactly located and to be extracted has long been a matter of debate [18]. Whereas it was already suggested by Goldstein [19] that CT could quantitatively account for Smoorenburg's measured pitch-shift effects, here we provide for the first time the reproduction and a quantitative explanation of the second pitch-shift effects on the basis of a biomorphic cochlea model. We find that the key factors for the striking correspondence between our modeling results and the biological data are the correct scaling of CT, from the feed-forward coupling and associated low-pass filtering implemented in our cochlea. Goldstein proposed an abstract spectrum-based pitch estimator [19], where the super-threshold frequencies across the channels would have to be summed, omitting frequencies from 2000 Hz upwards because of the limited frequency resolution of the auditory system. We would also be able to reproduce the pitch-shift results in this way. With the concept of a local waveform-based pitch [9] that we have used for our study, it is, however, not the limited frequency resolution, but the low-pass filtering that naturally filters out the highest frequencies, that is essential for the correct second pitch-shifts. We find an accurate reproduction of Smoorenburg's psychoacoustic data for a variable extraction region, shifting monotonically with the primary frequencies. As an advantage of our interpretation, the specific location may be either automatically determined (e.g. by the lowest CT) or, alternatively, guided by attention. While the present work fully establishes the perceived second pitch-shift on the basis of biophysical insight on the level of the cochlea, for an explanation of psychoacoustic pitch perception from physics, still the gap from continuous cochlear dynamics to the discrete world of auditory nerve spikes needs to be crossed. By the addition of inner hair and auditory nerve cell to the cochlea, we recently succeeded in modeling a full peripheral auditory system based on basic biophysical principles [20]. Using this approach, we were able to demonstrate the intriguing fact that despite the many transductions and transformations the signal undergoes from the continuous world of basilar membrane dynamics in the cochlea to the discrete world of neuronal spiking at the end of the auditory nerve, the information available at the level of the cochlea is transmitted without apparent loss. In particular, the auditory system seems to take extreme care to transmit the perceived pitch (including the second pitch-shift) across the different stages by exploiting stochastic resonance at the synaptic interface between inner hair and auditory nerve cells [20]. With these two steps, human psychoacoustic pitch perception can be fully explained by cochlear biophysics. In this sense, the original dream of Seebeck, Ohm and Helmholtz has finally come true.

References

[1] A. Cheveigné, "Pitch perception models," in *Pitch* (C. Plack, R. Fay, A. Oxenham, and A. Popper, eds.), vol. 24 of *Springer Handbook of Auditory Research*, pp. 169–233, Springer, New York, 2005.

[2] J. F. Schouten, "De toonhoogtegewaarwording," *Philips Technisch Tijdschr.*, vol. 5, pp. 298–306, 1940.

[3] E. de Boer, "Pitch of inharmonic signals," *Nature*, vol. 178, pp. 535–536, 1956.

[4] J. F. Schouten, R. J. Ritsma, and B. L. Cardozo, "Pitch of the residue," *J. Acoust. Soc. Am.*, vol. 34, no. 9B, pp. 1418–1424, 1962.

[5] D. R. Chialvo, O. Calvo, D. L. Gonzalez, O. Piro, and G. V. Savino, "Subharmonic stochastic synchronization and resonance in neuronal systems," *Phys. Rev. E*, vol. 65, p. 050902, May 2002.

[6] G. F. Smoorenburg, "Pitch perception of two-frequency stimuli," *J. Acoust. Soc. Am.*, vol. 48, no. 4B, pp. 924–942, 1970.

[7] J. Licklider, "A duplex theory of pitch perception," *Cell. Mol. Life. Sci.*, vol. 7, pp. 128–134, 1951.

[8] J. Goldstein and N. Kiang, "Neural correlates of the aural combination tone 2f1- f2," *Proc. IEEE*, vol. 56, no. 6, pp. 981–992, 1968.

[9] S. Martignoli and R. Stoop, "Local cochlear correlations of perceived pitch," *Phys. Rev. Lett.*, vol. 105, p. 048101, Jul 2010.

[10] L. Robles, M. A. Ruggero, and N. C. Rich, "Two-tone distortion on the basilar membrane of the chinchilla cochlea," *J. Neurophysiol.*, vol. 77, no. 5, pp. 2385–2399, 1997.

[11] R. Meddis, L. P. O'Mard, and E. A. Lopez-Poveda, "A computational algorithm for computing nonlinear auditory frequency selectivity," *J. Acoust. Soc. Am.*, vol. 109, no. 6, pp. 2852–2861, 2001.

[12] J. L. Goldstein, "Auditory nonlinearity," *J. Acoust. Soc. Am.*, vol. 41, no. 3, pp. 676–699, 1967.

[13] V. M. Eguíluz, M. Ospeck, Y. Choe, A. J. Hudspeth, and M. O. Magnasco, "Essential nonlinearities in hearing," *Phys. Rev. Lett.*, vol. 84, pp. 5232–5235, May 2000.

[14] S. Camalet, T. Duke, F. Jülicher, and J. Prost, "Auditory sensitivity provided by self-tuned critical oscillations of hair cells," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 97, no. 7, pp. 3183–3188, 2000.

[15] A. Kern and R. Stoop, "Essential role of couplings between hearing nonlinearities," *Phys. Rev. Lett.*, vol. 91, p. 128101, Sep 2003.

[16] R. Stoop and A. Kern, "Two-tone suppression and combination tone generation as computations performed by the hopf cochlea," *Phys. Rev. Lett.*, vol. 93, p. 268103, Dec 2004.

[17] S. Martignoli, J.-J. van der Vyver, A. Kern, Y. Uwate, and R. Stoop, "Analog electronic cochlea with mammalian hearing characteristics," *Appl. Phys. Lett.*, vol. 91, pp. 064108 –064108–3, aug 2007.

[18] R. Plomp, "Pitch of complex tones," *J. Acoust. Soc. Am.*, vol. 41, no. 6, pp. 1526–1533, 1967.

[19] J. L. Goldstein, "An optimum processor theory for the central formation of the pitch of complex tones," *J. Acoust. Soc. Am.*, vol. 54, no. 6, pp. 1496–1516, 1973.

[20] S. Martignoli, F. Gomez, and R. Stoop submitted (2013).