Local Cochlear Correlations of Perceived Pitch

Stefan Martignoli* and Ruedi Stoop†

Institute of Neuroinformatics, University of Zurich and ETH Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

(Received 29 October 2009; revised manuscript received 17 March 2010; published 20 July 2010)

Pitch is one of the most salient attributes of the human perception of sound, but is still not well understood. This difficulty originates in the entwined nature of the phenomenon, in which a physical stimulus as well as a psychophysiological signal receiver are involved. In an electronic realization of a biophysically detailed nonlinear model of the cochlea, we find local cochlear correlates of the perceived pitch that explain all essential pitch-shifting phenomena from physical grounds.

DOI: 10.1103/PhysRevLett.105.048101 PACS numbers: 87.85.fk, 43.66.+y, 87.19.lt

Pitch is one of the three fundamental auditory attributes of sounds, along with loudness and timbre. How the pitch is extracted from a sound has been a physics puzzle for centuries [1], starting in modern times with a controversy among Seebeck, Ohm, and von Helmholtz [2]. From linear models of the cochlea, the pitch frequency should equal an ingredient of the stimulus frequencies. For harmonic sounds, which are a superposition of frequencies $\{kf_0\}$, $k \in \mathbb{N}$ the harmonic number, the fundamental frequency f_0 itself is indeed the perceived pitch. If, however, from such a signal the lowest frequency components are removed, the pitch remains at f_0. This intriguing observation has been termed the problem of the "missing fundamental."

One explanation offered was that in the cochlea, f_0 is reintroduced, by some nonlinear distortion of the cochlea [3]. Alternatively, it was proposed that the temporal signal envelope is relevant to the pitch [4]. The pitch-shifting experiments of de Boer [5] and Schouten *et al.* [6] overthrew both explanation attempts. They used signals of the form $f(t) = \frac{A}{2}[1 + \cos(2\pi f_{\mathrm{mod}} t)] \times \sin(2\pi f_{\mathrm{car}} t)$ which, by trigonometric expansion, contain only the frequencies $\{f_{\mathrm{car}} - f_{\mathrm{mod}}, f_{\mathrm{car}}, f_{\mathrm{car}} + f_{\mathrm{mod}}\}$ [cf. Fig. 1]. The advantage of such a signal is that upon an appropriate shift $\delta f'$ from $f_{\mathrm{car}} \rightarrow f'_{\mathrm{car}} = f_{\mathrm{car}} + \delta f'$, the generically inharmonic sounds generated can be made harmonic: $\{f'_{\mathrm{car}} - f_{\mathrm{mod}}, f'_{\mathrm{car}}, f'_{\mathrm{car}} + f_{\mathrm{mod}}\} = \{(k-1)f_0, kf_0, (k+1)f_0\}$. When f_{car} is shifted by an arbitrary amount δf from the latter, the key observation is that the human perceived pitch f_{pp} (also called the "residue pitch") differs from f_{mod}, fails to match any frequency component of the input signal, but shifts with a dependence both on δf and on the harmonic number $k \in \mathbb{N}$ of the carrier frequency. Motivated by these results, the perceived pitch was postulated to follow the formula

$$f_{pp} = f_0 + \frac{\delta f}{k} \tag{1}$$

[5], which now is known as "de Boer's (first) pitch-shifting rule." As a consequence, for fixed k, as a function of δf, f_{pp} should follow a straight line of slope $1/k$. This is, however, only approximately the case. The systematic deviations from de Boer's rule are called the "second pitch-shifting effects" (SPSE) [5–7]. Deviations manifest in two major ways. For two- and three-tone sounds, the slopes for fixed $k \geq 6$ are consistently larger ([6,8], Fig. 2 and Figs. 3–4, respectively) and for four-tone sounds for $k = 2, 3$ smaller, than what is predicted by de Boers rule [9]. Moreover, if at constant carrier frequency f_{car} the modulation frequency f_{mod} is increased, a systematic decrease in f_{pp} is observed ([5,6], Fig. 3 of [6], e.g.). f_{pp} even depends slightly on the sound intensity [10]. For obtaining an idea of these deviations, see our Figs. 2 and

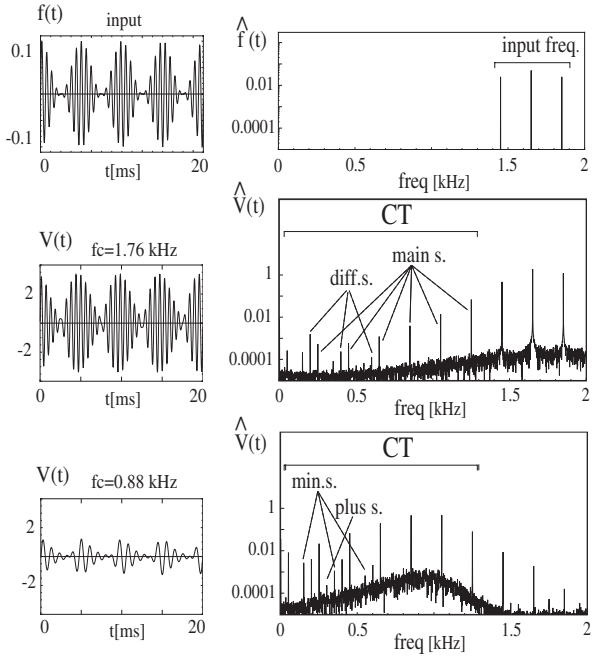

FIG. 1. Cochlea response $V(t)$ to an inharmonic AM tone $f(t, A = 0.1, f_{\mathrm{car}} = 1.65$ kHz, $f_{\mathrm{mod}} = 0.2$ kHz), waveforms $f(t)$, $V(t)$ (time origins artificially aligned) and their Fourier transforms. Stimulation (top line, three frequencies) and two frequency channels labeled by their characteristic frequencies f_c, revealing at $f_c = 0.88$ kHz a strong CT-generated signal that classically should be absent.

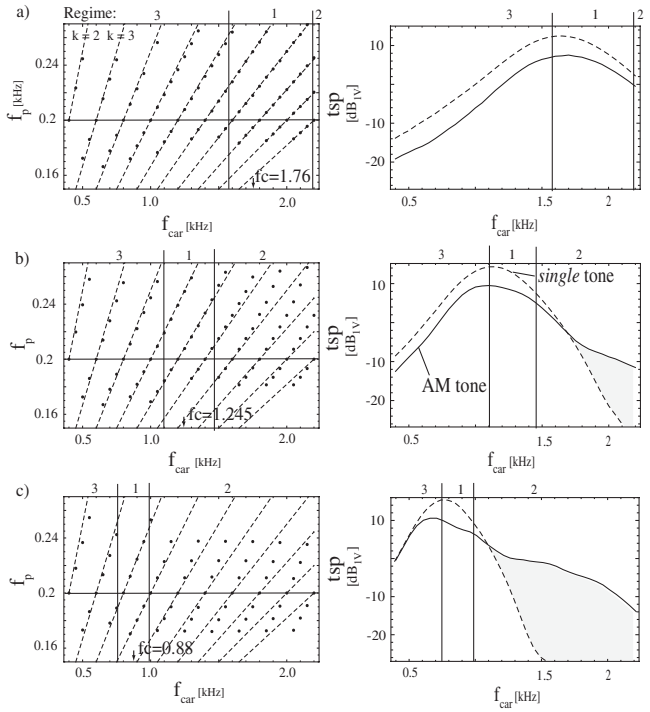

FIG. 2. Left panels: $f_p(x)$ for $f(t, A = 0.1, f_{\mathrm{mod}} = 0.2\ \mathrm{kHz}, f_{\mathrm{car}})$ measured at the cochlea sections (a) $f_c = 1.76$ kHz, (b) $f_c = 1.245$ kHz and (c) $f_c = 0.88$ kHz. From left to right: increasing harmonic numbers k. Dashed lines: de Boer's rule for fixed k. The horizontal line through $f_p = 0.2$ kHz: purely harmonic sounds providing no second effect ($\delta f = 0$). Right panels: Total signal power (tsp) of AM vs single tones (identical input power $-20\ \mathrm{dB}_{1V}$, $f_{\mathrm{single}} = f_{\mathrm{car}}$). Labels: three regimes (*s*. text), shading: CT-generated signal surplus.

3. In these figures, mostly not f_{pp}, but the corresponding f_p that we derive from the cochlea's local biophysics and compares very well with f_{pp}, are displayed. A recent theoretical result from nonlinear dynamics provides an interpretation of de Boer's rule from this point of view [7,11]. A nonlinear excitable system driven by N quasi-periodic frequencies with equal spacing f_0, where the lowest frequency is of the form $f_1 = nf_0 + \delta f$, has been shown [11] to respond with a (stochastic) resonance at the frequency

$$f_{qp} = f_0 + \frac{\delta f}{n + (N-1)/2}, \tag{2}$$

which for $N = 3$ recovers de Boers rule, since $n + 1 = k$ is the harmonic number of the carrier frequency.

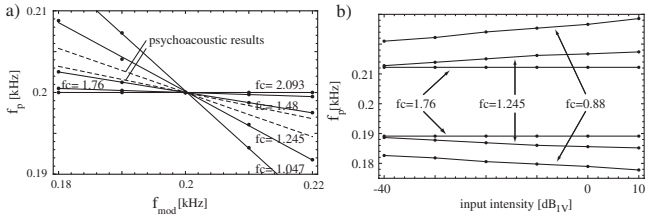

FIG. 3. (a) f_{mod} is varied at fixed $f_{\mathrm{car}} = 2.0$ kHz, $A = 0.1$. Dots: Local pitch measurements at five sections. Dashed lines: Experimental data from two subjects [6]. Regime 1: For section $f_c = 2.093(\sim f_{\mathrm{car}})$ kHz, a zero slope as predicted by de Boers rule emerges. Regime 2: Down the cochlea $f_c < f_{\mathrm{car}}$, the slope increases (*s*. text), (b) Loudness dependence of $f_p(x)$ for an AM tone ($f_{\mathrm{car}} = 1.7$ kHz, $f_{\mathrm{mod}} = 0.2$ kHz). Regime 1: $f_c = 1.76$ kHz (close to f_{car}), reveals zero dependence. Regime 2: For $f_c < f_{\mathrm{car}}$, increasing deviation with increasing distance to f_{car}.

While SPSE thus have remained an issue of speculations, we will show here that they are not primarily a consequence of the processing steps higher up the brain. In order to deal with the pitch issue for general sounds and for the signal within the cochlea, a basic physics definition of pitch must be used. The most useful for comparing physiological and electronic to psychoacoustic measurements is $f_p = 1/T$ [6,10], where T is the "most frequent time interval between signal peaks." If amplitude-modulated (AM) signals pass through the cochlea, they are changed in a nontrivial manner, due to the highly nonlinear nature of the cochlear amplification process [12,13]. As a consequence, we deal with measurements of "local pitches" $f_p(x)$, that depend strongly on the point of measurement x along the cochlear duct or its electronic analogue. The purpose of this Letter is to investigate whether these local pitches $f_p(x)$ could be at the origin of SPSE. Such a program has eluded an experimental investigation up to now. The three fundamental nonlinear aspects of physiologically measurable hearing [12,14] are the large amplification of faint sounds with a compressive nonlinearity of strong sounds [14] and, as a consequence of the nonlinear amplification and for this study most importantly, combination tone (CT) generation [15] and (multitone) suppression [16]. A recent electronic realization [17] of a biophysically detailed [18] model of the cochlea [19,20], which reproduces faithfully all essential physiological measurements (cf. supplementary material [21]) [13], allows us to access the transformation an auditory signal undergoes within a Hopf type cochlea. At the heart of this model is an active amplification implemented by a relaxation oscillator close to a Hopf bifurcation [22]. Systems below the bifurcation point work naturally as small-signal amplifiers [23,24], in a narrow band around the characteristic frequency f_c. In bullfrog hair cells, the Hopf nature of the amplification was experimentally corroborated [25]. For optimally reproducing mammalian hearing characteristics, strong coupling between the outer hair cells as the Hopf amplifiers, and the viscous fluid is essential [13,17,18]. The electronic cochlea is implemented as an array of discrete sections, where each section models an extended region along the basilar membrane by means of the Hopf equation: $\dot{\mathbf{z}} = (\mu + j)\omega_{\mathrm{ch}}\mathbf{z} - \omega_{\mathrm{ch}}|\mathbf{z}|^2\mathbf{z} - \omega_{\mathrm{ch}}\mathbf{F}(t)$, $\mathbf{z} \in \mathbb{C}$. Here, the vectors of the input $\mathbf{F}(t)$ and output $\mathbf{z}$ are complex variables (j is the imaginary unit), and $f_c = \omega_{\mathrm{ch}}/2\pi$ is the characteristic frequency of the section, μ is the tunable parameter that defines each

section's distance from the Hopf bifurcation point. Technically speaking, μ determines a section's gain and quality factor [17]. In each section, the Hopf amplifier is followed by a section-specific 6th-order Butterworth (low pass) filter modeling the viscous losses. For the results presented below, we used a cascade of 16 sections with logarithmically spaced characteristic frequencies f_c (4 sections per octave) and $\mu = -0.2$ for all sections. The resolution of temporal f_p measurements is limited by the sampling rate of 80 kHz, which results in a pitch frequency error of $\Delta f_p \sim 5 \times 10^{-4}$ kHz at 0.2 kHz.

Local temporal pitch.—Measurements taken at different sections confirm that the signal characteristics are strongly changed within the cochlea (see Fig. 1 for an inharmonic AM input). The waveform or Fourier spectra pairs clearly witness the salient nonlinear effects at work for complex sounds: the amplitudes of the forcing frequencies are reduced (compared to single tones) and integer combinations of the forcing frequencies (CT) emerge [15,26,27]. Low-frequency CT propagate further down the cochlea, where they are finally amplified at the corresponding places $f_c = f_{\mathrm{CT}}$ (see the responses at the section $f_c = 0.88$ kHz in Fig. 1). Four series of CT frequencies, ordered by the measured intensities, can be distinguished: The main series: $nf_0 + \delta f$, the difference tone series: nf_0, minus series: $nf_0 - \delta f$ and the plus series: $nf_0 + 2\delta f$, with $n < k - 1 \in \mathbb{N}$. For harmonic sounds ($\delta f \equiv 0$), the four series coincide. At places with $f_c \gg f_0$, the response is almost entirely determined by the main series (see $f_c = 1.76$ kHz in Fig. 1); the other series play a role only in the vicinity of f_0. Frequencies larger than f_c are dissipated in the passive parts of a cochlea section. The responses at $f_c < (k - 1)f_0 + \delta f$ therefore originate from CT that are successively amplified along the cascade (see Fourier transforms at $f_c = 0.88$ kHz in Fig. 1). Below, detailed measurements confirm the profound effect that CT contributions have for shaping the signal before it leaves the cochlea.

Measured local pitch.—Following the above paradigm, f_{car} is varied between 0.4 and 2.2 kHz, for fixed $f_{\mathrm{mod}} = 0.2$ kHz. Measurements at three cochlear locations, see Fig. 2, yield local pitches $f_p(x)$ that are close to the psychoacoustic experiments [6] and, in particular, share the perceived SPSE. For a specific frequency $f_{\mathrm{car}} = f_c$, the local $f_p(x)$ will coincide with f_p (assuming now a continuous cochlea). From the prediction made by de Boer's rule, to the right (left) hand side, the slopes of $f_p(x)$ increase (slightly decrease), consistently with results from psychoacoustical pitch-shift experiments for two-, three- and four-tone inputs [6,8,9]. From the total signal power of a complex AM sound vs that of a single tone, the decisive role of CT in shaping the signal is evident. We observe that the power peaks of the single tones obey a slight left shift on the stimulus intensity as in the biological example [13] and that f_c is exactly underneath the peak for very weak sounds at -70 dB_{1V} (not shown). Since AM tones are subject to both suppression and CT generation, we distinguish three amplification regimes. The border between Regimes 1 and 3 is given by the power peak of the local response. For $f_c = 0.88$ kHz, we observe a small shift between the AM and single tone maxima, due to the much narrowed amplification profile at this frequency that fails to amplify the higher frequency of the AM tone. Here, we take the single tone maximum as the reference. Left to this point on the f_{car}-axis, for AM tones, suppression is the dominant shaping force (Regime 3). The border between Regimes 1 and 2 are given where the effects of suppression and CT-generation balance. These borders mark roughly an area where the deviation from de Boer's rule is within $\pm 10\%$. In Regime 2, CT production and successive amplification leads to a massive increase of the total signal power if compared to single tones: Frequency channels that would be silent are activated by CT generation.

In order to further compare $f_p(x)$ to psychoacoustical measurements f_{pp} [6], in Fig. 3(a) f_{mod} is varied from 0.18 to 0.22 kHz, for fixed $f_{\mathrm{car}} = 2.0$ kHz. A situation very similar to that of the first experiment for single local $f_p(x)$ emerges: At $f_c(x)$ close to the frequencies of the input, the slope is zero (Regime 1), as is expected from de Boer's rule. As f_c decreases—the signal propagates down the cochlea—the measured negative slope increases (Regime 2). This effect increases as we move further down the cochlea. Between $x(f_c = 1.245\ \mathrm{kHz})$ and $x(f_c = 1.48\ \mathrm{kHz})$, the local pitch $f_p(x)$ corresponds to the perceived pitch f_{pp}. As the last effect, we confirm that the correct loudness dependence of $f_p(x)$ is obtained. The obtained results displayed in Fig. 3(b) are fully compatible with the exhibited regimes: No pitch-shift for f_c close to f_{car} (Regime 1), increased slope with increasing distance for $f_c < f_{\mathrm{car}}$ (Regime 2). The underlying mechanism of this dependence is the intensity variation of CT in dependence on the input sound level [15,27]. Thus, the small shift in $f_p(x)$ (and, analogously, of f_{pp}) in dependence of the input strength is fully explained by CT production.

From local pitch to perceived pitch.—The global perception is made on the basis of local pitches f_p [7,10,28–30]. In the nonlinear dynamics approach of [7,28], the assumption was that at the place with the largest response, the output should be dominated by the two lowest input frequencies, leading to a resonance with $N = 2$ in Eq. (2). Our Fig. 2 shows that main series CT trigger large deviations from de Boers rule. These CT have a frequency spacing of f_{mod}, a situation that is equivalent to an experiment with several input frequencies, with effective harmonic number $n < k - 1$. Equation (2) now correctly predicts increased slopes of f_p. We find that the responses start at the resonance place by three input frequencies following de Boer's rule (Regime 1), which then are modified by subsequently generated CT (Regime 2). Thus, the conjecture that CT generation in general is responsible for the second effect [7,28] is fully corroborated on the local level. If the psychoacoustical pitch is computed by a means of a summation over all frequency channels [29,30], the

addition of responses, weighted by the power of the responses, will naturally lead to the perceived SPSE. For high values of $k \geq 6$, the summation will increase the slope of f_{pp}. In Regime 2, this slope increases with the distance of places with $f_c < f_{\mathrm{car}}$. For very low pitches (i.e. for $k < 3$ in the presented experiments), the summed response will lead to a slightly smaller slope than the one predicted by de Boer's rule (Regime 3). From psychoacoustical studies involving two different stimuli of slightly differing perceived pitches, a pitch 'dominance region' has been conjectured [8,10]. Physiological experiments on the firing statistics of auditory nerve fibers [30] support this hypothesis and suggest that this may be due to phase-locking between auditory nerve neurons [31] and the cochlear output. Following this line of reasoning, Refs. [7,28] offer a convincing perspective of how from the local pitch $f_p(x)$ the perceived pitch f_{pp} might be extracted. For our Fig. 3(a), an estimate of the dominance region would be between 1.245 and 1.48 kHz, which is slightly elevated, but still compatible with estimates by psychoacoustical theories (about $4 \times f_0 = 0.8$ Hz [10], respectively, 5–6 times f_0 [8]).

Using a local pitch concept, our study demonstrates that local correlates of SPSE may exist on the cochlear level that are the result of the nonlinear cochlea's CT production. Preliminary results in our detailed modeling study indicate that the observed essential signal properties are maintained in the auditory nerve. Local pitches offer the possibility of computing and grouping simultaneously the pitches of multiple auditory objects, an ability that is crucial when listening, e.g., to an orchestra. The binaural pitch described in Ref. [11], although apparently of an entirely different physical nature, appears to offer an alternative or additional strategy for achieving this. We expect, however, that in both cases the final percepts will be extracted along parallel lines [32,33]. Our work sheds a new light on the classical Seebeck-Ohm-Helmholtz dispute. Whereas the Hopf amplification is organized similarly to the von Helmholtz's tonotopic principle, the local nonlinearly generated waveform more closely resembles Seebeck's temporal code [34]. For the global percept, both aspects, x and $f_p(x)$ are important, which, to some extent, may reconcile the views of von Helmholtz and Seebeck.

This work was supported by the Swiss SNF (Grant No. 200021-122276).

*mstefan@ini.phys.ethz.ch
†ruedi@ini.phys.ethz.ch

[1] A. de Cheveigné, in *Pitch, Neural Coding and Perception*, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper (Springer, New York, 2005), p. 169.
[2] R. S. Turner, Br. J. Hist. Sci. **10**, 1 (1977).
[3] H. L. F. von Helmholtz, *Die Lehre von den Tonempfindungen* (Vieweg, Braunschweig, 1863).
[4] J. C. R. Licklider, Experientia **7**, 128 (1951).
[5] E. de Boer, Nature (London) **178**, 535 (1956).
[6] J. F. Schouten, R. J. Ritsma, and B. L. Cardozo, J. Acoust. Soc. Am. **34**, 1418 (1962).
[7] J. H. E. Cartwright, D. L. González, and O. Piro, Phys. Rev. Lett. **82**, 5389 (1999).
[8] G. F. Smoorenburg, J. Acoust. Soc. Am. **48**, 924 (1970).
[9] A. Gerson and J. L. Goldstein, J. Acoust. Soc. Am. **63**, 498 (1978).
[10] R. J. Ritsma, in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (A. W. Sifthoff, Leiden, 1970), p. 250.
[11] D. R. Chialvo, O. Calvo, D. L. Gonzalez, O. Piro, and G. V. Savino, Phys. Rev. E **65**, 050902 (2002).
[12] L. Robles and M. A. Ruggero, Physiol. Rev. **81**, 1305 (2001).
[13] R. Stoop, T. Jasa, Y. Uwate, and S. Martignoli, Sensors **7**, 3287 (2007).
[14] M. A. Ruggero, Curr. Opin. Neurobiol. **2**, 449 (1992).
[15] L. Robles, M. A. Ruggero, and N. C. Rich, J. Neurophysiol. **77**, 2385 (1997).
[16] M. A. Ruggero, L. Robles, and N. C. Rich, J. Neurophysiol. **68**, 1087 (1992).
[17] S. Martignoli, J.-J. van der Vyver, A. Kern, Y. Uwate, and R. Stoop, Appl. Phys. Lett. **91**, 064108 (2007).
[18] A. Kern and R. Stoop, Phys. Rev. Lett. **91**, 128101 (2003).
[19] M. O. Magnasco, Phys. Rev. Lett. **90**, 058101 (2003).
[20] T. Duke and F. Jülicher, Phys. Rev. Lett. **90**, 158101 (2003).
[21] See supplementary material at http://link.aps.org/supplemental/10.1103/PhysRevLett.105.048101 for a comparison of the electronic Hopf cochlea with the essential biophysical measurements of the mammalian cochlea.
[22] V. M. Eguíluz, M. Ospeck, Y. Choe, A. J. Hudspeth, and M. O. Magnasco, Phys. Rev. Lett. **84**, 5232 (2000).
[23] K. Wiesenfeld and B. McNamara, Phys. Rev. Lett. **55**, 13 (1985).
[24] B. Derighetti, M. Ravani, R. Stoop, P. F. Meier, E. Brun, and R. Badii, Phys. Rev. Lett. **55**, 1746 (1985).
[25] P. Martin and A. J. Hudspeth, Proc. Natl. Acad. Sci. U.S.A. **98**, 14 386 (2001).
[26] F. Jülicher, D. Andor, and T. Duke, Proc. Natl. Acad. Sci. U.S.A. **98**, 9080 (2001).
[27] R. Stoop and A. Kern, Phys. Rev. Lett. **93**, 268103 (2004).
[28] J. H. E. Cartwright, D. L. Gonzalez, and O. Piro, Proc. Natl. Acad. Sci. U.S.A. **98**, 4855 (2001).
[29] R. Meddis and M. J. Hewitt, J. Acoust. Soc. Am. **89**, 2866 (1991).
[30] P. A. Cariani and B. Delgutte, J. Neurophysiol. **76**, 1698 (1996).
[31] J. O. Pickles, *An Introduction to the Physiology of Hearing* (Emerald Group, Bingley, 2008), 3rd ed.
[32] P. Balenzuela and J. García-Ojalvo, Chaos **15**, 023903 (2005).
[33] A. Lopera, J. Buldú, M. Torrent, D. Chialvo, and J. García-Ojalvo, Phys. Rev. E **73**, 021101 (2006).
[34] A. Seebeck, Ann. Phys. Chem. **53**, 417 (1841).