

How we hear what is not there: A neural mechanism for the missing fundamental illusion

Dante R. Chialvo

Department of Physiology, Northwestern University Medical School, Chicago, Illinois 60611

(Received 1 July 2003; accepted 18 August 2003; published 16 October 2003)

How the brain estimates the pitch of a complex sound remains unsolved. Complex sounds are composed of more than one tone. When two tones occur together, a third lower pitched tone is often heard. This is referred to as the “missing fundamental illusion” because the perceived pitch is a frequency (fundamental) for which there is no actual source vibration. This phenomenon exemplifies a larger variety of problems related to how pitch is extracted from complex tones, music and speech, and thus has been extensively used to test theories of pitch perception. A noisy nonlinear process is presented here as a candidate neural mechanism to explain the majority of reported phenomenology and provide specific *quantitative* predictions. The two basic premises of this model are as follows: (I) The individual tones composing the complex tones *add linearly producing peaks of constructive interference* whose amplitude is always insufficient to fire the neuron (II): The spike threshold is reached only with noise, which naturally selects the maximum constructive interferences. The spacing of these maxima, and consequently the spikes, occurs at a rate identical to the perceived pitch for the complex tone. Comparison with psychophysical and physiological data reveals a remarkable quantitative agreement not dependent on adjustable parameters. In addition, results from numerical simulations across different models are consistent, suggesting relevance to other sensory modalities. © 2003 American Institute of Physics. [DOI: 10.1063/1.1617771]

When two tones occur together, such as in a complex sound, a third lower pitched tone is often heard. This is referred to as the “missing fundamental illusion” because the perceived pitch is a frequency (fundamental) for which there is no actual source vibration. How the brain estimates the pitch is a controversial issue which remains unsolved despite extensive work. A nonlinear model is proposed here in which a noisy detection produces neural spikes, occurring at a rate identical to the perceived pitch for the complex tone. The comparison with psychophysical and physiological data reveals a remarkable quantitative agreement not dependent on adjustable parameters. In addition, results from numerical simulations across different models are consistent, suggesting relevance to other sensory modalities.

following premises. First, interference between intervening pure tones produces peaks of larger amplitude at the time points of maximum constructive interference. Second, neural spikes can occur selectively at these peaks if there is noise in the process. Noise in the firing dynamics of primary afferent auditory neurons is a realistic assumption supported by previously reported experimental evidence.^{6,7} This paper demonstrates precise coincidence between the peaks of maximum interference and the pitch that humans report in comparable psychophysical experiments. Furthermore, the predominant interspike intervals in cat auditory nerve discharges in response to complex sounds are consistent with this model. In the following section an illustrative example of the basic problem is setup and the model introduced.

I. INTRODUCTION

Pitch is a subjective attribute of a sound by which we give it a position on a scale from low to high. There is no objective caliper for pitch, but humans can discriminate two sounds with very similar pitches. Our appreciation of music and speech's intelligibility depend on exquisite pitch discrimination. However, its neural mechanism is contentious (for reviews see Refs. 1–3). Neither proponents of peripheral mechanisms^{4–7} nor defendants of cortical processes^{8,9} have gathered enough quantitative experimental evidence to confirm their theories. In this paper, the author proposes a simple and quantifiable, physiologically plausible neural mechanism able to extract values which fit perceived pitch reported by human listeners. The proposed mechanism is based on the

II. COMPLEX SOUNDS AND THE MISSING FUNDAMENTAL ILLUSION

Consider the simple case of a complex tone composed of two pure tones of equal amplitudes and frequencies multiple of a third one. A pure tone is one that is described by a sinusoid having a constant amplitude and single frequency. Complex tones are constructed by adding pure tones of frequencies $f_1 = kf_0$, $f_2 = (k+1)f_0 \dots f_n = (k+n-1)f_0$. Figure 1(A) shows an example of two frequencies producing a “missing fundamental illusion,”^{10,11} so-called because the perceived pitch corresponds to a third frequency f_0 for which there is no actual vibration. Despite the absence of energy at f_0 the harmonic tone complex exhibits large amplitude peaks [asterisks in Fig. 1(A)] spaced at intervals equal to the inverse of f_0 . These peaks are interpreted here as the result of a constructive interference between the constituting tones.

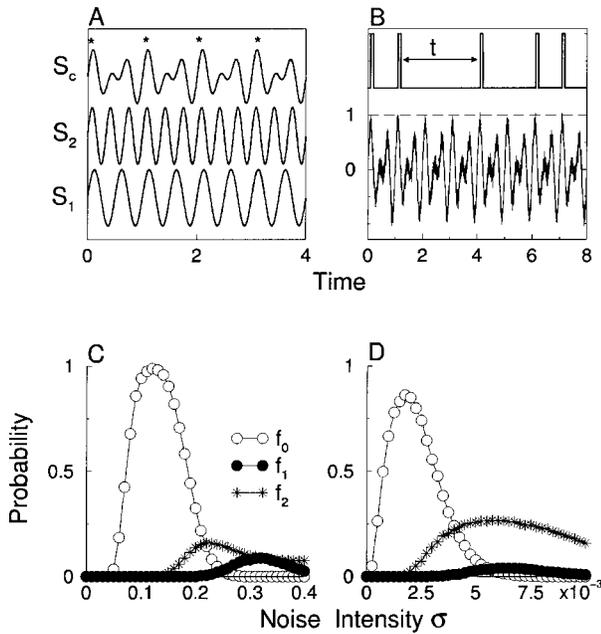


FIG. 1. Robust nonlinear stochastic detection of the missing fundamental f_0 . (A) Example of a complex sound S_c built by adding two sinusoidal S_1 and S_2 frequencies: $f_1 = kf_0$ (bottom) and $f_2 = (k+1)f_0$ (middle). Specifically $x(t) = (A_1 \sin(f_1 2\pi t) + A_2 \sin(f_2 2\pi t))/2$, with $A_1 = A_2 = 1$; $k=2$ and $f_0=1$ in this case. The peaks (asterisks) exhibited by S_c result from the constructive interference between S_1 and S_2 at the period of the missing fundamental f_0 , a frequency that is absent. (B) The noisy neuron model (Ref. 16): The peaks of the signal shown in (A) can be detected reliably by a noisy threshold, generating interspike intervals (“ t ”) close to (or to integer multiples of) the fundamental period. (C) and (D) The most probable interspike interval corresponds with the missing fundamental. Simulations were conducted with varying amounts of noise intensity added to S_c and the interspike intervals [“ t ” in (B)] were computed. From those data, the ratio of all spikes with time “ t ” equal ($\pm 5\%$ tolerance) to the period of the missing fundamental f_0 , was computed and plotted (open circles) as a function of noise intensity. The same calculation was done to determine the spike ratios spaced at the period of S_1 (filled circles, f_1) and with those with the period of S_2 (stars, f_2). Thus the plot shows the probability of observing a spike with an instantaneous frequency similar to f_0 , f_1 , and f_2 . In both neuron models the largest probability corresponds to interspike intervals close to the missing fundamental period (i.e., intervals $\sim 1/f_0$). (C) and (D) show results from simulations of the simple neuron model (Ref. 16) and the FitzHugh–Nagumo model (Ref. 17), respectively. Results from simulations using the Hodgkin and Huxley neuron model (Ref. 18) (not shown) are analogous to the ones illustrated here. The FitzHugh–Nagumo equations model generic excitable dynamics seen in biological and physical systems. The equations are $\varepsilon dv/dt = v(v-0.5)(1-v) - w + I + x(t) + \xi(t)$ and $dw/dt = v - w - b$. The fast variable v corresponds to the membrane voltage, and w is the variable associated with the slow refractoriness, $x(t)$ is the external stimulus and b , ε , and I are model parameters. The term $\xi(t)$ is a zero-mean Gaussian white noise of intensity σ and correlation time $t_c = 10^{-3}$ (Ref. 17). Guided by previous results, these simulations employed the parameter values $b=0.15$, $\varepsilon=0.005$, $I=0.04$ for the model and $A_1 = A_2 = 0.04$ and $f_0 = 0.125$ for the input signal $x(t)$.

This process was used first by Young in his explanation of Newton rings¹² more than a century ago. Amplitude builds up at intervals close to the least common multiple of the interfering periodicities (or in terms of frequencies to the greatest common divisor). This observation predicts that the main features of the interference will be so regardless of the relative phases of the individual tones or the number of harmonics, which could change the shape but not the spacing of the largest peaks. This prediction holds also for tones of dissimilar but sufficient amplitudes.

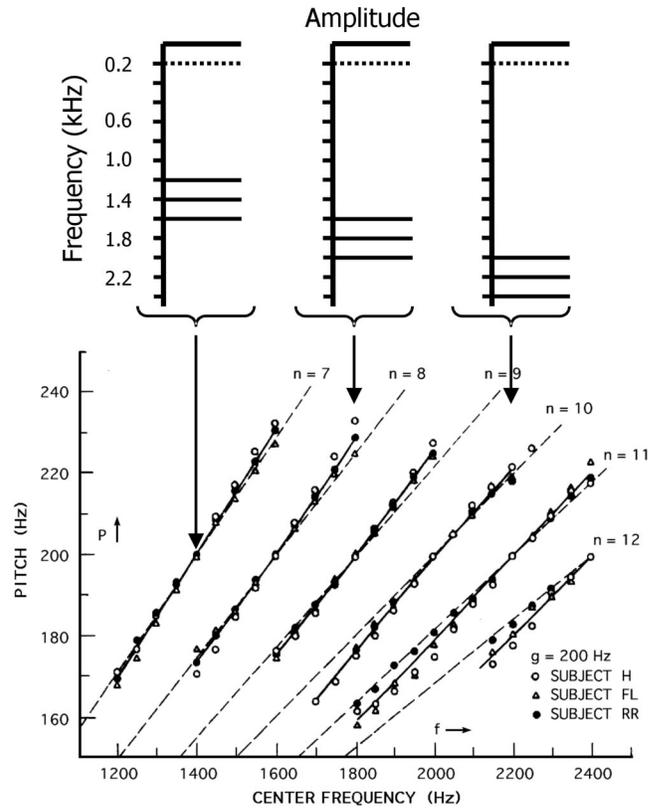


FIG. 2. Results from Schouten’s experiments (modified from Ref. 11) demonstrate that equidistant tones do not produce constant pitch. The authors used sounds composed of three equispaced frequencies with center values in the range of 1.2 to 2.4 kHz. Top diagrams depict three examples of the line amplitude spectra of the complex sounds used (with center frequencies of 1.4, 1.8, and 2.2 kHz, dotted lines correspond to the missing component f_0). The bottom graph indicates with symbols (open or filled circles and triangles) the pitch heard by the three subjects for each complex sound. The dashed lines show that a $1/n$ function consistently underestimates the linear relationship between frequency and pitch shift (in the original figure $g = 200$ Hz corresponds to “ f_0 ” and “ n ” is equivalent to “ $k+1$ ” in our notation).

This simple model proposed as a physiologically plausible mechanism to detect these peaks assumes the following.

- (1) The complex tone S_c (or some transformation of) is the input to a neuron which produces a membrane potential excursion that, because of its *low amplitude*, cannot fire a spike.
- (2) Noise added to S_c induces spikes preferentially at the interference peaks.

These assumptions imply that the single neuron “detects” only statistically the occurrence of peaks, as not all signal peaks produce neuronal spikes [Fig. 1(B)]. This neuronal subthreshold character of the deterministic signal S_c is not to be confused with the perceptual suprathreshold character of the sound, because the average over many neurons in parallel makes perfectly possible the perception of this, otherwise unperceptible, sound. Peak detection is optimized at some noise intensity.^{13–15} Across a relatively broad range of noise intensity, the majority of neural spikes are spaced at intervals corresponding to the missing fundamental frequency f_0 . The proportions of spikes spaced by a time equal

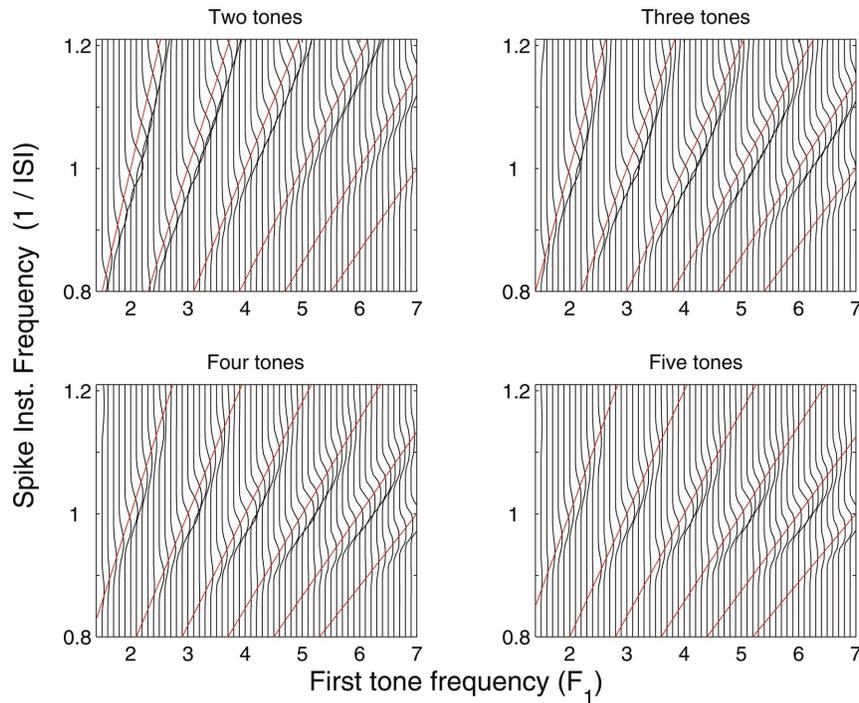


FIG. 3. (Color) Model responses to mistuned complex tones: Signals composed of multiple (two to five) sinusoidal tones were used as inputs to drive the simple neuron model receiving additional noise. As the frequency of the equidistant input tones increased the neural spiking rate favors values along specific lines resembling Schouten’s pitch shift (Ref. 11) and predicted by our theory. Shown are interspike interval histograms (or probability density functions) of the model spikes. Similar to Fig. 2, the vertical axis is the instantaneous frequency of the model spikes (1/interspike intervals, which should correlate with pitch) and the horizontal axis is the first driving frequency f_1 for which the computation was done. Superimposed straight lines are the theoretical expected resonance [from Eq. (3)], which show remarkable agreement with the experimental data. Because pitch is usually reported in frequency units, the interspike intervals are transformed here into instantaneous frequency to facilitate the comparison with experimental data.

to $1/f_0$ ($\pm 5\%$ tolerance) as a function of noise intensity (open circles) are shown with the ratio of spike times at intervals corresponding to the input frequencies f_1 (closed circles) and f_2 (stars) in Fig. 1(C). Note that relatively high noise levels are needed to evoke spikes at rates similar to the actual input frequencies. Similar computations using the more elaborate FitzHugh–Nagumo neuron (FHN) model,¹⁷ produce identical results [Fig. 1(D)]. The results in Figs. 1(C) and 1(D) resemble those described for the stochastic resonance phenomenon,^{19–21} a process where the output of some nonlinear system becomes more coherent with the input signal for some value of the noise intensity. However, in the missing fundamental case, there is an optimum noise intensity for which the system emits the majority of spikes at a frequency that is *not* in the input. Analogous to psychoacoustic experiments the neuron’s strongest resonance occurs at the missing frequency which is *not* in the input, a “ghost” stochastic resonance.^{14,15}

As shown, a noisy nonlinear process, static (i.e., simple threshold model) or dynamical (i.e., FHN model), suffices to extract the missing fundamental frequency. However, these results are only representative of harmonic tones, i.e., when f_1 and f_2 are both integer multiples of the missing fundamental. In the following, the more general case of inharmonic tones is considered.

III. COMPLEX INHARMONIC SOUNDS

As f_0 is not only the greatest common divider of f_1 and f_2 , but is also their difference, it was long assumed that the

brain extracted it by a computation of this difference $f_1 - f_2$. If so, all complex sounds should produce a constant pitch sensation equal to the difference Δf of the constituting tones. To the contrary, as shown in Fig. 2, Schouten *et al.*¹¹ demonstrated that when the tone frequencies of a complex sound composed of three equispaced tones are shifted linearly, subjects report linear shifts in perceived pitch. Therefore, the brain must be computing something other than the difference Δf .

The experimental conditions of Schouten *et al.*¹¹ can be replicated in the context of the present proposal by redefining the intervening frequencies of the neuron model input $x(t)$ as

$$f_1 = kf_0 + \Delta f, \quad f_2 = (k+1)f_0 + \Delta f, \\ \dots f_n = (k+n-1)f_0 + \Delta f. \quad (1)$$

Numerical simulations^{13–15} showed that, similar to the case shown in Fig. 1, a great majority of the neuron spikes occurred (“resonated”) at particular sets of intervals. More importantly, when the input frequency was shifted, the location of this resonance shifted linearly as $f_p = f_0 + \Delta f / (k + 1/2)$ for $n=2$ and $f_p = f_0 + \Delta f / (k+1)$ for $n=3$. The prediction for the neuronal response to complex stimuli composed of an arbitrary number N of frequencies can be generalized,

$$kf_0 + \Delta f, \quad (k+1)f_0 + \Delta f, \quad \dots (k+N-1)f_0 + \Delta f \quad (2)$$

for which the resonance will occur at frequencies given by

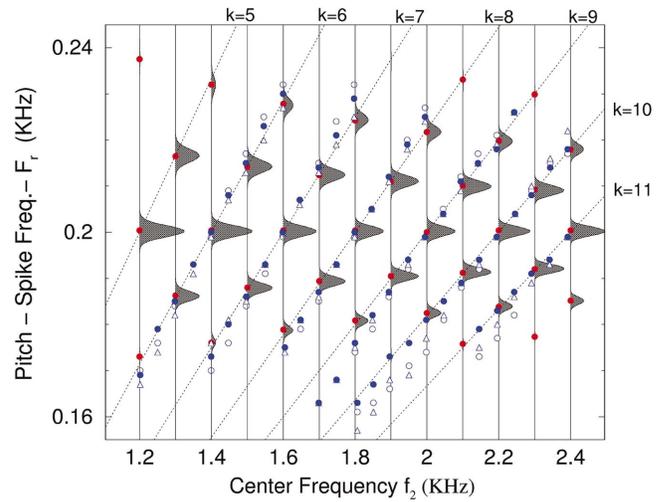
$$f_r = f_0 + \frac{\Delta f}{k + (N-1)/2}. \tag{3}$$

Four representative examples of numerical results using complex tones composed of 2, 3, 4, and 5 sinusoids (noise variance $\sigma = 0.1$, the threshold $U_{th} = 1$ and $f_0 = 1$) are shown in Fig. 3. The graphs represent the observed proportion of spikes with a given instantaneous firing frequency f_r (1/ISI in the ordinate) as a function of the frequency (f_1 , abscissa) of the lowest component of the driving signal. Theoretical predictions of Eq. (3) (lines) superimpose exactly with the simulation results. Equation (3) predicts that for any combination of sinusoids the neural responses will be represented by only one of two diagrams shifted horizontally by some multiple of f_0 . That is, tones composed of even numbers and of odd numbers of sinusoids produce spikes spaced at intervals with statistics similar to the ones in the left panels and those in the right panels, respectively. Responses always fall on lines with slopes of $2/3, 2/5, 2/7$, etc., when input is composed of an even number of tones, and of $1/2, 1/3, 1/4$, etc., when input is composed of an odd number of tones. In these simulations the noise intensity relative to the distance between the signal and the threshold is the relevant parameter. Changing the noise intensity makes more or less evident the resonance [Figs. 1(C) and 1(D)] but does not change its location. Such parameter independence is critical to the relevance of this model.

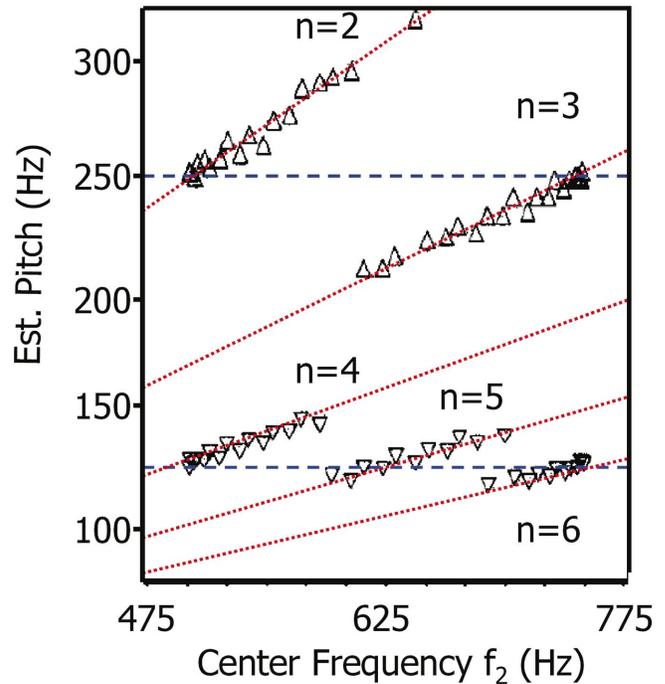
Simulations with parameters analogous to Schouten *et al.*¹¹ were conducted. The results are presented in Fig. 4(a) and superimposed onto the data from the psychoacoustic experiments. Notice the remarkable agreement between the data sets. The model spikes fit the pitch frequencies reported by human subjects. As there are no adjustable parameters in this theory, this fit is not the result of any adjustments to enhance agreement. Like psychoacoustic pitch perception, the predictions from Eq. (3) (dotted lines) are also best at low frequencies. Moreover, there is a good agreement between the theory-based simulation data and the physiological data points [Fig. 4(b)] of pitches estimated from the predominant interspike interval in the discharge patterns of cat auditory nerve fibers in response to complex tones.⁷

IV. DISCUSSION

Here we have discussed the model with respect to the quantitative aspects of the missing fundamental and pitch shift phenomena. However, other features of these phenomena are also consistent with this model. Multiple percepts can be elicited by one complex sound (e.g., multiple data points per center frequency as in Fig. 2). This ‘‘ambiguity’’ of pitch may be explained by the stochastic character of this proposed detection mechanism. There is always a more *probable* spike frequency and other less frequent spike rates [see histograms in Fig. 4(a)]. Pitch strength is then proportional to spike density at that rate. Another noteworthy qualitative aspect is the region of parameter (frequency and pitch) space in which the model spikes occur in response to complex tones. The slope of the lines where the neuron spikes fall decrease



(a)



(b)

FIG. 4. (Color) (a) There is remarkable agreement of the theoretical lines and numerical results from the model with data from psychophysical experiments. Schouten’s experimental pitch results (Ref. 11) (same symbols as in Fig. 2) are superimposed onto the neuronal interspike interval histograms (Spike Freq) obtained from similar simulations as presented in Fig. 3. The theoretical expectations (F_r) from Eq. (3) (with $N=3$) are plotted with dotted lines. The agreement between data and theory is parameter independent. (b) The theoretical expectation is superimposed here onto the experimental results of Cariani and Delgutte (1996) (Ref. 7). The physiological pitches estimated from the highest peak of the interspike interval distribution in response to two variable (500–750 Hz) carrier AM tones with modulation frequencies (horizontal dashed lines) of 125 Hz (downward triangles) or 250 Hz (upward triangles) fit the predictions of Eq. (3) (the diagonal dotted lines; $N=3$).

as the frequencies of the input tones increase [see Fig. 3 and Eq. (3)], limiting the range (of frequencies and pitches) where the phenomena of low pitch can be detected by the proposed neural mechanism. These boundaries correspond

well with those of the psychoacoustic (“existence”) area where complex tones can be perceived as having a definitive pitch.¹⁰

This theory can further accommodate the situation in which two tones enter through different ears. In this case, a (weak) low pitch percept also exists suggesting a central neural substrate for this binaural phenomenon. The spike trains from each cochlear nerve could add and interfere, and subsequently be detected by the noisy threshold of some other neuron. The only substantial difference is that spikes, instead of sinusoids, are added but all other dynamics and quantitative aspects remain equivalent.

The emphasis here is that a neuron can judge pitch by a simple nonlinear noisy transformation in the time domain which mimics the autocorrelation function of the input. The assumption that the pitch information is in the temporal envelope of the signal is well in line with old ideas. However, its simplicity and quantitative predictions contrast with previous proposals suggesting complicated mechanisms mediated by relatively sophisticated structures not as yet identified including delay lines;⁴ oscillators in combination with integration circuits,²² neural networks,⁹ and timing nets,² as well as other,^{23,24} more abstract alternatives.

In conclusion, very important qualitative and quantitative aspects of the missing fundamental problem and of complex tones pitch perception can be understood by considering the linear interference of individual tones being nonlinearly detected by a noisy threshold. In this view, the perceived pitch is associated with the response patterns of generic excitable systems. A general expression relating mistuning and pitch shift agrees extremely well with psychophysical and physiological data. These theoretical results are largely model-independent, and exactly reproduce without fitting parameters the paradigmatic pitch-shift findings of Schouten *et al.*,¹¹ as well as the Cariani and Delgutte physiological data.⁷ Finally, this argument further suggests that similar dynamics should be observed in other neural and sensory scenarios.^{25,26}

ACKNOWLEDGMENTS

Thanks to Dr. Oreste Piro who inspired a great deal of the initial work. A Visiting Professorship from the Physics Department of Universitat de les Illes Balears, Mallorca, Spain (summer of 2003) and the support from the Veterans Administrations GLAHS and the NIH are also recognized.

¹S. Greenberg, J.T. Marsh, W.S. Brown, and J.C. Smith, “Neural temporal coding of low pitch. I. Human frequency following responses to complex tones,” *Hear. Res.* **25**, 91–114 (1987).

²P.A. Cariani, “Temporal coding of periodicity pitch in the auditory system: An overview,” *Neural Plast.* **6**, 147–172 (1999).

³M.J. Tramo, P.A. Cariani, B. Delgutte, and L.D. Braida, “Neurobiological foundations for the theory of harmony in western tonal music,” *Ann. N.Y. Acad. Sci.* **930**, 92–116 (2001).

⁴J.C.R. Licklider, “A duplex theory of pitch perception,” *Experientia* **7**, 128–134 (1951).

⁵R. Meddis and J. Hewitt, “Virtual pitch and phase sensitivity of a computer model of the auditory periphery I: Pitch identification,” *J. Acoust. Soc. Am.* **89**, 2866–2882 (1991).

⁶P.A. Cariani and B. Delgutte, “Neural correlates of the pitch of complex tones. I. Pitch and pitch salience,” *J. Neurophysiol.* **76**, 1698–1716 (1996).

⁷P.A. Cariani and B. Delgutte, “Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch,” *J. Neurophysiol.* **76**, 1717–1734 (1996).

⁸J. Goldstein, “An optimum processor theory for the central formation of pitch of complex tones,” *J. Acoust. Soc. Am.* **54**, 1496–1516 (1973).

⁹M. Cohen, S. Grossberg, and L. Wyse, “A spectral network model of pitch perception,” *J. Acoust. Soc. Am.* **98**, 862–879 (1995).

¹⁰E. de Boer, “On the ‘residue’ and auditory pitch perception,” in *Handbook of Sensory Physiology*, edited by W.D. Keidel and W.D. Neff (Springer-Verlag, Berlin, 1976), Vol. 17, pp. 479–583.

¹¹J.F. Schouten, R.J. Ritsma, and B.L. Cardozo, “Pitch of the residue,” *J. Acoust. Soc. Am.* **34**, 1418–1424 (1962).

¹²G. Gamow, *The Great Physicists from Galileo to Einstein* (Dover, New York, 1961).

¹³D.R. Chialvo, O. Calvo, D.L. Gonzalez, O. Piro, and G.V. Savino, “Subharmonic resonance and synchronization in neuronal systems,” *Phys. Rev. E* **65**, 050902 (2002).

¹⁴O. Calvo and D.R. Chialvo, “Ghost stochastic resonance on an electronic circuit,” *Electronics Letters* (submitted).

¹⁵J.M. Buldu, D.R. Chialvo, C.R. Mirasso, M.C. Torrent, and J. Garcia-Ojalvo, “Ghost resonance in a semiconductor laser with optical feedback,” *Europhys. Lett.* (in press).

¹⁶Z. Gingl, L.B. Kiss, and F. Moss, “Non-dynamical stochastic resonance—Theory and experiments with white and arbitrarily colored noise,” *Europhys. Lett.* **29**, 191–196 (1995).

¹⁷A. Longtin and D.R. Chialvo, “Stochastic and deterministic resonances for excitable systems,” *Phys. Rev. Lett.* **81**, 4012–4016 (1998).

¹⁸A.L. Hodgkin and A. Huxley, “A quantitative description of membrane current and its application to conduction and excitation in nerve,” *J. Physiol. (London)* **117**, 500–544 (1952).

¹⁹K. Wiesenfeld and F. Moss, “Stochastic resonance and the benefits of noise: From ice ages to crayfish and SQUIDS,” *Nature (London)* **373**, 33–36 (1995).

²⁰A. Bulsara and L. Gammaitoni, “Tuning in to noise,” *Phys. Today* **49**(3), 39–47 (1996).

²¹L. Gammaitoni, P. Hanggi, P. Jung, and F. Marchesoni, “Stochastic resonance,” *Rev. Mod. Phys.* **70**, 223–287 (1998).

²²G. Langner, “Neural processing and representation of periodicity pitch,” *Acta Oto-Laryngol., Suppl.* **532**, 68–76 (1997).

²³J.H. Cartwright, D.L. Gonzalez, and O. Piro, “Pitch perception: A dynamical-systems perspective,” *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4855–4859 (2001).

²⁴F. Julicher, D. Andor, and T. Duke, “Physical basis of two-tone interference in hearing,” *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9080–9085 (2001).

²⁵F.A. Kingdom and D.R. Simmons, “The missing-fundamental illusion at isoluminance,” *Perception* **27**, 1451–1460 (1998).

²⁶K. Fujii, S. Kita, T. Matsushima, and Y. Ando, “The missing fundamental phenomenon in temporal vision,” *Psychol. Res.* **64**, 149–154 (2000).