



Regularity of Spike Trains and Harmony Perception in a Model of the Auditory System

Yu. V. Ushakov* and A. A. Dubkov†

Radiophysics Department, Nizhny Novgorod State University, 23 Gagarin Avenue, 603950 Nizhny Novgorod, Russia

B. Spagnolo‡

*Dipartimento di Fisica, Group of Interdisciplinary Physics, Università di Palermo and CNISM,
Viale delle Scienze edificio 18, I-90128 Palermo, Italy*

(Received 13 April 2011; published 2 September 2011)

Spike train regularity of the noisy neural auditory system model under the influence of two sinusoidal signals with different frequencies is investigated. For the increasing ratio m/n of the input signal frequencies (m, n are natural numbers) the linear growth of the regularity is found at the fixed difference ($m - n$). It is shown that the spike train regularity in the model is high for harmonious chords of input tones and low for dissonant ones.

DOI: [10.1103/PhysRevLett.107.108103](https://doi.org/10.1103/PhysRevLett.107.108103)

PACS numbers: 87.19.lt, 02.50.Ga, 05.40.-a, 89.70.-a

A mammal's auditory system represents a very effective sound perception apparatus with high-level characteristics of sensitivity, recognition of speech, musical chords, sound sources location, etc. [1]. As a result, it has attracted scientific attention since the time of Pythagoras until the present [2]. The increased interest in the auditory system is due to stability of its functionality against omnipresent noise of relatively high intensities, especially in neural ensembles of the brain [3]. In spite of many successful studies of the auditory periphery [2,4,5] and the corresponding pathways in the brain [6], some, at first glance, simple psychoacoustic effects of sound perception have not been explained for many years, which indicates an insufficient understanding of the functionality principles of the auditory analyzer, especially its neural part. Such effects are very attractive for physicists, because they seem explainable using relatively simple models as in Refs. [5,7,8]. In addition, achievements in the auditory system investigation provide an understanding of the other, less studied sensory systems, which exhibit the analogous principles of conversion of environment stimuli into the neural spike trains [9].

One of the most difficult open questions in hearing theory [10] relates to perception of very simple musical chords—pairs of pure sinusoidal tones. The question is, why do the chords with commensurable frequencies, referred to as consonant chords, sound pleasant (harmonious, stable), whereas the chords with incommensurable frequencies, referred to as dissonant ones, sound unpleasant (inharmonious, unstable)? It has been shown rather convincingly [11–14] that the behavioral preference of consonant chords is due to some basic principles of neural functionality and is common for birds, rodents, monkeys, and human. However, the question remains in the following form: which characteristics of physical processes in the brain reflect the differences between consonance and dissonance perception? The quantitative description of these differences has been obtained in

psychoacoustic experiments with people [15], in which subjects were asked to judge the “pleasantness” of various sounding chords on a limited scale of natural numbers. Then the scores for each chord were averaged, and in this way the table of “consonance level” values for a set of chords was obtained. In Refs. [6,16], the subjective “consonance level” has been associated with so-called pitch salience. Pitch is a subjective place of a perceived complex sound on the frequency scale. In a simple case, if a sound consists of 2 or 3 pure tones, the pitch is perceived as some additional, virtual tone. In the mentioned papers the pitch salience is defined as a height of a certain peak of a complicated cumulative quantity obtained from a large array of spike trains. At the moment, it is impossible to say that the pitch salience is a clear physical characteristic of some concrete signal. In other words, it provides a very good phenomenological description of consonance and dissonance perception, but hides its underlying physical basis.

In this Letter we try to answer the above question by showing which characteristic of the physical processes involved in brain functioning discriminates between consonance and dissonance. Specifically, we present the results of the consideration of spike train regularity at the output of the auditory system's neural model as a simple parameter of the neural signal showing the difference between the feeling of harmony and disharmony while listening to tone dyads. The regularity is introduced (see below) as the quantity linearly connected with informational entropy due to a natural framework for many problems in biological signal processing provided by information theory (see, e.g., Ref. [17], and references therein). We show that consonant chords influencing the auditory system produce regular spike trains at the system's output, in contrast to dissonant chords, which result in irregular spike trains.

The key element of the cochlea in the inner ear of mammals is the basilar membrane, which performs the

sound Fourier transform with a good precision [1,5]. As a result, different spectral components of the input signal, i.e., different oscillating parts of the basilar membrane, act upon different sensory neurons (sensors). Sensors transform mechanical oscillations into spike trains sent to the brain. Because we restrict our analysis by two spectral harmonics (simple chords of tone pairs), it is sufficient to consider the model with two sensors at the input (see Fig. 1). The sensors N_1, N_2 are subjected to the mixture of subthreshold sinusoidal signals with different frequencies and statistically independent additional white Gaussian noises. The sum of weighted sensors' spike trains summed with the third statistically independent white Gaussian noise is sent to the interneuron N_3 . The output spike train of the interneuron is the main object of investigation. Each neuron is modeled by the simple nonlinear model referred to as the noisy leaky integrate-and-fire neuron [18]. A similar system was used to study the ghost stochastic resonance phenomenon and binaural pitch perception [19]. The detailed description of the model, methods of analysis, and obtained analytical expressions can be found in Ref. [20]. In that paper the hidden Markov behavior [21] of interneuron's spiking has been theoretically revealed, analyzed, and proved by direct numerical simulation of Langevin equations. For each state of the hidden Markov chain (HMC) the first passage time probability density (FPTPD) for the passage of the interneuron's threshold of spike generation has been found. For commensurable input frequencies ($\Omega_1/\Omega_2 = m/n$, where m, n are natural numbers), all FPTPDs consist of peaks, and each peak corresponds to switching into some existing state of the HMC. Thus, the element of the HMC's transition matrix is obtained as follows: $\pi_{ij} = \int_{(i \rightarrow j)} \rho^{(i)}(t) dt$, where $\rho^{(i)}(t)$ is the FPTPD of the interneuron in the i th state, and $(i \rightarrow j)$ is the interval, in which the peak of $\rho^{(i)}(t)$, corresponding to switching into a state j , is situated.

In the paper [20] the average FPTPD—the interspike interval distribution (ISID) of a spike train at the output of N_3 was studied. It has been found that for small numerator

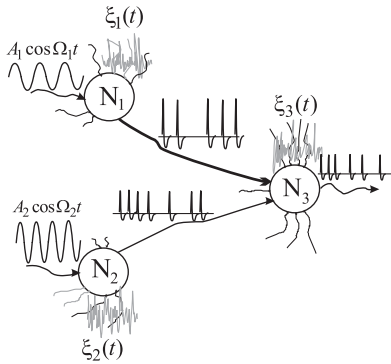


FIG. 1. The investigated model. N_1, N_2 are the sensors, N_3 is the interneuron. $\xi_1(t), \xi_2(t)$, and $\xi_3(t)$ are the statistically independent white Gaussian noises.

and denominator of the frequency ratio m/n , e.g., $2/1, 3/2, 4/3$ (consonant ratios in music), the ISID consists of well-shaped peaks. On the other hand, if m, n are big numbers, e.g., $m/n = 16/15, 45/32$ (dissonant ratios), then the ISID is blurred, which means a less regular spike train than in the previous, consonant case. However, a quantitative estimation of the spike train regularity has not been performed.

With this aim, here, starting from the HMC's transition matrix we calculate the informational entropy H of the interneuron's spike train using the Shannon's formula ([22], I.7)

$$H = - \sum_{i=0}^{M-1} p_i \sum_{j=0}^{M-1} \pi_{ij} \log_2 \pi_{ij}, \quad (1)$$

where p_i is the probability of state i , which can be obtained from the $\{\pi_{ij}\}$ matrix ([22], I.5), and M is the whole number of states of the HMC. As a result, the set of entropy values, for the ordered sequence of m/n ratios (Farey sequence), is found and shown in the inset of Fig. 2.

To characterize the regularity of the spike trains we introduce the spike regularity measure R as

$$R(m/n) = H_{\max} - H(m/n), \quad (2)$$

where H_{\max} is the maximal entropy value over all considered m/n ratios. Obviously, R is defined up to a multiplicative constant, because the minimal R is always zero: $R_{\min} = H_{\max} - H_{\max} = 0$, and the maximal one is the difference between maximal and minimal entropies: $R_{\max} = H_{\max} - H_{\min}$, whereas H_{\max}, H_{\min} are found only for the finite set of m/n ratios and could be different, e.g., for a reduced or extended one. The true values of H_{\max}, H_{\min} are unknown for the system, but they do not influence on the below consideration.

As it is easily seen, the dependence $R(m/n)$ (see Fig. 2) corroborates the hypothesis of the connection between the harmony perception and highly regular spike trains in neural ensembles of the auditory system [20]. Indeed, the regularity R (the entropy H) is high (low) for small natural numbers m, n (namely, $m, n < 10$); i.e., the investigated system produces a regular output spike train under the influence of consonant chords at the input.

In this Letter, we focus on the following property of the $R(m/n)$ dependence: R grows linearly with increasing ratio m/n at fixed difference $(m - n)$ (Fig. 2, bold solid lines). First of all, it is rather unexpected that the informational characteristic of a highly nonlinearly transformed signal demonstrates such a simple linear dependence on the parameters of the input.

It is also interesting to note that this behavior of the regularity is very similar to the well-known first pitch-shift effect [23] in the psychoacoustics: the linear growth of pitch for the linear upward shift of frequencies of sounding tones at a given difference between the frequencies. Indeed, as per studies [7,24], the pitch corresponds to a

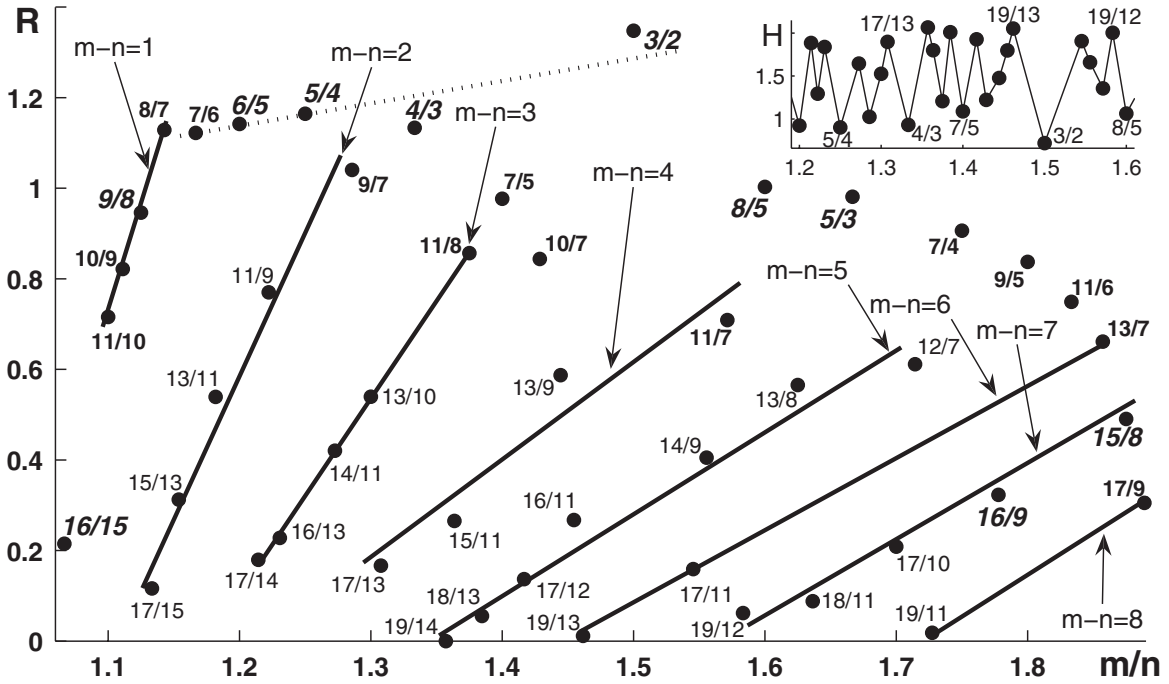


FIG. 2. Regularity of the interneuron's spike train depending on the frequency ratio of input sinusoidal signals $m/n = \Omega_1/\Omega_2$. Near each point there is the exact ratio m/n . The bold solid lines approximate the locus of the $R(m/n)$ points for constant differences ($m - n$). The dashed line is the example of locus change for $m - n = 1$ (see text for clarification). The constant frequency of the second sensor's sinusoidal signal is $\Omega_2 = 0.4$ rad/sec. The first sensor's frequency varies: $\Omega_1 = \frac{m}{n}\Omega_2$. The noise intensity is the same for all three noise sources: $D = 1.6 \times 10^{-3}$. Inset: informational entropy of the interneuron's spike train versus the frequency ratio of input sinusoidal signals m/n .

fundamental frequency of sounding tones, which in the presented case is

$$\omega_p = \Omega_1/m = \Omega_2/n, \quad (3)$$

because $\Omega_1/\Omega_2 = m/n$, i.e., $\Omega_1 = m\omega_p$ and $\Omega_2 = n\omega_p$. Figure 2 is obtained at $\Omega_2 = \text{const}$ and increasing Ω_1 . Thus, for the sake of comparison of R and ω_p , we are interested in the dependence of ω_p on the ratio m/n at the constant difference ($m - n$), which is simply derived from Eq. (3)

$$\omega_p = \frac{\Omega_2}{m-n} \left(\frac{m}{n} - 1 \right). \quad (4)$$

For a set of ($m - n$) values the dependence $\omega_p(m/n)$ shows a fan of straight lines, which, on the face of it, is similar to the $R(m/n)$ behavior. However, these lines are close to those in Fig. 2 only if the correction parameters are introduced into Eq. (4): $\omega_p = \alpha\Omega_2(m/n - \beta)/(m - n)$, where β depends on ($m - n$), so, there is not the direct connection between pitch and regularity. The necessity of α and β can be established from the following simple reasoning. Note, the common period of $\cos\Omega_1 t$ and $\cos\Omega_2 t$ is $2\pi m/\Omega_1 = 2\pi n/\Omega_2$, which is equal to $2\pi/\omega_p$ [see Eq. (3)]. Consequently, the pitch value ω_p is equal to the repetition frequency of the sum $\cos\Omega_1 t + \cos\Omega_2 t$. Of course, the interneuron N_3 sums not sinusoids,

but corresponding spike trains at its input. However, we should just say that a structure of the input spike train of N_3 on average repeats with the frequency ω_p . It is clear, the shorter an averaged repeated structure of the spike train at the input is, the more regular the output spike train will be. At the same time, we understand that not only the length, but also a complexity of this repeated input structure influences the output regularity. Moreover, this complexity depends on the noise intensity and mechanisms of signal transformation by the system, but not on the pitch. That is why the dependence $\omega_p(m/n)$ [Eq. (4)] is unlikely to coincide precisely with the $R(m/n)$ dependence (Fig. 2).

Nevertheless, we see that the relation between pitch and regularity should exist, and the observed qualitative correspondence between the obtained dependence $R(m/n)$ and the dependence $\omega_p(m/n)$, confirmed in experiments [23], proves the feasibility of the model under investigation. In some sense, the regularity embraces both the pitch value (periodicity of a spike train) and the pitch salience (evidence of the periodicity). Thus, the use of the regularity value R as a measure of the "consonance level" may have a number of advantages in comparison with the use of the pitch salience. First, regularity is a clear physical quantity of a concrete spike train. Second, an R value can be obtained directly from a spike train by the calculation of informational entropy

using one of many developed approaches [25]. Third, obtaining a regularity value does not require determination of a pitch value, which is a problem in a case of unknown or too complex input sound, e.g., a voice of a human. We suppose also that an experimental confirmation of the plots shown in Fig. 2 can be very fruitful for a neurophysiological application. For example, the discovery of brain regions, where the property of the spike train regularity shown in Fig. 2 is well expressed for various chords of pure tones, could help in the understanding of how pleasant or unpleasant sounds are perceived by mammals, which are more complex than the musical chords.

Actually, in Fig. 2 there are a number of points, which are noticeably shifted down from the bold lines corresponding to differences ($m - n$) for these points (see, e.g., 7/6, 6/5, 5/4, 4/3, and 3/2). This could be qualitatively explained as follows. As it has been said, the regularity increases due to increasing both of pitch value ω_p and pitch salience. When $m, n > 10$, the pitch salience increases rather quickly, for decreasing m, n , due to the separation of overlapping FPTPD's peaks in each state of HMC. Then, approximately at $m, n \leq 7$, all the peaks become separate [20], so the increase of the pitch salience slows down, whereas the increase of ω_p remains the same. Thus, in the behavior of $R(m/n)$ we observe a slope change. For example, at $m - n = 1$ (Fig. 2), the bold line switches to the dashed one. The analytical description of this phenomenon is in progress.

In summary, based on the theory developed in Ref. [20], we have presented the procedure of informational entropy estimation for the non-Markov spike train at the output of the auditory system model, which can be described by the hidden Markov chain. Based on the entropy dependence on the frequency ratio of input sinusoidal signals, the high or low regularity level of the auditory system's spike trains has been suggested as an indicator of feeling of harmony during sound perception or disharmony, respectively. From the physics viewpoint, in the Letter we have proposed the quantitative description of the distortion of a sinusoidal oscillations sum by the noisy threshold system. It is shown that, even for well-distinguishable frequencies of oscillations, a small shift of one of the frequencies, changing the ratio $\Omega_1/\Omega_2 = m/n$, may result in an appreciable either increase or decrease of regularity of the output impulse sequence (jumps between bold lines in Fig. 2).

This work was supported by the Federal purpose-oriented program "Scientific and Teaching staff of innovative Russia" 2009–2013 (State Contract No. II457) and the Italian Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR).

*yuriy.v.ushakov@gmail.com

†dubkov@rf.unn.ru

‡bernardo.spagnolo@unipa.it

- [1] D. Benson, *Music: A Mathematical Offering* (Cambridge University Press, Cambridge, 2006).
- [2] H.L.F. Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (Dover, New York, 1954).
- [3] H.E. Plesser and S. Tanaka, *Phys. Lett. A* **225**, 228 (1997).
- [4] G. von Békésy and W.A. Rosenblith, in *Handbook of Experimental Psychology*, edited by S.S. Stevens (Wiley & Sons, New York, 1951), pp. 1075–1115.
- [5] V.M. Eguiluz, M. Ospeck, Y. Choe, A.J. Hudspeth, and M.O. Magnasco, *Phys. Rev. Lett.* **84**, 5232 (2000).
- [6] G.M. Bidelman and A. Krishnan, *J. Neurosci.* **29**, 13 165 (2009).
- [7] D.R. Chialvo, O. Calvo, D.L. Gonzalez, O. Piro, and G.V. Savino, *Phys. Rev. E* **65**, 050902(R) (2002).
- [8] Jan-Moritz P. Franosch, Richard Kempter, Hugo Fastl, and J. Leo van Hemmen, *Phys. Rev. Lett.* **90**, 178103 (2003).
- [9] J.G. Nicholls *et al.*, *From Neuron to Brain* (Sinauer Associates, Sunderland, MA, 2001), 4th ed.
- [10] I. Shapira Lots and L. Stone, *J. R. Soc. Interface* **5**, 1429 (2008); B. Heffernan and A. Longtin, *J. Neurosci. Methods* **183**, 95 (2009).
- [11] E.E. Hannon and L.J. Trainor, *Trends Cogn. Sci.* **11**, 466 (2007).
- [12] S. Koelsch *et al.*, *J. Cogn. Neurosci.* **12**, 520 (2000).
- [13] M.J. Tramo *et al.*, *Ann. N.Y. Acad. Sci.* **930**, 92 (2006).
- [14] Y.I. Fishman *et al.*, *J. Neurophysiol.* **86**, 2761 (2001).
- [15] R. Plomp and W.J.M. Levelt, *J. Acoust. Soc. Am.* **38**, 548 (1965); A. Kameoka and M. Kuriyagawa, *J. Acoust. Soc. Am.* **45**, 1451 (1969).
- [16] P. Cariani, in *Proceedings of the 8th International Conference on Music Perception and Cognition, Evanston, IL*, edited by S.D. Lipscomb *et al.* (Causal Productions, Adelaide, Australia, 2004), pp. 310–314.
- [17] A.R. Bulsara and A. Zador, *Phys. Rev. E* **54**, R2185 (1996); A.P. Nikitin, N.G. Stocks, R.P. Morse, and M.D. McDonnell, *Phys. Rev. Lett.* **103**, 138101 (2009).
- [18] A.N. Burkitt, *Biol. Cybern.* **95**, 97 (2006).
- [19] O. Calvo and D.R. Chialvo, *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **16**, 731 (2006); P. Balenzuela and J. Garcia-Ojalvo, *Chaos* **15**, 023903 (2005).
- [20] Y.V. Ushakov, A.A. Dubkov, and B. Spagnolo, *Phys. Rev. E* **81**, 041911 (2010).
- [21] R.L. Stratonovich, *Conditional Markov Processes and Their Application to the Theory of Optimal Control* (Elsevier, New York, 1968).
- [22] C.E. Shannon, *Bell Syst. Tech. J.* **27**, 379 (1948).
- [23] J.F. Schouten *et al.*, *J. Acoust. Soc. Am.* **34**, 1418 (1962).
- [24] J.H.E. Cartwright, D.L. Gonzalez, and O. Piro, *Phys. Rev. Lett.* **82**, 5389 (1999).
- [25] J.D. Victor, *Biol. Theory* **1**, 302 (2006).