Harmony Perception and Regularity of Spike Trains in a Simple Auditory Model

B. Spagnolo[*], Yu.V. Ushakov[†] and A. A. Dubkov[†]

[*]*Dipartimento di Fisica, Group of Interdisciplinary Physics and CNISM, Viale delle Scienze, ed. 18, I-90128 Palermo, Italy*
[†]*Lobachevsky State University, Radiophysics Faculty, 23 Gagarin Ave., 603950 Nizhni Novgorod, Russia*

Abstract. A probabilistic approach for investigating the phenomena of *dissonance* and *consonance* in a simple auditory sensory model, composed by two sensory neurons and one interneuron, is presented. We calculated the interneuron's firing statistics, that is the interspike interval statistics of the spike train at the output of the interneuron, for consonant and dissonant inputs in the presence of additional "noise", representing random signals from other, nearby neurons and from the environment. We find that blurry interspike interval distributions (ISIDs) characterize *dissonant* accords, while quite regular ISIDs characterize *consonant* accords. The informational entropy of the non-Markov spike train at the output of the interneuron and its dependence on the frequency ratio of input sinusoidal signals is estimated. We introduce the regularity of spike train and suggested the high or low regularity level of the auditory system's spike trains as an indicator of feeling of *harmony* during sound perception or *disharmony*, respectively.

INTRODUCTION

The perception and processing of environmental complex signals resulting from the combination of two or more input periodical signals are still an open problem for physicists and physiologists. In particular, the precise neural and physiological bases for our perception of musical consonance and dissonance are still largely unknown [1] – [3]. Although there is no single musical definition, *consonance* is usually referred to as the pleasant stable sound sensation produced by certain combinations of two tones played simultaneously. Conversely, *dissonance* is the unpleasant unstable sound heard with other sound combinations [4]. The dominant and the oldest theory of consonance and dissonance is that of Pythagoras ($570 - 495$ BC). He observed that the simpler the frequency ratio between two tones [1], the more consonant they will be perceived. Example: the consonant octave is characterized by a $1/2$ frequency ratio between two tones, while the dissonant semitone is characterized by a $15/16$ ratio. In 1843 Georg Ohm first proposed that the ear works as a Fourier analyzer [5]. In the same period,

---

[1] *Pure tone* is a single frequency tone with no harmonic components, or *overtones*. *Complex tone* is a combination of the fundamental frequency tone together with its harmonic components. Sounds produced from musical instruments are complex tones.

August Seebeck noticed the "*missing fundamental*" pitch perception [6]: a stimulus with a severely attenuated lowest component is subjectively assigned the same pitch as one with the lowest component at full strength [2]. In this work, after shortly reviewing the first physical theory on *consonance* and *dissonance* of von Helmholtz, two recent theoretical approaches (ghost stocastic resonance and nonlinear synchronization of oscillators), and the *pitch shift effect* related to the fundamental experiment on *virtual pitch* perception, we review our theoretical probabilistic approach to the statistics of *consonance* and *dissonance* musical accords by a simple auditory sensory model.

Helmholtz's theory and pitch perception

In 1877, Helmholtz analyzed the phenomenon of consonance and dissonance in the more general context of complex tones and proposed the "beat theory" [7]. When two complex tones are played together as an interval [3], the harmonics of each tone are present in the stimulus arriving at the ear of the listener. For some combinations (simple ratio n/m) the harmonic frequencies match, for others (complicated ratio n/m) they do not. As the frequency ratio n/m becomes more "complicated", the two tones share fewer common harmonics and there is an increase in harmonics pair slightly mismatched in frequency which give unpleasant beating sensation. In other words, the dissonance is proportional to the number of frequency components present in the two complex tones that produce beats. In Table 1 is shown the ordering of consonances for two tone intervals as accepted in the Western musical culture in decreasing order of "perfection" from most consonant to most dissonant [7]. The third column lists the frequency ratios of the two tones, and the fourth column lists $\Delta\Omega$, the width of the stability interval.

**TABLE 1.** Ordering of consonances for two–tone intervals from most dissonant (down) to most consonant (up)

interval	name	interval ratio	$\Delta\Omega$	Consonance
absolute consonances	unison	1 : 1	0.075	↑
	octave	1 : 2	0.023	↑
perfect consonances	fifth	2 : 3	0.022	↑
	fourth	3 : 4	0.012	↑
medial consonances	major sixth	3 : 5	0.010	↑
	major third	4 : 5	0.010	↑
imperfect consonances	minor third	5 : 6	0.010	↑
	minor sixth	5 : 8	0.007	↑
dissonances	major second	8 : 9	0.006	↑
	major seventh	8 : 15	0.005	↑
	minor seventh	9 : 16	0.003	↑
	minor second	15 : 16	-	↑
				Dissonance

---

[2] *Pitch* is the perceived fundamental frequency of a tone. *Pitch salience* is the strenght of tone sensation.
[3] *Interval* in music theory is the difference in pitch between the fundamental frequencies of two tones.

In the following Fig. 1, three different intervals are shown, namely whole–tone, perfect $5th$, unison (from bottom to top). We can see that the unison matches exactly, and this interval is considered to be the most consonant. Next, the perfect $5th$ shows some matched and some mismatched frequencies. The whole tone shows a mismatch for all frequencies. These frequencies can be close enough together so that discernible beats can result. As a result, the whole–tone interval is more dissonant than the perfect $5th$, which in turn is more dissonant than the octave that is more dissonant than the unison. As one proceeds down the Fig. 1, the number of mismatched harmonics increases and so does the dissonances. In beat theory of Helmholtz therefore, mismatched harmonics are considered the cause of the dissonance. Intervals were consonant if there were no or few beats between the partials. For dissonant intervals, the partials of different tones were so close together in frequency that the beating between them was perceived as dissonance [7].
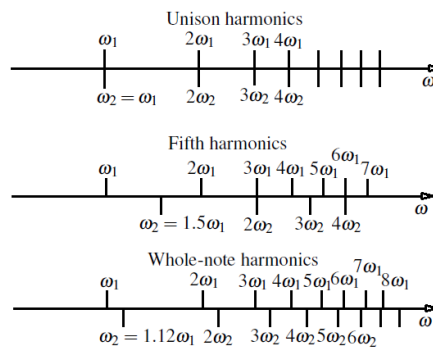
FIGURE 1. Three different intervals, namely whole–tone, perfect $5th$, unison (from bottom to top).

Pitch is a very fundamental concept in music. In fact, music is essentially a variation in loudnesses, pitches, and timbres as a function of time [8]. The official definition of pitch is "that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from high to low" [9]. Pitch is a subjective place of a perceived complex sound on the frequency scale: pitch represents the perceived frequency of a sound. Pitch may be quantified as a frequency, but pitch is not a purely objective physical property, it is a subjective psychoacoustic attribute of sound. A high pitch ($> 2kHz$) will be perceived to be getting higher if its loudness is increased. A low pitch ($< 2kHz$) will be perceived to be going lower with increasing loudness. This is called Stevens's rule [10], the pitch of a pure sinusoidal tone depends not only on its frequency but also on its intensity. Complex tones evoke pitch sensations which are often determined exclusively by overtones. However, how the brain estimates the pitch of complex sounds, formed by a combination of pure tones, remains a controversial issue [11] – [14]. Another important quantity in music is the *pitch salience*, which represents the evidence of the periodicity of some spike train, being the pitch value the periodicity itself. In other words, pitch salience is the probability of noticing a tone, the clarity or strength of tone sensation. The estimated salience, or relative strength, of the strongest pitch of complex tones, that is the maximum salience, is an estimation of the perceived consonance.

A complex tone composed of two sine waves of 900 and 1200 Hz gives rise to three pitches: two spectral pitches at 900 and 1200 Hz, due to the physical frequencies of the pure tones, and the combination tone at 300 Hz, corresponding to the repetition rate of

the waveform. This is the so called *missing fundamental* frequency, which is the greatest common divisor of the frequencies present in the input sound.

Perception of concurrent combinations of tones is central to physiological theories of musical harmony and melody. In fact, perception of consonance in music involves *sensory* and *perceptual* processes that are relatively independent of context, as well as *cognitive* processes depend on musical context [15]. When a harmonic interval is played, neurons throughout the auditory system that are sensitive to one or more frequencies (partials) contained in the interval respond by firing action potentials. For consonant intervals, the fine timing of auditory nerve fiber responses contains strong representations of harmonically related pitches implied by the interval and all or most of the partials can be resolved by finely tuned neurons throughout the auditory system. By contrast, dissonant intervals evoke auditory nerve fiber activity that does not contain strong representations of constituent notes or related bass notes. Moreover, many partials are too close together to be resolved. Consequently, they interfere with one another, cause coarse fluctuations in the firing of peripheral and central auditory neurons, and give rise to perception of roughness and dissonance [11].

It is important to distinguish between musical *consonance/dissonance*: a given sound evaluated within a *musical context*, and psychoacoustic, or *sensory consonance/dissonance*: a given sound evaluated in isolation. *Musical consonance/dissonance* is culturally determined: variation across cultures and historical periods. Judgments of *sensory consonance/dissonance* are culturally invariant and largely independent of musical training, involving basic auditory processing mechanisms. Moreover, rodents, birds, monkeys, and human infants discriminate isolated musical chords on the basis of sensory consonance and dissonance similarly to expert human listeners and experienced musicians [1]. We will consider in this work the *just intonation* [4] musical accords, that is the *sensory consonance/dissonance*.

Ghost Stochastic Resonance

For harmonic complex sound signals, whose constituent frequencies are multiple integers of a fundamental frequency, the perceived pitch is the fundamental, even if that frequency is not spectrally present in the input signal. This is known as *missing fundamental illusion*. Recently, a mechanism for the perception of pitch has been proposed on the basis of the so called *ghost stochastic resonance* (GSR) [16] – [20]. According to the proposed mechanism, a neuron responds optimally to the missing fundamental of a harmonic complex signal for an appropriate level of noise. The main ingredients are: (i) a linear interference between the individual tones, producing peaks of constructive interference at the fundamental frequency (ghost frequency), whose amplitude is not suitable to trigger the neuron; (ii) a nonlinear threshold that detects those peaks with the help of a suitable amount of noise.

In the following Fig. 2 it is shown a complex sound $s_c(t)$ obtained by adding two

---

[4] *The just intonation* tuning is the basic scaling method in which the frequencies of notes are related by ratios of integers.

sinusoidal signals $s_1(t)$ and $s_2(t)$. The constructive interference between $s_1(t)$ and $s_2(t)$ gives rise to the peaks (asterisks in the figure) in $s_c(t)$ at the period of the missing fundamental ω_0. These peaks together with a noise signal can be detected by a nonlinear threshold (see Fig. 2B). In fact, the complex tone s_c is the input to a neuron which produces a membrane potential excursion that, because of its low amplitude, cannot fire a spike. When noise is added to s_c, it induces spikes with high probability at the interference preaks. Moreover, peak detection is optimized at some noise intensity [16, 17].
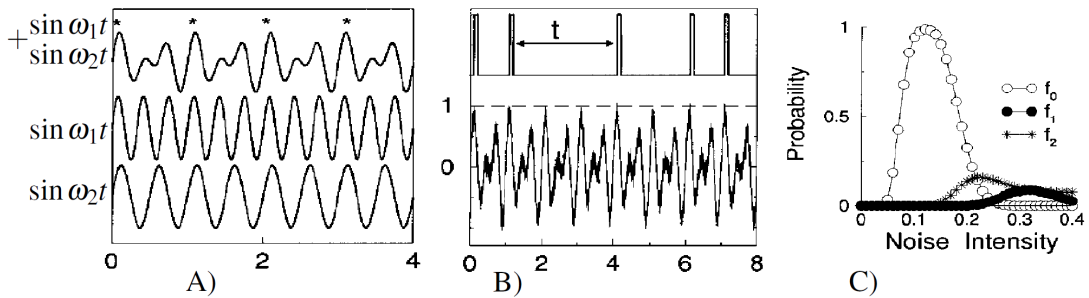
FIGURE 2. A) A complex sound s_c obtained by adding two sinusoidal signals with frequencies $\omega_1 = (k+1)\omega_0$ and $\omega_2 = k\omega_0$, namely $s_c(t) = s_1(t) + s_2(t) = a_1 sin(\omega_1 t) + a_2 sin(\omega_2 t)$. Here $a_1 = a_2 = 1, k = 2, \omega_0 = 1$. The peaks (asterisks) exhibited by s_c result from constructive interference between $s_1(t)$ and $s_2(t)$. B) The peaks of $s_c(t)$ shown in A) can be detected by a nonlinear threshold by adding a noise signal, generating interspike intervals "t" close to, or to integer multiples of, the fundamental period. C) The most probable interspike interval corresponds with the *missing fundamental* (here $f_0 = \omega_0/2\pi, f_1 = \omega_1/2\pi, f_2 = \omega_2/2\pi$).

The GSR mechanism was extended to describe a higher level of perception processing: the binaural pitch perception in Refs. [18, 19]. Two different neurons, at a different auditory channel, receive one single component of the complex signal each, and their output spike trains drive a third neuron that processes the information. This processing neuron responds preferentially at the ghost frequency and the response is optimized by synaptic noise.

Nonlinear synchronization theory of musical consonance

A nonlinear synchronization theory of *consonance* that goes beyond the linear beating theory of Helmholtz was recently proposed in Ref. [2]. This theory is based on the mode locking properties of simple dynamical models of pulse-coupled neurons. The mode locking describes the phenomenon where the frequencies of two oscillators remain in a given ratio for some finite range of parameters. When the oscillators, that is the periodically firing neurons, adjust their frequency to maintain the same ratio, this is a signature of nonlinear synchronization. For example we have one–to–one (1 : 1) mode locking if one neuron fires at a frequency which is synchronized with that of the second neuron. If the first neuron fires only once for every two firing of the second neuron, we have a 1 : 2 mode locking and so on, in general we have $n : m$ mode locking (with n and m integers). By using a simple scheme of two mutually coupled neural oscillators, the

authors show that the mode-locked states ordering gives precisely the standard ordering of *consonance* [3].

The authors of Ref. [2] analyze the dynamics of two coupled leaky integrate-and-fire neuron models, with mutual excitatory coupling, by finding that the mode locking ratios n/m are ordered according to the "Farey sequence", which orders all rational fractions n/m in the interval [0,1] according to their increasing denominators m [21]. By plotting the ratio of actual firing frequencies as a function of the ratio of natural intrinsic frequencies of the two coupled oscillators, they reproduce the so-called "Devil's Staircase", with flat steps corresponding to different mode-locked states. This is a universal feature of driven coupled oscillators [22]. The width of each step, that is of the mode-locked interval, is an indicator of the structural stability of the synchronization. It is therefore possible to order the mode-locked states by their stability index, by finding a correspondence with the theoretical ordering of musical intervals according to their *consonance* evaluation. The steps decrease in width as higher integers occur in their fractional representation of the mode locking (see Fig. 3). Heffernan and Longtin in Ref. [3] analyzed in detail the same model of Ref. [2] by considering different values of coupling between the oscillators. They found that the ordering of mode locked states is not universal, but depends on the coupling strength. Moreover, the noise jitters the spike times and mode locked patterns, but the overall shape of the firing mode lockings is preserved.
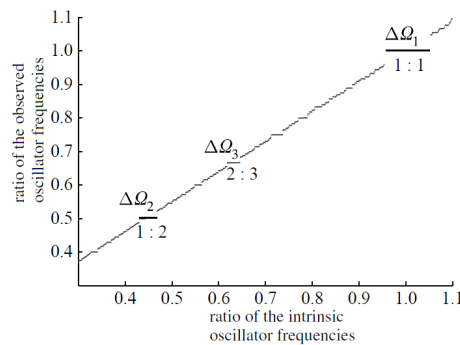

FIGURE 3. The ratio of the observed oscillator frequencies when coupled as a function of the ratio of the oscillator's natural intrinsic frequencies.

Pitch shift effect

Almost all musical sounds are complex tones that consist of a lowest frequency component, or fundamental, together with higher frequency overtones. The fundamental plus the overtones are together called partials. The first perceptual theories considered pitch to arise at a peripheral level in the auditory system [5, 6, 7, 23, 24], while experiments have shown that pitch processing of complex tones is carried out before the primary auditory cortex [25]. The ability of the auditory system to perceive the fundamental frequency of a sound even when this frequency is removed from the stimulus is an interesting phenomenon related to the pitch of complex sounds. This capability is known as "*residue perception*", "*virtual pitch*" or *missing fundamental*, and consists of the perception of a pitch that cannot be mapped to any frequency component of the stimulus

(see Fig. 4). According to the Helmholtz theory, the missing fundamental can be obtained by the difference combination tone between two sources with two frequencies. However, Schouten *et al.* [23] found in their crucial experiment that the behavior of the residue cannot be described by a difference combination tone. By shifting all the partials by the same amount Δf (see Fig. 4c), the complex is no longer harmonic, the difference combination tone remains unchanged, and the same should thus be true of the residue. Instead, it is found that the perceived pitch also shifts, showing a linear dependence on Δf (Fig. 4d). This phenomenon is known as the first *pitch-shift effect*, and has been accurately measured in many psychoacoustic experiments [26]. The fundamental experiment of Ref. [23] was accurately described in terms of generic attractors of nonlinear dynamical systems, by modeling the auditory system as a generic nonlinear forced oscillator [27].
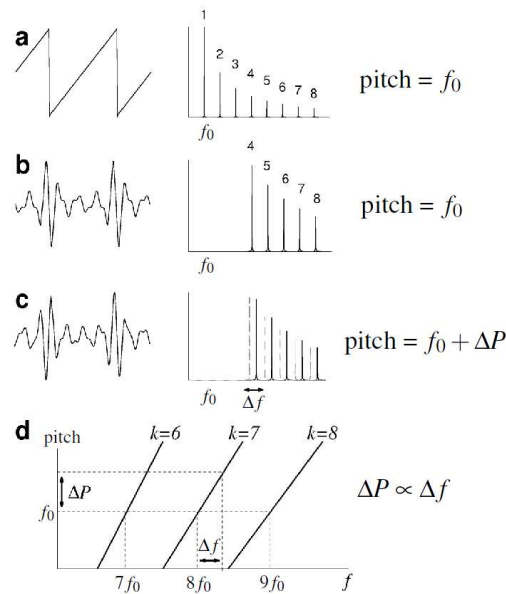

FIGURE 4. (a) A harmonic complex tone. The overtones are successive integer multiples $k = 2, 3, 4, ..$ of the fundamental f_0 that determines the pitch. (b) Another harmonic complex tone with the fundamental and the first few higher harmonics removed. The pitch remains the same and equal to the *missing fundamental*. This pitch is known as *virtual* or *residue pitch*. (c) An anharmonic complex tone, where the partials, which are no longer harmonics, are obtained by a uniform shift Δf of the previous harmonic case (shown as dashed line). Although the difference combination tones between successive partials remain unchanged and equal to the missing fundamental, the pitch shifts by a quantity ΔP that depends linearly on Δf. (d) Pitch as a function of the central frequency $f_c = (k+1)f_0 + \Delta f$ of a three component complex tone, namely $kf_0 + \Delta f$, $(k+1)f_0 + \Delta f$ and $(k+2)f_0 + \Delta f$. This is the *pitch shift effect*, shown here for $k = 6, 7$, and 8 (see Ref. [27]).

PROBABILISTIC APPROACH

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 [28, 29]. 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) , which send their output of spike trains to the interneurons. 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. 5). 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, which is an internal neuron connecting sensory neurons to other neurons within the same region of the brain. The output spike train of the interneuron is the main object of investigation [30, 31].
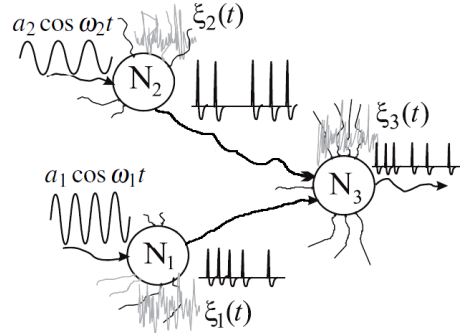
FIGURE 5. The investigated model. N_1 and N_2 are the sensory neurons, driven by subtreshold sinusoidal signals with different frequencies. Spike trains of sensors are received by the interneuron N_3. $\xi_1(t)$, $\xi_2(t)$, and $\xi_3(t)$ are the statistically independent white Gaussian noises.

Each neuron is modeled by the simple nonlinear model referred to as the noisy leaky integrate-and-fire neuron [32]. We analyze the probability distribution of interspike intervals (ISIDs) of the output signal of the interneuron by assuming to know the ISIDs of the output signals of the two sensory neurons $\rho_1(t)$ and $\rho_2(t)$. We reduce the number of events for which the interneuron can fire to four main scenarios because all other events have a very negligible probability to happen in comparison with the previous four. In this way we are able to calculate the first passage time distribution at the output of the interneuron $\rho_3(t)$, using conditional probabilities and first passage time distributions at the output of sensory neurons. Moreover, for periodical input signal at the sensors with frequency ratio m/n we obtain (m+n-1) different patterns of input spike trains for the interneuron, with different ISIDs at its output. The final interspike interval density of the interneuron $\rho_{out}(T)$ is obtained by averaging the first passage time density $\rho_3(t)$ over all different states (m+n-1) of the interneuron. We show how a complex input composed of two harmonic signals is transformed by the proposed simple sensory system into different types of spike trains, depending on the ratio of input frequencies. Looking for the differences in the statistical sense, we find out that the output ISIDs for some combinations of frequencies, corresponding to consonant accords, have more regular pattern, while inharmonious signals, corresponding to dissonant accords, show less regular spike trains and blurry ISIDs. This difference indicates that consonant accords are higher stable, with respect to the noise environment, in comparison with the dissonant accords in the processing of information throughout the auditory system.

Model

As a neuron model for our sensory system (see Fig. 5) we consider the Leaky Integrate-and-Fire (LIF) model. Therefore, the set of stochastic differential equations describing our system is

$$
\begin{cases}
\dot{v}_1 = -\mu_1 v_1 + A_1 \cos(\Omega_1 t) + \sqrt{D_1}\xi_1(t), \\
\dot{v}_2 = -\mu_2 v_2 + A_2 \cos(\Omega_2 t) + \sqrt{D_2}\xi_2(t), \\
\dot{v}_3 = -\mu_3 v_3 + k_1 s_1(t) + k_2 s_2(t) + \sqrt{D_3}\xi_3(t),
\end{cases} \tag{1}
$$

where $v_i(t)$ and μ_i stand for the membrane potential and the relaxation parameter, respectively, and the subscript i labels the different neurons, with $i = 1,2$ representing the two input sensory neurons (N_1 and N_2) and $i = 3$ (N_3) denoting the processing interneuron. A_i and Ω_i (with $i = 1,2$) are the amplitude and the frequency of the corresponding harmonic input of the sensors. We consider that the three neurons have different synaptic connections, they are not subject to the same background noise and the three noise sources $\xi_i(t)$ are independent of each other. Therefore, in Eqs. (1), the three white Gaussian noise terms $\xi_i(t)$ ($i = 1,2,3$) are uncorrelated and with the usual statistical properties $\langle \xi_i(t) \rangle = 0$ and $\langle \xi_i(t)\xi_j(t') \rangle = \delta(t-t')\delta_{ij}$. D_i is the noise intensity in each neuron. In Eq. (1) $s_i(t) = \sum_{j=0}^{N_i(t)} \delta(t - t_{ij}), i = 1,2$ are the spike trains generated by the sensors and received by the interneuron as input, k_i ($i = 1,2$) are the coupling coefficients. Spikes are modelled by Dirac δ-functions. The LIF model doesn't comprise any mechanism of spike generation. When the membrane potential v_i reaches the threshold value v_{th}, the neuron is said to fire a spike, and v_i is reset to its initial value v_i^0. In particular, the input spikes at the interneuron, coming from the sensory neurons, can produce spikes or jumps in the membrane potential of the interneuron, depending on whether or not they are suitable to fire the interneuron.

All simulation and theoretical results presented in this work are obtained using the following set of values of system parameters, namely $\mu_1 = \mu_2 = 1$, $\mu_3 = 0.3665$, $D_1 = D_2 = D_3 = 1.6 \cdot 10^{-3}$, $k_1 = k_2 = 0.98$, $v_1^0 = v_2^0 = 0$, $v_3^0 = -1$, and $v_{th} = 1$, unless stated otherwise. The *refractory period* T_{ref} of the output interneuron is introduced explicitly as the time at which the membrane potential reaches the level $v_3 = -0.1$, that is $T_{ref} = 6.28$. The first two equations of system (1) describe the Ornstein-Uhlenbeck processes with harmonic driving forces. For the Ornstein-Uhlenbeck neuronal model, the ISID was obtained analytically with different approaches in Refs. [33, 34]. This distribution, which coincides with the first passage time probability distribution related to the firing event of sensory neurons, is our starting point to obtain the ISID at the output of the interneuron.

It is important to note here that the ISIDs at the output of two sensors are non-Poissonian (see Fig. 6b). These spike trains are the input of the third neuron, and as a consequence the dynamics of the membrane potential of the interneuron is non-Markovian. The output of the interneuron is shown in Fig. 6c. In order to perform this analysis we use three main assumptions: (i) The input harmonic signals are *subthreshold* for the sensors, that is the signal $A_i \cos(\Omega_i t)$ is not able to bring the membrane potential of the ith sensor above the threshold in the absence of noise ($D_i = 0$). This means absence of spikes at the output of the sensors. (ii) Only one spike can be generated at each period

of the harmonic driving force, and, at the same time, the spiking on each period is the most probable situation (see Fig. 6a). This means that the relaxation times of sensors are smaller than the periods of the sinusoidal signals. (iii) Each of coupling coefficients k_i is less than the threshold value of the membrane potential v_{th}. It means that any separate incoming spike (see Fig. 6c) evokes a subthreshold impulse of the membrane potential of the interneuron $v(t)$, i.e. spike generation is impossible without noise. At the same time, the sum of the two coupling coefficients is greater than v_{th}.
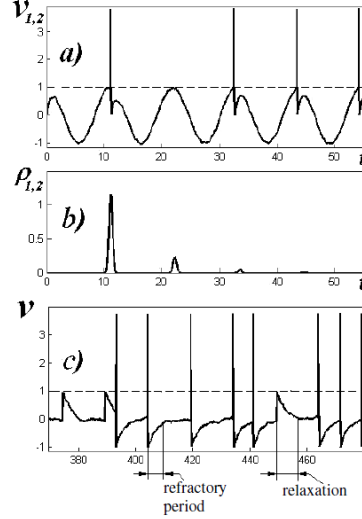
FIGURE 6. (a) Typical behavior of the membrane potential $v_i(t)$ of sensory neurons versus time for a noise realization. (b) ISI distribution of the sensory neurons. The highest probability of a spike after $t = 0$ is near one period of external force ($t = 10.47$). The probability of firing after two, three, etc. periods decreases exponentially. (c) Typical behavior of the membrane potential $v_3(t)$ of the interneuron versus time for the same noise realization. Here are well visible the refractory state (*ref*), characterized by the refractory time T_{ref}, and the noisy background (*bg*) during the relaxation time T_{relax}.

Therefore, we can evaluate the probability $\Delta P_3(t) = \rho_3(t)\Delta t$ that the interneuron N_3 fires in the short time interval $(t, t + \Delta t)$, by considering the occurrence of the following events:

1. receiving a *separate firing spike* from the sensory neuron N_1;
2. receiving a *separate firing spike* from the sensory neuron N_2;
3. receiving a firing spike from the neuron N_1 on the *background* of the membrane potential relaxing, after the jump due to the spike from the N_2 neuron, towards the zero value; in other words, sensor N_2 causes the jump and then sensor N_1 the spike;
4. receiving a firing spike from the neuron N_2 on the *background* of the membrane potential relaxing, after the jump due to the spike from the N_1 neuron, towards the zero value; in other words sensor N_1 causes the jump and then sensor N_2 the spike.

We neglect the contribution of multiple jump events to fire the interneuron and the noise-induced spike events occurring during the relaxation of the membrane potential after a jump, because they have very negligible probability to happen in comparison with the previous four, with the chosen range of system parameters. The four described scenarios exclude each other, so they are mutually exclusive events. As a result, accord-

ing to the formula of total probability we have to add up all probabilities of the above mentioned events.

Interspike interval distributions

Now we calculate numerically the interspike interval distributions of the interneuron for two groups of *consonant* and *dissonant* accords by numerical simulations of Eqs. (1) (see Figs. 7 and Fig. 8).
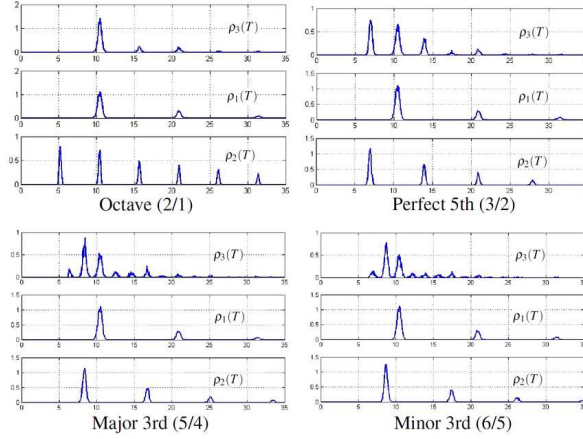
FIGURE 7. ISI distributions of the *consonant* accords: octave $(2/1)$, perfect $5th$ $(3/2)$, major $3rd$ $(5/4)$, and minor $3rd$ $(6/5)$. All curves are obtained through the direct numerical simulation of the Eqs. (1).

We note the very regular behavior of the patterns of $\rho_{out}(T)$ in all the *consonant* accords considered, and the very rich pattern with many peaks in the major $3rd$ $(5/4)$ and minor $3rd$ $(6/5)$ accords.
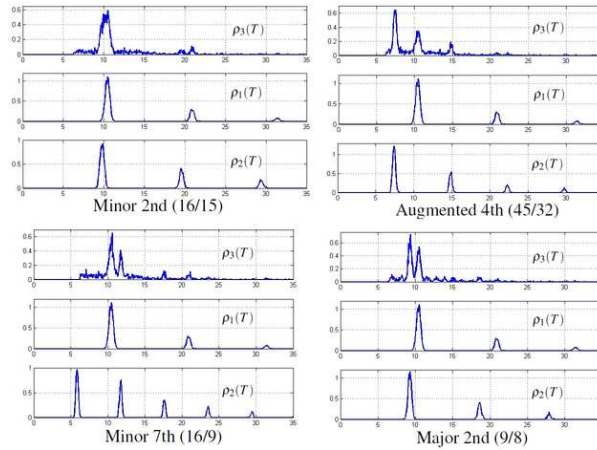
FIGURE 8. ISI distributions of the *dissonant* accords: major $2nd$ $(9/8)$, minor $7th$ $(16/9)$, minor $2rd$ $(16/15)$, and augmented $4th$ $(45/32)$. All curves are obtained through the direct numerical simulation of the Eqs. (1).

The ISIDs of *dissonant* accords are blurry with respect to the ISIDs of the *consonant* accords. This means that we can consider the ISID as an investigative tool to discriminate between *consonant* and *dissonant* accords. In fact higher are the integers m, n less regular and blurry are ISIDs, while lower are the integers more regular are the ISIDs.

REGULARITY

Each different state of the interneuron belongs to a hidden Markov chain (HMC). For each state of the HMC we are able to calculate the First Passage Time Probability Density (FPTPD) for the passage of the interneuron's threshold of spike generation by the theoretical approach presented in the previous section (see Ref. [30] for details). For input frequencies with ratio $(\Omega_1/\Omega_2 = m/n)$, 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: $\pi_{ij} = \int\limits_{(i \to j)} \rho^{(i)}(t)dt$, where $\rho^{(i)}(t)$ is the FPTPD of the interneuron in the i-th state, and $(i \to j)$ is the interval, in which the peak of $\rho^{(i)}(t)$, corresponding to switching into a state j, is situated.

Starting from the HMC's transition matrix we calculate the specific informational entropy H of the interneuron's spike train using the Shannon's formula [31]

$$H = -\sum_{i=0}^{M-1} p_i \sum_{j=0}^{M-1} \pi_{ij} \log_2 \pi_{ij}, \tag{2}$$

where p_i is the probability of state i, which can be obtained from the $\{\pi_{ij}\}$ matrix, and M is the whole number of states of the HMC.

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

$$R(m/n) = H_{max} - H(m/n), \tag{3}$$

where H_{max} is the maximal entropy value over all considered m/n ratios. Obviously, R is defined up to a multiplicative constant [31]. In Fig. 9 the dependence $R(m/n)$ corroborates the hypothesis of the connection between the harmony perception and highly regular spike trains in neural ensembles of the auditory system [30]. Indeed, the regularity R (the entropy H) is high (low) for small integers m, n (namely, $m, n < 10$), i.e. the investigated system produces a regular output spike train under influence of consonant accords at the input. R grows linearly with increasing ratio m/n at fixed difference $(m - n)$ (Fig. 9, bold solid lines).

We 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. Therefore, because the pitch is a proxy of the regularity, the observed qualitative correspondence between the obtained dependence $R(m/n)$ and the dependence $f_p(m/n)$, confirmed in experiments [23] (see also Fig. 4), 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
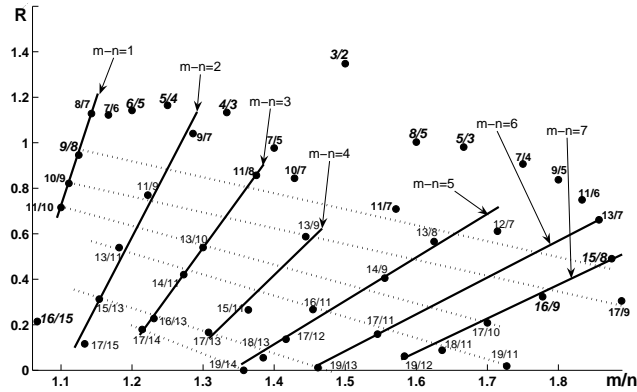
FIGURE 9. Regularity of the interneuron's spike train depending on the frequency ratio of input sinusoidal signals $m/n = \Omega_1/\Omega_2$. The bold solid lines approximate the locus of the $R(m/n)$ points for constant differences $(m-n)$. The noise intensity is the same for all three noise sources.

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. Firstly, regularity is a clear physical quantity of a concrete spike train. Secondly, an R value can be obtained directly from a spike train by calculation of specific informational entropy [31]. Thirdly, obtaining of 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 the Fig. 9 can be very fruitful for neurophysiological applications. For example, the discovery of brain regions where property of the spike train regularity could help to understand how pleasant or unpleasant are perceived by a mammal sounds, which are more complex than simple musical accords.

CONCLUSIONS

With our simple model of the auditory system, we are able to discriminate between *consonant* and *dissonant* accords by analyzing the first passage time probability distributions at the output of the interneuron. Blurry ISIDs characterize *dissonant* accords, while quite regular ISIDs characterize *consonant* accords (Figs. 7, 8). We have calculated the informational entropy for the non-Markov spike train at the output of the auditory system model, and introduced the regularity of spike train. The high or low regularity level of auditory system's spike trains has been suggested as an indicator of feeling of harmony during sound perception or disharmony, respectively. By considering an extension of this simple model to a more complex realistic auditory system, composed of many sensory neurons and different layers, we should be able to know at which extent the *dissonant* accords will "survive", against the *consonant* ones, in the noisy neural environment of the brain.

ACKNOWLEDGMENTS

REFERENCES

1. Y. I. Fishman *et al.*, *J. Neurophysiol* **86**, 2761–2788 (2001).
2. I. Shapira Lots and Lewi Stone, *J. R. Soc. Interface* **5**, 1429–1434 (2008).
3. B. Heffernan, A. Longtin, *J. Neurosci. Meth.* **183**, 95–106 (2009).
4. R. Plomp and W. J. M. Levelt, *J. Acoust. Soc. Am.* **38**, 548–560 (1965).
5. G. S. Ohm, *Ann. Phys. Chem.* **59**, 513–565 (1843).
6. A. Seebeck, *Ann. Phys. Chem.* **60**, 449–481 (1843).
7. H. L. F. Helmholtz *On the sensations of tone as a physiological basis for the theory of music*, Dover, New York, 1954. Translation of Fourth German edition (1877) by Alexander J. Ellis, originally published in 1885.
8. A. J. M. Houtsma, *Music Perception* **1** (3), 296–307 (1984).
9. *American Standards Association: Acoustic Terminology*, S1.1, New York 1960.
10. S. S. Stevens, *J. Acoust. Soc. Am.* **6**, 150–154 (1935).
11. M. J. Tramo, P. A. Cariani, B. Delgutte, and L. D. Braida, *Ann. N. Y. Acad. Sci.* **930**, 92–116 (2001).
12. C. J. Plack, A. J. Oxenham, *The psychophysics of pitch*, in *Pitch: Neural coding and perception*, C. J. Plack, A. J. Oxenham, R. R. Fay and A. N. Popper eds, New York, Springer-Verlag, 2005, pp. 7Ű-55.
13. A. de Cheveigné, *Pitch perception models*, in *Pitch: Neural coding and perception*, C. J. Plack, A. J. Oxenham, R. R. Fay and A. N. Popper eds, New York, Springer-Verlag, 2005, pp. 169Ű-233.
14. D. Wile, E. Balaban, *PLoS ONE* **2**(4): e369 (2007).
15. P. Cariani, in *Proc. 8th Int. Conf. on Music Perception and Cognition, Evanston, IL*, edited by S. D. Lipscomb *et al.* (Causal Productions, Adelaide, Australia), 2004 pp. 310–314.
16. D. R. Chialvo, O. Calvo, D. L. Gonzalez, O. Piro, and G. V. Savino, *Phys. Rev. E* **65**, 050902(R)(4) (2002).
17. D. R. Chialvo, *Chaos* **13**, 1226–1230 (2003).
18. P. Balenzuela and J. García-Ojalvo, *Chaos* **15**, 023903(8) (2005).
19. A. Lopera, J. M. Buldu, M. C. Torrent, D. R. Chialvo, and J. Garcia-Ojalvo, *Phys. Rev. E* **73**, 021101(6) (2006).
20. P. Balenzuelaa, H. Braun, and D. R. Chialvo, *Contemporary Physics* **53**(1), 17Ű-38 (2012).
21. L. Glass, and M. C. Mackey, *From clocks to chaos*, (Princeton University Press, Princeton, 1988).
22. H. G. Schuster, *Deterministic Chaos*, VCH-Verlag, Weinheim, 1994.
23. J. F. Schouten, R. J. Ritsma and B. L. Cardozo, *J. Acoust. Soc. Am.* **34**, 1418–1424 (1962).
24. R. Plomp, *J. Acoust. Soc. Am.* **41**, 1526–1533 (1967).
25. C. Pantev, M. Hoke, B. Lütkenhöner, and K. Lehnertz, *Science* **246**, 486–488 (1989).
26. W. M. Hartmann, *J. Acoust. Soc. Am.* **100**, 3491–3502 (1996).
27. J. H. E. Cartwright, D. L. Gonzalez, and O. Piro, *Phys. Rev. Lett.* **82**, 5389–5392 (1999).
28. D. Benson, *Music: a Mathematical Offering*, Cambridge University Press, Cambridge, 2006.
29. V. M. Eguiluz, M. Ospeck, Y. Choe, A. J. Hudspeth, and M. O. Magnasco, *Phys. Rev. Lett.* **84**, 5232–5235 (2000).
30. Y. V. Ushakov, A. A. Dubkov, and B. Spagnolo, *Phys. Rev. E* **81**, 041911(13) (2010).
31. Y. V. Ushakov, A. A. Dubkov, and B. Spagnolo, *Phys. Rev. Lett.* **107**, 108103(4) (2011).
32. A. N. Burkitt, *Biol. Cybern.* **95**, 97–112 (2006).
33. H. E. Plesser, S. Tanaka, *Phys Lett A* **225**, 228–234 (1997); H. E. Plesser, T. Geisel, *Phys Rev E* **63**(6), 031916 (2001).
34. C.F. Lo and T.K. Chung, *First Passage Time Problem for the Ornstein-Uhlenbeck Neuronal Model*, King et al. (Eds.), ICONIP 2006, Part I, Springer-Verlag Berlin Heidelberg, Lecture Notes in Computer Science vol. **4232**, pp. 324-331.