The auditory perception of pitch and illusory phase-locking

Dante R.. Chialvo
Department of Physiology, University of California Los Angeles.
Los Angeles, California, 90024, USA.
dchialvo@ucla.edu – www.bol.ucla.edu/~dchialvo

ABSTRACT

Since the earliest interest of Pythagoras and despite the gargantuan literature the problem of how the pitch of complex sounds is extracts remains unsolved. Pitch is a subjective attribute of a sound sensation by which it may be ordered on a scale from low to high. While the pitch of a tuning fork is objective (i.e., its single frequency) most natural sounds are combination of different frequencies. How the brain assigns a single definite pitch to the sound emitted by a string that is plucked in the middle, which contains many frequencies? The purpose of this work is to show that a simple nonlinear neuronal mechanism can account for all relevant features described for the auditory pitch perception. We considered recently [3], a simple toy model of the first neuron of the auditory system. Briefly, the system is a non-dynamical threshold device [4] that reduces the most elementary aspects known to exist in neurons, to a set of rules comparing the input signal *x(t)* with a threshold U_{th} like: $x(t) > U_{th}$ or $< U_{th}$. Anytime *x(t)* crosses the fixed threshold U_{th}=1, a "spike" is emitted. The only quantity of relevance for the problem here are the inter- spike intervals (ISI). The model is driven with combinations of sinusoidal of various frequencies, i.e., simulating complex sounds used in experiments [1,2,5]:

$$x(t) = A\left(\sin(f_1)2\pi t + \sin(f_2)2\pi t + \dots \sin(f_n)2\pi t\right)\frac{1}{n} + \varepsilon(t) \qquad (1)$$

where $f_1 = kf_0$, $f_2=(k+1)f_0$,... $f_n=(k+n-1)f_0$ and $k > 1$. The term $\varepsilon(t)$ is white noise with zero mean a Gaussian distribution with variance σ . The region of parameters of concern here is A small (respect to U_{th}) such that the deterministic forcing alone is not enough to cross the threshold and fire a spike. The range of σ values need to be explored on the region where the timing of the spikes is expected that will become more coherent with the input signal. Notice that the deterministic terms in Eq. 1 represent complex tones where the fundamental f_0 is absent. Since it is well known in psycho-acoustic that sound tones of this form are perceived with a pitch equal to f_0, i.e., the "missing" fundamental, numerical experiments were aimed to investigate the model response to these signals. Choosing initially signals composed by two periodic terms (f_1=2Hz and f_2=3Hz, A= .9 in this example) and increasing noise it was found: I) there is a wide range of noise for which the neuron spikes are spaced preferably ~ $1/f_0$ (Fig. 1). Thus, analogous to the psycho-acoustic experiments the neuron' strongest resonance occurs for 1 Hz, a frequency that is missing in the input. II) This resonance to the "missing" f_0 is not analogous to the difference f_2 - f_1. This was found by studying the response to inharmonic signals constructed by shifting all components of the harmonic signal by the same amount Δf :

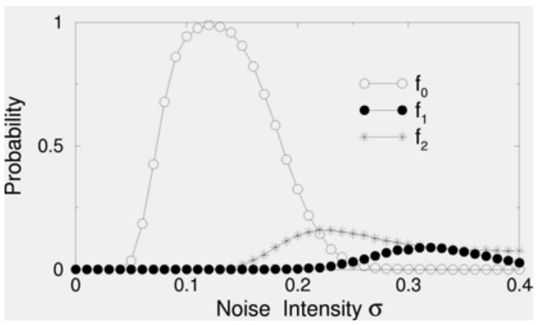

Fig. 1: Neuron response to two-frequencies tones for increasing noise. $1/f_0$ are the most probable inter-spike intervals.

$$f_1 = kf_0 + \Delta f\,, \quad f_2 = (k+1)f_0 + \Delta f\,, \quad \dots fn = (k+n-1)f_0 + \Delta f \qquad (2)$$

It this case, despite that the difference $f_n - f_{n-1}$ remains constant, the location of the strongest resonance (i.e. the largest peak in Fig. 1) was found to shift linearly as f_p=f_0 + /(*k*+1/2) for *n*=2 and f_p = f_0 + /(*k*+3/2) for *n*=3. Furthermore, it is possible a generalization predicting the response of the neuron to stimuli composed of N sinusoidal signals of frequencies:

$$kf_0 + \Delta f, \quad (k+1)f_0 + \Delta f, \quad ...(k+N-1)f_0 + \Delta f \tag{3}$$

the resonances will be expected to occur at frequencies given by:

$$f_r = f_0 + \frac{\Delta f}{k + (N-1)/2} \tag{4}$$

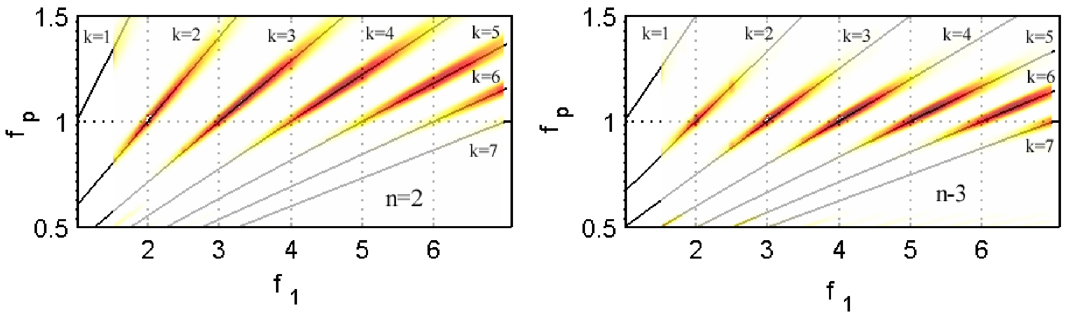

Fig. 2: Neuron response to 2- and 3-frequencies tones (redrawn from [3]).

Figure 2 shows the numerical results using signals with n=2 and n=3 (left and right panels) frequencies signals (using noise variance σ = 0.1). The graphs represent the probability (in grayscale) of observing spikes with a given instantaneous firing frequency f_p (ordinate) as a function of the frequency f_1, (abscissa) of the lowest of two (or three) components of the driving signal. The lines are the theoretical predictions of Eq. 4 that over-imposes exactly with the simulation results. Note that the neural response to any combinations of sinusoids will produce only one of these two diagrams: tones composed of *even* numbers of sinusoids will produce neural responses with statistics as the ones in the left panel; tones with *odd* numbers will produce responses like those in the right panel. Two specific quantitative predictions agree with published reports in the field of hearing: First, the slopes illustrated in Fig. 2 match perfectly the pitch reported by individuals listening complex tones [5], as well as with neuro-physiological data collected from auditory neurons in response to two-frequencies sounds [1,2]. Second, listeners often report more than one value of pitch for a constant f_1, especially for high k values, a fact that can be immediately identified in Fig. 2. The litmus test for models attempting to explain pitch perception is their response to periodically amplitude modulated (AM) white noise, which is known to be perceived as having a definite pitch. This is naturally replicated in this model because firings occur at the envelope of the AM noise and thus coincides with the perception subjects report. The model results are formally interpreted in [3] as a constructive interference between the sinusoids in the input signal, which is then nonlinearly detected by the neuron noisy threshold. The intervals between these peaks of constructive interference changes with Δf, giving rise to the illusion of a phase-locking between frequencies occurring "inside" the system. Instead, it follows from these results that the perceived pitch can be the non-linear detection of a linear interference happening outside our senses.

Keywords: Auditory physiology, Pitch perception, Stochastic Resonance.

REFERENCES

[1] Cariani P.A. and Delgutte B., *J. Neurophysiol.* **76**, 1698-1716, 1996.
[2] Cariani P.A. and Delgutte B., *J. Neurophysiol.* **76**, 1717-1734, 1996.
[3] Chialvo D.R., Calvo O., Gonzalez D.L., Piro O., Savino, G.V. Subharmonic resonance and synchronization in neuronal systems. *Phys. Rev. Letters,* (2001 Submitted).
[4] Gingl Z, Kiss L.B., Moss F., *Europhys. Lett.*, **29**, 19, 1995.
[5] Schouten J.F., Ritsma R.J. and Cardozo B.L., *J. Acoust. Soc. Am.* **34**, 1418, 1962.