Strong competition between tonotopic neural ensembles explains pitch-related dynamics of auditory cortex evoked fields

A. Tabas, A. Rupp & E. Balaguer-Ballester

Introduction

The latency of the N100m transient component of the magnetic auditory evoked fields presents a widely reported correlation with the perceived pitch. This high correlation has been robustly replicated in the literature for a range of different stimuli, indicating that the neural generator of the N100m, located in antero-lateral Heschl's Gyrus (alHG), has an important role in cortical pitch processing. In this work, we introduce a realistic model of pitch processing in auditory cortex; which provides a novel explanation of the neural mechanisms underlying the N100m responses associated with the perceived pitch.

The model

Subcortical input was simulated using a detailed model of the peripheral auditory system generating realistic auditory nerve spike trains [7] followed by a *delay-and-multiply* processing carried out by chopper neurons in cochlear nucleus and coincidence detector units in the inferior colliculus [3]. Chopper neurons systematically delay input spike trains by $\{\delta t_n\}_1^N$, whilst coincidence detector units spike for such specific delays of the auditory nerve fibres. This operation yields a series of $N = 300$ frequency channels characterised by the delays δt_n. Channel n activates when the stimulus' waveform contains a periodicity with frequency $f_0 = 1/\delta t_n$. Channels corresponding to lower harmonics of the peridocities of the stimulus (i.e. channels characterised by delays $\delta t_n = 1/kf$ with $k = 1, 2, \dots$) are also coactivated after the delay-and-multiply process. Figure 1 shows the subcortical inputs elicited by three different tonal stimuli with the same pitch ($f = 250$ Hz, $\delta = 4$ ms).
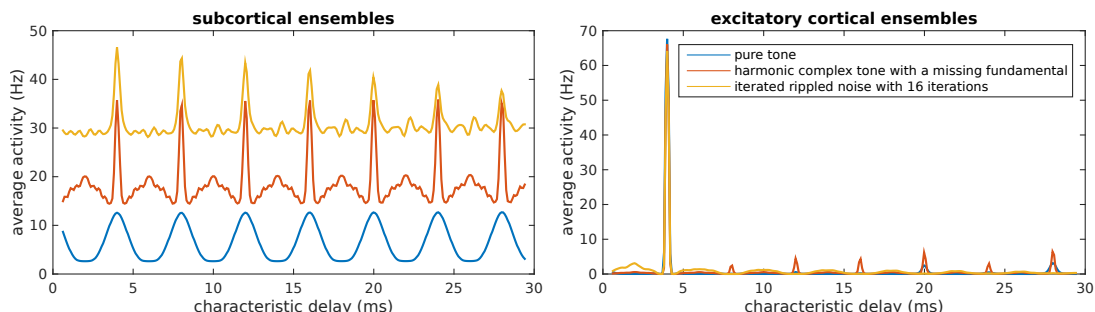

Figure 1: left) Average of the subcortical input generated by the model for a pure tone, a harmonic complex tone, and an iterated rippled noise of the same fundamental pitch $f = 250$ Hz. right) Average cortical output for the same stimuli. Plots show the activity per ensemble averaged at $100 - 200$ ms after onset.

The cortical model consist of a series of $N = 300$ cortical microcolumns described as sets of two neural ensembles: one excitatory (H_n^e) and one inhibitory (H_n^i; see Figure 2). An excitatory ensemble in one of such blocks n receives realistic input from the nth subcortical channel. A large activation in a column is typically associated with a fundamental pitch of δt_n [4].

Excitatory ensembles connect to both excitatory and inhibitory ensembles of adjacent blocks; whereas inhibitory ensembles connect globally with other inhibitory and excitatory populations. Crucially, inhibitory-to-excitatory connections are stronger when they link a population encoding the period δt_n with a population encoding any of its lower harmonics $k\delta t_n$ (see full connectivity matrices in Figure 2); in agreement with reported data on cortical connectivity in mammals [8]. This setting facilitates the inhibition of such low harmonics elicited during the peripheral processing as will be discussed next.
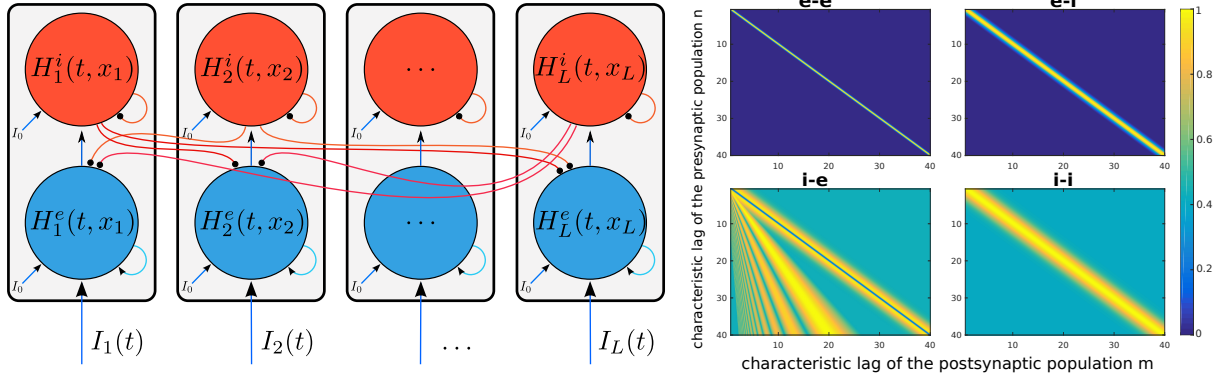
Figure 2: left) Basic schematics of the model. right) Excitatory-to-excitatory (e-e), excitatory-to-inhibitory (e-i), inhibitory-to-excitatory (i-e) and inhibitory-to-inhibitory (i-i) connectivity matrices.

Ensembles are modelled using a neural rate model [6] together with a mean-field approximation, using empirically shaped transfer functions [5]. Excitatory connections consist of realistic NMDA- and AMPA-driven synapses. Inhibitory connections are only of the GABA$_A$ type. AMPA and GABA synapses were modelled using leaky integrators with instantaneous rising times, NMDA dynamics were modelled considering slow rising times [5]. Additive synaptic noise was introduced in the form of white noise in the gating variables. Time constants and conductivities were taken from the literature [5]. Conductivities were slightly tuned within the biophysical range to match the experimental observations.

Results

The model was tested on three families of stimuli typically eliciting N100m auditory cortex responses highly correlated with pitch: pure tones (PT), harmonic complex tones (HCT) and iterated rippled noises (IRN, consisting on the aggregation of delayed copies of a white noise delayed by a fixed amount δt [4]). HCTs typically evoke the pitch of the fundamental frequency f_0 of the harmonic mixture, even if f_0 itself is not present in the tone (phenomenon known as *virtual* pitch [3]). IRNs evoke a pitch equivalent to the inverse of the delay $1/\delta t$.

We considered a variable number of harmonics in the HCT (with and without missing fundamental) and IRNs of 8, 16 and 32 iterations; for a range of pitch values between 200 Hz and 1000 Hz in all stimuli. After an unstable transient response of around $100-150$ ms, the activity in the cortical ensembles systematically converged to a unimodal distribution centred on the population corresponding to the perceived fundamental (see an example for $f = 250$ Hz in Figure 1); fully in line with predictions of abstract pitch perception models from the literature [4]. Simulations were performed using the same parameters for all stimuli, with the only exception of the conductivity of the connection between subcortical and excitatory cortical ensembles J_{thal}, which was tuned for each of the three families of stimuli in order to compensate the large differences between the average activity elicited in the subcortical patterns (see Figure 1).

Auditory evoked fields were predicted by the activity dynamics of the excitatory pyramidal ensembles in the cortical model. Assuming that all microcolumns have the same orientation, the total dipolar moment elicited by the full cortical model is proportional to the aggregated activity across populations $m(t) = \sum_n H_n^e(t + \Delta t)$, where Δt accounts for the time elapsed from tone onset until the signal first arrives to cortex. Typically, $\Delta t \simeq 30 - 50$ms. To account for the trial to trial variability of the model, we further averaged the predicted moment across 10 runs $M(t) = \langle m(t) \rangle_{\text{runs}}$.

An example of the simulated fields is shown in Figure 3 for several stimuli. The resulting waveform components can be related with the evoked fields observed in MEG auditory experiments: the first large transient predicts the N100m component, whilst the sustained model response shows a good agreement with the sustained field.

In order to assess quantitatively the relation between the N100m and the model's output, we computed the latency of the component for pure tones and HCTs and compared them with available results in the literature [1, 2]. Results are shown in the Figure of the short abstract.
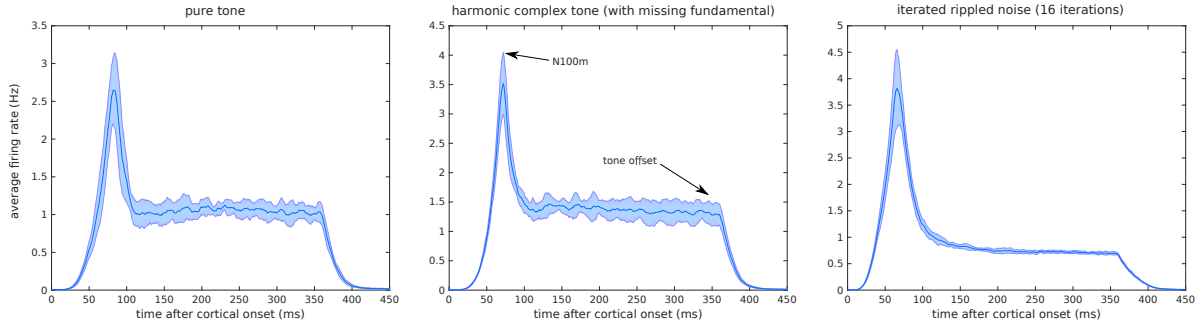
Figure 3: Simulations for $M(t)$ of a 16 iterations IRN with a delay of $2\,\mathrm{ms}$ (i.e. an equivalent pitch of $500\,\mathrm{Hz}$). Note that the proportionality factor relating $M(t)$ with the generated dipole moment should be negative.

Discussion

We present a novel model of pitch perception able to account for the pitch-related components of the auditory evoked fields. The model quantitatively explains the observed N100m transient neural response in a range of stimuli as a transient instability in the neural dynamics underlying pitch processing. The instability period begins at the cortical input onset i.e. when cortical ensembles start to integrate the subcortical activation patterns. Pyramidal neurons encoding the perceived pitch and lower harmonics become increasingly active, propagating forward activity to the inhibitory ensembles; which then feedback reduce the activation of excitatory ensembles encoding lower harmonics. Thus, the aggregated activity in the excitatory neurons shows a transient component that begins with the subcortical input onset, peaks when the inhibitory/excitatory input is balanced, and stabilises when the population encoding the perceived pitch is the more active one; accounting for the pitch perceived.

Importantly, we found that the latency of the N100m component directly stems from the time required by the model to achieve equilibrium after stimulus' onset. High-pitched sounds have a larger amount of lower harmonics represented in cortex that low-pitched ones, and thus they elicit bottom-up activation in more excitatory ensembles; which induce top-down inhibitory activity in a larger amount of inhibitory populations. Namely, high pitched sounds trigger top-down inhibition faster, thus explaining the observed dependency on pitch of the N100m's latency.

In other words, our model shows that N100m can be explained by transient dynamics of a winner-takes-all competition among balanced excitatory and inhibitory populations tonotopically distributed; endowing the model with a flexibility mechanism which explains the perception of pitch.

References

[1] See main abstract.

[2] See main abstract.

[3] See main abstract.

[4] See main abstract.

[5] See main abstract.

[6] Srdjan Ostojic, and Nicolas Brunel. From spiking neuron models to linear-nonlinear models. *PLoS Comput Biol*, 7(1):e1001056, 2011.

[7] Muhammad S A Zilany, Ian C Bruce, and Laurel H Carney. Updated parameters and expanded simulation options for a model of the auditory periphery. *JASA*, 135(January):283–286, 2014.

[8] Xiaoqin Wang. The harmonic organization of auditory cortex. *Frontiers in systems neuroscience*, 7(December):114, 2013.