

# Competition between cortical ensembles explains pitch-related dynamics of auditory evoked fields

Alejandro Tabas, André Rupp, and Emili Balaguer-Ballester

Faculty of Science and Technology, Bournemouth University, BH12 5BB Poole, UK  
Biomagnetism Section, Im Neuenheimer Feld 400, 69120 Heidelberg, Germany  
Bernstein Center for Computational Neuroscience, Heidelberg/Mannheim, Germany  
`atabas@bournemouth.ac.uk`

**Abstract.** The latency of the N100m transient component of the magnetic auditory evoked fields presents a widely reported correlation with perceived pitch. These observations have been robustly reproduced in the literature for a number of different stimuli, indicating that the neural generator of the N100m has an important role in cortical pitch processing. In this work, we introduce a realistic cortical model of pitch perception revealing, for the first time to our knowledge, the mechanisms responsible for the observed relationship between the N100m and the perceived pitch. The model describes the N100m deflection as a transient state in cortical dynamics that starts with the incoming of a new subcortical input, holds during a winner-takes-all ensemble competition, and ends when the cortical dynamics reach equilibrium. This model qualitatively predicted the latency of the N100m of three families of stimuli.

**Keywords:** cortical dynamics, auditory evoked fields, N100m, pitch perception, perceptual integration, multi-attractor systems

## 1 Introduction

Auditory evoked fields (AEFs) observed in MEG experiments systematically present a transient deflection known as the N100m, elicited around 100 ms after tone onset in the antero-lateral Heschl's Gyrus. The exact N100m's latency is correlated with the perceived pitch of a wide range of stimuli [5, 7, 8], suggesting that the cortical source of the transient component has an important role on the processing of pitch in auditory cortex [9]. However, the biophysical substrate of the relationship between pitch decoding and the N100 morphology remains an enigma.

Existing models of pitch, focused on perceptual phenomena, do not explain the mechanisms generating cortical evoked fields during pitch processing in biophysical detail (e.g. [1, 6]). Cortical models of the evoked fields, like the Dynamic Causal Models [3], often assume an unrealistic cortical input (e.g. white noise) and are thus unable to reproduce stimulus-driven properties of the AEF. In this work, and for the first time to our knowledge, we introduce a model of interacting neural ensembles describing how stimulus-dependent cortical pitch processing gives rise to the observed human neuromagnetic responses. Specifically, we

focus on the N100m transient dynamics and its peak latency. Our conclusion is that the N100m reflects a decoding process occurring in the onset of the stimuli, that can be described as a competition between neural ensembles sensitive to different pitch values.

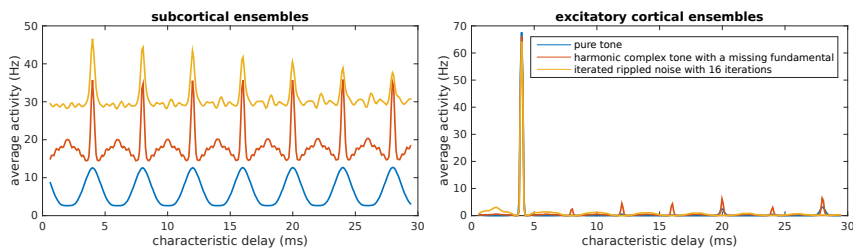
## 2 The model

Subcortical input was simulated using a realistic model of the peripheral auditory system generating realistic auditory nerve spike trains [14] followed by a *delay-and-multiply* processing carried out by chopper neurons in cochlear nucleus and coincidence detector units in the inferior colliculus [6]. The spike trains generated by the peripheral system, represented by the probability of spiking  $p(t)$ , are phase-locked to the waveform of the stimulus, thus preserving all the periodicities of the sound. 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. The final subcortical output  $A_n(t)$  represents a leaky-integration of the coincidence detectors output as follows:

$$\tau_n^{sc} \dot{A}_n(t) = -A_n(t) + p(t)p(t - \Delta t) \quad (1)$$

Lag-dependent time constants  $\tau_n^{sc}$  were taken from the literature [11].

The formulation in Equation 1 yields a series of  $N = 300$  channels characterised by the chopper delays  $\delta t_n$ . Channel  $n$  activates when the stimulus' waveform presents a periodicity with frequency  $f_0 = 1/\delta t_n$ . Channels corresponding to lower harmonics of the periodicities 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 ( $f_0 = 250$  Hz,  $\delta = 4$  ms).

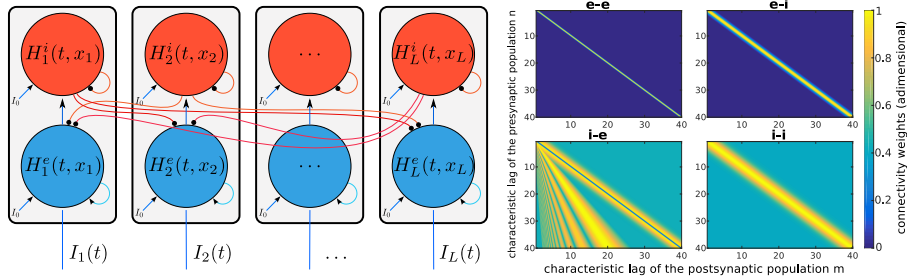


**Fig. 1.** Left) Average of the subcortical input generated by the model for a pure tone, a harmonic complex tone, and an iterated rippled noise. All stimuli have the same  $f_0 = 250$  Hz. Right) Average cortical output for the same stimuli. Plots show the activity per ensemble averaged at 100 – 200 ms after onset.

The cortical model consists 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  $n$ th subcortical channel. A large activation in a column is typically associated with a fundamental pitch of  $\delta t_n$  [1].

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  $\delta t_n$  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 [10]. This setting facilitates the inhibition of low harmonics elicited during the peripheral processing as will be discussed next.



**Fig. 2.** Left) Basic schematics of the model. Each block represents a cortical microcolumn. Each column consist of excitatory pyramidal neurons (blue) and inhibitory interneurons (red). Right) Connectivity weights between the ensembles of the model. Excitatory-to-excitatory (e-e), excitatory-to-inhibitory (e-i), inhibitory-to-excitatory (i-e) and inhibitory-to-inhibitory (i-i) connectivity matrices. Note that, although the excitation is local, inhibition is induced globally.

Ensembles are modelled using a neural rate model with a mean-field approximation, where empirically shaped transference functions  $\phi^{e,i}(I)$  are [13]:

$$\tau^{\text{pop}}(t) \dot{H}_n^{e,i}(t) = -H_n^{e,i} + \phi^{e,i}(I_n^{e,i}(t)) \quad (2)$$

with

$$\phi^{e,i}(I) = \frac{a^{e,i}I - b^{e,i}}{1 - e^{-d^{e,i}(a^{e,i}I - b^{e,i})}} \quad (3)$$

Excitatory connections consist of NMDA- and AMPA-driven synapses. Inhibitory connections are only of the GABA<sub>A</sub> type. AMPA and GABA synapses were modelled using leaky integrators with instantaneous rising times [2]:

$$\dot{S}_n^j(t) = -\frac{S_n^j(t)}{\tau_j} + H_n^{e,i}(t), \quad j = \text{AMPA, GABA} \quad (4)$$

NMDA dynamics were modelled considering slow rising times [2]:

$$\dot{S}_n^{\text{NMDA}}(t) = -\frac{S_n^{\text{NMDA}}(t)}{\tau_{\text{NMDA}}} + \gamma(1 - S_n^{\text{NMDA}}(t))H_n^{e,i}(t) \quad (5)$$

Additive synaptic noise was introduced in the form of white noise in the gating variables  $S_n^j$ . Subcortical input was driven by NMDA and AMPA dynamics according to Equations 4 and 5, using the ensemble firing rates of the coincidence detectors  $A_n(t)$  as gate triggers. Thus, the total synaptic input for the excitatory populations can be written as follows:

$$\begin{aligned} I_n^e(t) &= J^{\text{NMDA,th}} S_n^{\text{NMDA,th}}(t) + J^{\text{AMPA,th}} S_n^{\text{AMPA,th}}(t) \\ &+ \sum_k C_{n,k}^{ee} (J^{\text{NMDA}} S_k^{\text{NMDA}}(t) + J^{\text{AMPA}} S_k^{\text{AMPA}}(t)) \\ &- \sum_k C_{n,k}^{ie} J^{\text{GABA}} S_k^{\text{GABA}}(t) \end{aligned} \quad (6)$$

In equation 6, the first two terms correspond to the subcortical input, the third term accounts for cortical excitatory input, and the last term accounts for cortical inhibitory inputs. The conductivities  $J^{\text{NMDA,th}}$ ,  $J^{\text{AMPA,th}}$ ,  $J^{\text{NMDA}}$ ,  $J^{\text{AMPA}}$ , and  $J^{\text{GABA}}$  were taken from the literature [13] and slightly tuned within the biophysical range to match the experimental observations.

Synaptic inputs for the inhibitory populations follow a similar pattern:

$$\begin{aligned} I_n^i(t) &= \sum_k C_{n,k}^{ei} (J^{\text{NMDA}} S_k^{\text{NMDA}}(t) + J^{\text{AMPA}} S_k^{\text{AMPA}}(t)) \\ &- \sum_k C_{n,k}^{ii} J^{\text{GABA}} S_k^{\text{GABA}}(t) \end{aligned} \quad (7)$$

The connectivity matrices used in Equations 6 and 7 ( $\mathbf{C}^{ee}$ ,  $\mathbf{C}^{ei}$ ,  $\mathbf{C}^{ie}$ , and  $\mathbf{C}^{ii}$ ) are depicted in Figure 2. The connectivity patterns were designed ad-hoc, always following biophysical constraints defined in the literature [10].

Neural adaptation in cortex was modelled as an effective negative input current  $B_n^{e,i}(t)$  in the neural ensembles  $H_n^{e,i}$  [4]. Adaptation effective currents followed leaky-integrator-like dynamics:

$$\tau_{\text{adapt}} \dot{B}_n^{e,i}(t) = -B_n^{e,i}(t) + \theta_{\text{adapt}} H_n^{e,i}(t) \quad (8)$$

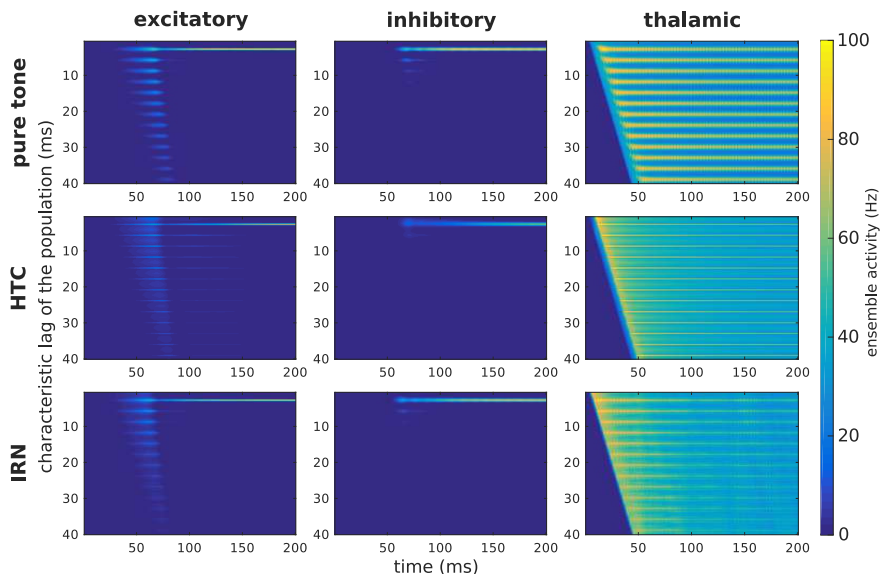
Adaptation parameters were identical in both excitatory and inhibitory ensembles.  $\tau_{\text{adapt}}$  was chosen from the literature [4] and  $\theta_{\text{adapt}} \ll 1$  such that the effect of adaptation is only noticeable under high firing rate regimes.

### 3 Results

The model was tested using 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 iteratively lagged copies of a white noise with a fixed delay  $\delta t$ ). 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* [6]). 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 for all stimulus types. 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 Figures 1 and 3), fully in line with predictions of abstract pitch perception models from the literature [1].



**Fig. 3.** Time evolution of the activity of the ensembles for three different stimuli with the same pitch ( $f = 333$  Hz). From left to right, plots show the temporal evolution of the  $N = 300$  ensembles ( $y$ -axis) in each of the different groups: excitatory populations  $H^e(t)$ , inhibitory populations  $H^i(t)$ , and subcortical populations  $A(t)$ . From top to bottom, each row shows results for: a pure tone, a harmonic complex tone with 10 harmonics, and an iterated rippled noise with 32 iterations.

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  $J_{\text{thal}}$ , 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. Auditory evoked fields are

typically represented using equivalent dipoles that model neural activity in a localised cortical area. Dipoles in auditory cortex are usually fitted using band-passed MEG fields averaged along a few hundreds of trials. In order to predict the elicited fields, we assumed that all microcolumns in our model present the same orientation. Then, the total dipolar moment elicited by the cortical model is proportional to the aggregated activity across populations  $m(t) = \sum_n H_n^e(t + \Delta t)$ , where  $\Delta t$  accounts for the time elapsed from tone onset until the signal first arrives in primary auditory cortex ( $\Delta t \simeq 30\text{--}50\text{ms}$ ). To account for the trial to trial variability of the model, we further averaged the predicted dipole moment across 10 runs  $M(t) = \langle m(t) \rangle_{\text{runs}}$ .

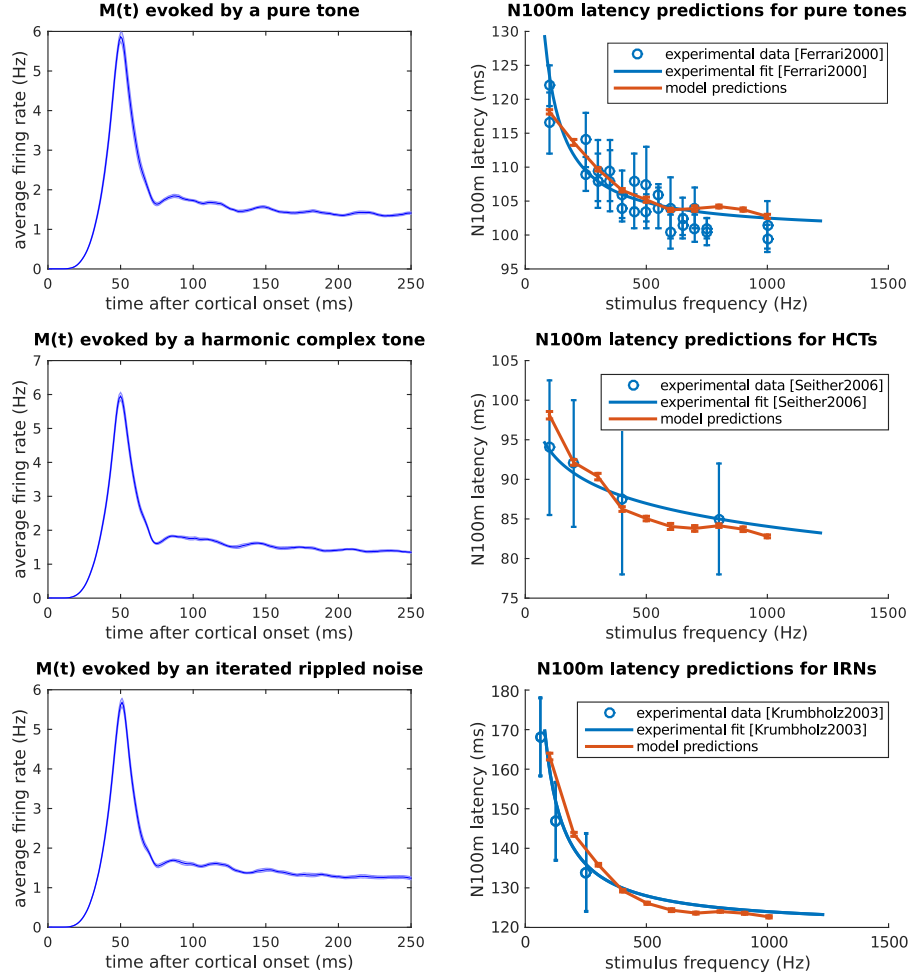
An example of the simulated fields is shown in Figure 4 for several stimuli. The resulting waveform components can be related with the evoked fields observed in MEG auditory experiments: the first large negative transient predicts the N100m component, whilst the sustained model response shows a good 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 [5, 7, 8]. Results are shown in Figure 4. Good agreements between the model’s response and the experimental data were generally observed in the range  $f_0 \sim 150\text{--}2000\text{ Hz}$ . Specifically, latency predictions over 1000 Hz were all quite similar, consistently with experimental observations.

Stimuli presenting fundamental or effective frequencies under  $f_0 = 150\text{ Hz}$  yielded an overly late predicted N100m. This is due to intrinsic limitations of the peripheral model, that does not present cochlear channels solving frequencies under  $f = 125\text{ Hz}$  [14]. Stimuli with  $f_0 > 2000\text{ Hz}$  failed to yield satisfactory perceptual outputs, as a reflection of the limit for phase-locking in the peripheral auditory system [14].

## 4 Discussion and Conclusions

We introduced a biophysical model of cortical responses related to pitch processing. The model accounts for the pitch-related components of auditory evoked fields for the first time to our knowledge, and 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; whose 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 perceived pitch. A balanced excitation and



**Fig. 4.** Simulations of the auditory evoked fields evoked by pure tones, harmonic complex tones, and iterated rippled noises. In the left column, we show an example of the  $M(t)$  for each of the families of stimuli with a pitch of 500 Hz (blue shades represent the standard deviation). In the right column, we show the derived predictions of the N100m latency and the observed experimental data for a range of fundamental frequencies.

inhibition setting such as the one shown in this model has been found to underlie cognitive flexibility [12].

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 typically have a larger amount of lower harmonics represented in cortex than 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.

*Conclusion.* This study shows that N100m morphology associated with pitch perception can be explained by transient dynamics of a winner-takes-all competition among balanced, excitatory and inhibitory populations, tonotopically distributed in cortex. In conclusion, we suggest that these characteristics provide a specific mechanism which enables aHG ensembles to process pitch.

## References

- Balaguer-Ballester, E., Clark, N.R., Coath, M., Krumbholz, K., Denham, S.L.: Understanding pitch perception as a hierarchical process with top-down modulation. *PLoS Computational Biology* 5(3), e1000301 (2009)
- Brunel, N., Wang, X.J.: Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *Journal of Computational Neuroscience* 11(1), 63–85 (2001)
- Daunizeau, J., David, O., Stephan, K.E.: Dynamic causal modelling: A critical review of the biophysical and statistical foundations. *NeuroImage* 58(2), 312–322 (2011)
- Gerstner, W., Kistler, W.M., Naud, R., Paninski, L.: *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition*. Cambridge University Press, 1st edn. (2014)
- Krumbholz, K., Patterson, R., Seither-Preisler, A., Lammertmann, C., Lütkenhöner, B.: Neuromagnetic evidence for a pitch processing center in Heschl’s gyrus. *Cerebral Cortex* 13(7), 765–772 (2003)
- Meddis, R., O’Mard, L.P.: Virtual pitch in a computational physiological model. *The Journal of the Acoustical Society of America* 120(6), 3861 (2006)
- Roberts, T.P.L., Ferrari, P., Stufflebeam, S.M., Poeppel, D.: Latency of the auditory evoked neuromagnetic field components. *Journal of clinical Neurophysiology* 17(2), 114–129 (2000)
- Seither-Preisler, A., Patterson, R., Krumbholz, K., Seither, S., Lütkenhöner, B.: Evidence of pitch processing in the N100m component of the auditory evoked field. *Hearing research* 213(1-2), 88–98 (2006)
- Tabas, A., Siebert, A., Supek, S., Pressnitzer, D., Balaguer-Ballester, E., Rupp, A.: Insights on the Neuromagnetic Representation of Temporal Asymmetry in Human Auditory Cortex. *Plos One* 11(4), e0153947 (2016)
- Wang, X.: The harmonic organization of auditory cortex. *Frontiers in Systems Neuroscience* 7, 114 (2013)
- Wiegand, L.: Searching for the time constant of neural pitch extraction. *The Journal of the Acoustical Society of America* 109(3), 1082–1091 (2001)
- Wimmer, K., Compte, A., Roxin, A., Peixoto, D., Renart, A., de la Rocha, J.: Sensory integration dynamics in a hierarchical network explains choice probabilities in cortical area MT. *Nature Communications* 6, 6177 (2015)
- Wong, K.F., Wang, X.J.: A recurrent network mechanism of time integration in perceptual decisions. *The Journal of Neuroscience* 26(4), 1314–28 (2006)
- Zilany, M.S.A., Bruce, I.C., Carney, L.H.: Updated parameters and expanded simulation options for a model of the auditory periphery. *The Journal of the Acoustical Society of America* 135, 283–286 (2014)