

# A realistic model of pitch perception explains the N100m morphology evoked by chords – Details on methods

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

## The model

### Architecture

The main contribution of this work is a model of pitch consisting in three processing steps: peripheral, subcortical, and cortical (see Figure 1A). The peripheral processing was reproduced using a realistic model of the peripheral auditory system [2]. The subcortical and cortical processing were implemented as follows.

### Subcortical cortex

The model of the peripheral auditory system simulates the spiking patterns of the auditory nerve  $p(t)$  [2]. This representation is transformed into a spectro-temporal pattern by means of an autocorrelation-like process implemented in the cochlear nucleus and inferior colliculus [1]. The process can be represented as a delay-and-multiply operation [1]:

$$A_l(t) = p(t)p(t-l) \quad (1)$$

The output of this process is parametrised by the delay or lag  $l$ . This pattern is then integrated by a leaky integrator with a lag-dependent time constant [5]. This integration yields to a stable representation of the nerve activity  $I_l(t)$ .

### Cortical Integration

These spectro-temporal patterns are then further integrated in cortex. The integrator is a network of cortical ensembles consisting on  $N$  selective populations, parametrised by the lag  $l$  and receiving the direct input  $I_l(t)$  from the subcortical processing; and two non-selective populations, one excitatory and one inhibitory [3].

Each of the populations can be described as a network of spiking neurons simplified using a mean-field approximation [6]. By assuming a constant activity in the non-selective population, a linear transfer function for the ensemble of interneurons, and that the fast gating variables (GABA and AMPA) of the populations reach equilibrium much faster than the slow (NMDA) gating variables, the phase space of model can be further simplified to a set of  $N$  dynamic equations [3, 4]. This simplified system is represented as a set  $N$  selective ensembles showing mutual effective inhibitions and recurrent self-excitations (see Figure 1A) [3, 4]:

$$\dot{S}_l = -\frac{S_l}{\tau} + \gamma(1 - S_l)H(x_l) + \xi_l \quad (2)$$

$$H(x_l) = \frac{f(x_l)}{1 - e^{-df(x_l)}}, \quad f(x) = ax + b \quad (3)$$

$$x_l = w J_N S_l + G J_N \sum_m C_{lm} S_m + g I_l + I_0 \quad (4)$$

where  $x_l$  is the input current for the ensemble  $l$ ,  $H(x)$  represents transfer function of the ensembles and  $S_l$  reflects the evolution of the gating variables. The effective input current  $x_l$  consists of self-excitation

$w J_N$ , an effective inhibitory current  $G J_N \sum_m C_{lm}$ , the selective input  $I_l$  and a background current  $I_0$ . The gain of the input  $g$ , the factor  $G < 0$ , and the connectivity matrix  $C_{ij}$  are free parameters of the model. The time constant  $\tau \sim 100$  ms can be tuned to some extent. The rest of the parameters were chosen as in [4].

## The connectivity matrix

The mutual inhibition between the ensembles ( $C_{ij}$  in Equation 2) plays a crucial role in pitch processing. Consider the case of a simple stimulus such as a pure tone. The activation  $I_l(t)$  after subcortical processing of such stimulus shows a high activation for the lag  $l = L$  corresponding to the frequency of the stimuli, but also for the lags corresponding to the harmonics of that frequency [1], such as  $l = 2L$ ,  $l = 3L, \dots$ . However, previous studies on perceptual integration suggest that the output of the perceptual model should consist on a winning population encoding the perceptual outcome of the processing [2, 3]. To obtain such output we designed a connectivity matrix taking into account the behaviour of the harmonics.

Note that  $C_{ij}$  encodes the strength of the inhibition applied by the population  $j$  to the population  $i$ . If the population  $j$ , encoding the lag  $l_j$ , receives a large input, we can inhibit the activation of higher harmonics using a periodic function as follows:

$$\psi(l_i, l_j) = \left(1 - \cos\left(2\pi \frac{l_i}{l_j} + \pi\right)\right)^\alpha, \quad \frac{l_i}{l_j} \geq \frac{3}{2} \quad (5)$$

where the domain has been chosen such that the self-inhibition of the population is zero.

In the other hand, the population  $j$  can receive a large input for having a characteristic lag  $l$  being a higher harmonic of a lower pitch. We can use that information to improve convergence by inhibiting the populations not encoding the possible frequencies generating such harmonic  $l_i = 1/2l_j, l_i = 1/3l_j, \dots$ . This can be implemented using a similar periodic function:

$$\hat{\psi}(l_i, l_j) = \left(1 - \cos\left(2\pi \frac{l_j}{l_i}\right)\right)^{\hat{\alpha}}, \quad \frac{l_i}{l_j} \leq 1 \quad (6)$$

In practice, we do not need to take into account all the harmonics, which can actually rise some issues in the case of Equation 6 for  $l_i \ll l_j$ . Thus, we restricted the domains of Equations 5 and 6 to consider only up to four harmonics (see Figure 1B). If we now set  $\phi(l_i, l_j)$  and  $\hat{\phi}(l_i, l_j)$  equal to zero out of such domains, we can summarise the connectivity matrix as follows:

$$C_{ij} = a_j \psi(l_i, l_j) + \hat{a}_j \hat{\psi}(l_i, l_j) + c_0, \quad i \neq j \quad (7)$$

where  $c_0$  is an offset inhibition acting over the whole network and  $\{a_j, \hat{a}_j\}$  are normalisation constants chosen such that an uniform input would inhibit it equally all the populations:

$$\sum_j C_{ij} = 1 \quad (8)$$

The diagonal of the matrix,  $C_{ii}$ , is set to zero.

## Experimentation

### Perception

The perceptual dimension of the model was trained using pure tones. This kind of stimuli elicit a well known perceptual response [7] that can be easily matched with the predictions of our model. Specifically, we matched the normalised activity of the ensembles after the stabilisation of the neural dynamics (ca. 70 ms) with the pitch of the stimuli. A perfect solution for a given pitch shows a null activation for all the ensembles except for the population characterising the given pitch. The difference between the expected outcome was used as an error function in a gradient-descend methodology to fit the free parameters of the connectivity matrix:  $\alpha$ ,  $\hat{\alpha}$  and  $c_0$ . The resulting configuration, shown in Figure 1B, was further validated using harmonic complex tones with missing fundamental [1].

## Electrophysiology

If, as we claim, the last stage of our model is located in primary auditory cortex (lateral Heschl’s gyrus), we should expect to find a correlation between the measured evoked fields in humans and the aggregated dynamics of the ensembles.

To test this hypothesis we used MEG recordings of neuromagnetic fields evoked by dyads, a kind of chord consisting in two notes. For our experimentation we used three different dyads generated using iterated ripple noises [7]. The dyads consisted on an unison (with a frequency ratio of 1:1 between the two tones), a perfect fifth (3:2 ratio) and a tritone (45:32 ratio). The fields evoked by the three stimuli were recorded on 20 subjects with 250 repetitions per dyad. Neuromagnetic data was fitted to a four dipoles model (see, for example, [8]) localising the origin of the N100m in Heschl’s gyrus.

The fitted MEG signal evoked by the unison dyad was compared with the prediction of the model generated for the same stimulus. This comparison was used to heuristically tune three free parameters of the ensemble model: the gain of the selective input  $g$ , the multiplicative factor of the connectivity matrix  $G$  and the time constant of the neural populations  $\tau$ . The resulting parametrisation was then validated using the data of the other two dyads (see the results in the figure in the main abstract).

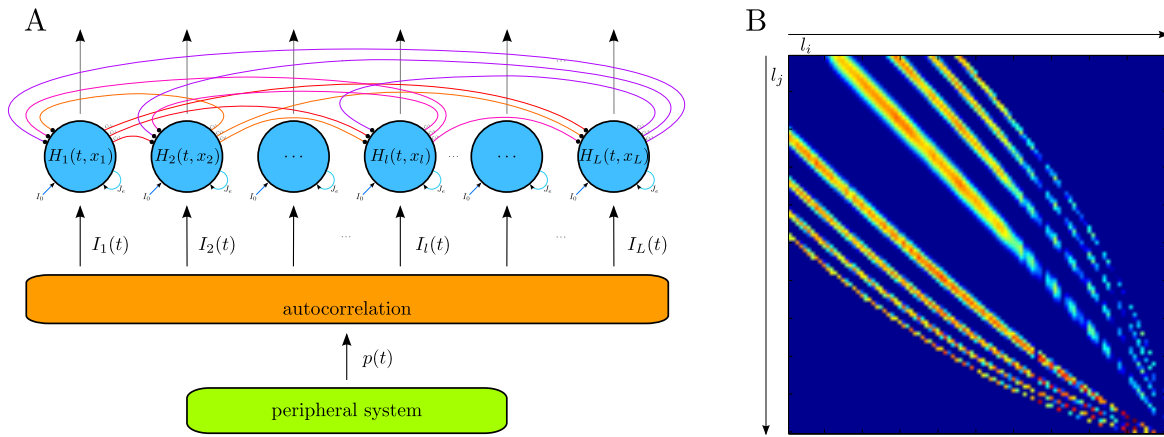


Figure 1: A) Schematics of the model B) Connectivity matrix. Values in deep blue are zero, red is the maximum. The figure displays the logarithm of the actual matrix to accentuate the structure of the matrix.

## References

- [1] R. Meddis and L. P. O’Mard. Virtual pitch in a computational physiological model. *The Journal of the Acoustical Society of America*, 120(6):3861, 2006.
- [2] See entry on Abstract.
- [3] See entry on Abstract.
- [4] See entry on Abstract.
- [5] L. Wiegrebe. Searching for the time constant of neural pitch extraction. *The Journal of the Acoustical Society of America*, 109(3):1082–1091, Mar. 2001.
- [6] X.-J. Wang. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, 36:955–968, 2002.
- [7] B. C. J. Moore. *An Introduction to the Psychology of Hearing*. Firth edit edition, 2004.
- [8] K. Krumbholz and R. Patterson. Neuromagnetic evidence for a pitch processing center in Heschl’s gyrus. *Cerebral Cortex*, 13(7):765–772, 2003.