Hierarchical Processing of Temporal Asymmetry in Human Auditory Cortex

Alejandro Tabas
Faculty of Science and Technology, Bournemouth University;
atabas@bournemouth.ac.uk

Emili Balaguer-Ballester
Faculty of Science and Technology, Bournemouth University;
Berstein Center for Comp. Neuroscience

Daniel Pressnitzer
Dép. d'études cognitives, Ecole normale supérieure; CNRS

Anita Siebert
Bruker BioSpin MRI GmbH

André Rupp
Dep. of Neurology, University of Heidelberg

***Abstract*—Communication sounds are typically asymmetric in time and human listeners are highly sensitive to short-term temporal asymmetry. Nevertheless neurophysiological correlates of perceptual asymmetry remain largely elusive to current approaches. Physiological recordings suggest that perceptual asymmetry is based on multiple scales of temporal integration within the auditory processing hierarchy. To test this hypothesis, we used magneto-encephalographic recordings to perform a model-driven analysis of auditory evoked fields (AEF) elicited by asymmetric sounds characterised by rising or decreasing envelopes (ramped and damped, respectively), using a hierarchical model of pitch perception with top-down modulation. We found a strong correlation between the perceived salience of ramped and damped stimuli and the AEFs, as quantified by the amplitude of the N100m component. Furthermore, the N100m magnitude is closely mirrored by a hierarchical model with stimulus-driven temporal integration windows of auditory nerve activity patterns. This strong correlation of AEFs, perception and modelling suggests that temporal asymmetry is processed in a hierarchical manner where integration windows are top-down modulated.**

I. Introduction

Sounds like speech and music are typically asymmetric in time. The term *temporal asymmetry* [1] has been used to describe sounds or individual periods of sounds that display different attack and decay times. These differences influence perceptual timing [2], pitch [3] and loudness [4], crucial factors in auditory perception.

Ramped and *damped* stimuli [1] introduce a systematic approach to study the effect of asymmetries in human auditory perception. These stimuli consist of a pure sinusoid multiplied either by a periodic rising exponential function (ramped) or a periodic decaying exponential function (damped). Ramped and damped sinusoids evoke two different perceptual components: ramped sounds are perceived as continuous tones with the pitch of the carrier whereas the repetitive streams of damped sinusoids are perceived as a drumming sound with a less salient carrier. However, the long-term Fourier energy spectra are identical in both stimuli. For this reason, traditional models of auditory perception, essentially based on extracting the auditory nerve periodicities on a fixed time window (e.g. [5]), cannot explain these perceptual differences.

Recent models proposed that pitch is processed in a hierarchical manner [6], [7] in line with observations in functional magnetic resonance imaging studies (e.g. [8]). Within this processing hierarchy, top-down modulation plays an essential role in the rapid perception of sounds, as proposed by the reversed hierarchical theory of perception (RHT, [9], [10]) or the closely related predictive coding principles of hierarchical generative models [11].

Essentially, top-down processing provides the auditory system for access to fine-grained stimulus detail represented in brainstem when expectancies of the pitch generated in high areas are violated by the following bottom-up predictions of pitch [7]. In this model, top-down modulation is implemented as a stimulus-driven adaptation of the temporal integration window in order to explain the balance between temporal integration and high resolution of the auditory system: While long integration time windows are necessary for a wide range of perceptual phenomena (e.g. [12]), short windows are needed for explaining rapid fluctuations of pitch (e.g. [13]). Consistently, the top-down model for pitch perception introduced by Balaguer-Ballester et al. [7] achieves balance between perceptual integration and resolution by enabling the auditory system to detect quick stimulus variations while producing a stable pitch prediction. Temporally asymmetric sounds are a challenging testbed for this pitch model, given the subtle differences between stimulus waveform leading to a significantly different pitch perception.

In the current study we investigated the neuromagnetic representation of the auditory perceptual asymmetry by considering the N100m deflection of the auditory evoked field (AEF), a well-known transient neuromagnetic response elicited 100 ms after the tone onset (see Figure 4). This deflection arises from multiple sources of auditory cortex, lateral Heschl's gyrus and planum temporale [14], [15]. Moreover, the N100m latency in the antero-lateral Heschl's gyrus has been associated with the perceived pitch [15] and salience [16] of the stimuli. Therefore, perceptual differences between ramped and damped sounds might be explained through the differences in the N100m morphology.

During the study, we performed a model-driven analysis of the N100m evoked by ramped and damped sounds in human listerners in order to decipher the representation of temporal asymmetry at the cortical level of auditory processing. We show that N100m amplitude and latency can be accurately predicted using a hierarchical model of pitch with top-down modulation, suggesting the importance of stimulus-dependent

effective integration windows for processing auditory temporal asymmetry.

II. Model description

A hierarchical model of interacting neural ensembles incorporating a top-down modulation process was used to analyse the MEG recordings (aka hierarchical *generative* pitch perception model, HGPM). In this section we roughly review the principles of the model. Details are described in depth in [7]. Software for the model is freely available in http://sourceforge.net/projects/topdownpitchmodel/.

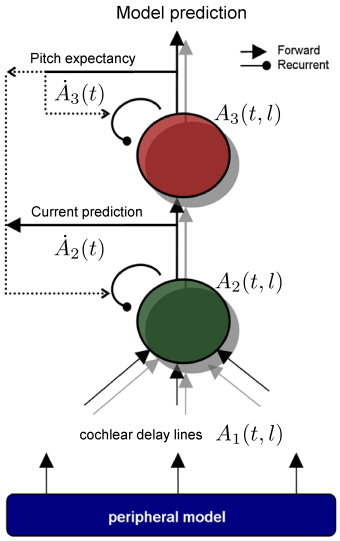

Fig. 1. Schematic view of the hierarchical generative pitch perception model. Round circles represent the two integration stages in the model. Recurrent lines reflect top-down mechanisms tuning the effective integration windows.

The pitch model receives its input from a realistic model of the peripheral auditory areas, which includes a dual-resonance filter at the basilar membrane passed then though a hair cell transduction model e.g. [17] that simulates the auditory-nerve spike probabilities $p(t)$ at each instant t for a given stimulus. This input is fed into a cascade of three ensemble models A_1, A_2 and A_3. The output of the first stage represents the probability of generating two spikes delayed by a certain lag l or *cochlear delay lines*:

$$A_1(t,l) = p(t)\,p(t-l)$$

The values for the lag l in which $A_1(t,l)$ reaches its maxima represent the pitch value in the so-called autocorrelation models of pitch [5], [18]. However, autocorrelation models failed to explain a large range of pitch phenomena, including the differences in the perception of ramped and damped sounds, requiring a more realistic processing [7]. In the considered model this is solved using a leaky integration process. Such a proceess is implemented in the superior two layers of the model as a cascade of two neural ensembles with top-down recurrent connections (see details in Figure 1) which control the size of the integration windows as a function of the perceived pitch.

Each of the two neural populations integrate the activity from the previous stage:

$$\tau_n \cdot \dot{A}_n(t,l) = -A_n(t,l) - \Psi_n(A_n(t,l), A_{n-1}(t,l))$$

where $n = 2,3$ and τ_n represents the the characteristic processing time of the population. The *activation functions* Ψ_n are time-dependent multiplicative gains:

$$\Psi_n(A_n, A_{n-1}) = \frac{\omega_n(t)}{\lambda_n(t)}\,A_n - \left(\frac{\omega_{n-1}(t)}{\lambda_{n-1}(t)} + 1\right) A_{n-1}$$

with $\omega_1(t)/\lambda_1(t) = 0$.

Crucially, for $n = 2,3$ the gains $\omega_n(t)/\lambda_n(t)$ show an adaptive behaviour that modulates the size of the integration window. In simple terms, the gain at n become large when the prediction of the stage n (i.e. the value for the lag l in which $A_n(t,l)$ reach its maxima) presents an unstable behaviour (for instance, in the interface of two different tones during a melody), therefore reducing the size of the integration window and *resetting* the information previously integrated at this stage. On the contrary, when a continuous mismatch is observed, the gains are reduced, enlarging the integration window, until a stable pitch is achieved. Once the pitch prediction is stable, the size of the integration window decays again, preparing the population for sudden changes in the stimuli. Full details of this mechanism can be found in [7].

This model has close parallels with Hierarchical Generative Models of neural ensembles trained using Bayesian inference [19], where the amplitude of the top layer $A_3(t,l)$ represents the probability distribution of the inverse of the pitch l at each instant t. Therefore the final pitch value prediction of the model at each instant is given by the inverse of the value of l for which $A_3(t,l)$ reaches its maximum [7].

Moreover, the three stages of the model represent, in a simplified fashion, physiological steps of auditory processing: the first model stage is assumed to represent peripheral areas, the second stage would correspond to sub-thalamic neural populations [17] and the third stage can be located more centrally in the brain. The model is consistent with the available neuroimaging data: a sustained pitch response (SPR) in lateral Heschl's gyrus has been shown to adapt to the recent temporal context of a pitch sequence, enhancing the response to rare and brief events [20]. In fact, a smoothed version of the model's top layer response derivative correlated remarkably well with a transient neuromagnetic response in Heschl's gyrus (termed Pitch Onset Response [15]). This parallelism suggests that further correlations between the top layer of the model and physiological measurements could be found, as we will demonstrate in the following MEG study.

III. Materials and analyses

A. Experimental set

Nineteen normal hearing subjects without any history of audiological or neurological deficits participated in the study. All subjects were familiar with MEG recordings and psychoacoustic procedures. Measurements were approved by a local

ethics committee and conducted with an informed consent of each subject.

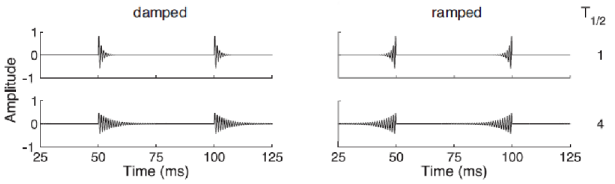

Fig. 2. Waveforms of damped (left) and ramped (right) tones for two of the different $T_{1/2}$ used in the experiments (time in ms).

We used ramped and damped sinusoids as stimuli (see Figure 2) generated according to the specifications by Patterson et al. [1] using a 1000 Hz carrier. The length of one cycle was set to 50 ms to ensure that the discontinuity in the envelope at the end of each modulation cycle occurs at an upward-going zero-crossing of the carrier. Half-life times ($T_{1/2}$) of the exponential modulator were 0.5 ms, 1 ms, 4 ms, 16 ms and 32 ms, respectively. To obtain approximately constant loudness for all conditions the amplitude was normalised by a factor proportional to the square root of the stimulus $T_{1/2}$ [1]. Stimuli were presented diotically at an intensity level of 65 dB. The order of the stimuli was randomised.

We concatenated 20 modulation periods adding up to a total duration of 1 s train stimuli. Inter-stimulus interval was set to 1.0 s−1.1 s. The session contained 120 repetitions for each condition.

B. Perceptual Measurements

Psychoacoustic measurements of the paired comparison task were carried out separately using the identical temporally asymmetric sounds as in the MEG experiments. The stimuli were presented in a single block of trials, presenting all possible combinations of pairs of non-identical stimuli (45) twice, such that a specific pair was presented in both orders. Thus, the psychoacoustic test consisted of 90 trials. For each trial, listeners had to indicate in a two-alternative task which sound of the pair was more tonal. After a training session, blocks were run just once. A scale for the relative pitch salience was derived from the results of the paired comparison experiment using the Bradley-Terry-Luce (BTL) method [21]. This method allows to order the carrier saliences of the temporally asymmetric stimuli on a perceptual scale.

C. Data Recording and Processing

The gradient of the magnetic fields were acquired with a Neuromag 122 whole-head MEG system inside of a magnetically shielded room. Subjects were exposed to the acoustical stimuli while watching a silent film of their own choice. MEG preprocessing was performed using standard procedures in the field [22]. Neuromagnetic fields were averaged over an epoch from −500 ms to 1500 ms after tone onset (See Figure 4). Spatio-temporal analysis was performed using a two-dipole model with one dipole in each hemisphere [22].

D. Model-driven analysis

Our analysis was performed for the ten (five different $T_{1/2}$ for each, ramped and damped, shapes) stimuli considered in the experimentation. For each of them, we matched the response of the model's top layer at the pitch value prediction, $A_3(t, l_{\text{pred}})$, to the amplitude of N100m MEG recordings of all the subjects. For the fitting, we proposed a linear relationship between the amplitude of the model and the amplitude of the MEG signal. This linear mapping is meant to reflect monotonic transformations occurring between the neuroelectric activity and the signal captured in the scalp [23].

N-fold cross validation was used to robustly compute the parameters of such linear transformation: we performed an individual fitting for each of the $N = 19$ subjects in the experimentation; each of those fittings was tested over the signal of the remaining $N - 1$ subjects, yielding to a total of $N(N - 1)$ tests per stimuli, enabling a robust statistical assessment.

IV. Results

A. Perceptual results

Pitch salience values are presented in Figure 3a as a function of the stimulus' envelope's $T_{1/2}$. Larger $T_{1/2}$ produced a higher pitch salience for both ramped and damped sounds. Ramped tones were generally judged as more salient than their damped counterparts. However, approaching the extremal half-life time values 0.5 ms and 32 ms the curves overlap, showing a maximum difference for the critical value $T_{1/2} = 4$ ms. The relatively small size of the standard errors indicate a noticeable strong agreement across subjects.

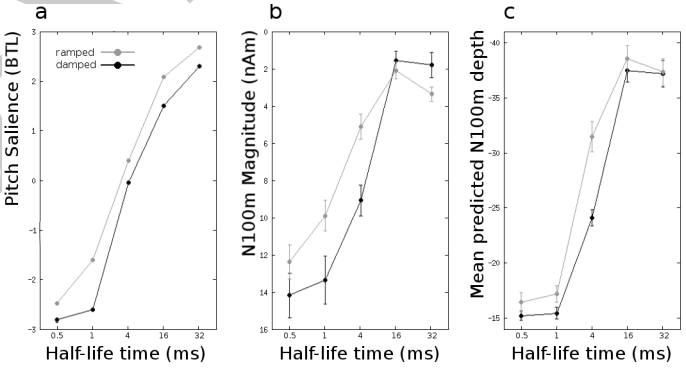

Fig. 3. Comparison of (a) pitch salience, (b) measured N100m amplitude and (c) HGPM prediction of the N100m for ramp (light grey) and damp (dark grey) stimuli for different $T_{1/2}$.

B. Analysis of the MEG recordings

The stimuli of the experimentation evoked transient cortical responses and sustained and steady state fields as shown in Figure 4. N100m sources were localised at the border between the lateral Heschl's gyrus and planum temporale (left: $x = -48(\pm 10)$, $y = -27(\pm 6)$, $z = 9(\pm 8)$; right: $x = 50(\pm 6)$, $y = -22(\pm 10)$, $z = 9(\pm 5)$; brackets indicate standard error). The peak amplitude of all conditions increased with the $T_{1/2}$ of the stimuli (Figure 3b), whereas N100m latencies decreased with the increasing $T_{1/2}$.

C. Model-driven analysis

Predictions obtained with the model were fully in line with the magnitude and latency of the N100m. To measure

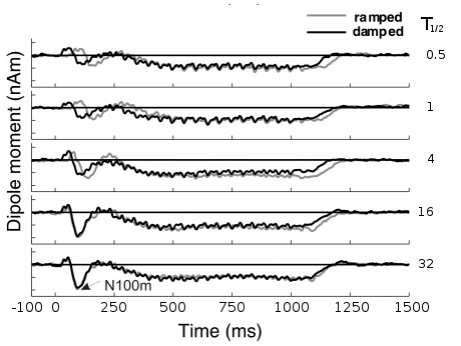

Fig. 4. Evoked neuromagnetic fields to ramped and damped sinusoids, shown as the grand mean source waveforms over subjects and hemispheres for stimuli with different envelope's $T_{1/2}$.

the fitting of the recordings with the prediction we computed the Pearson correlation coefficients and the root-mean-square errors (RMSE) between the two signals. Relatively high Pearson coefficients and small RMSEs were observed, with only a small bias error for short $T_{1/2}$, indicating that the model has a high ability to emulate the cortical MEG response to the given stimuli. Therefore we used the model to simulate the amplitude of the N100m characterising the ramp and damp stimuli. These predictions are shown in Figure 3c for the 10 different stimuli used in the experiment.

V. Conclusions

In this study we have analysed auditory evoked fields using a model-driven approach, comparing three different measures: the N100m observed in the MEG recordings, the N100m prediction of the hierarchical pitch perception model with top-down modulation and the perceived pitch salience for ramp and damp sounds with different envelopes, parametrised by the decaying half-life time $T_{1/2}$.

The main finding of this work, is that these three measures (N100m, pitch salience and model prediction) are closely related to each other. This suggests that the N100m component magnitude provides a physiological representation of temporal asymmetry in lateral Heschl's gyrus. From Figure 3 also follows that processing differences between ramped and damped stimuli are maximum for $T_{1/2} = 4\,\text{ms}$. This observation is in agreement with previous studies considering ramped and ramped stimuli (e.g. [1] or [24]).

Furthermore, the accuracy of the model predictions, robustly cross-validated across a large population, suggests that temporal asymmetry encoding is mediated by a hierarchical processing with top-down driven integration windows. These results provide further evidence for stimulus-specific temporal integration, which is sensitive to subtle differences in input stimuli and can thus potentially explain temporal asymmetry in auditory perception.

References

[1] R. Patterson, "The sound of a sinusoid: Spectral models," *The Journal of the Acoustical Society of America*, vol. 96, no. 3, pp. 1409–1418, 1994.

[2] J. W. Gordon, "The perceptual attack time of musical tones." *The Journal of the Acoustical Society of America*, vol. 82, no. 1, pp. 88–105, Jul. 1987.

[3] W. Hartmann, "The effect of amplitude envelope on the pitch of sine wave tones," *The Journal of the Acoustical Society of America*, vol. 63, no. 4, pp. 1105–1113, 1978.

[4] G. C. Stecker and E. R. Hafter, "An effect of temporal asymmetry on loudness," *The Journal of the Acoustical Society of America*, vol. 107, no. 6, pp. 3358–3368, 2000.

[5] E. Balaguer-Ballester, S. L. Denham, and R. Meddis, "A cascade autocorrelation model of pitch perception." *The Journal of the Acoustical Society of America*, vol. 124, no. 4, pp. 2186–95, Oct. 2008.

[6] S. Kumar, W. Sedley, K. V. Nourski, H. Kawasaki, H. Oya, R. D. Patterson, M. A. H. III, K. J. Friston, and T. D. Griffiths, "Predictive Coding and Pitch Processing in the Auditory Cortex." *J. Cognitive Neuroscience*, vol. 23, no. 10, pp. 3084–3094, 2011.

[7] E. Balaguer-Ballester, N. R. Clark, M. Coath, K. Krumbholz, and S. L. Denham, "Understanding pitch perception as a hierarchical process with top-down modulation." *PLoS computational biology*, vol. 5, no. 3, p. e1000301, Mar. 2009.

[8] R. D. Patterson, S. Uppenkamp, I. S. Johnsrude, and T. D. Griffiths, "The processing of temporal pitch and melody information in auditory cortex." *Neuron*, vol. 36, no. 4, pp. 767–76, Nov. 2002.

[9] M. Ahissar, M. Nahum, I. Nelken, and S. Hochstein, "Reverse hierarchies and sensory learning." *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 364, no. 1515, pp. 285–99, Feb. 2009.

[10] S. Hochstein and M. Ahissar, "View from the Top : Hierarchies and Reverse Hierarchies Review," *Neuron*, vol. 36, no. 5, pp. 791–804, 2002.

[11] A. M. Bastos, W. M. Usrey, R. a. Adams, G. R. Mangun, P. Fries, and K. J. Friston, "Canonical microcircuits for predictive coding." *Neuron*, vol. 76, no. 4, pp. 695–711, Nov. 2012.

[12] C. J. Plack and L. J. White, "Perceived continuity and pitch perception," *The Journal of the Acoustical Society of America*, vol. 108, no. 3 Pt 1, pp. 1162–1169, 2000.

[13] I. Nelken, "Processing of complex sounds in the auditory system." *Current opinion in neurobiology*, vol. 18, no. 4, pp. 413–7, Aug. 2008.

[14] B. Lütkenhöner and O. Steinsträter, "High-precision neuromagnetic study of the functional organization of the human auditory cortex," *Audiology and Neurotology*, vol. 3, no. 2-3, pp. 191–213, 1998.

[15] K. Krumbholz and R. Patterson, "Neuromagnetic evidence for a pitch processing center in Heschl's gyrus," *Cerebral Cortex*, vol. 13, no. 7, pp. 765–772, 2003.

[16] A. Seither-Preisler, R. Patterson, K. Krumbholz, S. Seither, and B. Lütkenhöner, "Evidence of pitch processing in the N100m component of the auditory evoked field." *Hearing research*, vol. 213, no. 1-2, pp. 88–98, Mar. 2006.

[17] R. Meddis and L. P. OMard, "Virtual pitch in a computational physiological model," *The Journal of the Acoustical Society of America*, vol. 120, no. 6, p. 3861, 2006.

[18] R. Meddis and L. O'Mard, "A unitary model of pitch perception." *The Journal of the Acoustical Society of America*, vol. 102, no. 3, pp. 1811–20, Sep. 1997.

[19] K. Friston, "A theory of cortical responses." *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 360, no. 1456, pp. 815–36, Apr. 2005.

[20] A. Gutschalk, R. D. Patterson, M. Scherg, S. Uppenkamp, and A. Rupp, "The effect of temporal context on the sustained pitch response in human auditory cortex." *Cerebral cortex (New York, N.Y. : 1991)*, vol. 17, no. 3, pp. 552–61, Mar. 2007.

[21] H. David, *The method of paired comparisons*. New York: Oxford University Press, 1963.

[22] M. Scherg, "Fundamentals of dipole source potential analysis," *Advanced in Audiology*, vol. 6, pp. 40–69, 1990.

[23] S. Williamson and L. Kaufman, "Biomagnetism," *Journal of Magnetism and Magnetic Materials*, vol. 22, pp. 129–201, 1981.

[24] D. Pressnitzer, I. M. Winter, and R. D. Patterson, "The responses of single units in the ventral cochlear nucleus of the guinea pig to damped and ramped sinusoids." *Hearing research*, vol. 149, no. 1-2, pp. 155–66, Nov. 2000.