

# Insights on the neuromagnetic representation of temporal asymmetry in human auditory cortex

---

**Tabas, Alejandro; Siebert, Anita; Supek, Selma; Pressnitzer, Daniel; Balaguer-Ballester, Emil; Rupp, Andre**

Source / Izvornik: **PLoS One, 2016, 11**

**Journal article, Published version**

**Rad u časopisu, Objavljena verzija rada (izdavačev PDF)**

<https://doi.org/10.1371/journal.pone.0153947>

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:217:268887>

Rights / Prava: [Attribution 4.0 International](#)/[Imenovanje 4.0 međunarodna](#)

Download date / Datum preuzimanja: **2024-12-20**



Repository / Repozitorij:

[Repository of the Faculty of Science - University of Zagreb](#)



RESEARCH ARTICLE

# Insights on the Neuromagnetic Representation of Temporal Asymmetry in Human Auditory Cortex

Alejandro Tabas<sup>1\*</sup>, Anita Siebert<sup>2</sup>, Selma Supek<sup>3</sup>, Daniel Pressnitzer<sup>4</sup>, Emili Balaguer-Ballester<sup>1,5‡</sup>, André Rupp<sup>6‡</sup>

**1** Faculty of Science and Technology, Bournemouth University, Bournemouth, England, United Kingdom, **2** Institute of Pharmacology and Toxicology, University of Zurich, Zürich, Zürich, Switzerland, **3** Department of Physics, Faculty of Science, University of Zagreb, Zagreb, Croatia, **4** Département d'Études Cognitives, École Normale Supérieure, Paris, France, **5** The Bernstein Center for Computational Neuroscience Heidelberg-Mannheim, Mannheim, Baden-Württemberg, Germany, **6** Department of Neurology, Heidelberg University, Heidelberg, Baden-Württemberg, Germany

‡ These authors are joint last authors on this work.

\* [atabas@bournemouth.ac.uk](mailto:atabas@bournemouth.ac.uk)



CrossMark  
click for updates

 OPEN ACCESS

**Citation:** Tabas A, Siebert A, Supek S, Pressnitzer D, Balaguer-Ballester E, Rupp A (2016) Insights on the Neuromagnetic Representation of Temporal Asymmetry in Human Auditory Cortex. PLoS ONE 11 (4): e0153947. doi:10.1371/journal.pone.0153947

**Editor:** Warren H Meck, Duke University, UNITED STATES

**Received:** November 24, 2015

**Accepted:** April 6, 2016

**Published:** April 20, 2016

**Copyright:** © 2016 Tabas et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data is available at the following link in the OpenScience Framework: [osf.io/zvy2a](https://osf.io/zvy2a).

**Funding:** AT receives funding from the Bournemouth University Doctoral Fund ([bournemouth.ac.uk](http://bournemouth.ac.uk)) and the Santander Research Mobility Award program ([santander.co.uk](http://santander.co.uk)). AR and AS were funded by the Deutsche Forschungsgemeinschaft ([dfg.de](http://dfg.de)) Grant Ru-652/1-3. SS was supported by the Croatian Ministry of Science and Technology ([public.mzos.hr](http://public.mzos.hr)) Grant 0119265. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Communication sounds are typically asymmetric in time and human listeners are highly sensitive to this short-term temporal asymmetry. Nevertheless, causal neurophysiological correlates of auditory perceptual asymmetry remain largely elusive to our current analyses and models. Auditory modelling and animal electrophysiological recordings suggest that perceptual asymmetry results from the presence of multiple time scales of temporal integration, central to the auditory periphery. To test this hypothesis we recorded auditory evoked fields (AEF) elicited by asymmetric sounds in humans. We found a strong correlation between perceived tonal salience of ramped and damped sinusoids and the AEFs, as quantified by the amplitude of the N100m dynamics. The N100m amplitude increased with stimulus half-life time, showing a maximum difference between the ramped and damped stimulus for a modulation half-life time of 4 ms which is greatly reduced at 0.5 ms and 32 ms. This behaviour of the N100m closely parallels psychophysical data in a manner that: i) longer half-life times are associated with a stronger tonal percept, and ii) perceptual differences between damped and ramped are maximal at 4 ms half-life time. Interestingly, differences in evoked fields were significantly stronger in the right hemisphere, indicating some degree of hemispheric specialisation. Furthermore, the N100m magnitude was successfully explained by a pitch perception model using multiple scales of temporal integration of auditory nerve activity patterns. This striking correlation between AEFs, perception, and model predictions suggests that the physiological mechanisms involved in the processing of pitch evoked by temporal asymmetric sounds are reflected in the N100m.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

Waveforms of sound sources like speech and music are typically asymmetric in time. The term *temporal asymmetry* [1] has been used to describe auditory stimuli that display different *attack* (sound onset) and *decay times* (sound offset). For example, striking cymbals produce a rapid attack followed by an exponential decay in the waveform amplitude, whereas bowing the same instrument results in a more gradual attack. Thus, temporal asymmetry influences the timbre of a stimulus considerably [2]. The temporal envelope also contributes substantially to the identification of instruments: when instruments producing sounds of asymmetrically shaped temporal envelopes are played backwards, humans often fail to identify the instrument [3]. Furthermore, it is well known that differences in attack and decay times affect perceptual timing [4–6] and duration [7], pitch [8], and loudness [9].

Ramped and damped stimuli [1, 10] enable us to study temporal asymmetry in a systematic fashion. This family of stimuli consists of a sinusoid multiplied either by a periodically rising (*ramped*) or decaying (*damped*) exponential function (see Fig 1). Thus, stimuli present two different periodicities that are perceived simultaneously: the periodicity of the carrier (the fundamental frequency of the pure tone before modulation) and the *envelope's* one (the periodicity of the modulation pattern). Ramped and damped sinusoids evoke different perceptions: ramped sounds are perceived as continuous tones with the pitch of the carrier, whereas repetitive streams of damped sinusoids are perceived as a drumming sound with a lower pitch salience.

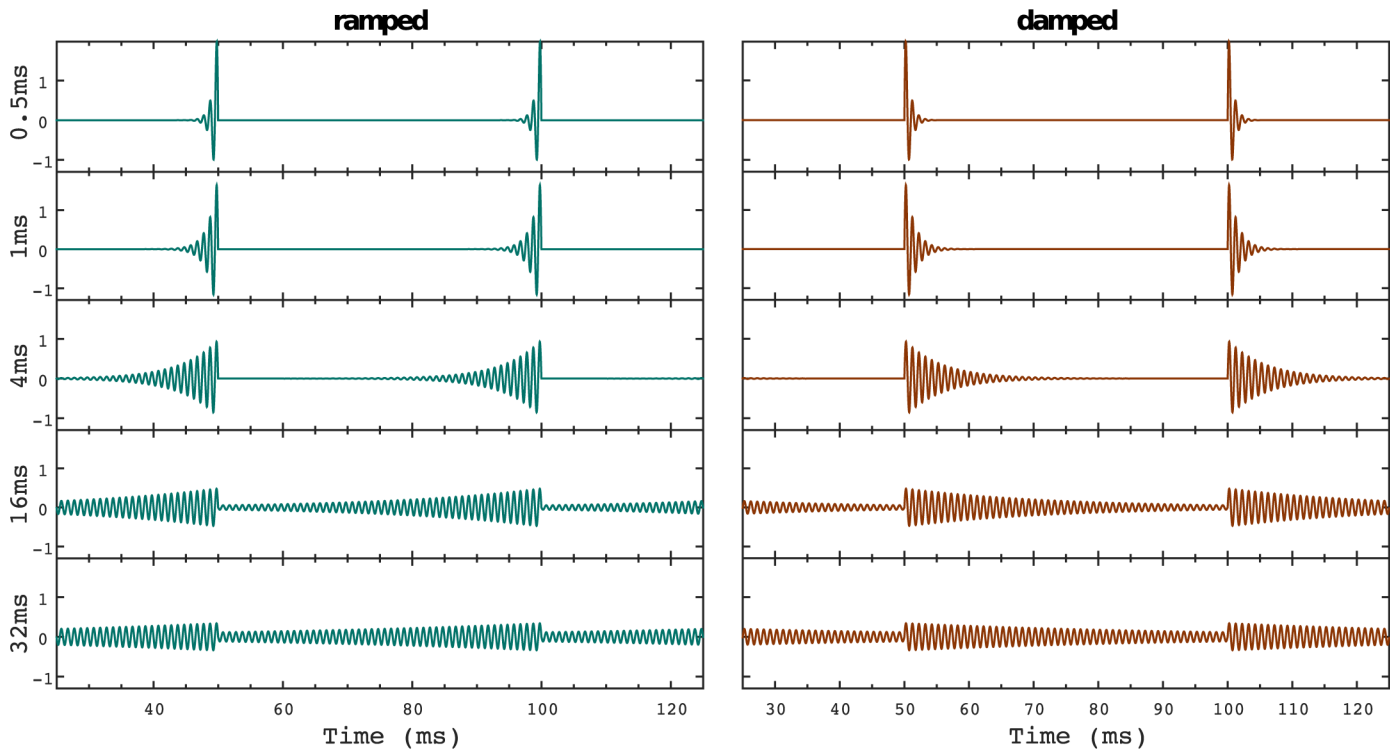
These stimuli pose an interesting problem for the understanding of temporal processing in the auditory system because their long-term Fourier spectra are identical. Hence, models of auditory perception, essentially based on extracting the auditory nerve periodicities on a fixed, and often long, time window (see, e.g., [11–13]) cannot fully explain such perceptual differences.

Here we propose that models incorporating stimulus-dependent adaptive processing of the auditory nerve activity patterns provide an insight into perceptual asymmetry phenomena. One of the earliest models of this kind is the Auditory Image Model (AIM) [14], which simulates the representation of sounds beyond the auditory nerve using an adaptive mechanism for temporal integration called *strobed temporal integration*. This non-linear transform converts the activity pattern of the auditory nerve into the so-called *stabilised auditory image* (SAI), which correlates to the perceived pitch and salience of ramped and damped sounds [15].

More recently, empirical and modelling studies [16–18] offered further evidence of the existence of a stimulus-driven adaptation of the temporal integration window. In a recent model of pitch perception [17] this adaptation was proposed to explain that, while long integration windows are necessary to understand a wide range of perceptual phenomena (e.g. [19–21]), short integration windows are necessary to identify quick variations in the inputs stimuli on the millisecond range [22]. This balance between perceptual integration and resolution was achieved by a top-down modulation process which is sensitive to quick stimulus variations, such as those occurring in temporally asymmetric sounds [17].

Neurophysiological responses to ramped and damped sinusoids have been analysed both in subcortical and in cortical structures (e.g., in ventral cochlear nucleus [23], in inferior colliculus [24] and in primary auditory cortex [25]). Taken together, these electrophysiological studies demonstrate the consistency of the temporal asymmetry in single-unit responses with the perceptual asymmetry. However, those animal studies did not attempt to identify the causal physiological correlate of the perception elicited by damped and ramped sounds in human listeners [24].

In the present work, we combine non-invasive magnetoencephalography (MEG) and perceptual studies in human listeners with pitch perception models in order to better understand



**Fig 1. Waveforms of the ramped and damped sinusoids.** Ramped (left) and damped (right) sinusoidal waves with half-life times ( $T_{1/2}$ ) of 0.5, 1, 4, 16, and 32 ms used in the experiment. Note the two periodicities present in the stimuli corresponding to the carrier (1000 Hz) and the repetition period (20 Hz) of the ramped/damped modulation.

doi:10.1371/journal.pone.0153947.g001

the processing of asymmetric sounds. We identified a neuromagnetic representation of auditory perceptual asymmetry in the morphology of the N100m deflection of the auditory evoked fields (AEF). The N100m is a well-known transient neuromagnetic response elicited 100 ms after the tone onset. This deflection arises from multiple sources in auditory cortex, lateral Heschl's gyrus and planum temporale [26, 27]. As the N100m is sensitive to the intensity [28, 29] and the rise-time of the sound [30], it is often regarded as an energy-onset response. However, there is evidence that the deflection is also sensitive to the other stimulus features, such as spectral composition [31], pure tone frequency [32, 33], fundamental frequency of harmonic tones [34] and temporal pitch extraction [27, 35, 36]. Furthermore, amplitudes of the N100m increase with pitch salience [27, 37] and fMRI studies show a correlate of the pitch salience with BOLD-responses in the non-primary auditory cortex [38]. Models with multiple dipoles have been successfully used to separate specific energy and pitch responses [39].

In the present study we propose that N100m morphology reflects the processing of temporal asymmetry in auditory cortex. We hypothesise that amplitude differences observed in the transient response can be explained using adaptive, stimulus-dependent, windows of integration. Towards this goal, we first demonstrate the correlation between N100m amplitude and the perceived asymmetry in the salience of ramped and damped sinusoids. Second, we show how pitch perception models using adaptive integration windows can account for the processing mechanisms underlying the N100m deflection during temporal asymmetry perception. Our results suggest that the auditory system is capable to discern those two different sounds by continuously adapting the integration window of perceptual integration. Moreover, data also

shows that temporal asymmetry is much more strongly represented in the right hemisphere than in the left hemisphere.

## Materials and Methods

### Experimentation

The study and all the measurements were approved by the ethics committee of the Heidelberg University's Medical School and conducted with written informed consent of each subject.

**Subjects.** 13 subjects were included in the perceptual study (aged between 24 and 37 years old) and 27 subjects participated in the neurophysiological experiment (aged 22-44 years old). All subjects reported normal hearing and had no history of audiological or neurological deficits. All of them were familiar with MEG recordings and psychoacoustic procedures. Measurements were approved by the local ethics committee and conducted with informed consent of each subject.

**Stimuli.** Experimental stimuli were ramped and damped sinusoids (see Fig 1) generated according to the parameter specifications described in [1] using a 1000 Hz carrier and an exponential amplitude envelope given by:

$$E(t) = \sqrt{\frac{1}{T_{1/2}}} e^{-\frac{t \log 2}{T_{1/2}}} \quad (1)$$

Fig 1 illustrates two cycles of the modulated sinusoids. Stimuli consisted of a total concatenation of 20 cycles, as has been commonly done in human psychophysics and animal recordings using ramped and damped sinusoids [1, 23]. 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, so all stimuli present the same onset phase. Therefore, stimuli duration added up to a total of 1 s.

Half-life times ( $T_{1/2}$ ) of the modulator were 0.5 ms, 1 ms, 4 ms, 16 ms and 32 ms, respectively. To obtain approximately constant loudness for all conditions and minimise undesirable artefacts on the neuromagnetic signal, the amplitude was normalised by a factor proportional to the square root of the stimulus half life time [1].

**Perceptual measurements.** Psychoacoustic measurements of the paired comparison task were carried out using the temporally asymmetric sounds described above. Sounds were delivered through K240-DF headphones (AKG Acoustics, Vienna, Austria) at a level of 65 dB (SPL). Stimuli were presented in a single block of trials for each part of the experiment and listener. In each block, all possible combinations of pairs of non-identical stimuli (45) were presented in both orders. Thus, the psychoacoustic test consisted of 90 trials per block. For each trial, listeners had to indicate in a two-alternative task without feedback 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 [40]. This method allows to order the carrier salience of the temporally asymmetric stimuli on a perceptual scale. To analyse the results, we used the temporal asymmetry index defined in Eq (2) with  $x = S(AI_S)$ , where  $S$  denotes the relative pitch salience measured with the BTL method.

**Neuromagnetic data recording and processing.** Stimuli were presented diotically at an intensity level of 65 dB SPL using ER-3 transducers (Etymotic Research, Inc., Elk Grove Village, IL) connected to 90 cm plastic tubes and foam ear pieces. The sampling rate was set to 48 kHz. The order of the stimuli was randomized. ISIs were randomised between 1.0–1.1 s. The MEG session consisted of 120 trials for each condition.

Gradients of the magnetic field were acquired with a Neuromag 122 whole-head MEG system (Elekta Neuromag Oy, Helsinki, Finland) inside of a magnetically shielded room (IMEDCO, Hägendorf, Switzerland). Subjects sat in an upright position and watched a silent film of their own choice. Since neural mechanisms underlying pitch processing seem to evoke equivalent fields on attentive and inattentive subjects [27, 41], we chose to separate the psychophysical task from the MEG recordings in order to maximise the number of trials per session. Note that animal recordings on the same stimuli (e.g. [23]) were performed under anaesthesia and obviously also without a task.

The sampling rate was 1000 Hz and a bandwidth ranging from 0.01 Hz to 330 Hz. Auditory evoked fields were averaged over an epoch of from -500 ms to 1400 ms. Off-line averaging with artefact monitoring was performed using BESA 5.1 software (BESA Software, Gräfelfing, Germany). Epochs containing signals exceeding an absolute level of 8000 fT/cm and a gradient of 800 fT/cm per sample were discarded automatically, resulting in about 5% rejection rate. The baseline was calculated over the 100 ms interval prior to tone onset.

T1-weighted magnetic resonance images (MRI) were obtained from 10 of the listeners on a Magnetom Symphony 1.5 Tesla scanner (Siemens, Erlangen, Germany). Scans were performed in 176 sagittal slices yielding an isotropic voxel size of 1 mm<sup>3</sup>. Three-dimensional reconstructions were computed using the BrainVoyager software (version 4.4, Brain Innovation, Maastricht, The Netherlands). Dipole positions for these subjects were co-registered onto the individual MRI and then transformed into the standard space of Talairach [42] to illustrate the location of the generators (see, for instance, [43]). Since MRI images were not available for the remaining subjects, the spherical model was used without co-registration for 17 of the listeners. This method typically yields accurate locations for the N100m dipoles.

**Neuromagnetic data analysis.** In order to setup a model for the N100m we applied a spatio-temporal model [41] with one equivalent dipole per hemisphere. Dipole fits were based on the pooled 16 ms and 32 ms ramped and damped conditions since these sounds elicited a clear tonal percept. Fits were performed using unfiltered data and the fitting interval was about 30 ms around the peak of the N100m for each subject. A symmetry constraint was applied in 8 of 27 subjects. No further constraints concerning orientation or location of the equivalent dipoles were used. This method provided stable models for all subjects and was used as a spatio-temporal filter to derive the source waveforms of all 10 conditions. A principal component was computed over the last 100 ms of the epoch for each condition in order to compensate for drift artefacts [44]. This procedure, applied to each subject, assumed that the N100m response is evoked by the same generators in the auditory cortex; i.e. that the location and orientation of the equivalent dipole remain constant. This assumption is reasonable for this family of stimuli, since they all evoked the same pitch value.

Identical procedures were followed to compute an equivalent dipole model for the sustained field (SF), but the asymmetry constraint was applied only in 1 of the 27 subjects. The interval used to fit the SF dipoles covered the DC portion of the field, spanned from 500 ms to 1000 ms after tone's onset.

In order to quantify ramped/damped asymmetry we used the temporal asymmetry index ( $AI_x$ ) [14]:

$$AI_x = \frac{(x_r - x_d)}{(x_r + x_d)} \quad (2)$$

where  $x_r$  and  $x_d$  denote the magnitude associated with ramped and damped stimuli respectively. To quantify the asymmetric behaviour in the amplitude of the N100m, we used Eq (2) with  $x = M$ , the amplitude of such component in the measured evoked fields.

Individual source waveform estimates were used to assess the N100m difference between conditions. Peak amplitudes were assessed using such averaged waveforms. Critical  $t$ -intervals were computed using the resulting distribution of the minima and the surrounding points in a 15 ms interval for each subject. A similar procedure was used to assess the properties of the sustained field, pooling the data points on the interval spanned between 800 ms and 1000 ms after tone's onset.

## Modelling

We attempted to model the relationship between the dynamics of the N100m and perception by using two complementary models of pitch perception employing stimulus-dependent integration windows. Software for both models is freely available in <http://www.pdn.cam.ac.uk/groups/cnbh/research/aim.php> for the AIM and in <http://sourceforge.net/projects/topdownpitchmodel/> for the GPM.

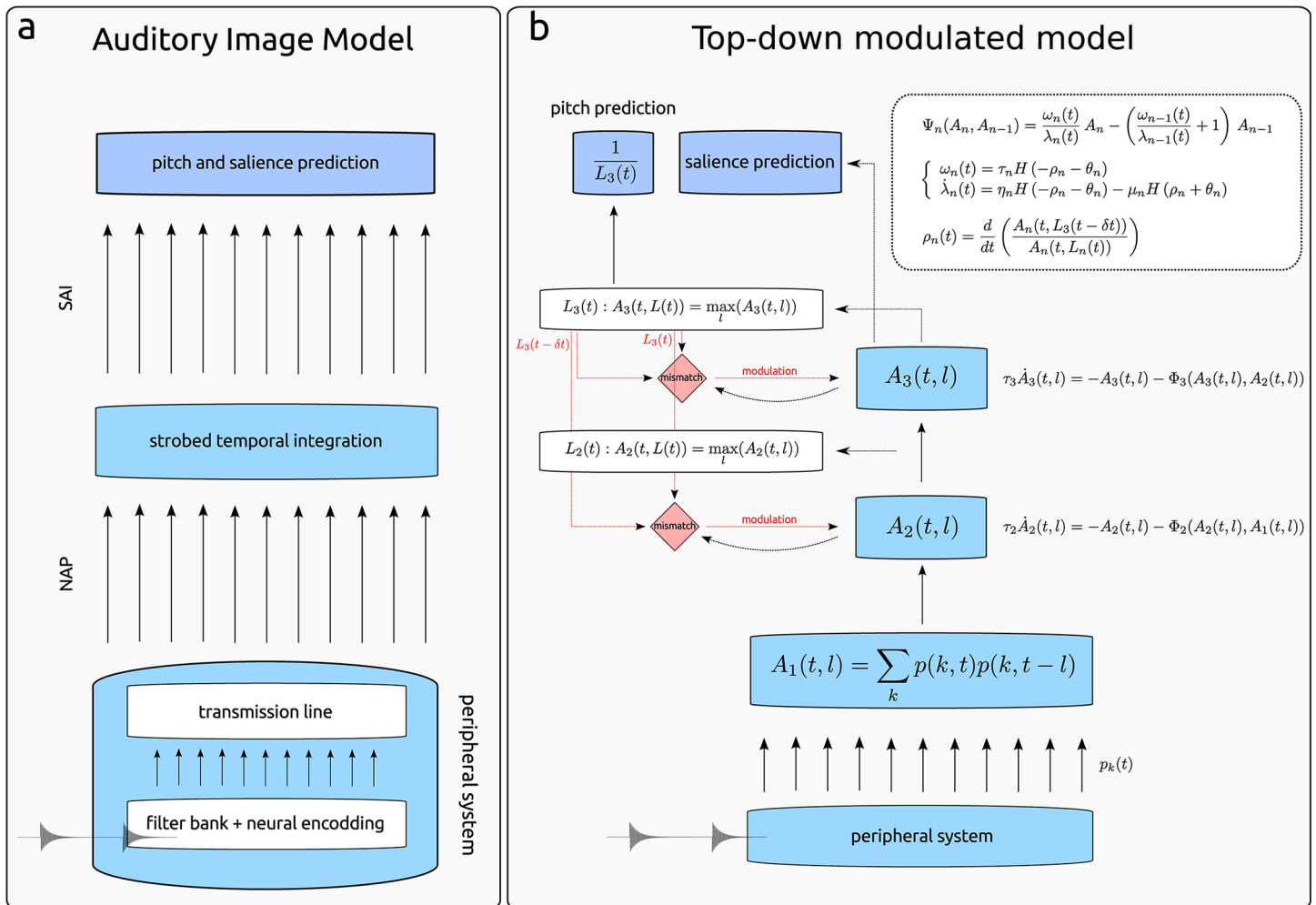
**Auditory Image Model.** The Auditory Image Model (AIM) [14] consists of three sequential transforms associated with three different processing stages of the ascending auditory pathway, two at the peripheral auditory system and one at a central stage as illustrated in Fig 2a. AIM was originally designed to simulate a highly idealized neural representation of auditory stimuli, assumed to underlie the first conscious awareness of sounds [14].

The first stage of AIM uses a non-linear transmission-line filter bank, accounting for the spectral analysis performed in the cochlea in the range of 100–10000 Hz [45]. The simulations in the present study were carried out using 100 channels. During the second stage, the basilar membrane motion is converted into a multi-channel pattern which simulates the neural activity (NAP) of the auditory nerve (shown for ramped and damped sounds in Fig 3a). In the third stage of AIM, the spike probability  $p(t, k)$  for each channel  $k$  at each time point  $t$  is transformed into an interpretable representation: the *stabilised auditory image* (SAI).

This last transformation is carried out by means of a mechanism called *strobed temporal integration*, operating independently in each cochlear channel by transforming a spike-train signal into a time-interval signal. When a pulse is detected in the spike train of each channel (i.e. when the value of the signal exceeds some adaptive threshold which is asymmetric in time, see Fig 3b) the signal is copied point by point into the buffer. This mapping continues until a new pulse exceeds the adapted threshold. Then a new strobe is triggered and the signal is transferred to the buffer. The buffer decays within 30 ms. This decay allows the system to respond to rapid stimulus changes. The integrated SAI provides complete information about the perceived pitch and its strength. Time-interval repetitions along time are represented in the first peak in the SAI, whose position in the time-interval space represents the perceived pitch of the stimulus [46]. Similarly, the ridge height is related to its pitch strength [47]. Therefore, we can use the mean value across cochlear channels of the height of the first peak of the SAI to extract a prediction of the perceptual pitch salience from the model. Note that, as a consequence of the peripheral preprocessing and the adaptive strobed integration, the SAI is not a simply spectral decomposition of the waveform of the stimuli, but the result of an elaborated nonlinear transformation that reflects the pitch elicited by the stimuli [46].

A key feature of the peak detection during the strobing process is that the threshold is adaptive, such that the rising envelopes of ramped sounds provide multiple snapshots of activity in a channel whereas decaying envelopes as given by damped sounds exhibit just a relative small number of strobos. This effect is illustrated in Fig 3b and 3c, which show the different strobing for a sinusoid with  $T_{1/2} = 4\text{ms}$  and the resulting SAI with a much larger peak height for the ramped sound.

**Top-down modulated model.** A hierarchical model of interacting neural ensembles incorporating a top-down modulation process (top-down modulated model of pitch perception, or



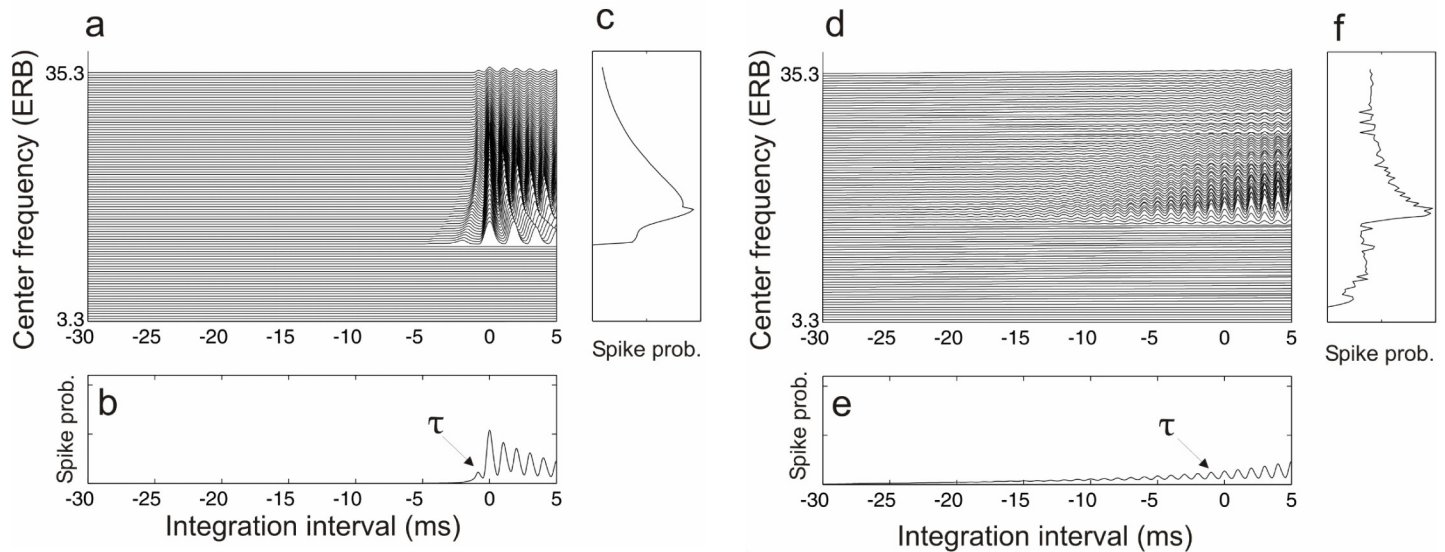
**Fig 2. Schematic diagram of the Auditory Image Model and the Top-down Modulated Hierarchical Model of Pitch.** a) Schematic view of the Auditory Image Model (AIM) [14]. In the first stage, peripheral auditory filters transform the input waveform into a multi-channel representation of basilar membrane motion. The next stage applies a hair cell model and converts this motion into a neural activity pattern in the auditory nerve (NAP). In the final stage, this signal is used to produce a stabilised representation of the stimuli by means of strobed temporal integration. The output of this process is termed the stabilised auditory image (SAI) of the input stimulus. b) Schematic view of the top-down modulated Hierarchical Generative Model of pitch perception (GPM) [17]. The peripheral processing is similar to the one in AIM (bottom). The next step consists of a coincidence detection process of auditory nerve activity patterns for different cochlear delay lines  $l$ ,  $A_1(t, l)$ . Further processing is carried out by two consecutive ensemble models  $A_2$  and  $A_3$  performing leaky integrations of input activity using time-varying integration windows. Such ensembles correspond putatively to pre-thalamic and central auditory areas. A top-down, stimulus-dependent mechanism modulates the size of the effective integration windows of bottom-up information.

doi:10.1371/journal.pone.0153947.g002

*generative* pitch model in short, GPM [17, 48]) was used for further analysing the role of adaptive integration windows in the perception of ramped and damped sinusoids.

Similarly to AIM [14], and to the so-called autocorrelation models of pitch [11, 13], the top-down model receives its input from the hair cell transduction model [49], which generates the auditory-nerve spike probabilities  $p(t, k)$  as a function of time  $t$  in each cochlear frequency channel  $k$ . The GPM consists of a cascade of three layers of activation with time-dependent outputs  $A_1$ ,  $A_2$  and  $A_3$ . The output of the first stage represents the probability of generating two





**Fig 3. Output of the auditory image model for the  $T_{1/2} = 4\text{ms}$  ramped and damped sinusoids.** Auditory Image Models' output for damped (a–c) and ramped (d–f) trains ( $T_{1/2} = 4\text{ms}$ ) at the time point of the same envelope height. Panels a) and d) show the stabilized auditory image (SAI) over time in each cochlear frequency channel. Panels c) and f) represent the spike probability averaged over time. Panels b) and e) show the summarised activity of all channels in the auditory image. The integration interval is the inverse of the carrier frequency applied [11], thus it shows a peak at  $\tau = -1\text{ms}$  in the figure. The height of this peak predicts the perceived carrier salience.

doi:10.1371/journal.pone.0153947.g003

spikes delayed by a certain lag  $l$  across all channels:

$$A_1(t, l) = \sum_k p(t, k)p(t - l, k) \quad (3)$$

The sum of this quantity for the stimulus onset  $t = 0$  to  $t$  weighted by an exponential decay function renders the summarized autocorrelation function (SACF) [11]. The value for the lag  $l$  where  $\text{SACF} = \sum_t A_1(t, l)$  reaches its maximum represents the pitch value in autocorrelation models [11, 13], whilst the pitch strength is often represented in the difference between  $\text{SACF}(t, l_{\max})$  and the value of  $\text{SACF}(t, l)$  at the second highest lag. However, these models fail to explain a large range of pitch phenomena [17] requiring a more realistic processing. In this model, this is solved using a leaky integration process implemented in the superior two layers, endowing a top-down mechanism in order to control the size of the integration windows (see Fig 2b). The integrators are implemented as a cascade of two highly idealised neural ensemble models [17, 50] with top-down recurrent connections modulating the size of the integration windows.

The activity at the second processing stage  $A_2(t, l)$  (see Fig 2b) is computed as a nonlinear leaky integrator of the activity at the previous stage  $A_1(t, l)$ , using a lag-dependent short time constant  $2\text{ms} \leq \tau_2 \leq 100\text{ms}$  [51]. This activity represents the firing rate of a set of auditory nerve fibres receiving inputs from different delays  $l$ . Overall, the output of this stage simply represents a periodicity extraction averaged along channels using a short exponential decay, like the one used in [11]. This stage mirrors processing carried out by sub-thalamic neural populations [49, 52, 53].

The subsequent, last third stage  $A_3(t, l)$  implements a low-pass filter of short-term periodicities encoded in  $A_2(t, l)$  using a long time scale  $\tau_3$  (typically,  $\tau_3 \geq 100\text{ms}$ ) and a nonlinear activation function which is briefly discussed in the next section. This processing is assumed to be

located more centrally in the brain. This kind of hierarchical architecture embodying multiple time scales is fully in line with observations of functional magnetic resonance imaging studies (e.g. [54, 55]).

Both integration stages are implemented as simple time-varying exponential averages:

$$A_n(t, l) = A_n(t - \Delta t, l)e^{-\frac{\Delta t}{E_n(t)}} + \frac{1}{g_n(t)} \frac{\Delta t}{\tau_n} A_{n-1}(t, l) \Delta t \quad n = 2, 3 \quad (4)$$

where  $\Delta t$  is the time step of the integration,  $g_n(t)$  is a normalisation factor and  $E_n(t)$  is the effective integration window of the  $n$ th stage, represented as the instantaneous exponential decay rate of the response at the  $n$ th integration stage ( $E_n(t) \leq \tau_n$  for  $n = 2, 3$  and  $E_1 \equiv \tau_1 = 1$ ).

Similarly to AIM, the lag in which the output at the final processing stage  $A_3(t, l)$  is maximum will be denoted as  $L_n(t)$  throughout the work, and will be referred to as the lag prediction. Therefore,  $1/L_3(t)$  represents the predicted pitch at time  $t$ . Equivalently, we define the expected pitch as the pitch prediction at the previous time step  $1/L(t-\Delta t)$  [17].

Crucially, the effective integration  $E_n(t)$  windows are not static. Instead they are adaptive and top-down driven, which permits to detect unexpected changes in the input stimulus (such as the offset of a tone in a sequence). In AIM (see [Methods](#)), information about past events is integrated until the auditory image is *stable*, and then the adaptation is performed with an exponential decay across time. Consistently, in the GPM model, the integration windows decay rapidly during periods where either there is a sudden discrepancy between the pitch prediction  $1/L_n(t)$  and expectation  $1/L_n(t-\Delta t)$  or there is a long sustained period with no discrepancies between them (see next section and [17] for further details).

**Parallels with neural ensemble models.** The idealized GPM can also be understood in terms of neural ensemble models. Taking the limit  $\Delta t \rightarrow dt$ , the modulated cascade of integrators is equivalent to a hierarchy of neural ensemble models of the Wilson-Cowan type [56]:

$$\tau_n \dot{A}_n(t, l) = -A_{n-1}(t, l) - \Psi_n(A_n(t, l), A_{n-1}(t, l)) \quad (5)$$

with the following activation function:

$$\Psi_n(A_n(t, l), A_{n-1}(t, l)) = \frac{\omega_n(t)}{\lambda_n(t)} A_n(t, l) - \left(1 + \frac{\omega_{n-1}(t)}{\lambda_{n-1}(t)}\right) A_{n-1}(t, l) \quad (6)$$

The gains  $\frac{\omega_n(t)}{\lambda_n(t)}$  in the activation function are modulated by the top-down mechanism, and at the same time modulate the effective integration windows in [Eq \(4\)](#):

$$E_n(t) = \frac{\tau_n}{1 + \frac{\omega_n(t)}{\lambda_n(t)}} \quad (7)$$

In the absence or deactivation of the top-down mechanism,  $\frac{\omega_n(t)}{\lambda_n(t)} = 0$  and the integration windows are set at a fixed time. The top-down mechanism gets activated when a mismatch between the expectation and prediction of pitch occurs, by setting the gains to a positive value and thus decreasing the size of the integration windows. Full details of the mechanism can be found in the original publication [17].

In summary, the shape of this model preserved certain constraints established in neural ensemble theory. This model has been shown capable of explaining a wide range of pitch perception phenomena, including the balance between temporal integration and resolution of pitch perception [17]. Thus, it is worth investigating whether it can predict effects of temporal asymmetry like AIM [14].

**Top-Down modulation and the N100m.** The GPM approach has been shown to be consistent with available neuroimaging data associated to the perception of Iterated Ripple Noise pitch [27]. More precisely, the derivative of the model output at the predicted pitch  $L_3(t)$ ,  $A_3(t, L_3)$ , was closely correlated with the latency of the N100m component of the evoked responses in antero-lateral Heschl's Gyrus (see [17, 27] for details).

Hence, in the present study, we evaluated the capacity of this model for further explaining electrophysiological results by comparing the dynamics of the top layer neural ensemble, representing activity in auditory cortex, with the morphology of the N100m response evoked by the each of the ramped and damped stimuli. The analysis was performed for the 10 stimuli considered in the experimentation (five different  $T_{1/2}$  for each, ramped and damped envelope; see Fig 1). For each of the sounds, we matched the response of the model's top layer at the pitch value prediction  $A_3(t, L_3(t))$  to the amplitude of the evoked response within a time window of 50 ms surrounding the N100m peak. To fit the peak, we proposed a linear relationship between the amplitude of the model and the amplitude of the MEG signal (see e.g. [57]).

$N$ -fold cross validation was used to robustly compute the parameters of the transformation: we performed an individual linear fitting for each of the  $N = 27$  subjects in the experimentation. Then, parameters of the linear fits were fixed and tested using the evoked fields of the remaining  $N - 1$  subjects, yielding to a total of  $N(N - 1) = 702$  cross-validation folds per stimuli. This procedure enabled a highly robust statistical assessment.

## Statistical testing

Correlations shown in the Results section were computed using the Pearson's coefficient.  $p$ -values were obtained using non-parametric Wilcoxon rank-sum tests, since samples were generally non-Gaussian distributed (normality was assessed according to  $\chi^2$  and nonparametric Lilliefords tests, and accepted at  $p < 0.001$ ).

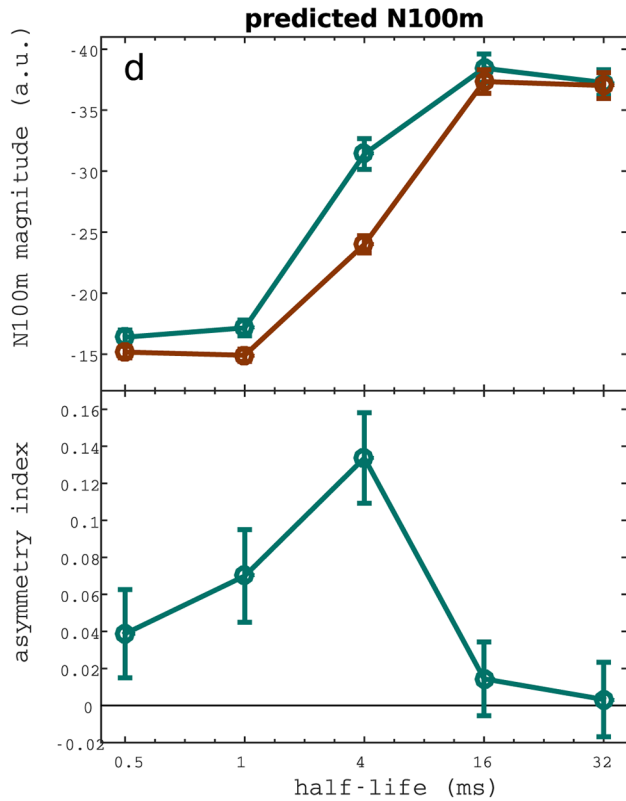
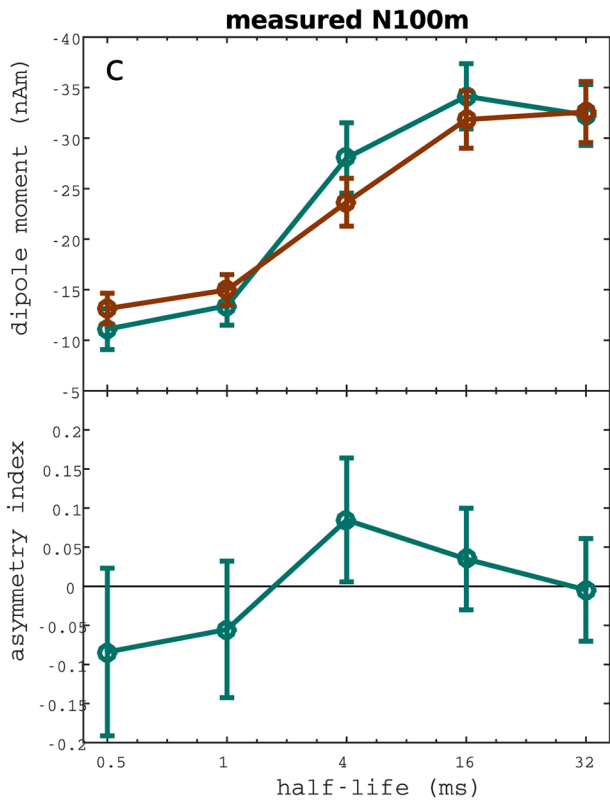
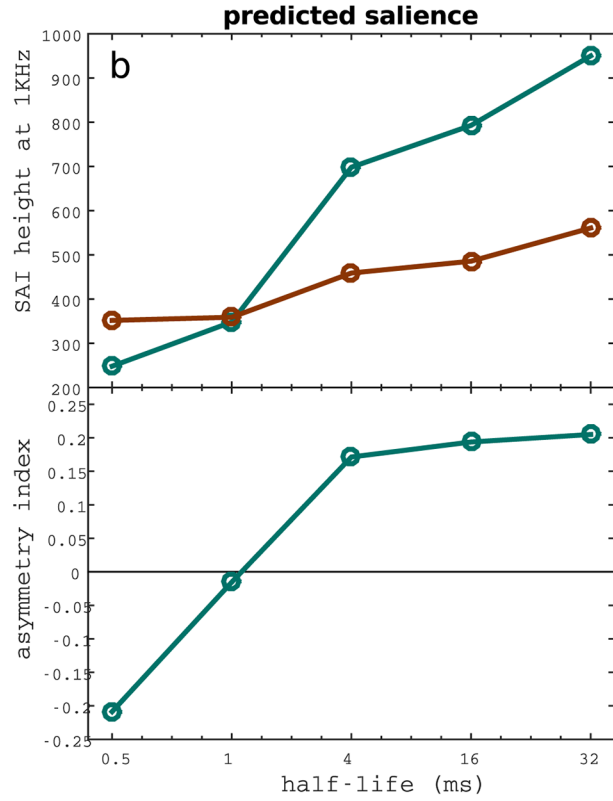
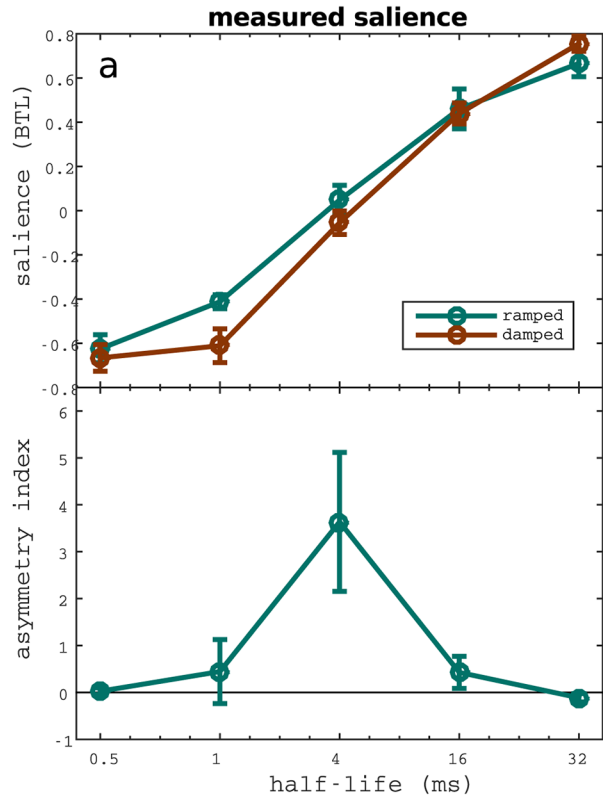
## Results

### Experimental results

**Psychoacoustics and evoked responses.** Fig 4a shows perceptual responses as a function of the stimulus' envelope's  $T_{1/2}$ . Pitch salience increased with  $T_{1/2}$  values for both, ramped and damped sounds, but the pitch of the ramped tones was generally judged as more salient than the pitch of their damped counterparts. This difference reached significance for the critical value  $T_{1/2} = 4\text{ms}$  ( $p = 0.0077$ ,  $n = 13$ ) and for  $T_{1/2} = 1\text{ms}$  ( $p < 0.001$ ,  $n = 16$ ). The difference is attenuated and remains not significant for the rest of the conditions. This behaviour is also reflected in the salience asymmetry index  $AI_p$  (see Fig 4a). Note that the behaviour of the temporal asymmetry index is not well defined over values near zero, which accentuates the difference between ramped and damped at 4 ms. For that reason, statistical significance was not measured using the temporal asymmetry index but rather using the raw BTL perceptual data.

**Neuromagnetic data.** In the next step we analysed the neuromagnetic data in order to find a correspondence between the perceived salience and the morphology of the N100m deflection. Approximate Talairach-coordinates [42] for the sources of the N100m as given by the standard BESA spherical model were localised in lateral Heschl's gyrus (left:  $x = -48 \pm 1$ ,  $y = -24 \pm 2$ ,  $z = 0 \pm 2$ ; right:  $x = 49 \pm 1$ ,  $y = -23 \pm 2$ ,  $z = 0 \pm 2$ ).

The cortical responses are summarised in Fig 5. Source waveforms showed a prominent N100m followed by a large sustained field. Due to the envelope structure of the ramped sounds, latencies of the corresponding responses were delayed in comparison to their damped counterparts.



**Fig 4. Comparison of the perceived salience, N100m magnitude, and the prediction of the two models of pitch.** Perceptual and neuromagnetic results for each of the five pairs of ramp and damp stimuli. The corresponding temporal asymmetry indices are drawn at the bottom of each plot (see Eq (2)). (a) Perceived salience estimated by the BTL method and averaged across subjects ( $N = 13$ ). (b) SAI mean ridge height at the frequency of the carrier (1 kHz). Ridge height was used to predict the perceived salience of the stimuli [14]. (c) Magnitude of the N100 component averaged across subjects. (d) Top-down modulated model's predictions for the amplitude of the N100m peak, computed as a linear transform of the derivative of the activation of the top layer population evaluated at the winning frequency. The linear relationship was cross-validated across subjects (see Methods), yielding to a total of 702 predictions. The figure shows the average of the predictions. Significant correlations were found between perceived saliency 4a) and N100m magnitude (4c); between the perceptual observations 4a AIM responses (4b) and between the N100m magnitude 4c) and GPM predictions (4d). Error bars represent SME.

doi:10.1371/journal.pone.0153947.g004

N100m amplitudes were assessed using the averaged responses across hemispheres. As shown in Fig 4c, the peak amplitude increased with the  $T_{1/2}$  of the stimuli for all conditions and was significant for the transition from  $T_{1/2} = 1$  ms to higher half-life values (ramped:  $p = 0.0003$ ,  $n = 837$ ; damped:  $p = 0.0039$ ,  $n = 837$ ) and for the transition from  $T_{1/2} = 4$  ms to higher half-life times in the damped case ( $p = 0.0146$ ,  $n = 837$ ). Consistently with perceptual results, ramped tones evoked larger N100m than damped ones, with a maximal difference at the critical value of  $T_{1/2} = 4$  ms ( $p = 0.0008$ ,  $n = 837$ ). Accordingly, the corresponding temporal asymmetry indices  $AI_M$  (also shown in Fig 4c), exhibited a maximum for  $T_{1/2} = 4$  ms, fully in line with the perceptual results shown in Fig 4a.

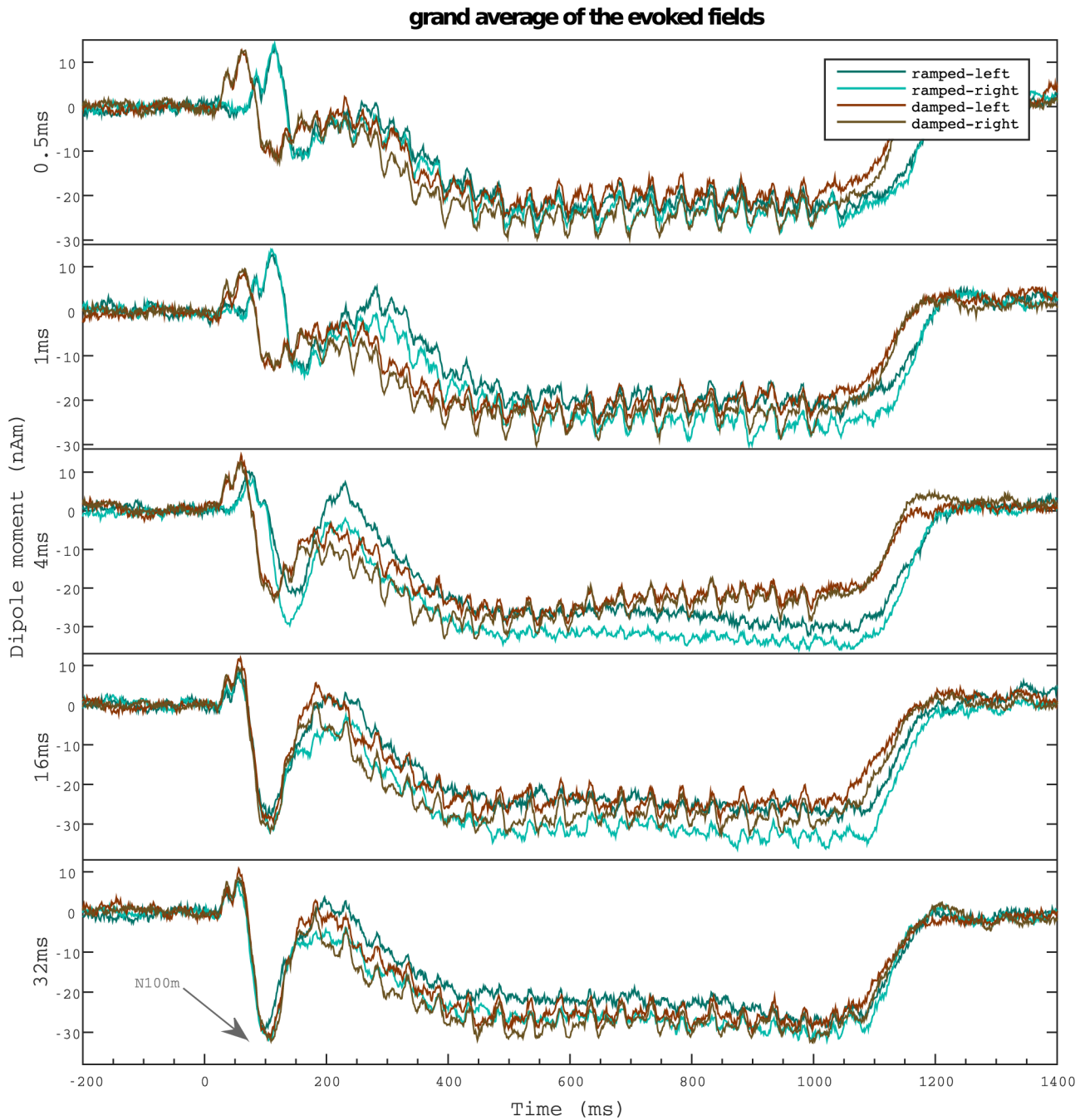
According to the standard BESA spherical model, sources of the sustained field were located more medially but near the N100m sources, in full agreement with previous studies in pitch [39] (left:  $x = -45 \pm 1$ ,  $y = -23 \pm 2$ ,  $z = 3 \pm 2$ ; right:  $x = 45 \pm 1$ ,  $y = -20 \pm 2$ ,  $z = 2 \pm 2$ ). Waveform morphologies were similar to the fields observed in the N100m model in all the conditions and thus they are not shown in a separate plot.

Sustained fields' behaviour mimicked the trends of the N100m. Significant correlations were found between SF's average depth and N100m amplitude (ramped:  $R = 0.9247$ ,  $p = 0.0245$ ; damped:  $R = 0.9744$ ,  $p = 0.049$ ). Correspondingly, damped responses were shallower than ramped responses for the five half-life times ( $p < 0.0001$ ,  $n = 5427$ ). SF's depth also increased with the  $T_{1/2}$  of the stimuli for all conditions, and it was significant for the transition from  $T_{1/2} = 0.5$  ms to higher half-life values (ramped:  $p = 0.0244$ ,  $n = 5427$ ; damped:  $p < 0.0001$ ,  $n = 5427$ ); for the transition from  $T_{1/2} = 1$  ms to higher half-life times (ramped and damped:  $p < 0.0001$ ,  $n = 837$ ); and for the transition from  $T_{1/2} = 4$  ms to higher half-life times in the damped case ( $p < 0.0001$ ,  $n = 5427$ ).

**Correlation between neuromagnetic and perceptual responses.** Taken together, these results show a high correlation between the magnitudes of N100m and the relative perceived carrier salience. This linear correlation was quantitatively measured using the Pearson's correlation coefficient between the BTL salience scores and the magnitude of the N100m for ramped ( $R = -0.9597$ ,  $p = 0.0097$ ) and damped ( $R = -0.9867$ ,  $p = 0.0018$ ) stimuli.

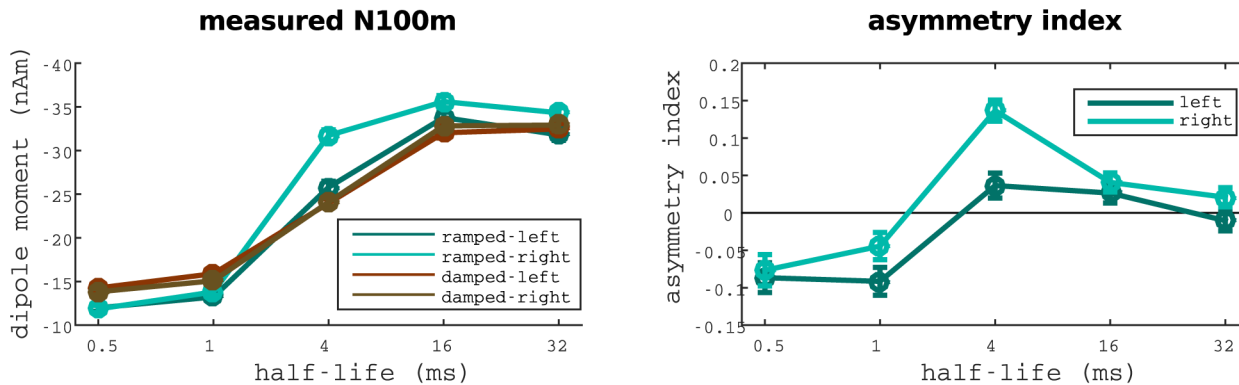
**Inter-hemispheric differences.** Strong differences between hemispheres were observed in the evoked fields for the  $T_{1/2} = 4$  ms condition (see Fig 5). Specifically, the difference between ramped and damped sinusoids was much larger in the right than in the left hemisphere. To assess the size of the effect, we analysed the magnitude of N100m evoked in each hemisphere separately (see Fig 6). Strikingly, the difference between the N100m evoked by ramped and damped for the critical half-life of 4 ms was only significant in the right hemisphere (right:  $p < 0.0001$ ,  $n = 837$ , left:  $p = 0.7124$ ,  $n = 837$ ); whilst differences between fields evoked by sinusoids modulated with different half-life values were similar in both hemispheres.

Correspondingly, we computed the difference between the N100m's magnitude in left and right hemispheres for all the stimuli. The hemispheric asymmetry was, again, only significant for the  $T_{1/2} = 4$  ms ramped sinusoid ( $p < 0.0001$ ,  $n = 837$ ).



**Fig 5. Auditory fields evoked by the ramped and damped sinusoids.** Grand mean source waveforms for the five different conditions of ramped and damped sinusoids. Average was taken over subjects ( $n = 27$ ) for both hemispheres. The magnitude of the N100m increases for rising  $T_{1/2}$  values of the stimuli. Note the maximal difference between ramped and damped sinusoids in the right hemisphere for the  $T_{1/2} = 4$ ms condition.

doi:10.1371/journal.pone.0153947.g005



**Fig 6. Inter-hemispherical differences observed between the fields evoked by ramped and damped sinusoids.** Comparison between the fields evoked in left and right hemispheres for ramped and damped stimuli. N100m's magnitude is plotted in the left panel. Corresponding asymmetry indices are displayed in the right panel.

doi:10.1371/journal.pone.0153947.g006

The sustained field showed similar hemispheric behaviour as the N100m amplitude. Correlations between these two magnitudes in each hemisphere were very high for ramped sinusoids (left:  $R = 0.9926$ ,  $p = 0.0008$ ; right:  $R = 0.9959$ ,  $p = 0.0003$ ), and smaller but still significant for the damped stimuli (left:  $R = 0.9251$ ,  $p = 0.0243$ ; right:  $R = 0.9322$ ,  $p = 0.0210$ ). Responses in the right hemisphere were generally larger than in the left hemisphere in all conditions ( $p < 0.0001$ ,  $n = 5427$ ).

### Model simulations

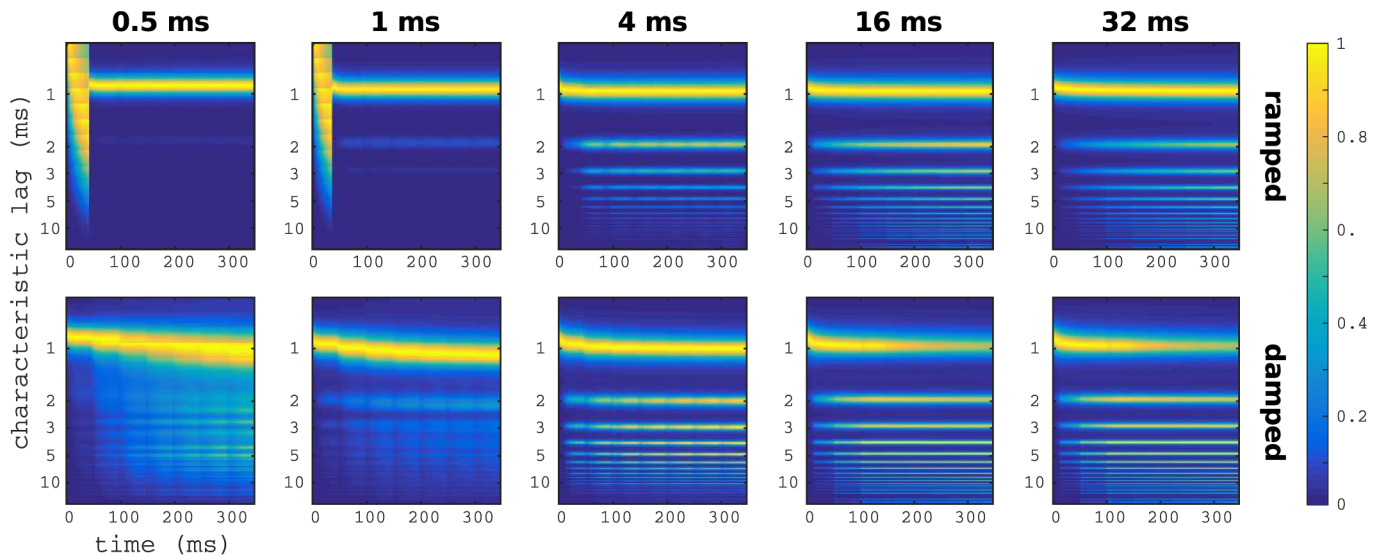
In this subsection we compare the simulation output of the models and compare these patterns with the psychoacoustic and neuromagnetic results.

**Simulations with AIM and perception.** The Auditory Image Model successfully accounted for the carrier salience for ramped and damped stimuli (see Fig 4b) as evidenced from the high correlation with the measured perceptual trends shown in Fig 4a (ramped:  $R = 0.978$ ,  $p < 0.05$ ; damped  $R = 0.978$ ,  $p < 0.05$ ). However, the temporal asymmetry index did not show a high amplitude with the perceptual results for large  $T_{1/2}$ , predicting larger differences than observed in the experimentation. Still, AIM was able to predict the perception elicited by ramped and damped stimuli, suggesting that the strobed integration process effectively amplifies differences in responses at compared to the pattern at the level of the auditory nerve. The next question we will address is whether we can find a functional explanation to this effect in terms of top-down modulatory processes.

**Top-down modulated model and neuromagnetic data.** As a complementary analysis, the GPM model was used to predict the evoked response in the neighbourhood of the N100m deflection. Interestingly, this model provides a phenomenological explanation of the processing in central auditory stages in terms of top-down modulatory effects.

First, we computed the raw output of the model for the set of ramped and damped stimuli. As expected, and in agreement with AIM results, the model output shows a pronounced peak of activation in the 1 ms lag (corresponding to the frequency of the carrier). Moreover, perceptual differences are noticeable between ramped and damped stimuli, and between stimuli with different  $T_{1/2}$  (see Fig 7).

The GPM enables us to analyse correlations between the models' ensemble dynamics and neuromagnetic data (see Methods). An example of such quantitative prediction is shown in Fig 8



**Fig 7. GPM raw output for the ramped and damped stimuli.** Heat maps show the evolution in time ( $x$ -axis) of the activity of the different ensembles ( $y$ -axis) in the third layer of the GPM model for ramped (top) and damped (bottom) sinusoids with different  $T_{1/2}$ . In all cases, after a small period of instability, there is a maximum centred in the ensemble characterized by  $\delta t = 1$  ms, the frequency of the carrier sinusoid. Qualitative differences are noticeable between the output of ramped and damped stimuli, and also between stimuli with different envelope time constants.

doi:10.1371/journal.pone.0153947.g007

for a ramped sound modulated by an envelope with  $T_{1/2} = 0.5$ ms. In the figure, the prediction is compared with the grand average of the auditory evoked fields. The simulation closely resembles the trend of the recorded activity, in particular with regards to the magnitude and latency of the N100m. More generally, the two histograms in Fig 8 show a summary of the model prediction fittings with MEG responses for all cross-validation combinations (see Methods). Although simulations do not show, in general, a close agreement with the observed overall waveform of the neuromagnetic recordings, small root mean square values were observed, with only a small bias error for short  $T_{1/2}$ .

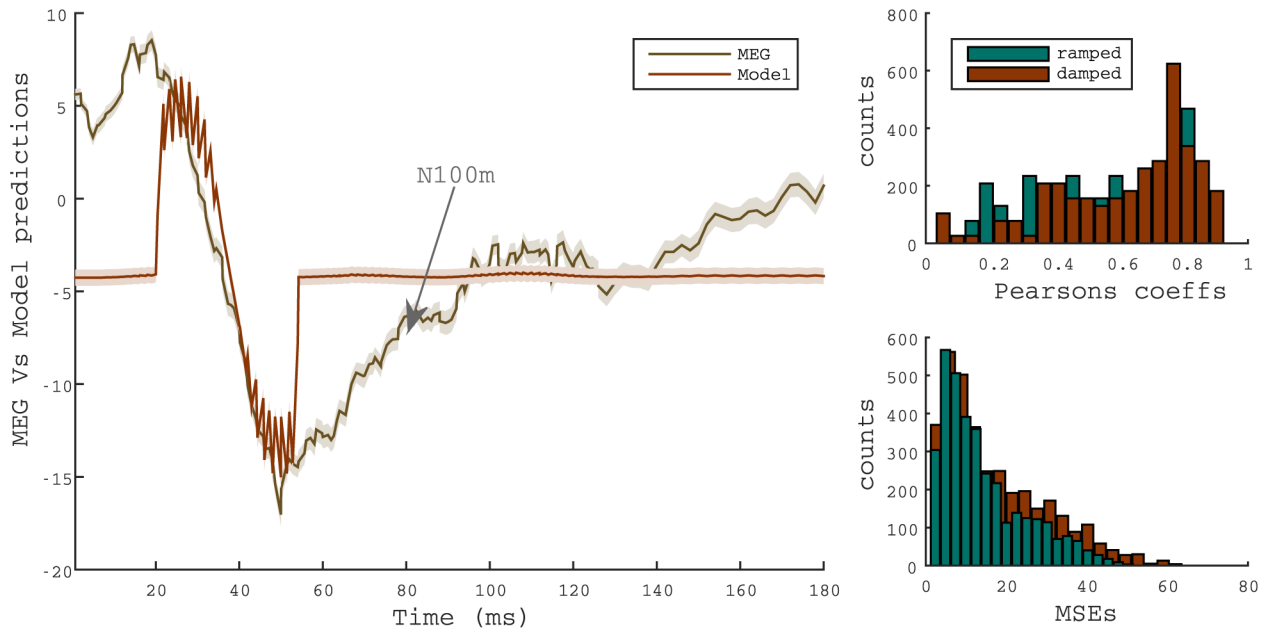
A systematic analysis of the N100m magnitude predictions for all stimuli is shown in Fig 4d. Consistently with perceptual results, differences between model simulations for ramped and damped stimuli are highly significant for a  $T_{1/2} = 4$ ms stimulus ( $p < 0.0001$ ,  $n = 702$ ). Moreover, results in Fig 4d show a strong linear correlation with the magnitude of the N100m observed in the auditory evoked fields (see Fig 4c) for both, ramped ( $R = 0.9972$ ,  $p = 0.0002$ ) and damped ( $R = 0.9899$ ,  $p = 0.012$ ) stimuli.

To test whether the adaptation of the temporal window of integration is necessary to successfully predict the N100m amplitude, we tried to replicate the previous results using an auto-correlation model without top-down modulation [13], effectively equivalent to the top-down modulated model introduced in Methods with static rather than adaptive integration windows  $E_n(t)$ . The analysis failed to produce significant results (see Fig 9), indicating that the top-down has a crucial role in the N100m dynamics elicited by this family of stimuli.

## Discussion

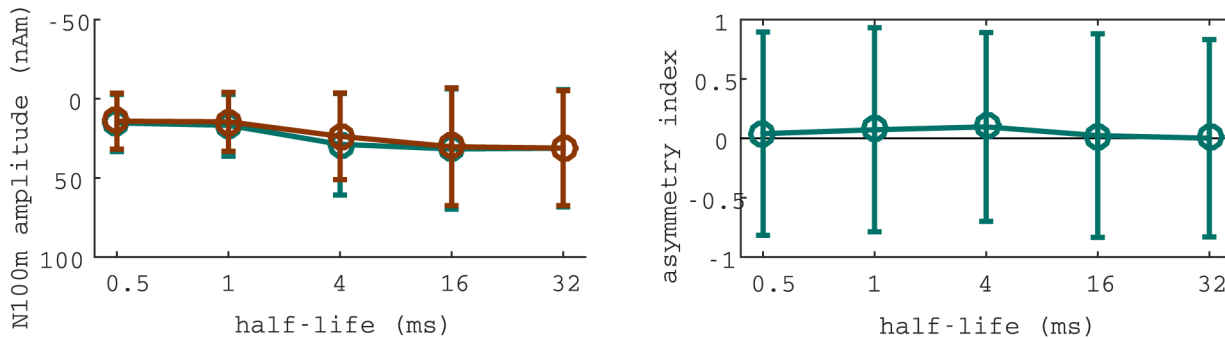
The aim of this study was to characterise the neuromagnetic representation of auditory temporal asymmetry in human auditory cortex and to compare these neurophysiological responses with perceptual data and computer simulations of perceived pitch. We found that the N100m magnitude was closely correlated with perceived pitch salience. Furthermore, N100m





**Fig 8. Summary of the statistics of the fit between the N100m transient and the output of GPM.** Left panel: Example of the model response derivative, normalized to the amplitude of the recording, for a ramped stimulus ( $T_{1/2} = 0.5\text{ms}$ ) and the corresponding recordings, averaged across right and left hemispheres and participants. Transparent shadows represent standard deviations. Right panel: Histograms of the Pearson's correlation coefficient and root-mean-square errors corresponding to the fittings between the GPM prediction and MEG recordings in an interval of 50 ms around the N100m peak. Each value corresponds to a single cross-validation instance for ramped and damped stimuli.

doi:10.1371/journal.pone.0153947.g008

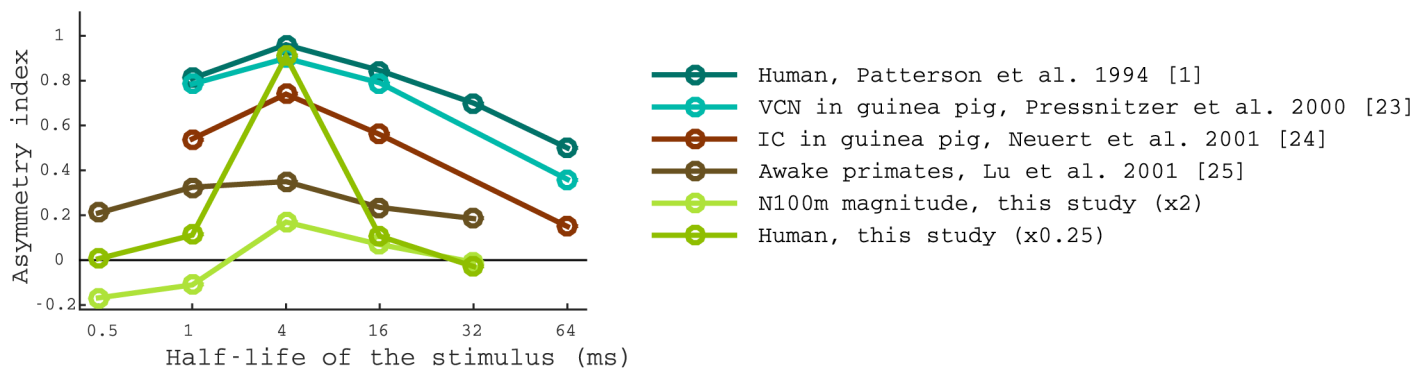


**Fig 9. Autocorrelation model's predictions for the amplitude of the N100m peak.** Predictions were computed following the same procedure as in the analysis of the top-down modulated model (see Fig 4d). Predictions of the autocorrelation model do not show statistically significant correlations with the N100m values or the perceptual predictions. Moreover, the predicted amplitudes elicited by ramped and damped sinusoids with  $T_{1/2} = 4\text{ms}$  are not significantly different in this analysis.

doi:10.1371/journal.pone.0153947.g009

amplitudes were closely related to the computer simulations of the classical Auditory Image Model [14] as well as the hierarchical top-down modulated model of pitch (GPM) [17]. The latter enabled us to provide a phenomenological understanding of bottom-up and top-down processes which may underlie the neural coding of perceived temporal asymmetry.

The present study extends the work of Patterson and colleagues [14] by analysing the auditory evoked fields elicited by the same set of ramped and damped in human listeners. We observed that the amplitude of N100m increased with stimulus'  $T_{1/2}$  for all conditions and both hemispheres, thus providing a neurophysiological correlate of the actual strength of the



**Fig 10. Comparison between our results and previously reported measures of the perceptual asymmetry between ramped and damped sinusoids.** Comparison between asymmetry preference of ventral cochlear nucleus [23], inferior colliculus [24], cortical neurons [25], human psychophysical performance in discriminating the ramped and damped sinusoids in A1 [1], the N100m magnitude temporal asymmetry, and psychophysical perceptual asymmetry measured in this work. Multiplicative factors (2 and 0.25, respectively) were applied to rescale the results of our study in order to improve visualisation. Note that the absolute values of the indices depend on the individual scale of each quantity.

doi:10.1371/journal.pone.0153947.g010

tonal component. The morphology of the N100m source waveforms strongly varied as a function of the temporal features of the envelope (see Fig 5), especially for the critical  $T_{1/2} = 4$  ms pair of stimuli, again in full agreement with subject’s perceived tonality.

It is important to note that subcomponents of the N100m exhibit different temporal integration times [39, 58]. However, the location of the N100m sources found in this work are located in aHGC, and hence we can safely assume that we assessed pitch related generators, as reported in humans [27, 39] and animal studies (e.g. [59]).

It is also noticeable that we observed a tight correlation between the psychophysical data, which was based on judgements of 20 modulation cycles lasting 1000 ms overall, and the N100m, peaking at about 100 ms after sound onset. Presumably, the N100m reflect mechanisms affecting only the beginning of the sound, revealing that processes occurring at the onset of the stimuli are crucial for the decoding of temporal asymmetry.

The observed results are also in agreement with the model simulations. For instance, a closer look at the summary SAI of the simulations for these stimuli (see Fig 3) reveals a steep increase in the height of the first peak for the ramped sound which indicates a specific carrier salience extraction. The simulations performed with the stimuli with longer  $T_{1/2}$  (i.e. 16 ms and 32 ms) showed that damped sounds also elicit an increase of the first peak height.

We observed that ramped stimuli are associated with a stronger tonal percept, particularly for half life times of  $T_{1/2} = 1–16$ ms, in agreement with the landmark study from Patterson and colleagues [14]. Moreover, we found that the maximal asymmetry between the perceived salience of ramped and damped stimuli occurs at  $T_{1/2} = 4$ ms (see Fig 4a). This result is fully in line with previous studies on perceptual asymmetry as shown in Fig 10.

Temporal asymmetry indices for the rest of  $T_{1/2}$  values vary across studies. There are two potential reasons for such a variability. First, the N100m amplitude is larger for  $T_{1/2}$  (e.g., 0.5 ms and 1 ms stimuli) and the transient often does not reach a sharp maximum, which hampers the identification of the minimum. Speculatively, this variability in the N100m may underlie part of the perceptual variability. Second, the tonal sensation of stimuli with short  $T_{1/2}$  stimuli is very weak due to the presence of the simultaneous *drumming* sensation that occludes pitch. This might explain also the different shape of the psychometric curve obtained in [1].

Responses to ramped and damped sinusoids modulated with a 4 ms envelope’s time constant are significantly different between hemispheres. Specifically, responses to ramped and

damped stimuli were largely different in the right hemisphere, but statistically indistinguishable in the left hemisphere (see Fig 6). Moreover, hemispheric differences were not observed in the N100m evoked by any of the remaining 9 conditions. This finding indicates a lateralisation of the mechanisms responsible for temporal asymmetry processing at time scales of about 4 ms.

Time-scale specific hemispheric specialisation has been reported before in connection to language [60] and is the target of the *asymmetric sampling in time* (AST) theory [61]. Based on a large amount of experimental evidence on previous literature, AST assumes that the right hemisphere responds preferably to processes requiring longer time scales, whilst the left hemisphere responds preferably to short modulations. However, further investigations are needed to investigate the specific relationship of temporal integration processes and the AST model. Our robust finding of an asymmetry for sounds with 4 ms half-life time indicates that related sounds with slightly different envelopes and durations might be used to further specify auditory processing of the left and right hemisphere.

Our modelling results further suggest that the N100m is related to pitch decoding, as frequently reported in the literature (e.g. [27]). In this work we emphasised that the adaptive processing as implemented in both models is a key to understand the perception of asymmetric sounds and the observed differences in the N100m morphology. Although the spectral analysis on the basilar membrane and the neural transduction process enhance temporal asymmetry to a certain extent [15], this enhancement is indeed not sufficient to explain perceptual effects [15].

Autocorrelation models [11, 13] have been shown to be very successful in pitch extraction of complex tones, but stimulus-dependent temporal integration was required to explain how the auditory system furnishes the balance between temporal resolution and robust pattern recognition [17].

In contrast, the two idealised computational models considered in this study were able to amplify this temporal asymmetry and successfully predict the perceived differences between ramped and damped stimuli (see Fig 4b and 4d). Furthermore, the top-down model accurately predicted the magnitude of the evoked N100m. This result, robustly cross-validated across a large set of samples, suggests that temporal asymmetry encoding may be also mediated by a hierarchical process with top-down driven stimulus-specific integration windows.

However, a more detailed identification of the biophysical processes underlying such stimulus-dependent temporal integration is out of the scope of this study. Our hypothesis is that pitch integration is drawn on the basis of a harmonic pattern of connectivity in aHG [62]. Another potential contributor to the rapid detection of auditory stimuli is neuromodulation [63], a very recent and interesting hypothesis which has not been analysed yet using non-invasive recordings in human subjects.

In summary, the current study provides further evidence that the N100m magnitude indicates the presence of a neurophysiological mechanism encoding pitch saliency in auditory temporal asymmetry, and suggest that pitch salience asymmetry can only be explained by means of adaptive windows of temporal integration. This process seems to be an important component in the perception of natural communication sounds, whose onsets often exhibit complex temporal and spectral changes within the first milliseconds [36, 37] like the ramped and damped sinusoids.

## Acknowledgments

A.T. receives funding from the Bournemouth University Doctoral Fund and the Santander Research Mobility Award program. The study was further supported by the Croatian Ministry

of Science and Technology (grant 0119265) and the Deutsche Forschungsgemeinschaft Grant Ru-652/1-3.

## Author Contributions

Conceived and designed the experiments: AS AR. Performed the experiments: AS AR AT. Analyzed the data: AT EB AR. Contributed reagents/materials/analysis tools: EB DP. Wrote the paper: AT AR EB SS.

## References

1. Patterson R. The sound of a sinusoid: Spectral models. *The Journal of the Acoustical Society of America*. 1994; 96(3):1409–1418. doi: [10.1121/1.410286](https://doi.org/10.1121/1.410286)
2. Rosen S, Howell P. Plucks and bows are not categorically perceived. *Perception & Psychophysics*. 1982; 31(5):462–476. doi: [10.3758/BF03204856](https://doi.org/10.3758/BF03204856)
3. Paquette C, Peretz I. Role of familiarity in auditory discrimination of musical instrument: a laterality study. *Cortex; a journal devoted to the study of the nervous system and behavior*. 1997; 33(4):689–96. doi: [10.1016/S0010-9452\(08\)70726-3](https://doi.org/10.1016/S0010-9452(08)70726-3)
4. Morton J, Marcus S, Frankish C. Perceptual centers (P-centers). *Psychological Review*. 1976; 83(5):405–408. doi: [10.1037/0033-295X.83.5.405](https://doi.org/10.1037/0033-295X.83.5.405)
5. Vos J, Rasch R. The perceptual onset of musical tones. *Perception & psychophysics*. 1981; 29(4):323–35. doi: [10.3758/BF03207341](https://doi.org/10.3758/BF03207341)
6. Gordon JW. The perceptual attack time of musical tones. *The Journal of the Acoustical Society of America*. 1987; 82(1):88–105. doi: [10.1121/1.395441](https://doi.org/10.1121/1.395441) PMID: [3624645](https://pubmed.ncbi.nlm.nih.gov/3624645/)
7. Schlauch RS, Ries DT, DiGiovanni JJ. Duration discrimination and subjective duration for ramped and damped sounds. *The Journal of the Acoustical Society of America*. 2001; 109(6):2880–2887. doi: [10.1121/1.1372913](https://doi.org/10.1121/1.1372913) PMID: [11425130](https://pubmed.ncbi.nlm.nih.gov/11425130/)
8. Hartmann W. The effect of amplitude envelope on the pitch of sine wave tones. *The Journal of the Acoustical Society of America*. 1978; 63(4):1105–1113. doi: [10.1121/1.381818](https://doi.org/10.1121/1.381818) PMID: [649869](https://pubmed.ncbi.nlm.nih.gov/649869/)
9. Stecker GC, Hafter ER. An effect of temporal asymmetry on loudness. *The Journal of the Acoustical Society of America*. 2000; 107(6):3358–3368. doi: [10.1121/1.429407](https://doi.org/10.1121/1.429407) PMID: [10875381](https://pubmed.ncbi.nlm.nih.gov/10875381/)
10. Irino T, Patterson R. A time-domain, level-dependent auditory filter: The gammachirp. *The Journal of the Acoustical Society of America*. 1997; 101(1):412–419. doi: [10.1121/1.417975](https://doi.org/10.1121/1.417975)
11. Meddis R, O'Mard L. A unitary model of pitch perception. *The Journal of the Acoustical Society of America*. 1997; 102(3):1811–20. doi: [10.1121/1.420088](https://doi.org/10.1121/1.420088) PMID: [9301058](https://pubmed.ncbi.nlm.nih.gov/9301058/)
12. Cheveigne AD. Pitch perception models. In: *Pitch: Neural Coding and perception*. Springer; 2005. p. 169–233.
13. Balaguer-Ballester E, Denham SL, Meddis R. A cascade autocorrelation model of pitch perception. *The Journal of the Acoustical Society of America*. 2008; 124(4):2186–95. doi: [10.1121/1.2967829](https://doi.org/10.1121/1.2967829) PMID: [19062858](https://pubmed.ncbi.nlm.nih.gov/19062858/)
14. Patterson R, Allerhand M, Giguere C. Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *The Journal of the Acoustical Society of America*. 1995; 98(4):1890–4. doi: [10.1121/1.414456](https://doi.org/10.1121/1.414456) PMID: [7593913](https://pubmed.ncbi.nlm.nih.gov/7593913/)
15. Patterson RD, Irino T. Modeling temporal asymmetry in the auditory system. *The Journal of the Acoustical Society of America*. 1998; 104(5):2967–79. doi: [10.1121/1.423879](https://doi.org/10.1121/1.423879) PMID: [9821341](https://pubmed.ncbi.nlm.nih.gov/9821341/)
16. Kumar S, Sedley W, Nourski KV, Kawasaki H, Oya H, Patterson RD, et al. Predictive Coding and Pitch Processing in the Auditory Cortex. *J Cognitive Neuroscience*. 2011; 23(10):3084–3094. doi: [10.1162/jocn\\_a\\_00021](https://doi.org/10.1162/jocn_a_00021)
17. Balaguer-Ballester E, Clark NR, Coath M, Krumbholz K, Denham SL. Understanding pitch perception as a hierarchical process with top-down modulation. *PLoS computational biology*. 2009; 5(3): e1000301. doi: [10.1371/journal.pcbi.1000301](https://doi.org/10.1371/journal.pcbi.1000301) PMID: [19266015](https://pubmed.ncbi.nlm.nih.gov/19266015/)
18. Kumar S, Stephan KE, Warren JD, Friston KJ, Griffiths TD. Hierarchical processing of auditory objects in humans. *PLoS computational biology*. 2007; 3(6):e100. doi: [10.1371/journal.pcbi.0030100](https://doi.org/10.1371/journal.pcbi.0030100) PMID: [17542641](https://pubmed.ncbi.nlm.nih.gov/17542641/)
19. Hall J, Peters R. Pitch for nonsimultaneous successive harmonics in quiet and noise. *The Journal of the Acoustical Society of America*. 1981; 69(2):509–13. doi: [10.1121/1.385480](https://doi.org/10.1121/1.385480) PMID: [7462473](https://pubmed.ncbi.nlm.nih.gov/7462473/)

20. Plack CJ, White LJ. Perceived continuity and pitch perception. *The Journal of the Acoustical Society of America*. 2000; 108(3 Pt 1):1162–1169. doi: [10.1121/1.1287022](https://doi.org/10.1121/1.1287022) PMID: [11008817](https://pubmed.ncbi.nlm.nih.gov/11008817/)
21. Grose JH, Hall JW, Buss E. Virtual pitch integration for asynchronous harmonics. *The Journal of the Acoustical Society of America*. 2002; 112(6):2956–2961. doi: [10.1121/1.1514934](https://doi.org/10.1121/1.1514934) PMID: [12509016](https://pubmed.ncbi.nlm.nih.gov/12509016/)
22. Suied C, Agus TR, Thorpe SJ, Pressnitzer D. Auditory gist: Recognition of very short sounds from timbre cues. *Journal of the Acoustical Society of America*. 2014; 135(3):1380–1391. doi: [10.1121/1.4863659](https://doi.org/10.1121/1.4863659) PMID: [24606276](https://pubmed.ncbi.nlm.nih.gov/24606276/)
23. Pressnitzer D, Winter IM, Patterson RD. The responses of single units in the ventral cochlear nucleus of the guinea pig to damped and ramped sinusoids. *Hearing research*. 2000; 149(1-2):155–66. doi: [10.1016/S0378-5955\(00\)00175-1](https://doi.org/10.1016/S0378-5955(00)00175-1) PMID: [11033255](https://pubmed.ncbi.nlm.nih.gov/11033255/)
24. Neuert V, Pressnitzer D, Patterson RD, Winter IM. The responses of single units in the inferior colliculus of the guinea pig to damped and ramped sinusoids. *Hearing research*. 2001; 159(1-2):36–52. doi: [10.1016/S0378-5955\(01\)00318-5](https://doi.org/10.1016/S0378-5955(01)00318-5) PMID: [11520633](https://pubmed.ncbi.nlm.nih.gov/11520633/)
25. Lu T, Liang L, Wang X. Neural representations of temporally asymmetric stimuli in the auditory cortex of awake primates. *Journal of neurophysiology*. 2001; 85(6):2364–80. PMID: [11387383](https://pubmed.ncbi.nlm.nih.gov/11387383/)
26. Lütkenhöner B, Steinsträter O. High-precision neuromagnetic study of the functional organization of the human auditory cortex. *Audiology and Neurotology*. 1998; 3(2-3):191–213. doi: [10.1159/000013790](https://doi.org/10.1159/000013790) PMID: [9575385](https://pubmed.ncbi.nlm.nih.gov/9575385/)
27. Krumbholz K, Patterson R. Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex*. 2003; 13(7):765–772. doi: [10.1093/cercor/13.7.765](https://doi.org/10.1093/cercor/13.7.765) PMID: [12816892](https://pubmed.ncbi.nlm.nih.gov/12816892/)
28. Rapin I, Schimmel H, Tourk LM, Krasnegor NA, Pollak C. Evoked responses to clicks and tones of varying intensity in waking adults. *Electroencephalography and Clinical Neurophysiology*. 1966; 21(4):335–344. doi: [10.1016/0013-4694\(66\)90039-3](https://doi.org/10.1016/0013-4694(66)90039-3) PMID: [4162205](https://pubmed.ncbi.nlm.nih.gov/4162205/)
29. Beagley H, Knight J. Changes in auditory evoked response with intensity. *J Laryngol Otol*. 1967; 81(8):861–873. doi: [10.1017/S0022215100067815](https://doi.org/10.1017/S0022215100067815) PMID: [6036752](https://pubmed.ncbi.nlm.nih.gov/6036752/)
30. Biermann S, Heil P. Parallels between timing of onset responses of single neurons in cat and of evoked magnetic fields in human auditory cortex. *Journal of neurophysiology*. 2000; 84(5):2426–39. PMID: [11067985](https://pubmed.ncbi.nlm.nih.gov/11067985/)
31. Roberts T, Ferrari P. Latency of the auditory evoked neuromagnetic field components: stimulus dependence and insights toward perception. *Journal of Clinical Neurophysiology*. 2000; 17(2):114–29. doi: [10.1097/00004691-200003000-00002](https://doi.org/10.1097/00004691-200003000-00002) PMID: [10831104](https://pubmed.ncbi.nlm.nih.gov/10831104/)
32. Jacobson GP, Lombardi DM, Gibbens ND, Ahmad BK, Newman CW. The effects of stimulus frequency and recording site on the amplitude and latency of multichannel cortical auditory evoked potential (CAEP) component N1. *Ear and Hearing*. 1992; 13(5):300–306. doi: [10.1097/00003446-199210000-00007](https://doi.org/10.1097/00003446-199210000-00007) PMID: [1487089](https://pubmed.ncbi.nlm.nih.gov/1487089/)
33. Roberts TP, Poeppel D. Latency of auditory evoked M100 as a function of tone frequency. *Neuroreport*. 1996; 7(6):1138–1140. doi: [10.1097/00001756-199604260-00007](https://doi.org/10.1097/00001756-199604260-00007) PMID: [8817518](https://pubmed.ncbi.nlm.nih.gov/8817518/)
34. Ragot R, Lepaul-Ercole R. Brain potentials as objective indexes of auditory pitch extraction from harmonics. *Neuroreport*. 1996; 7(4):905–909. doi: [10.1097/00001756-199603220-00014](https://doi.org/10.1097/00001756-199603220-00014) PMID: [8724670](https://pubmed.ncbi.nlm.nih.gov/8724670/)
35. Ritter S, Günter Dosch H, Specht HJ, Rupp A. Neuromagnetic responses reflect the temporal pitch change of regular interval sounds. *NeuroImage*. 2005; 27(3):533–43. doi: [10.1016/j.neuroimage.2005.05.003](https://doi.org/10.1016/j.neuroimage.2005.05.003) PMID: [15964207](https://pubmed.ncbi.nlm.nih.gov/15964207/)
36. 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*. 2006; 213(1-2):88–98. doi: [10.1016/j.heares.2006.01.003](https://doi.org/10.1016/j.heares.2006.01.003) PMID: [16464550](https://pubmed.ncbi.nlm.nih.gov/16464550/)
37. Seither-Preisler A, Krumbholz K, Lütkenhöner B. Sensitivity of the neuromagnetic N100m deflection to spectral bandwidth: a function of the auditory periphery. *Audiology and Neuro-Otology*. 2003; 8(6):322–337. doi: [10.1159/000073517](https://doi.org/10.1159/000073517) PMID: [14566103](https://pubmed.ncbi.nlm.nih.gov/14566103/)
38. Penagos H, Melcher JR, Oxenham AJ. A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *The Journal of neuroscience*. 2004; 24(30):6810–5. doi: [10.1523/JNEUROSCI.0383-04.2004](https://doi.org/10.1523/JNEUROSCI.0383-04.2004) PMID: [15282286](https://pubmed.ncbi.nlm.nih.gov/15282286/)
39. Gutschalk A, Patterson RD, Scherg M, Uppenkamp S, Rupp A. Temporal dynamics of pitch in human auditory cortex. *NeuroImage*. 2004; 22(2):755–766. doi: [10.1016/j.neuroimage.2004.01.025](https://doi.org/10.1016/j.neuroimage.2004.01.025) PMID: [15193604](https://pubmed.ncbi.nlm.nih.gov/15193604/)
40. David H. *The method of paired comparisons*. New York: Oxford University Press; 1963.
41. Okamoto H, Stracke H, Bermudez P, Pantev C. Sound processing hierarchy within human auditory cortex. *Journal of Cognitive Neuroscience*. 2011; 23(8):1855–63. doi: [10.1162/jocn.2010.21521](https://doi.org/10.1162/jocn.2010.21521) PMID: [20521859](https://pubmed.ncbi.nlm.nih.gov/20521859/)

42. Tailarach J, Tournoux P. Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system. Thieme Classics. Thieme Medical Pub; 1988.
43. Andermann M, van Dinther R, Patterson RD, Rupp A. Neuromagnetic representation of musical register information in human auditory cortex. *NeuroImage*. 2011; 57(4):1499–506. doi: [10.1016/j.neuroimage.2011.05.049](https://doi.org/10.1016/j.neuroimage.2011.05.049) PMID: [21640834](https://pubmed.ncbi.nlm.nih.gov/21640834/)
44. Berg P, Scherg M. A multiple source approach to the correction of eye artifacts. *Electroencephalography and clinical neurophysiology*. 1994; 90(3):229–241. doi: [10.1016/0013-4694\(94\)90094-9](https://doi.org/10.1016/0013-4694(94)90094-9) PMID: [7511504](https://pubmed.ncbi.nlm.nih.gov/7511504/)
45. Lopez-Poveda Ea, Meddis R. A human nonlinear cochlear filterbank. *The Journal of the Acoustical Society of America*. 2001; 110(6):3170–3118. doi: [10.1121/1.1416197](https://doi.org/10.1121/1.1416197)
46. Patterson R, Sheft S. A time domain description for the pitch strength of iterated rippled noise. *The Journal of the Acoustical Society of America*. 1996; 99(2):1066–1078. doi: [10.1121/1.414593](https://doi.org/10.1121/1.414593) PMID: [8609290](https://pubmed.ncbi.nlm.nih.gov/8609290/)
47. Yost W, Patterson R, Sheft S. A time domain description for the pitch strength of iterated rippled noise. *The Journal of the Acoustical Society of America*. 1996; 99(2):1066–78. doi: [10.1121/1.414593](https://doi.org/10.1121/1.414593) PMID: [8609290](https://pubmed.ncbi.nlm.nih.gov/8609290/)
48. Friston K. A theory of cortical responses. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*. 2005; 360(1456):815–36. doi: [10.1098/rstb.2005.1622](https://doi.org/10.1098/rstb.2005.1622) PMID: [15937014](https://pubmed.ncbi.nlm.nih.gov/15937014/)
49. Meddis R, O'Mard LP. Virtual pitch in a computational physiological model. *The Journal of the Acoustical Society of America*. 2006; 120(6):3861. doi: [10.1121/1.2372595](https://doi.org/10.1121/1.2372595) PMID: [17225413](https://pubmed.ncbi.nlm.nih.gov/17225413/)
50. Gerstner W, Kistler WM, Naud R, Paninski L. *Neuronal dynamics: from single neurons to networks and models of cognition*. 1st ed. Cambridge University Press; 2014.
51. Wiegrefe L. Searching for the time constant of neural pitch extraction. *The Journal of the Acoustical Society of America*. 2001; 109(3):1082–1091. doi: [10.1121/1.1348005](https://doi.org/10.1121/1.1348005) PMID: [11303922](https://pubmed.ncbi.nlm.nih.gov/11303922/)
52. Winter I. The neurophysiology of pitch. In: *Pitch: Neural Coding and Perception*. Springer; 2005. p. 99–146.
53. Fishbach A, Yeshurun Y, Nelken I. Neural model for physiological responses to frequency and amplitude transitions uncovers topographical order in the auditory cortex. *Journal of neurophysiology*. 2003; 90(6):3663–78. doi: [10.1152/jn.00654.2003](https://doi.org/10.1152/jn.00654.2003) PMID: [12944531](https://pubmed.ncbi.nlm.nih.gov/12944531/)
54. Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD. The processing of temporal pitch and melody information in auditory cortex. *Neuron*. 2002; 36(4):767–76. doi: [10.1016/S0896-6273\(02\)01060-7](https://doi.org/10.1016/S0896-6273(02)01060-7) PMID: [12441063](https://pubmed.ncbi.nlm.nih.gov/12441063/)
55. Patterson RD, Johnsrude IS. Functional imaging of the auditory processing applied to speech sounds. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*. 2008; 363(1493):1023–35. doi: [10.1098/rstb.2007.2157](https://doi.org/10.1098/rstb.2007.2157) PMID: [17827103](https://pubmed.ncbi.nlm.nih.gov/17827103/)
56. Gerstner W, Kistler W. *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge University Press; 2002.
57. Daunizeau J, Kiebel S, Friston K. Dynamic causal modelling of distributed electromagnetic responses. *NeuroImage*. 2009; 47(2):590–601. doi: [10.1016/j.neuroimage.2009.04.062](https://doi.org/10.1016/j.neuroimage.2009.04.062) PMID: [19398015](https://pubmed.ncbi.nlm.nih.gov/19398015/)
58. Alain C, Woods D, Covarrubias D. Activation of duration-sensitive auditory cortical fields in humans. *Electroencephalography and Clinical Neurophysiology*. 1997; 104(6):531–539. doi: [10.1016/S0168-5597\(97\)00057-9](https://doi.org/10.1016/S0168-5597(97)00057-9) PMID: [9402895](https://pubmed.ncbi.nlm.nih.gov/9402895/)
59. Bendor D, Wang X. Cortical representations of pitch in monkeys and humans. *Current opinion in neurobiology*. 2006; 16(4):391–9. doi: [10.1016/j.conb.2006.07.001](https://doi.org/10.1016/j.conb.2006.07.001) PMID: [16842992](https://pubmed.ncbi.nlm.nih.gov/16842992/)
60. Belin P, Zilbovicius M, Crozier S, Thivard L, Fontaine A, Masure MC, et al. Lateralization of speech and auditory temporal processing. *Journal of cognitive neuroscience*. 1998; 10(4):536–540. doi: [10.1162/089892998562834](https://doi.org/10.1162/089892998562834) PMID: [9712682](https://pubmed.ncbi.nlm.nih.gov/9712682/)
61. Poeppel D. The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*. 2003; 41(1):245–255. doi: [10.1016/S0167-6393\(02\)00107-3](https://doi.org/10.1016/S0167-6393(02)00107-3)
62. Wang X. The harmonic organization of auditory cortex. *Frontiers in Systems Neuroscience*. 2013; 7:114. doi: [10.3389/fnsys.2013.00114](https://doi.org/10.3389/fnsys.2013.00114) PMID: [24381544](https://pubmed.ncbi.nlm.nih.gov/24381544/)
63. Happel MFK, Deliano M, Handschuh J, Ohl FW. Dopamine-Modulated Recurrent Corticoefferent Feedback in Primary Sensory Cortex Promotes Detection of Behaviorally Relevant Stimuli. *The Journal of Neuroscience*. 2014; 34(4):1234–1247. doi: [10.1523/JNEUROSCI.1990-13.2014](https://doi.org/10.1523/JNEUROSCI.1990-13.2014) PMID: [24453315](https://pubmed.ncbi.nlm.nih.gov/24453315/)