



# The auditory P2 is influenced by pitch changes but not pitch strength and consists of two separate subcomponents

Kurt Steinmetzger<sup>a,b</sup>, André Rupp<sup>a</sup>

<sup>a</sup>Section of Biomagnetism, Department of Neurology, Heidelberg University Hospital, Heidelberg, Germany

<sup>b</sup>Tinnitus Center, Charité - Universitätsmedizin Berlin, Berlin, Germany

Corresponding Author: Kurt Steinmetzger ([kurt.steinmetzger@charite.de](mailto:kurt.steinmetzger@charite.de))

## ABSTRACT

The P2 component of the auditory evoked potential has previously been shown to depend on the acoustic stimulus properties and prior exposure to the materials. Here, we show that it is also affected by acoustic changes, as P2 amplitudes were strongly enhanced in response to voice pitch changes with a stepwise pattern compared to dynamic pitch changes typical for natural speech, and also reflected the magnitude of these pitch changes. Furthermore, it is demonstrated that neither the P2 nor any other component is affected by the harmonicity of the materials. Despite no prior exposure and a weaker pitch, artificially created inharmonic versions of the materials elicited similar activity throughout the auditory cortex. This suggests that so-called harmonic template neurons observed in animal studies are either absent or do not exist in sufficient number in the human auditory cortex to detect their activity extracranially. Crucially, morphology as well as scalp maps and source reconstructions of the EEG data showed that the P2 appears to consist of two separate subcomponents. While the “P2a” was localised to the auditory cortex, the subsequent “P2b” included generators spread across the auditory cortex and association areas. The two subcomponents thus likely reflect processing at different stages of the auditory pathway.

**Keywords:** speech, hearing, harmonicity, event-related potentials, EEG source reconstructions

## 1. INTRODUCTION

Although it is part of the P1-N1-P2 complex of transient responses and therefore ubiquitous in recordings of auditory evoked cortical activity, functional significance, cortical generators, and morphology of the auditory P2 component have remained elusive. Other components have been studied much more thoroughly, especially the preceding N1, partly because its large amplitude and short duration facilitate the localisation of its sources (e.g., [Krumbholz et al., 2003](#); [Näätänen & Picton, 1987](#); [T. P. Roberts & Poeppel, 1996](#)). Based on early findings, it has been assumed that the N1 is an index of any kind of acoustic change, whereas the P2 was thought to depend on the acoustic characteristics of the stimulus materials

([Hari et al., 1987](#); [Näätänen & Picton, 1987](#)). For example, while an N1 is elicited by both sound onsets and offsets, the P2 is only observed following sound onset ([Hari et al., 1987](#)). Subsequently, the P2 has been shown to increase with spectral complexity, with the largest amplitudes observed for sounds containing multiple adjacent harmonics, as is typical for musical and speech sounds ([Shahin et al., 2003, 2005](#)). Additionally, P2 amplitudes in response to harmonic sounds have been found to increase with repeated stimulus exposure, suggesting that the P2 also reflects familiarity with specific types of sounds ([MacLean et al., 2024](#); [Sheehan et al., 2005](#); [Tremblay et al., 2014](#)). However, recent findings showed that the P2 is also enhanced in response to pitch changes in speech

Received: 5 July 2023 Revision: 25 January 2024 Accepted: 8 April 2024 Available Online: 23 April 2024



© 2024 Massachusetts Institute of Technology.  
Published under a Creative Commons Attribution 4.0  
International (CC BY 4.0) license.

Imaging Neuroscience, Volume 2, 2024  
[https://doi.org/10.1162/imag\\_a\\_00160](https://doi.org/10.1162/imag_a_00160)

and music (Andermann et al., 2021; Steinmetzger, Megbel, et al., 2022; Steinmetzger, Meinhardt, et al., 2022), implying that it is involved in the processing of acoustic changes too. Although not the focus of the current study, higher-order cognitive processes such as expectancy and prior contextual beliefs also affect the P2 (e.g., Seidel et al., 2021; Sowman et al., 2012), in addition to stimulus-specific sensory processing.

Regarding the cortical generators of the P2, the findings have also been inconsistent. MEG studies using a dipole-based approach consistently localised the auditory P2 to the lateral part of Heschl's gyrus (HG), slightly anterior and medial to the N1, irrespective of whether speech or non-speech stimuli were used (Hari et al., 1987; Pantev et al., 1996; Ross & Tremblay, 2009; Tiitinen et al., 1999). However, intracerebral recordings (Godey et al., 2001) and recent fMRI-based dipole source localisations of MEG data (Benner et al., 2023) suggested separate sources in planum temporale (PT) as well as in planum polare (PP) anterior to HG. For EEG data obtained from unilateral cochlear implant (CI) users with preserved contralateral normal hearing, in contrast, a single dipole source of the P2 was localised to the planum temporale (Steinmetzger, Meinhardt, et al., 2022). Moreover, distributed MEG source reconstructions of the P2 revealed broadly distributed, right-lateralised activity in auditory areas in response to speech (Coffey et al., 2017). Lastly, some studies have also suggested that non-auditory cortical areas may at least partially be involved in generating the auditory P2 (Knight et al., 1980, 1988; Ponton et al., 2000).

In terms of the P2 morphology, an interesting feature is that it frequently contains two separate peaks, both at the scalp (Bertoli et al., 2011; Davis et al., 1966; Steinmetzger, Meinhardt, et al., 2022; Steinmetzger et al., 2020; Tremblay et al., 2014) and source levels (Andermann et al., 2017; Ross & Tremblay, 2009; Steinmetzger, Meinhardt, et al., 2022; Tiitinen et al., 1999). Although most of these studies did not explicitly discuss this characteristic, some have explicitly referred to it as “distinct second peak” (Ross & Tremblay, 2009) or “splitting P2” (Davis et al., 1966). While this feature does not appear to reflect the type of stimulus material used, there is some evidence that it is more pronounced in middle-aged and older subjects (Bertoli et al., 2011; Ross & Tremblay, 2009). However, it has not been investigated yet whether the two peaks might represent separate P2 subcomponents generated in different cortical areas and reflecting different functional processes, demonstrating how little is known about this component.

Prompted by the large, double-peaked P2s observed in response to voice pitch changes in our previous work (Steinmetzger, Megbel, et al., 2022; Steinmetzger et al., 2020), we here studied the P2 in more detail. To better

understand which stimulus features drive the P2, we compared the effects of pitch change magnitude and the type of pitch change. It has recently been shown that the P2 amplitude reflects the magnitude of pitch changes in musical sequences (Andermann et al., 2021), but it remains unclear if the P2 is also affected by the context in which these changes occur and how both effects compare. Specifically, participants were presented with sequences of speech-like sounds consisting of stimuli that either had a static pitch or dynamically varying pitch contours typical for natural speech, resulting in stepwise pitch changes confined to the transitions between stimuli or continuous pitch changes, respectively. Larger P2 amplitudes at stimulus onset were expected for stepwise pitch changes due to their greater saliency, despite a similar pitch change magnitude for both types of pitch changes.

Additionally, it was evaluated whether the P2 amplitude is affected by the harmonicity of the stimulus materials, that is, the property that the frequencies of the spectral components are integer multiples of the fundamental frequency ( $F_0$ ). As the P2 is enhanced for sounds with multiple harmonically related spectral components and also appears to reflect the familiarity with the materials, one would expect larger amplitudes for natural harmonic sounds as compared to artificially created inharmonic versions of these. Indeed, so-called harmonic template neurons, which preferentially fire in response to harmonic sounds, have been observed across the auditory cortex of marmoset monkeys (Feng & Wang, 2017; Wang, 2018) and to a lesser extent also in the rabbit mid-brain (Su & Delgutte, 2020). Yet, it is unclear whether such neurons also exist in sufficient number in the human auditory cortex to detect enhanced responses to harmonic sounds extracranially using EEG. Furthermore, it is unknown if the presence of these neurons is confined to the auditory cortex or whether harmonic sounds also elicit larger responses in auditory association cortex.

Regarding the cortical generators and morphology of the P2, we sought to determine if the two peaks in the sensor and source waveforms that were evident in several previous studies indeed represent separate P2 subcomponents. It was hence tested if the cortical sources of the two peaks differ and whether they are evoked at different stages of the auditory processing hierarchy. In contrast to most studies investigating the sources of the P2, distributed source reconstructions were used to be able to estimate the spatial extent of activity. Generally, the longer the latency of an auditory evoked response, the higher up the cortical hierarchy it is generated. The shortest latencies are usually observed in medial HG, that is, the primary auditory cortex (PAC), followed by secondary areas in lateral HG, and auditory association

areas adjacent to HG (Camalier et al., 2012; Godey et al., 2001; Nourski, 2017; Recanzone et al., 2000). We therefore expected that the first P2 peak might be generated in the lateral part of HG (Hari et al., 1987; Pantev et al., 1996; Ross & Tremblay, 2009; Tiitinen et al., 1999). In contrast, the sources of the second peak were assumed to include auditory association areas anterior and posterior to HG, that is, PP and PT, as reported in another group of studies concerned with localising the sources of the P2 (Benner et al., 2023; Coffey et al., 2017; Godey et al., 2001; Steinmetzger, Meinhardt, et al., 2022).

## 2. MATERIALS AND METHODS

### 2.1. Participants

Twenty subjects (9 females, 11 males; mean age 23 years, SD = 2.8 years) were tested and paid for their participation. They were all right-handed and reported no history of neurological or psychiatric illnesses. All participants used German as their main language and had audiometric thresholds of less than 20 dB hearing level (HL) at octave frequencies between 125 and 8000 Hz. All subjects gave written consent prior to the experiment, and the study was approved by the local research ethics committee (Medical Faculty, University of Heidelberg).

### 2.2. Stimuli

The stimulus materials were the same as in Steinmetzger et al. (2020), where the data were pooled across conditions for analysis. The experiment comprised five different stimulus conditions, four with discrete spectral components and speech-shaped noise. The stimuli with discrete spectral components were based on recordings from the EUROM database (Chan et al., 1995), consisting of five- to six-sentence passages read by 16 different male talkers. Using methods as previously described (Green & Rosen, 2013; Steinmetzger & Rosen, 2015), the  $F_0$  contours of the 16 passages were extracted and interpolated through unvoiced and silent periods to generate continuous  $F_0$  contours.

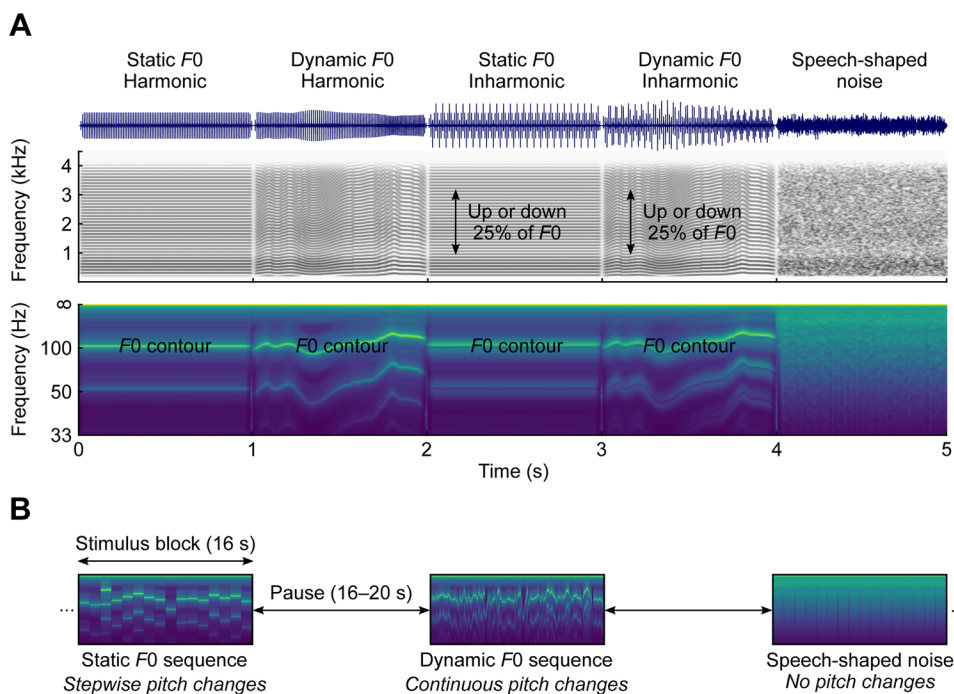
For the first stimulus condition (*Static  $F_0$  – Harmonic*), the log-transformed distribution of the  $F_0$  values for each individual talker was divided into 12 quantiles and used to generate a set of 192 1-s harmonic complex tones with static pitch contours (16 talkers x 12 quantiles). The complexes were synthesised with equal-amplitude components in sine phase and normalised to a median  $F_0$  of 100 Hz. To produce the second condition (*Dynamic  $F_0$  – Harmonic*), the 16 original pitch contours were used to generate harmonic complexes with dynamically varying pitch tracks. The first 12 s of each tone complex were

selected and divided into consecutive 1-s segments. For these two conditions, the frequencies of all component tones were integer multiples of the  $F_0$  and thus harmonically related. Additionally, inharmonic equivalents of the first two conditions were produced by shifting the frequencies of all component tones by 25% of the median  $F_0$  (*Static  $F_0$  – Inharmonic & Dynamic  $F_0$  – Inharmonic*). This procedure renders the stimuli inharmonic and reduces their pitch strength (B. Roberts & Brunstrom, 2001), but leaves all other acoustic properties largely unchanged (Steinmetzger & Rosen, 2023). The components were shifted by 25% as this value was shown to maximise the degree of inharmonicity for tone complexes with a fixed pitch (B. Roberts et al., 2010). For half the stimuli, the shift was applied upwards, for the other half it was applied downwards. A fifth condition (*Speech-shaped noise*), in which the stimuli contained no discrete spectral components and hence no pitch, was based on 192 different 1-s segments of white noise.

All stimuli had a sampling rate of 48 kHz and their spectra were shaped to have a similar long-term average speech spectrum, as described in Steinmetzger et al. (2020). After applying 25-ms Hann-windowed on- and off-set ramps, all stimuli were adjusted to have the same root-mean-square level. Example stimuli of all five conditions are shown in Figure 1A. For the waveforms depicted in the upper row, it is apparent that only the harmonic stimuli have periodic waveforms, while they are less regular for the inharmonic conditions, and completely aperiodic for speech-shaped noise. The narrow-band spectrograms in the middle row demonstrate that the spectra of the stimuli are indeed very similar, despite the markedly different waveforms. To visualise the different degrees of stimulus periodicity, spectrographic representations of summary autocorrelation functions (SACFs; Meddis & Hewitt, 1991; Meddis & O'Mard, 1997; for computational details see Steinmetzger et al., 2020) are shown in the bottom row. While the first peak in these SACF spectrograms represents the  $F_0$  contours of the stimuli, the height of this peak may be interpreted as a measure of pitch strength (Yost et al., 1996). In line with this notion, the peak around 10 ms is noticeably more pronounced for the harmonic stimuli compared to the inharmonic equivalents. Due to the lack of any temporal regularity, there was no such peak at all for speech-shaped noise.

### 2.3. Experimental design and procedure

The present experiment was originally designed as a simultaneous EEG and fNIRS study (Steinmetzger et al., 2020), and hence a block design was used to maximise the haemodynamic responses. Here, we re-analysed the EEG data, but omitted the fNIRS data because their



**Fig. 1.** Example stimuli and experimental design. (A) Waveforms, narrow-band spectrograms, and summary autocorrelation function (SACF) spectrograms showing the pitch contours for examples of the five stimulus types. (B) The individual 1-s stimuli were presented as continuous blocks. Sequences consisting of static F0 stimuli were characterised by stepwise pitch changes between stimuli, and those consisting of dynamic F0 stimuli exhibited continuous pitch changes. For speech-shaped noise sequences, there were no pitch changes at all.

limited depth resolution precludes fine-grained analyses of the activity emanating from deeper structures such as primary auditory cortex.

The individual 1-s stimuli in each condition were randomly concatenated into blocks consisting of 16 stimuli with no breaks in between and followed by pauses with random durations ranging from 16–20 s. Thus, the resulting stimulus blocks were continuous apart from the 25-ms on- and offset ramps applied to the individual stimuli. Each participant was presented with 12 blocks of each of the 5 stimulus conditions, adding up to a total duration of about 34 mins. The order of the blocks was randomised without any constraints. As the EEG data were analysed relative to the onset of the individual 1-s stimuli in each block, this design resulted in 192 trials per condition. As shown in Figure 1B, concatenating the harmonic or inharmonic stimuli with static F0s into blocks resulted in stepwise pitch changes at the onsets of the individual stimuli, while blocks consisting of harmonic or inharmonic stimuli with dynamic F0s were characterised by continuously changing, speech-like pitch contours. Blocks containing concatenated segments of speech-shaped noise were included as a control condition that had a similar spectral envelope but no pitch changes.

The experiment took place in a sound-attenuating and electrically shielded room, with the participant sitting in a comfortable reclining chair during data acquisition. There

was no behavioural task, but pauses were inserted about every 10 mins to ensure the vigilance of the subjects. The stimuli were presented with 24-bit resolution at a sampling rate of 48 kHz using an RME ADI-8 DS sound card (Haimhausen, Germany) and Etymotic Research ER2 earphones (Elk Grove Village, IL, USA) connected to a Tucker-Davis Technologies HB7 headphone buffer (Alachua, FL, USA). The presentation level was set to 70 dB SPL using an artificial ear (Brüel & Kjær, type 4157, Nærum, Denmark) and a corresponding measurement amplifier (Brüel & Kjær, type 2610, Nærum, Denmark).

## 2.4. EEG recording and analysis

Continuous EEG signals were recorded using a BrainVision actiCHamp system (Brain Products, Gilching, Germany) with 60 electrodes arranged according to the extended international 10-20 system. Four additional electrodes were placed around the eyes to record vertical and horizontal eye movements. The EEG data were recorded with an initial sampling rate of 500 Hz, an online anti-aliasing low-pass filter with a cut-off frequency of 140 Hz and were referenced to the right mastoid. The electrode positions of each subject were digitized with a Polhemus 3SPACE ISOTRAK II system before the experiment.

The data were pre-processed offline in the same way as in Steinmetzger et al. (2020) using FieldTrip (version

20180924; Oostenveld et al., 2011) and custom MATLAB code. The continuous waveforms were first segmented into epochs ranging from -0.3–1.1 s around stimulus onset. Next, the epochs were re-referenced to the mean of both mastoids and detrended as well as demeaned by removing a 1<sup>st</sup>-order polynomial. The epochs were then low-pass filtered (cut-off 15 Hz, 4<sup>th</sup>-order Butterworth, applied forwards and backwards), baseline corrected by subtracting the mean amplitude from -0.1–0 s before stimulus onset, and subsequently down-sampled to 250 Hz. After visually identifying and excluding bad channels (total = 4, max. 2 per subject), the data were decomposed into 20 principal components to detect and eliminate eye artefacts. After the 4 eye electrodes were removed from the data, epochs in which the amplitudes between -0.2–1 s around stimulus onset exceeded  $\pm 60 \mu\text{V}$  or the z-transformed amplitudes differed by more than 15 standard deviations from the mean of all channels were excluded from further processing. On average, 86% of the trials (830/960 per subject, min. 65% per subject) passed the rejection procedure. Lastly, bad channels were interpolated using the weighted average of the neighbouring channels; the data were re-referenced to the average of all 60 channels, and again baseline corrected from -0.1–0 s before stimulus onset.

Distributed source reconstructions of the resulting event-related potentials (ERPs) were computed using the MNE-dSPM approach implemented in Brainstorm (version 10-Jun-2022; Dale et al., 2000; Tadel et al., 2011). The electrode positions of each subject were co-registered to the ICBM152 MRI template by first aligning three external fiducial points (LPA, RPA, and Nz) and subsequently projecting the electrodes to the scalp of the template MRI. A Boundary Element Method (BEM) volume conduction model based on the ICBM152 template and the corresponding cortical surface (down-sampled to 15,000 vertices) were used as head and source models. The BEM head model was computed using OpenMEEG (version 2.4.1; Gramfort et al., 2010) and comprised three layers (scalp, outer skull, and inner skull) with 1082, 642, and 642 vertices, respectively. Linear MNE-dSPM solutions with dipole orientations constrained to be normal to the cortex were estimated for each subject and condition after pre-whitening the forward model with the averaged noise covariance matrix calculated from the individual trials in a time window from -0.2–0 s before stimulus onset. The default parameter settings for the depth weighting (order = 0.5, max. amount = 10), noise covariance regularisation (regularise noise covariance = 0.1), and regularisation parameter (SNR = 3) were used throughout.

Regarding the choice of auditory regions of interest (ROIs) from which the source waveforms were extracted, we opted for a simple scheme that takes the limited

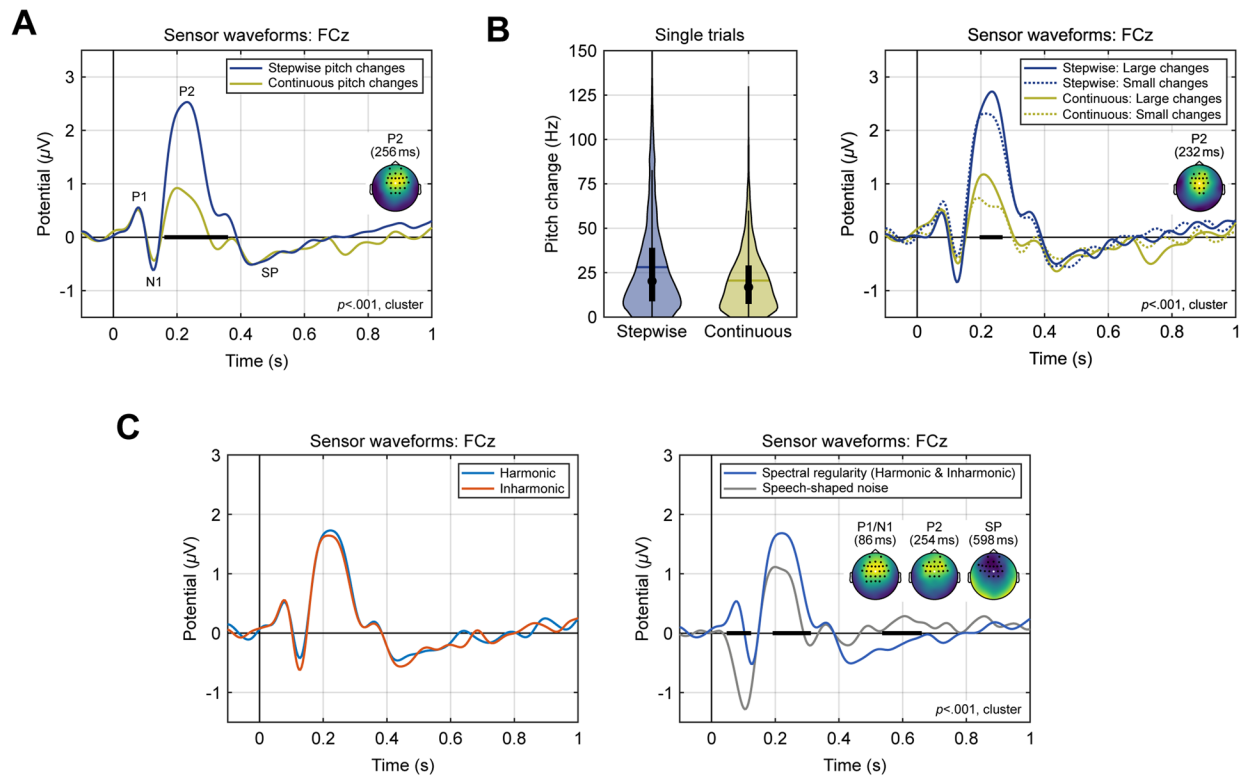
spatial resolution of EEG source reconstructions into account. Thus, a rather coarse macro-anatomical atlas that only distinguishes between gyri and sulci was used (Destrieux et al., 2010). As there is no strict correspondence between macro-anatomy, cytoarchitecture, and functional mapping results, a consensus concerning the organisation of the human auditory cortex is still lacking (Moerel et al., 2014; Saenz & Langers, 2014; Zachlod et al., 2020). The tonotopic organisation of auditory-sensitive areas extends well beyond HG, reaching anteriorly into PP and posteriorly into PT, demonstrating that PAC is not confined to HG (Moerel et al., 2014; Saenz & Langers, 2014). Yet, combined functional and micro-structural mapping results showed that frequency selectivity and myelination decrease when moving from HG into PT and PP, implying that the latter two regions are largely not part of PAC (Besle et al., 2019). For simplicity, we hence distinguished between a smaller core region [“auditory cortex”; HG & HS (Heschl’s sulcus)] and a surrounding larger region consisting of areas associated with higher-order auditory processing [“auditory association cortex”; PT, PP, & STS (superior temporal sulcus)].

### 3. RESULTS

#### 3.1. Sensor-level ERPs

In a first step, the scalp ERPs evoked by stepwise and continuous pitch changes were compared after pooling together the harmonic and inharmonic versions of both conditions. As shown in Figure 2A, stepwise pitch changes caused by transitions between stimuli with static  $F0$ s elicited markedly larger P2 amplitudes, as confirmed by a cluster-based permutation test ( $\sim 160$ – $352$  ms,  $t_{(\text{cluster})} = 4540.86$ ,  $p < 0.001^{***}$ ,  $d = 1.52$ ; Maris & Oostenveld, 2007). This test was based on sample-wise dependent-samples  $t$ -tests with a cluster-forming threshold of  $p < 0.05$  (two-sided), a minimum of 3 neighbouring electrodes per cluster, and 10,000 randomisations to determine the cluster  $p$ -values. The returned cluster had a fronto-central scalp distribution and included 24 electrodes at its midpoint (scalp map insert in Fig. 2A).

To estimate the effect of the pitch change magnitude on the P2 amplitudes, the stimulus sequences for each individual participant were then re-constructed and the pitch steps between successive stimuli were calculated using the SACF software described above. As the average magnitude was found to be somewhat larger for stepwise pitch changes (means = 28.1/20.6 Hz; medians = 16.8/20.2 Hz; see Fig. 2B for the distributions), the single trials were divided into subgroups above and below the average median across both conditions (18.5 Hz). Additionally, trials in the stepwise condition were excluded from analysis



**Fig. 2.** Sensor-level ERPs. Effects of pitch change type (A) and magnitude (B), as well as harmonicity and spectral regularity (C) on the P2 amplitude. ERPs traces are shown for electrode FCz, highlighted in the scalp maps. The thick horizontal black bars indicate significant time windows. In the scalp maps, the voltage of the second condition was subtracted from the first one, as indicated in the legends. The maps exhibit the voltage difference and the electrodes that were part of the respective cluster at its temporal midpoint. Violin plots in (B) show the distributions of pitch change magnitudes at the transitions between individual trials, along with ERPs after dividing the trials into subgroups with pitch change magnitudes above or below the overall median. SP, sustained potential.

if the magnitude exceeded the maximum of the continuous condition (129.6 Hz) to align the distributions. As illustrated in Figure 2B, trials with a magnitude above the median elicited larger P2 amplitudes for both types of pitch change (~196–268 ms,  $t_{(\text{cluster})} = 1055.1$ ,  $p < 0.001^{***}$ ,  $d = 1.61$ ; data pooled across stepwise and continuous conditions for testing). However, the duration as well as the P2 amplitude difference of this effect were much smaller compared to the effect of pitch change type (Fig. 2A).

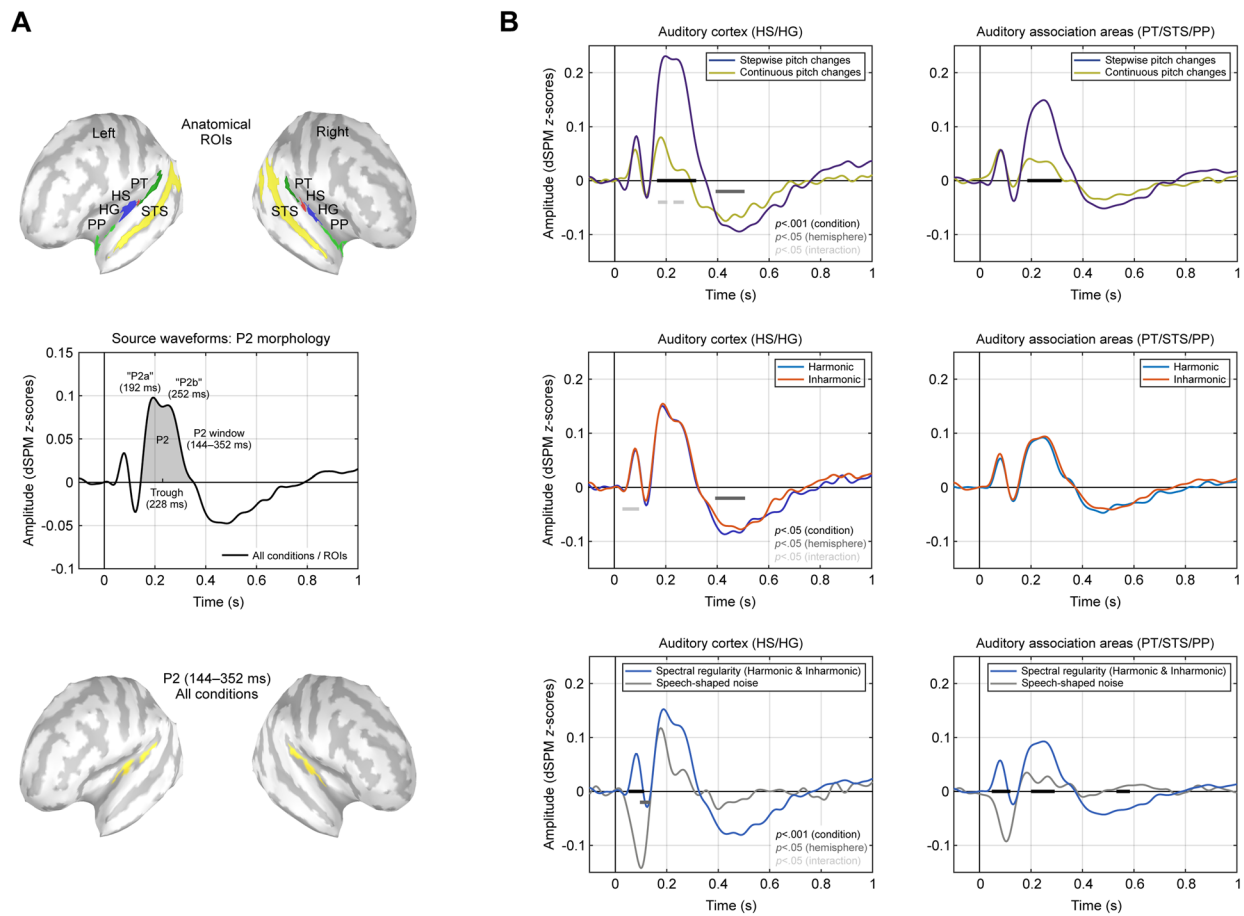
Next, the ERP data were analysed regarding potential effects of harmonicity by comparing all harmonic and inharmonic stimuli (i.e., the stepwise and continuous pitch change conditions were pooled together). Figure 2C shows that no such effects were evident throughout the entire duration of the stimuli, and the largest cluster returned had a  $p$ -value of 0.537.

In contrast, when comparing all stimuli with a regular spectral structure (i.e., all harmonic and inharmonic conditions pooled together) to a control condition comprising speech-shaped noise, cluster-based testing indicated three separate highly significant clusters, as shown on the right side of Figure 2C. These clusters

were due to the absence of a P1 and the larger N1 elicited by speech-shaped noise (~48–124 ms,  $t_{(\text{cluster})} = 3260.16$ ,  $p < 0.001^{***}$ ,  $d = 2.25$ ), and the increased P2 (~192–316 ms,  $t_{(\text{cluster})} = 2254.94$ ,  $p < 0.001^{***}$ ,  $d = 1.43$ ) and sustained potential amplitudes (SP; ~536–660 ms,  $t_{(\text{cluster})} = -2094.85$ ,  $p < 0.001^{***}$ ,  $d = 1.33$ ) evoked by stimuli with spectral regularity. All three clusters had a fronto-central scalp distribution and comprised at least 20 electrodes during the midpoint of the respective cluster time windows.

### 3.2. Source waveforms

As shown previously (Steinmetzger et al., 2020), the stimuli mainly evoked activity on the supratemporal plane and STS, as is typical for speech and speech-like stimuli (e.g., Belin et al., 2000). The source waveforms were hence extracted from a set of anatomical ROIs comprising regions along the supratemporal plane (PT, HS, HG, and PP) as well as STS, bilaterally (Fig. 3A, top panel), as specified in the Destrieux atlas (Destrieux et al., 2010) implemented in Brainstorm. The underlying distributed source reconstructions, however, were computed across



**Fig. 3.** Source waveforms. (A) Auditory anatomical regions of interest (ROIs) from which the source waveforms were extracted (top). Two-peaked morphology of the P2 component after averaging across all stimulus conditions and ROIs (middle), and corresponding source localisation across the entire P2 window (bottom). (B) Source waveforms for the contrasts of type of pitch change (top), harmonicity (middle), and spectral regularity (bottom). The waveforms are shown after averaging across ROIs belonging to the auditory cortex and auditory associations areas and were averaged across hemispheres. The thick horizontal bars indicate significant time windows. Effects shorter than 25 ms were omitted throughout. PT, planum temporale; HS, Heschl's sulcus; HG, Heschl's gyrus; PP, planum polare; STS, superior temporal sulcus.

the entire cortical surface, without applying any ROI-based spatial restrictions.

When averaged across all stimulus conditions and ROIs, the resulting source waveform was dominated by a large P2 (Fig. 3A, middle panel), same as the sensor waveforms. However, unlike the sensor-level ERPs, this source waveform exhibited a double-peaked morphology for the P2, with two separate peaks spaced approximately 60 ms apart (192/252 ms), hereafter referred to as “P2a” and “P2b”. We opted for a typical nomenclature based on morphology, where components are named based on their temporal sequence, unlike Benner et al. (2023) who referred to the anterior and posterior P2 sources as “P2” and “P2a”, respectively,

In a first step to identify the locations of the cortical generators of the P2, a source localisation across all stimulus conditions and the entire P2 window (144–352 ms)

was then computed (Fig. 3A, bottom panel). The areas showing the largest activity were located on the supra-temporal plane in both hemispheres, consistent with the set of ROIs. Here and in the remainder of the paper, the source maps were plotted such that activity beyond auditory areas was masked by adjusting the amplitude threshold and the minimum number of connected vertices accordingly.

Next, the source waveforms were averaged over ROIs belonging to the auditory cortex (HG & HS) and auditory association areas surrounding auditory cortex (PT, STS & PP), and analysed using the same three condition contrasts as before (Fig. 3B). These contrasts were statistically evaluated via dependent-samples *t*-tests (two-sided) for each time point from 0–1000 ms, with *p*-values determined by permutation testing (10,000 randomisations). The reported *t*-values represent the average over the

respective significant time window. To test for main effects of condition, the source waveforms were averaged across hemispheres, while main effects of hemisphere were evaluated by averaging across conditions. Interactions of condition and hemisphere were tested by comparing condition differences across hemispheres. Effects with a duration of less than 25 ms were considered as false positives and omitted throughout. In Figure 3B, the source waveforms are depicted after averaging over ROIs and hemispheres, but the waveforms for each individual ROI and hemisphere are provided in Figure 4.

The comparison of stepwise and continuous pitch changes (Fig. 3B, top row) showed that P2 amplitudes were significantly larger in both auditory cortex (164–320 ms,  $t_{(19)} = 4.94$ ,  $p < 0.001^{***}$ ,  $d = 1.26$ ) and association areas (184–316 ms,  $t_{(19)} = 4.93$ ,  $p < 0.001^{***}$ ,  $d = 1.23$ ) following stepwise changes. In the auditory cortex, a main effect of hemisphere (392–504 ms,  $t_{(19)} = 2.61$ ,  $p < 0.05^*$ ,  $d = 0.61$ ) furthermore indicated that the sustained potential was overall larger in the right hemisphere. In addition, the particularly large P2 amplitudes elicited by stepwise pitch changes in the right auditory cortex (Fig. 4A) resulted in a significant condition\*hemisphere interaction (168–204 & 228–268 ms,  $t_{(19)} = 2.39$ ,  $p < 0.05^*$ ,  $d = 0.61$ ).

For the harmonicity contrast (Fig. 3B, middle row), there were again no significant condition differences ( $p < 0.05$ ) for the P2 or any other component, neither in the auditory cortex nor association areas. Even when considering all ROIs separately, no significant main effects of condition were evident (Fig. 4B). However, a main effect of hemisphere (392–504 ms,  $t_{(19)} = 2.54$ ,  $p < 0.05^*$ ,  $d = 0.60$ ) indicated a larger sustained potential in the right auditory cortex, and a larger P1 for the inharmonic condition in the left auditory cortex resulted in a significant interaction (36–92 ms,  $t_{(19)} = 2.49$ ,  $p < 0.05^*$ ,  $d = 0.67$ ).

The contrast of spectral regularity and speech-shaped noise (Fig. 3B, bottom row), on the other hand, revealed highly significant condition differences during the P1/N1 period in both auditory cortex (52–112 ms,  $t_{(19)} = 8.96$ ,  $p < 0.001^{***}$ ,  $d = 2.25$ ) and association areas (48–120 ms,  $t_{(19)} = 8.23$ ,  $p < 0.001^{***}$ ,  $d = 2.16$ ), as well as during the P2 (200–292 ms,  $t_{(19)} = 5.39$ ,  $p < 0.001^{***}$ ,  $d = 1.35$ ) and SP windows (532–584 ms,  $t_{(19)} = 3.71$ ,  $p < 0.001^{***}$ ,  $d = 0.86$ ) in association areas. In line with the sensor-level results, P1, P2, and SP were thus larger in amplitude for the stimuli with spectral regularity, while the N1 was enhanced for speech-shaped noise. In addition, the source-level results showed that the P2 and SP effects emerged from the auditory association cortex, particularly STS (Fig. 4C). Furthermore, a main effect of hemisphere was observed for the N1 (96–136 ms,  $t_{(19)} = 2.59$ ,  $p < 0.05^*$ ,  $d = 0.61$ ), indicating larger amplitudes in the right auditory cortex in both conditions.

### 3.3. Comparison of the P2a and P2b subcomponents

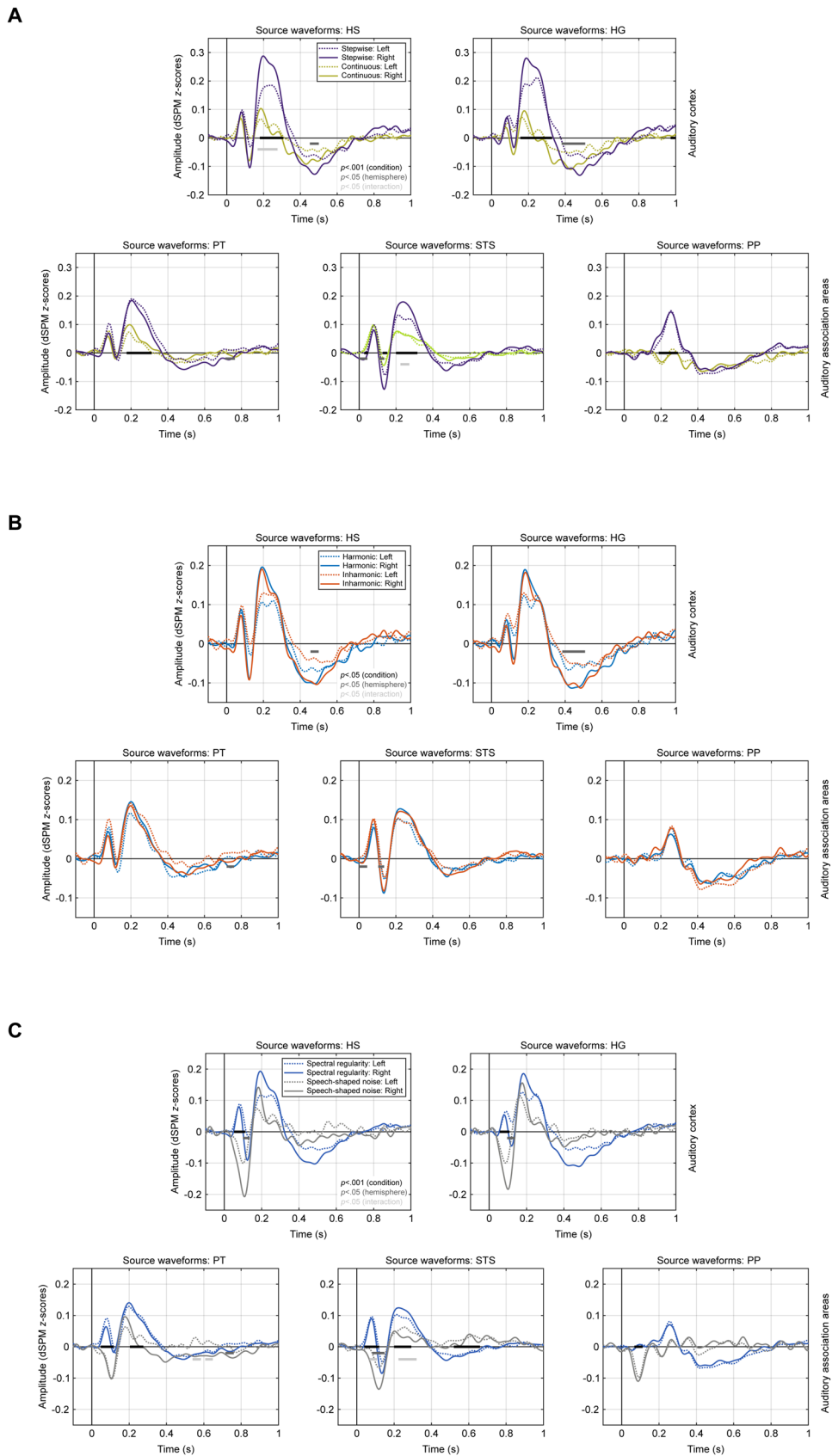
Since the largest P2 amplitudes were evoked by stepwise pitch changes, we first evaluated the source and scalp maps for these stimuli. As shown in Figure 5A, the time-averaged source activity for the P2a (144–228 ms) was greatest in the auditory cortex, whereas the generators of the P2b (228–352 ms) were more broadly distributed along the supratemporal planes. The corresponding scalp maps revealed a central scalp distribution for the P2a, while the P2b had a fronto-central and slightly right-lateralised topography. The time windows of the P2 subcomponents were derived by dividing the entire P2 window (144–352 ms) of the grand-average source waveform across all conditions and ROIs into segments before and after the trough at 228 ms that separates the two P2 peaks (Fig. 3A).

In Figure 5B, the averaged source amplitudes in response to stepwise pitch changes are shown for each ROI and hemisphere. A mixed-effects regression model with the fixed effects *component*, *region*, and *hemisphere*, and random intercepts for each subject confirmed that the P2a amplitude was significantly larger in the auditory cortex compared to auditory association areas (*component* \* *region* interaction:  $F_{(1,133)} = 5.03$ ,  $p = 0.027^*$ ; Tukey post-hoc contrast:  $t_{(133)} = 4.10$ ,  $p < 0.001^{***}$ ,  $d = 0.92$ ). On the contrary, no such difference was observed for the P2b ( $t_{(133)} = 1.46$ ,  $p = 0.486$ ,  $d = 0.32$ ), indicating a similar level of activity across regions.

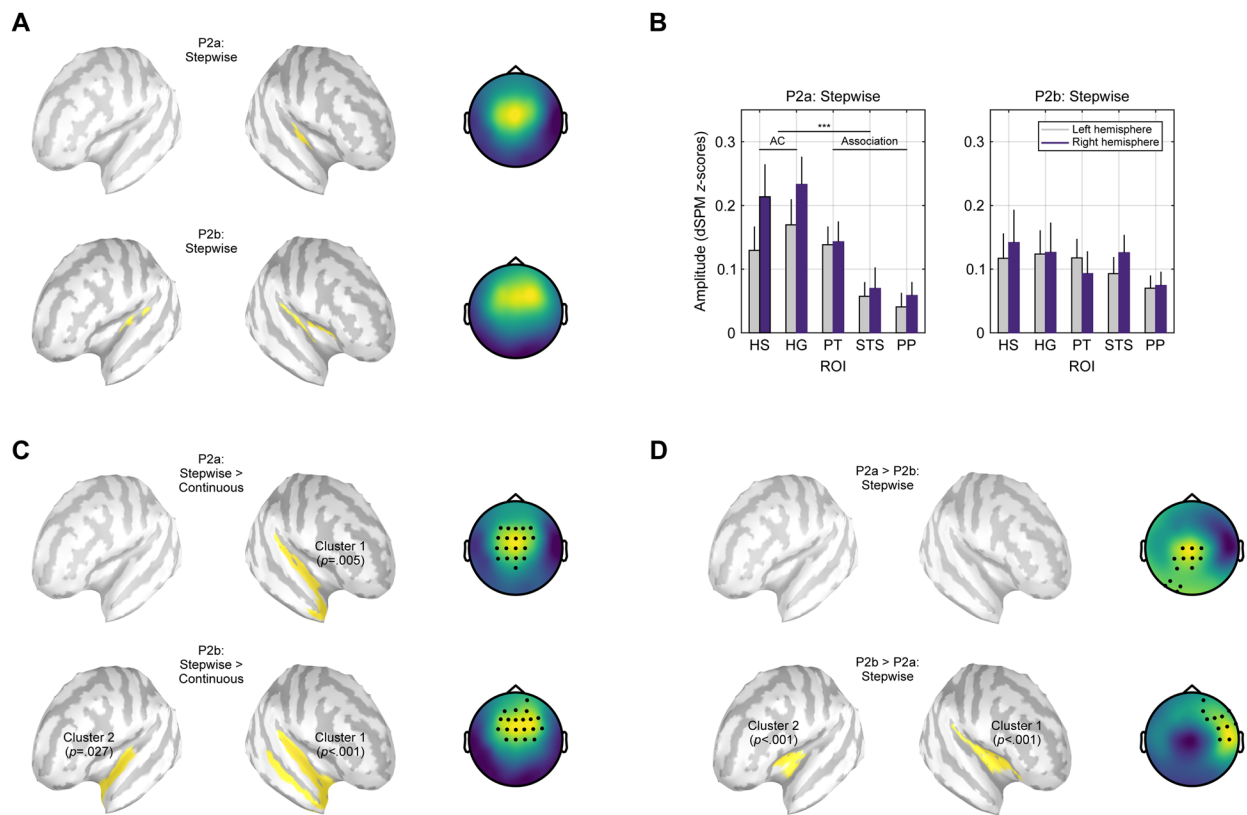
Next, the distributed source reconstructions and scalp distributions of the P2 subcomponents evoked by stepwise and continuous pitch changes were statistically compared (Fig. 5C). To identify auditory regions where stepwise changes elicited greater activity than continuous changes, cluster-based permutation tests were computed, for which the source amplitudes of each vertex on the cortical surface were averaged over the respective P2 time window. These tests were based on dependent-samples *t*-tests for each vertex with a cluster-forming threshold of  $p < 0.05$  (one-sided), a minimum of 3 neighbouring vertices per cluster, and 10,000 randomisations to determine the cluster *p*-values. Only clusters overlapping with the auditory ROIs are reported (Fig. 3A).

For the P2a window, a single significant cluster ( $t_{(\text{cluster})} = 890.98$ , size = 192 vertices,  $p = 0.005^{**}$ ,  $d = 1.77$ ) indicated that activity in response to stepwise pitch changes was stronger across the length of the right supratemporal plane. At the sensor level, this comparison revealed a large cluster in the central scalp region (19 electrodes,  $t_{(\text{cluster})} = 95.70$ ,  $p < 0.001^{***}$ ,  $d = 1.30$ ). For the P2b window, two separate clusters indicated greater activity following stepwise changes along the right supratemporal





**Fig. 4.** Source waveforms for the individual ROIs and hemispheres: type of pitch change (A), harmonicity (B), and spectral regularity (C). The structure and details of the figure are the same as for the source waveforms shown in Figure 3B.



**Fig. 5.** Comparison of P2a and P2b. (A) Cortical generators and scalp maps of the P2a and P2b subcomponents elicited by stepwise pitch changes. (B) Source amplitudes of the P2a and P2b elicited by stepwise changes, separately for each hemisphere and auditory ROI. (C) Source and scalp level statistical comparisons showing where P2a and P2b in response to stepwise changes were larger than for continuous changes. (D) Source and scalp level statistics comparing P2a and P2b evoked by stepwise changes.

plane as well as STS ( $t_{(\text{cluster})} = 1081.39$ , size = 319 vertices,  $p < 0.001^{***}$ ,  $d = 1.07$ ) and the anterior portion of the left supratemporal plane ( $t_{(\text{cluster})} = 457.94$ , size = 155 vertices,  $p = 0.027^*$ ,  $d = 0.90$ ). At the scalp level, this comparison resulted in a single significant cluster with a fronto-central distribution P2b (20 electrodes,  $t_{(\text{cluster})} = 99.76$ ,  $p < 0.001^{***}$ ,  $d = 1.26$ ). Despite the greater spatial extent, the locations of these clusters are in line with the generators of the P2a and P2b in response to stepwise changes (Fig. 5A). Whereas the first P2 subcomponent mainly originated from the auditory cortex, particularly in the right hemisphere, the second subcomponent showed a broader distribution including auditory and association areas in both hemispheres.

Finally, the P2a and P2b evoked by stepwise pitch changes were directly compared (Fig. 5D). For the P2b, activity around the left ( $t_{(\text{cluster})} = 439.86$ , size = 79 vertices,  $p < 0.001^{***}$ ,  $d = 1.58$ ) and particularly the right lateral sulcus ( $t_{(\text{cluster})} = 881.91$ , size = 142 vertices,  $p < 0.001^{***}$ ,  $d = 2.00$ ) was significantly larger than for the P2a, while no effects in the opposite direction were observed in auditory areas ( $p \geq 0.202$ ). To limit the spatial extent of the significant clusters, a cluster-forming

threshold of  $p < 0.001$  was applied. At the scalp level, the P2b amplitudes were significantly larger in the right fronto-temporal scalp region (11 electrodes,  $t_{(\text{cluster})} = 53.57$ ,  $p = 0.002^{**}$ ,  $d = 1.41$ ), whereas P2a amplitudes were larger at central and left posterior electrode sites (13 electrodes,  $t_{(\text{cluster})} = 51.70$ ,  $p = 0.002^{**}$ ,  $d = 1.31$ ). To quantify the difference of the scalp topographies of P2a and P2b, we furthermore computed their Global Map Dissimilarity (GMD; Brunet et al., 2011). The GMD can take values between 0 and 2, indicating that two maps are either identical or the inversion of each other. Here, the average GMD across all subjects was 0.79 (SD = 0.37) and significantly greater than 0 ( $t_{(19)} = 9.51$ ,  $p < 0.001^{***}$ ), confirming that the scalp maps of the two P2 subcomponents differed markedly.

## 4. DISCUSSION

### 4.1. Cortical generators and functional significance of the P2

The distributed ERP source reconstruction computed across all stimulus conditions and the entire P2 time

window revealed bilateral foci of activity in PT and around HG, extending anteriorly into PP (Fig. 3A). At the scalp level, this was reflected in a broad positive voltage deflection with a fronto-central distribution (Fig. 2). This present results are in line with previous data reporting separate sources of the P2 in the lateral part of the auditory cortex and anterior to it (Benner et al., 2023; Hari et al., 1987; Pantev et al., 1996; Ross & Tremblay, 2009; Tiitinen et al., 1999) as well as in PT (Benner et al., 2023; Godey et al., 2001; Steinmetzger, Meinhardt, et al., 2022). The PT has been argued to be the cortical site in which complex spectro-temporal patterns in auditory scenes are segregated and compared with learned representations (Griffiths & Warren, 2002). Hence, its involvement in the processing of the spectro-temporal modulations constituting voice pitch changes, as reflected by the P2, appears plausible from a functional point of view. Similarly, regions in the anterior to auditory cortex, particularly in the right hemisphere, have been shown to be crucially involved in the processing of pitch changes and melodies (Johnsrude et al., 2000; Patterson et al., 2002; Zatorre & Belin, 2001).

Regarding the acoustic stimulus properties, the current results have shown that the P2 amplitude is sensitive to the type and, to a lesser extent, the magnitude of voice pitch changes in sequences of speech-like sounds. The two types of voice pitch changes employed comprised stepwise and continuous changes, and the P2 was the only auditory ERP component affected by these changes. In sequences with stepwise pitch changes, the individual stimuli had a static pitch resembling monotonised speech and pitch changes were consequently restricted to the transitions between sounds. In contrast, sequences with continuous pitch changes were formed of sounds with dynamically varying pitch contours extracted from natural speech. Hence, there were pitch jumps between the individual stimuli as well as continuous pitch changes throughout the sequences. In both sequence types, other acoustic factors such as duration, level, and spectral envelope were kept constant. The driving factor behind the substantially larger P2 evoked by stepwise changes appears to be their greater saliency. The ongoing modulation of the pitch contours in sequences with continuous pitch changes likely resulted in a greater degree of neural adaptation compared to sequences with stepwise pitch changes.

The present results furthermore revealed no differences in P2 amplitude between harmonic stimuli and their inharmonic equivalents. This finding applies to all auditory cortical regions examined as well as all other ERP components besides the P2. The stimuli were rendered inharmonic by shifting all spectral components in frequency, a technique that maintains the presence and

regular spacing of spectral components and leaves the envelope modulations unaffected, but results in a weaker pitch. These properties, which have recently been verified by detailed acoustic analyses and psychoacoustic measurements (Steinmetzger & Rosen, 2023), make this stimulus type ideally suited for investigating potential effects of stimulus harmonicity. In the neurosciences, however, shifted inharmonic stimuli have previously only been used in animal studies. Invasive recordings from marmosets (Feng & Wang, 2017; Wang, 2018) and rabbits (Su & Delgutte, 2020) have provided evidence for the existence of so-called “harmonic template neurons” that show increased firing in response to harmonic sounds. Yet, at least in the core auditory cortex of marmosets (Feng & Wang, 2017), there appear to be relatively few such neurons. Assuming the same applies to the human auditory cortex, the current results suggest that it may not be possible to detect the responses of these neurons in non-invasive recordings due to their limited number.

Furthermore, the P2 has been shown to reflect both short- and long-term neuroplastic changes, as amplitudes were found to increase across experimental blocks and sessions using the same materials (MacLean et al., 2024; Sheehan et al., 2005; Tremblay et al., 2014), and speech as well as musical sounds elicited larger amplitudes in musicians compared to non-musicians (MacLean et al., 2024; Shahin et al., 2003). The long-term exposure to harmonic sounds in speech and music might thus have been expected to result in larger P2 amplitudes. However, the similar P2 amplitudes in response to the harmonic and inharmonic stimuli suggest that the latter were perceived as speech-like despite their unusual timbre.

The absence of an effect of harmonicity in the current study furthermore implies that the pitch strength of the stimulus materials *per se* does not affect auditory cortex activity. As can be seen in Figure 1A, and demonstrated in more detail in Steinmetzger and Rosen (2023), the pitch strength, or periodicity, of the inharmonic stimuli is markedly lower than that of their harmonic equivalents. This makes for a different reading of experiments that have investigated pitch-related responses in the auditory cortex. Several of the studies that reported enhanced activity in the “pitch centre” located at the anterolateral border of HG in response to sounds giving rise to a pitch percept contrasted pulse trains with regular and irregular spacing (Bendor & Wang, 2005, 2010; Gutschalk et al., 2002, 2004; Gutschalk & Uppenkamp, 2011). Yet, while irregular pulse trains do not evoke a clear pitch, this manipulation also results in stimuli without discrete spectral components. The same is true for studies using iterated rippled noise (IRN; Griffiths et al., 1998; Ritter et al., 2005). When the number of iterations in the construction of IRN materials is reduced, both the pitch and

spectral peaks dissipate. Lastly, the pitch strength of the materials has also been altered by either reducing the number of resolved harmonics (S. Norman-Haignere et al., 2013) or by using complex tones with only resolved or unresolved harmonics (Penagos et al., 2004). Crucially, none of the stimuli used in the above studies allowed for a manipulation of the pitch strength that is independent of the presence and number of spectral components, unlike the shifted inharmonic materials used in the present experiment. In agreement with this, Feng and Wang (2017) also did not report increased responses in the marmoset pitch centre for harmonic compared to shifted inharmonic stimuli. It is thus conceivable that the activity in the cortical pitch centre merely reflects the number of discrete, regularly spaced spectral components in the stimulus materials rather than their pitch.

When pooling the harmonic and inharmonic conditions together, however, the P2 was markedly larger compared to a control condition of speech-shaped noise that had no spectral regularity, that is, no discrete spectral components. This finding is consistent with fMRI results showing enhanced responses to harmonic sounds compared to spectrally matched noise across the human auditory cortex (S. Norman-Haignere et al., 2013; S. Norman-Haignere et al., 2019). Same as the P2, the P1 was larger for sounds with spectral regularity and, additionally, almost absent in response to speech-shaped noise. In turn, speech-shaped noise appeared to elicit a larger N1. However, due to the temporal overlap of these effects and the limited spatial resolution of EEG source reconstructions, it cannot be inferred whether these components were in fact larger or if they cancelled each other.

In general, surface-positive deflections such as the P2 are thought to originate from deeper cortical layers and thus primarily receive thalamic input, whereas surface-negative potentials like the N1 are assumed to be generated in superficial cortical layers with predominantly cortico-cortical input (Fernandez Pujol et al., 2023; Jones et al., 2007; Steinschneider et al., 2011, 2013). The thalamic input might suggest a sharper frequency tuning of evoked responses with positive surface polarity. However, the P2 is considered to result from the non-lemniscal auditory pathway, which exhibits broad frequency tuning and no tonotopic organisation, in contrast to P1 and N1 that are generated via the lemniscal pathway (Crowley & Colrain, 2004; Parras et al., 2017). Consistent with the idea of a broader frequency tuning of the underlying neuronal populations, the auditory P2 but not the N1 exhibited attention-related frequency-specific sharpening depending on the prior auditory context in an adaptation paradigm (de Boer & Krumbholz, 2018). In contrast, attention-related gain effects were found to be much stronger for the N1

compared to the P2 (de Boer & Krumbholz, 2018; Neelon et al., 2006). In the context of the present study, the broad frequency tuning of the neurons generating the P2 appears to be a crucial pre-requisite for the detection of distinct spectral components. This property might help explain the larger amplitudes in response to sounds with distinct spectral components compared to noise, as well as the previous finding that sounds containing a greater number of adjacent harmonics evoked larger P2 amplitudes (Shahin et al., 2005).

#### 4.2. The P2 can be partitioned into two distinct subcomponents

We furthermore examined if the two separate P2 peaks evident in the source waveforms are generated in different cortical areas. It was assumed that focussing on the large P2 evoked by stepwise voice pitch changes would enable robust source estimations due to the favourable signal-to-noise ratio. For the first subcomponent, termed P2a, activity in the auditory cortex was significantly stronger than in the surrounding auditory association cortex and a pronounced trend for more activity in the right hemisphere was observed (Fig. 5B). For the subsequent P2b, in contrast, a comparable degree of activity was evident across all auditory ROIs and activity levels were similar across hemispheres. The wide network of cortical regions involved in generating the first and particularly the second P2 subcomponent suggests that distributed source reconstructions might be better suited to identify the cortical generators of the P2 than classic dipole solutions. It should be emphasised that the observed P2 subcomponents were elicited by the acoustic stimulus features in a bottom-up manner and that further research is needed to determine if a similar partition is also evident when investigating effects of top-down processing on the P2.

The differences observed at the source level were also reflected in significant differences in the respective scalp topographies. Although both subcomponents exhibited a positive surface polarity, the P2a had a central scalp distribution, whereas the P2b showed a fronto-central and slightly right-lateralised distribution. The more anterior distribution of the P2b is consistent with its cortical generators, as the source maps showed significantly stronger activity in areas anterior to the auditory cortex than for the P2a (Fig. 5). In addition, the two distinct P2 subcomponents were also evident in the sensor waveforms, with central electrode sites (e.g., Cz) showing a clear P2a peak, while fronto-central channels such as Fz exhibited a pronounced P2b instead (Suppl. Fig. 1).

As is evident from the source waveforms in Figure 3B, all stimulus conditions evoked a discernible P2a in the

auditory cortex. Although the amplitude of this peak was markedly larger for stepwise changes, the first P2 subcomponent is thus elicited irrespective of the acoustic properties of the materials. The P2a might therefore represent an obligatory initial processing step, indicating that some form of change to the pitch or spectral structure of the stimuli has occurred. The second P2 subcomponent, in turn, was practically absent for all conditions except stepwise pitch changes, both in the auditory cortex and association areas. Hence, it is conceivable that a sufficiently large first P2a triggers additional processing on the next level of the cortical hierarchy, as reflected in the P2b. Besides the auditory cortex, generators of the P2b were localised to PT, a region associated with the processing of complex spectro-temporal patterns (Griffiths & Warren, 2002), and PP, located anterior to the auditory cortex. As several neuroimaging and lesion studies have shown (e.g., Johnsrude et al., 2000; Patterson et al., 2002; Zatorre & Belin, 2001), pitch changes in speech and music are preferentially processed in the right PP. Furthermore, the cortical generators of the P2b included STS, a region exhibiting voice-selective activity (e.g., Belin et al., 2000; Hickok & Poeppel, 2007), suggesting that the stimuli were classified as voice-like at this point.

In contrast to the current results, where the cortical generators of the P2 were confined to auditory areas, it has been claimed that the P2 may at least in part be generated in non-auditory cortical areas (Crowley & Colrain, 2004). The most substantial evidence for this assumption is provided by human lesions studies, which demonstrated that, contrary to the N1, the P2 amplitude evoked by tone bursts showed little reduction in patients with unilateral lesions of the posterior superior temporal gyrus (Knight et al., 1980, 1988). According to this view, the P2 reflects output of the mesencephalic reticular activation system (RAS) that responds to all sensory modalities, with the insular cortex as one possible non-auditory generator site (Ponton et al., 2000). Importantly, the ERP data in Ponton et al. (2000) provide clear evidence that a P2 at posterior scalp sites, that is, with non-auditory cortical sources, is present by the age of 5, while a P2 in the central scalp region with supposedly auditory generators only emerges several years later. This suggests that this non-auditory P2 may be yet another subcomponent of the P2, in addition to the auditory P2a and P2b subcomponents delineated in the present study.

#### DATA AND CODE AVAILABILITY

The stimuli and EEG data are available at <https://osf.io/tnfdg>, and the code used to process the data can be found at <https://osf.io/bnzmy>.

#### AUTHOR CONTRIBUTIONS

K.S.: Conceptualisation, Software, Formal analysis, Investigation, Data curation, Writing—Original Draft, Visualisation, and Funding acquisition. A.R.: Conceptualisation, Resources, Writing—Review & Editing, Supervision, Project administration, and Funding acquisition.

#### FUNDING

This work was supported by the Dietmar Hopp Stiftung (grant number 2301 1239).

#### DECLARATION OF COMPETING INTEREST

None of the authors has any competing interests to declare.

#### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available with the online version here: [https://doi.org/10.1162/imag\\_a\\_00160](https://doi.org/10.1162/imag_a_00160).

#### REFERENCES

- Andermann, M., Günther, M., Patterson, R. D., & Rupp, A. (2021). Early cortical processing of pitch height and the role of adaptation and musicality. *Neuroimage*, 225, 117501. <https://doi.org/10.1016/j.neuroimage.2020.117501>
- Andermann, M., Patterson, R. D., Vogt, C., Winterstetter, L., & Rupp, A. (2017). Neuromagnetic correlates of voice pitch, vowel type, and speaker size in auditory cortex. *Neuroimage*, 158, 79–89. <https://doi.org/10.1016/j.neuroimage.2017.06.065>
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403, 309–312. <https://doi.org/10.1038/35002078>
- Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, 436, 1161–1165. <https://doi.org/10.1038/nature03867>
- Bendor, D., & Wang, X. (2010). Neural coding of periodicity in marmoset auditory cortex. *Journal of Neurophysiology*, 103, 1809–1822. <https://doi.org/10.1152/jn.00281.2009>
- Benner, J., Reinhardt, J., Christiner, M., Wengenroth, M., Stippich, C., Schneider, P., & Blatow, M. (2023). Temporal hierarchy of cortical responses reflects core-belt-parabelt organization of auditory cortex in musicians. *Cerebral Cortex*, 33, 7044–7060. <https://doi.org/10.1093/cercor/bhad020>
- Bertoli, S., Probst, R., & Bodmer, D. (2011). Late auditory evoked potentials in elderly long-term hearing-aid users with unilateral or bilateral fittings. *Hearing Research*, 280, 58–69. <https://doi.org/10.1016/j.heares.2011.04.013>
- Besle, J., Mougin, O., Sánchez-Panchuelo, R.-M., Lanting, C., Gowland, P., Bowtell, R., Francis, S., & Krumbholz, K. (2019). Is human auditory cortex organization compatible with the monkey model? Contrary evidence from ultra-high-field functional and structural MRI. *Cerebral Cortex*, 29, 410–428. <https://doi.org/10.1093/cercor/bhy267>

- Brunet, D., Murray, M. M., & Michel, C. M. (2011). Spatiotemporal analysis of multichannel EEG: CARTOOL. *Computational Intelligence and Neuroscience*, 2011, 2. <https://doi.org/10.1155/2011/813870>
- Camalier, C. R., D'Angelo, W. R., Sterbing-D'Angelo, S. J., de la Mothe, L. A., & Hackett, T. A. (2012). Neural latencies across auditory cortex of macaque support a dorsal stream supramodal timing advantage in primates. *Proceedings of the National Academy of Sciences*, 109, 18168–18173. <https://doi.org/10.1073/pnas.1206387109>
- Chan, D., Fourcin, A., Gibbon, D., Granström, B., Huckvale, M., Kokkinas, G., Kvale, L., Lamel, L., Lindberg, L., & Moreno, A. (1995). EUROM—A spoken language resource for the EU. Paper presented at the Eurospeech.
- Coffey, E. B., Chepesiuk, A. M., Herholz, S. C., Baillet, S., & Zatorre, R. J. (2017). Neural correlates of early sound encoding and their relationship to speech-in-noise perception. *Frontiers in Neuroscience*, 11, 479. <https://doi.org/10.3389/fnins.2017.00479>
- Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology*, 115, 732–744. <https://doi.org/10.1016/j.clinph.2003.11.021>
- Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, 26, 55–67. [https://doi.org/10.1016/S0896-6273\(00\)81138-1](https://doi.org/10.1016/S0896-6273(00)81138-1)
- Davis, H., Mast, T., Yoshie, N., & Zerlin, S. (1966). The slow response of the human cortex to auditory stimuli: Recovery process. *Electroencephalography and Clinical Neurophysiology*, 21, 105–113. [https://doi.org/10.1016/0013-4694\(66\)90118-0](https://doi.org/10.1016/0013-4694(66)90118-0)
- de Boer, J., & Krumbholz, K. (2018). Auditory attention causes gain enhancement and frequency sharpening at successive stages of cortical processing—Evidence from human electroencephalography. *Journal of Cognitive Neuroscience*, 30, 785–798. [https://doi.org/10.1162/jocn\\_a\\_01245](https://doi.org/10.1162/jocn_a_01245)
- Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, 53, 1–15. <https://doi.org/10.1016/j.neuroimage.2010.06.010>
- Feng, L., & Wang, X. (2017). Harmonic template neurons in primate auditory cortex underlying complex sound processing. *Proceedings of the National Academy of Sciences*, 114, E840–E848. <https://doi.org/10.1073/pnas.1607519114>
- Fernandez Pujol, C., Blundon, E. G., & Dykstra, A. R. (2023). Laminar specificity of the auditory perceptual awareness negativity: A biophysical modeling study. *PLoS Computational Biology*, 19, e1011003. <https://doi.org/10.1371/journal.pcbi.1011003>
- Godey, B., Schwartz, D., De Graaf, J., Chauvel, P., & Liegeois-Chauvel, C. (2001). Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: A comparison of data in the same patients. *Clinical Neurophysiology*, 112, 1850–1859. [https://doi.org/10.1016/S1388-2457\(01\)00636-8](https://doi.org/10.1016/S1388-2457(01)00636-8)
- Gramfort, A., Papadopoulos, T., Olivi, E., & Clerc, M. (2010). OpenMEEG: Opensource software for quasistatic bioelectromagnetics. *Biomedical Engineering Online*, 9, 45. <https://doi.org/10.1186/1475-925X-9-45>
- Green, T., & Rosen, S. (2013). Phase effects on the masking of speech by harmonic complexes: Variations with level. *Journal of the Acoustical Society of America*, 134, 2876–2883. <https://doi.org/10.1121/1.4820899>
- Griffiths, T. D., Büchel, C., Frackowiak, R. S., & Patterson, R. D. (1998). Analysis of temporal structure in sound by the human brain. *Nature Neuroscience*, 1, 422–427. <https://doi.org/10.1038/1637>
- Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences*, 25, 348–353. [https://doi.org/10.1016/S0166-2236\(02\)02191-4](https://doi.org/10.1016/S0166-2236(02)02191-4)
- Gutschalk, A., Patterson, R. D., Rupp, A., Uppenkamp, S., & Scherg, M. (2002). Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. *Neuroimage*, 15, 207–216. <https://doi.org/10.1006/nimg.2001.0949>
- Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S., & Rupp, A. (2004). Temporal dynamics of pitch in human auditory cortex. *Neuroimage*, 22, 755–766. <https://doi.org/10.1016/j.neuroimage.2004.01.025>
- Gutschalk, A., & Uppenkamp, S. (2011). Sustained responses for pitch and vowels map to similar sites in human auditory cortex. *Neuroimage*, 56, 1578–1587. <https://doi.org/10.1016/j.neuroimage.2011.02.026>
- Hari, R., Pelizzone, M., Mäkelä, J., Hällström, J., & Leinonen, L. (1987). Neuromagnetic responses of the human auditory cortex to on- and offsets of noise bursts. *Audiology*, 26, 31–43. <https://doi.org/10.3109/00206098709078405>
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393–402. <https://doi.org/10.1038/nrn2113>
- Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, 123, 155–163. <https://doi.org/10.1093/brain/123.1.155>
- Jones, S. R., Pritchett, D. L., Stufflebeam, S. M., Hämäläinen, M., & Moore, C. I. (2007). Neural correlates of tactile detection: A combined magnetoencephalography and biophysically based computational modeling study. *Journal of Neuroscience*, 27, 10751–10764. <https://doi.org/10.1523/JNEUROSCI.0482-07.2007>
- Knight, R. T., Hillyard, S. A., Woods, D. L., & Neville, H. J. (1980). The effects of frontal and temporal-parietal lesions on the auditory evoked potential in man. *Electroencephalography and Clinical Neurophysiology*, 50, 112–124. [https://doi.org/10.1016/0013-4694\(80\)90328-4](https://doi.org/10.1016/0013-4694(80)90328-4)
- Knight, R. T., Scabini, D., Woods, D. L., & Clayworth, C. (1988). The effects of lesions of superior temporal gyrus and inferior parietal lobe on temporal and vertex components of the human AEP. *Electroencephalography and Clinical Neurophysiology*, 70, 499–509. [https://doi.org/10.1016/0013-4694\(88\)90148-4](https://doi.org/10.1016/0013-4694(88)90148-4)
- Krumbholz, K., Patterson, R., Seither-Preisler, A., Lammertmann, C., & Lütkenhöner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex*, 13, 765–772. <https://doi.org/10.1093/cercor/13.7.765>
- MacLean, J., Stirn, J., Sisson, A., & Bidelman, G. M. (2024). Short- and long-term neuroplasticity interact during the perceptual learning of concurrent speech. *Cerebral Cortex*, bhad543. <https://doi.org/10.1093/cercor/bhad543>
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164, 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>

- Meddis, R., & Hewitt, M. J. (1991). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *Journal of the Acoustical Society of America*, 89, 2866–2882. <https://doi.org/10.1121/1.400725>
- Meddis, R., & O'Mard, L. (1997). A unitary model of pitch perception. *Journal of the Acoustical Society of America*, 102, 1811–1820. <https://doi.org/10.1121/1.420088>
- Moerel, M., De Martino, F., & Formisano, E. (2014). An anatomical and functional topography of human auditory cortical areas. *Frontiers in Neuroscience*, 8, 225. <https://doi.org/10.3389/fnins.2014.00225>
- Nääätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, 24, 375–425. <https://doi.org/10.1111/j.1469-8986.1987.tb00311.x>
- Neelon, M. F., Williams, J., & Garell, P. C. (2006). The effects of auditory attention measured from human electrocorticograms. *Clinical Neurophysiology*, 117, 504–521. <https://doi.org/10.1016/j.clinph.2005.11.009>
- Norman-Haignere, S., Kanwisher, N., & McDermott, J. H. (2013). Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *Journal of Neuroscience*, 33, 19451–19469. <https://doi.org/10.1523/JNEUROSCI.2880-13.2013>
- Norman-Haignere, S., Kanwisher, N., McDermott, J. H., & Conway, B. R. (2019). Divergence in the functional organization of human and macaque auditory cortex revealed by fMRI responses to harmonic tones. *Nature Neuroscience*, 22, 1057–1060. <https://doi.org/10.1038/s41593-019-0410-7>
- Nourski, K. V. (2017). Auditory processing in the human cortex: An intracranial electrophysiology perspective. *Laryngoscope Investigative Otolaryngology*, 2, 147–156. <https://doi.org/10.1002/lio2.73>
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011, 156869. <https://doi.org/10.1155/2011/156869>
- Pantev, C., Eulitz, C., Hampson, S., Ross, B., & Roberts, L. (1996). The auditory evoked “off” response: Sources and comparison with the “on” and the “sustained” responses. *Ear and Hearing*, 17, 255–265. <https://doi.org/10.1097/00003446-199606000-00008>
- Parras, G. G., Nieto-Diego, J., Carbajal, G. V., Valdés-Baizabal, C., Escera, C., & Malmierca, M. S. (2017). Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nature Communications*, 8, 2148. <https://doi.org/10.1038/s41467-017-02038-6>
- Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36, 767–776. [https://doi.org/10.1016/S0896-6273\(02\)01060-7](https://doi.org/10.1016/S0896-6273(02)01060-7)
- Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *Journal of Neuroscience*, 24, 6810–6815. <https://doi.org/10.1523/JNEUROSCI.0383-04.2004>
- Ponton, C. W., Eggermont, J. J., Kwong, B., & Don, M. (2000). Maturation of human central auditory system activity: Evidence from multi-channel evoked potentials. *Clinical Neurophysiology*, 111, 220–236. [https://doi.org/10.1016/S1388-2457\(99\)00236-9](https://doi.org/10.1016/S1388-2457(99)00236-9)
- Recanzone, G. H., Guard, D. C., & Phan, M. L. (2000). Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. *Journal of Neurophysiology*, 83, 2315–2331. <https://doi.org/10.1152/jn.2000.83.4.2315>
- Ritter, S., Dosch, H. G., Specht, H.-J., & Rupp, A. (2005). Neuromagnetic responses reflect the temporal pitch change of regular interval sounds. *Neuroimage*, 27, 533–543. <https://doi.org/10.1016/j.neuroimage.2005.05.003>
- Roberts, B., & Brunstrom, J. M. (2001). Perceptual fusion and fragmentation of complex tones made inharmonic by applying different degrees of frequency shift and spectral stretch. *Journal of the Acoustical Society of America*, 110, 2479–2490. <https://doi.org/10.1121/1.1410965>
- Roberts, B., Holmes, S. D., Darwin, C. J., & Brown, G. J. (2010). Perception of concurrent sentences with harmonic or frequency-shifted voiced excitation: Performance of human listeners and of computational models based on autocorrelation. In E. A. Lopez-Poveda, A. R. Palmer, & R. Meddis (Eds.), *The Neurophysiological Bases of Auditory Perception* (pp. 521–531). Springer. [https://doi.org/10.1007/978-1-4419-5686-6\\_48](https://doi.org/10.1007/978-1-4419-5686-6_48)
- Roberts, T. P., & Poeppel, D. (1996). Latency of auditory evoked M100 as a function of tone frequency. *Neuroreport*, 7, 1138–1140. <https://doi.org/10.1097/00001756-199604260-00007>
- Ross, B., & Tremblay, K. (2009). Stimulus experience modifies auditory neuromagnetic responses in young and older listeners. *Hearing Research*, 248, 48–59. <https://doi.org/10.1016/j.heares.2008.11.012>
- Saenz, M., & Langers, D. R. (2014). Tonotopic mapping of human auditory cortex. *Hearing Research*, 307, 42–52. <https://doi.org/10.1016/j.heares.2013.07.016>
- Seidel, A., Ghio, M., Studer, B., & Bellebaum, C. (2021). Illusion of control affects ERP amplitude reductions for auditory outcomes of self-generated actions. *Psychophysiology*, 58, e13792. <https://doi.org/10.1111/psyp.13792>
- Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *Journal of Neuroscience*, 23, 5545–5552. <https://doi.org/10.1523/JNEUROSCI.23-13-05545.2003>
- Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., & Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport*, 16, 1781–1785. <https://doi.org/10.1097/01.wnr.0000185017.29316.63>
- Sheehan, K. A., McArthur, G. M., & Bishop, D. V. (2005). Is discrimination training necessary to cause changes in the P2 auditory event-related brain potential to speech sounds? *Cognitive Brain Research*, 25, 547–553. <https://doi.org/10.1016/j.cogbrainres.2005.08.007>
- Sowman, P. F., Kuusik, A., & Johnson, B. W. (2012). Self-initiation and temporal cueing of monaural tones reduce the auditory N1 and P2. *Experimental Brain Research*, 222, 149–157. <https://doi.org/10.1007/s00221-012-3204-7>
- Steinmetzger, K., Megbel, E., Shen, Z., Andermann, M., & Rupp, A. (2022). Cortical activity evoked by voice pitch changes: A combined fNIRS and EEG study. *Hearing Research*, 420, 108483. <https://doi.org/10.1016/j.heares.2022.108483>
- Steinmetzger, K., Meinhardt, B., Praetorius, M., Andermann, M., & Rupp, A. (2022). A direct comparison of voice pitch processing in acoustic and electric hearing. *NeuroImage: Clinical*, 103188. <https://doi.org/10.1016/j.nicl.2022.103188>
- Steinmetzger, K., & Rosen, S. (2015). The role of periodicity in perceiving speech in quiet and in background noise. *Journal of the Acoustical Society of America*, 138, 3586–3599. <https://doi.org/10.1121/1.4936945>

- Steinmetzger, K., & Rosen, S. (2023). No evidence for a benefit from masker harmonicity in the perception of speech in noise. *Journal of the Acoustical Society of America*, 153, 1064–1072. <https://doi.org/10.1121/10.0017065>
- Steinmetzger, K., Shen, Z., Riedel, H., & Rupp, A. (2020). Auditory cortex activity measured using functional near-infrared spectroscopy (fNIRS) appears to be susceptible to masking by cortical blood stealing. *Hearing Research*, 396, 108069. <https://doi.org/10.1016/j.heares.2020.108069>
- Steinschneider, M., Liégeois-Chauvel, C., & Brugge, J. F. (2011). Auditory evoked potentials and their utility in the assessment of complex sound processing. In J. A. Winer & C. E. Schreiner (Eds.), *The auditory cortex* (pp. 535–559). Springer. [https://doi.org/10.1007/978-1-4419-0074-6\\_25](https://doi.org/10.1007/978-1-4419-0074-6_25)
- Steinschneider, M., Nourski, K. V., & Fishman, Y. I. (2013). Representation of speech in human auditory cortex: Is it special? *Hearing Research*, 305, 57–73. <https://doi.org/10.1016/j.heares.2013.05.013>
- Su, Y., & Delgutte, B. (2020). Robust rate-place coding of resolved components in harmonic and inharmonic complex tones in auditory midbrain. *Journal of Neuroscience*, 40, 2080–2093. <https://doi.org/10.1523/JNEUROSCI.2337-19.2020>
- Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, 2011, 8. <https://doi.org/10.1155/2011/879716>
- Tiitinen, H., Sivonen, P., Alku, P., Virtanen, J., & Näätänen, R. (1999). Electromagnetic recordings reveal latency differences in speech and tone processing in humans. *Cognitive Brain Research*, 8, 355–363. [https://doi.org/10.1016/S0926-6410\(99\)00028-2](https://doi.org/10.1016/S0926-6410(99)00028-2)
- Tremblay, K. L., Ross, B., Inoue, K., McClannahan, K., & Collet, G. (2014). Is the auditory evoked P2 response a biomarker of learning? *Frontiers in Systems Neuroscience*, 8, 28. <https://doi.org/10.3389/fnsys.2014.00028>
- Wang, X. (2018). Cortical coding of auditory features. *Annual Review of Neuroscience*, 41, 527–552. <https://doi.org/10.1146/annurev-neuro-072116-031302>
- Yost, W. A., Patterson, R., & Sheft, S. (1996). A time domain description for the pitch strength of iterated rippled noise. *Journal of the Acoustical Society of America*, 99, 1066–1078. <https://doi.org/10.1121/1.414593>
- Zachlod, D., Rüttgers, B., Bludau, S., Mohlberg, H., Langner, R., Zilles, K., & Amunts, K. (2020). Four new cytoarchitectonic areas surrounding the primary and early auditory cortex in human brains. *Cortex*, 128, 1–21. <https://doi.org/10.1016/j.cortex.2020.02.021>
- Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, 11, 946–953. <https://doi.org/10.1093/cercor/11.10.946>