CORTEX 184 (2025) 79–95

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research Report

Hemispheric asymmetries in the auditory cortex reflect discriminative responses to temporal details or summary statistics of stationary sounds

Martina Berto [a,**], Patrick Reisinger [b], Emiliano Ricciardi [a], Nathan Weisz [b,c] and Davide Bottari [a,*]

[a] Molecular Mind Lab, IMT School for Advanced Studies Lucca, Italy
[b] Department of Psychology and Centre for Cognitive Neuroscience, Paris-Lodron-University of Salzburg, Austria
[c] Neuroscience Institute, Christian Doppler University Hospital, Paracelsus Medical University, Salzburg, Austria

ARTICLE INFO

Article history:
Received 12 October 2023
Reviewed: 27 February, 2024
Revised 26 April 2024
Accepted 9 September 2024
Action editor Ryo Kitada
Published online 7 January 2025

Keywords:
Auditory statistics
Computational models
Sound textures
Hemispheric asymmetry
Auditory functional specialization

ABSTRACT

The processing of stationary sounds relies on both local features and compact representations. As local information is compressed into summary statistics, abstract representations emerge. Whether the brain is endowed with distinct neural architectures predisposed to such computations is unknown. In this magnetoencephalography (MEG) study, we employed a validated protocol to localize cortical correlates of local and summary auditory representations, exposing participants to sequences embedding triplets of synthetic sound textures systematically varying for either local details or summary statistics. Sounds varied for their duration and could be short (40 ms) or long (478 ms) to favor change detections based on local or summary statistics, respectively. Results clearly revealed distinct activation patterns for local features and summary auditory statistics. Neural activations diverged in magnitude, spatiotemporal distribution, and hemispheric lateralization. The right auditory cortex, comprising both primary and neighboring temporal and frontal regions were engaged to detect sound changes in both local features (for short sounds) and summary statistics (for long sounds). Conversely, the left auditory cortex was not selective to these auditory changes. However, the ventro-lateral portion of left frontal lobe, a region associated with sound recognition, was engaged in processing changes in summary statistics at a long sound duration. These findings highlight the involvement of distinct cortical pathways and hemispheric lateralization for the computation of local and summary acoustic information occurring at different temporal resolutions.

Significant statement: We revealed hemispheric specializations for auditory computations at high (local) and low (summary statistics) temporal resolutions. The right hemisphere was engaged for both computations, while the left hemisphere responded more to summary statistics changes. These findings highlight the multifaceted functions of the right

* Corresponding author.
** Corresponding author.
E-mail addresses: martina.berto@imtlucca.it (M. Berto), davide.bottari@imtlucca.it (D. Bottari).
https://doi.org/10.1016/j.cortex.2024.09.020

hemisphere in capturing acoustic properties of stationary sounds and the left hemisphere's involvement in processing abstract representations.

1. Introduction

Sounds are processed at low and high temporal resolutions, depending on the amount of information they encompass. When sounds are short, the processing of fine-grained temporal modulations (local features) is crucial for detecting transient changes in the entering sound waves. As the amount of information increases, the system relies on compact summary representations. For stationary sounds (e.g., sound textures, such as fire, rain, and typewriting) we can mathematically describe such summary representations as a set of time-averaged statistics (McDermott & Simoncelli, 2011).

In this context, the term "statistics" refers to measurements extracted from a sound wave using an auditory model. These measurements include averages and variations in intensity at different frequencies, and changes in intensity over time and represent the outcome of basic auditory computations, the processes by which the brain analyzes and interprets sound signals. Average summary statistics approximate outcomes of computations occurring in the periphery of the system and are encoded by both primary and post-primary auditory cortices (Giordano et al., 2023; McDermott & Simoncelli, 2011). The amount of acoustic information is crucial for extracting summary statistics from local features; thus, the engagement of one computational mode (local features processing) or the other (summary statistics) strictly depends on sound duration (McDermott et al., 2013).

The role of local features and summary statistics in auditory processing was evaluated in humans through computational methods (Berto et al., 2021, 2022; McDermott & Simoncelli, 2011; McWalter & McDermott, 2019; Norman-Haignere & McDermott, 2018; Zuk et al., 2020). Specifically, systematic synthesis approaches provide means to create sounds that are progressively more distinguishable at high, but not low, temporal resolution, or vice-versa (McDermott et al., 2013), objectively disentangling the auditory processing based on local features from summary statistics.

Specifically, a previous study employed this approach to isolate neural signatures of local features and summary statistics of sound textures (Berto et al., 2023). The electroencephalogram (EEG) of participants was recorded while they were exposed to streams of synthetic sound excerpts comprising systematic changes in local features or summary statistics. To manipulate the reliability of local features or summary statistics, the duration of the sound excerpts could be short, medium, or long. Results showed a clear dissociation: greater evoked responses were measured for a change in local features compared to summary statistics when sounds were short, and the opposite when sounds were long, while no difference was observed at medium sound length. The significant effects emerged over similar, but not identical, scalp locations, possibly indicating different cortical sources. However, we did not compute normalizations of the electrical activity measured at the scalp level. Thus, it is possible that the different magnitudes of electrical activity across conditions might be biasing the observed topographic differences. Moreover, different oscillatory profiles were associated with local or summary changes, with changes in local features being encoded by faster oscillations than changes in summary statistics (Berto et al., 2023). Noteworthy, these results emerged without explicit tasks. Thus, allowing to hypothesize that sound discrimination occurring at high or low temporal resolutions can be automatically driven by selective stimulus properties, i.e., local features or summary statistics. For instance, the discrimination of brief sound excerpts may rely on broadband impulse amplitude modulations, while the discrimination of long sounds may rely on spectral content averaged over time.

While it was possible to distinguish local features or summary statistics computations from neural activity, it remained unknown which neural structures were specifically associated with the processing of fine-grained information compared to summary one. Here we specifically investigated whether we could replicate the observed dissociation on an independent sample of individuals and, most importantly, whether these mechanisms differ solely in timing and intensity (as shown in Berto et al., 2023) or arise from distinct brain regions. We tackled this research question by employing magnetoencephalography (MEG), which is known to have better spatial resolution compared to the EEG, to provide reliable source estimates of the automatic processing associated with local and summary acoustic information. Using a previously validated protocol (Berto et al., 2023), we presented short and long synthetic sounds as triplets, with the third sound systematically differing for its local features or summary statistics, depending on the experimental context. Noticeably, we utilized sound triplets to maintain consistency with a presentation scheme validated at the behavioral (Berto et al., 2021; McDermott et al., 2013) and electrophysiological (Berto et al., 2023) level and to investigate the brain correlates of automatic discrimination of novel sounds in an auditory stream pattern. By disrupting the repetition of the first two sounds (with the third one), we aimed to pinpoint neural mechanisms involved in detecting a variation in the incoming stimulation. Directly contrasting responses to the third sounds changing for their local features or summary statistics, we aimed to provide insights into which neural structures were associated with these auditory analyses and their processing dynamics.

Sound excerpts used to generate sound triplets were drawn from a large pool of synthetic sound textures (e.g., fire, rain, typewriting, waterfall, etc.) which ensured generalizing results across several sound categories. Triplets formed highly variable sound streams which allowed to control for irrelevant features by variability rather than elimination or homogenization.

This approach allowed measuring brain correlates of automatic discriminative responses to sound changes occurring at high or low temporal resolution. Leveraging the high temporal precision and spatial resolution of MEG, we aimed to find neural correlates for local and summary acoustic computations.

2. Material and methods

2.1. Participants

24 healthy individuals voluntarily participated in the investigations. Specifically, 8 females, 2 left-handed, with a mean age of 35.42 years (std = 13.37, range = 20–63). Prior to data collection, we did not have any specific hemispheric lateralization hypothesis related to these auditory processes, so we did not add constraints related to the handedness of participants. Providing that we triplicated the total number of trials per condition (from 216 in the original EEG study, Berto et al., 2023, to 648 in the current MEG) and that the SNR of the MEG is expected to be higher than EEG, leading to a generally cleaner signal, we deemed a sample size of 22 as adequate for the purpose of the study. All participants reported no neurological impairments and normal hearing, which was confirmed by performing an audiogram before the experiment. Participants signed informed consent and were paid 10€ per hour for their contribution. The study was approved by the Ethics Committee of the Paris-Lodron-University of Salzburg, in accordance with the Declaration of Helsinki.

2.2. Data exclusion

Two participants were excluded from the analysis, leading to a final sample size of 22 subjects, 7 females 7, 2 left-handed, with a mean age of 35.09 years (std = 13.92, range = 20–63). One participant was excluded due to technical problems during the recording; the other participant's data was rejected because of co-registration failure (for more details on the latter, see the Source Reconstruction paragraph). Left-handed participants were included as their activation patterns did not diverge from the others, as revealed by their data visualization. Additionally, performing data analysis excluding these participants provided the same insights (see Supplementary Information).

3. Stimuli

Experimental sounds were a sub-selection of the ones used in the previous EEG study (Berto et al., 2022) and behaviorally validated by Berto et al. (2021) with the same protocol by McDermott et al. (2013). They represented short snippets (either 40 or 478 ms long) extracted from 5s synthetic sound textures. Each synthetic sound was produced by using the computational Auditory Texture Model and the Sound Texture Synthesis Toolbox (http://mcdermottlab.mit.edu/downloads.html) designed by McDermott and Simoncelli (2011). The synthesis procedure is described below. The Dataset and Results, Code and Experimental protocol (including the stimuli) are available online at this link: http://doi.org/10.60817/Q5A6-H216.

3.1. Auditory Texture Model

A set of time-averaged summary statistics was extracted from 54 original recordings of 7s natural sound textures (i.e., rain, applause, waterfall; see the complete list in Table S1, in Supplementary Information). Statistics were measured using the auditory texture model as in McDermott and Simoncelli (2011). For clarity, we provide a summary of the model in the following paragraph. For a detailed description, see the original paper (McDermott & Simoncelli, 2011).

The model employed a filter bank cascade to process an input sound waveform, $x(t)$, representing an original recording of a sound texture, and extracted summary statistics from their outputs. The procedure can be divided into two processing stages.

First, to replicate the frequency analysis occurring in the cochlea, the input sounds were filtered into subbands using a bank of bandpass filters with varying center frequencies and bandwidths. The model employed 4th-order gammatone filters consisting of 32 zero-phase bandpass filters with center frequencies equally spaced on an equivalent rectangular bandwidth (ERB; Glasberg & Moore, 1990) scale between 20 and 10,000 Hz. The filters had a bandwidth of 3 db. The outcome of the filtering stage were cochlear subbands, the analytic signal (or fine structure) at each center frequency. The envelope of each subband was then computed via Hilbert transform. To emulate the non-linearity of the basilar membrane compression, the envelopes were elevated by a power of .3 (Ruggero, 1992). Overall, this represented the first processing stage resulting in the cochleagrams of sounds (examples of cochleagrams are displayed in Fig. 1A,B). For computational efficiency, the subband envelopes were then downsampled to 400 Hz.

In the second processing stage, each cochlear envelope was convolved with a second band of filters to obtain amplitude modulation rate subbands. These modulation filters consisted of 20 half-octave spaced bandpass filters (from .5 to 200 Hz) with a constant quality factor (Q) of 2 (for 3 dB bandwidths). This reflects the selectivity of the human auditory system, likely a result of thalamic processing (Dau et al., 1997).

Auditory texture statistics were extracted from the cochlear envelope subbands $x_k(t)$ and the modulation subbands, $b_{k,n}(t)$, where k and n indexed the cochlear and modulation channels, respectively.

The computed envelope statistics included three marginal moments: (i) the mean, (ii) the coefficient of variance, (iii) and the skewness. Respectively:

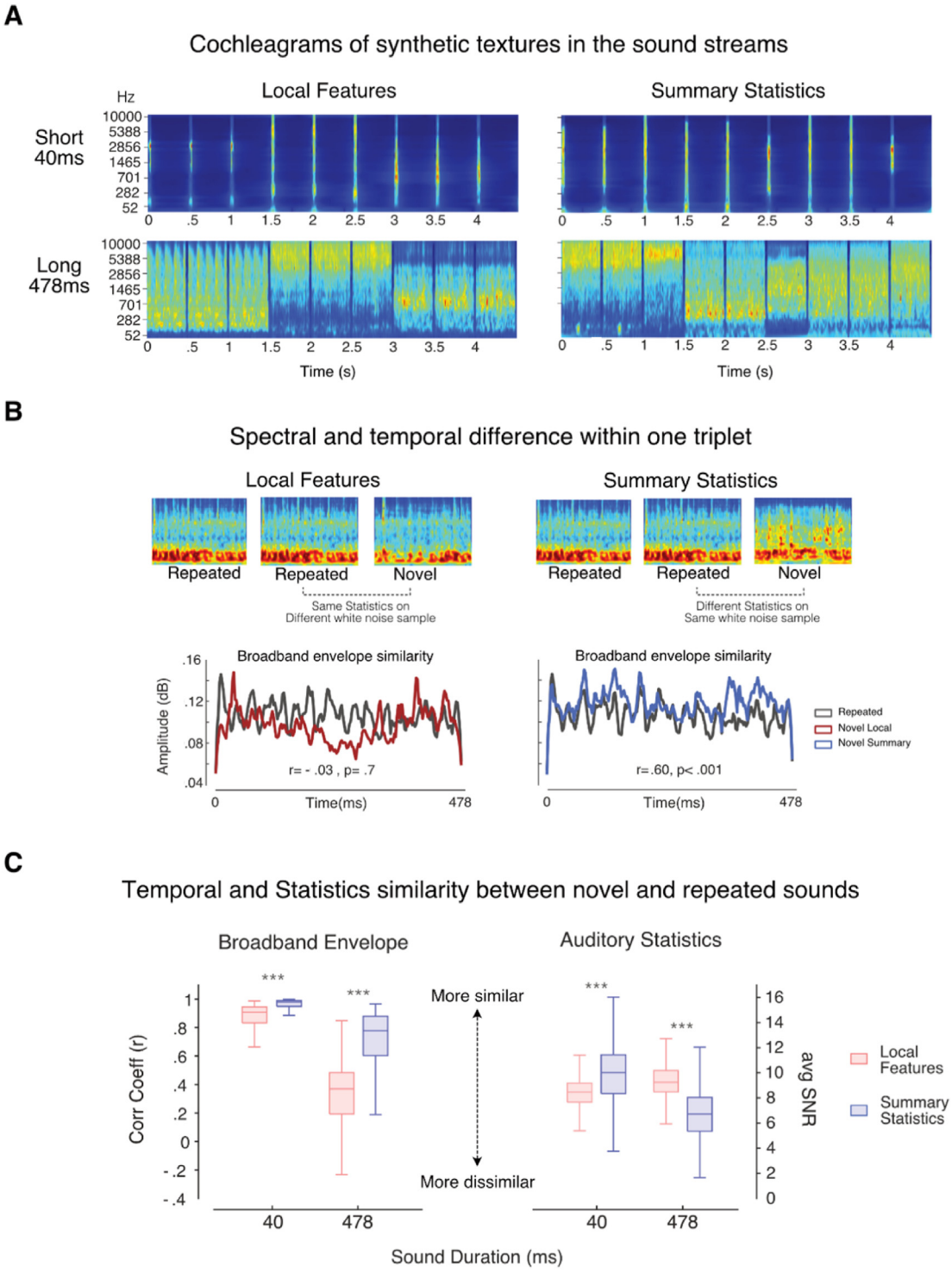

Fig. 1 – Stimuli and Experimental Protocol. (A) Cochleagrams of the first three sound pairs in a continuous stream presented to one participant. Cochleagrams were computed by filtering the sound excerpts through the Auditory Texture Model (McDermott & Simoncelli, 2011). (B) Spectral and temporal difference within a triplet. The top panel shows the cochleagrams of the sound excerpts contained in two example triplets (one for each experiment). In this example, in both experiments, the first two sounds (repeated) were a 478 ms excerpt from a white noise to which we imposed the auditory statistics of the sound texture "Motorcycle idling". In the Local features experiment, the third sound, (novel) was another exemplar of the "Motorcycle Idling" sound texture; thus, it is a different white noise sample to which we impose the same statistics. In the Summary Statistics experiment, the novel sound contains the statistics from the "Idling Boat" sound texture, but it is the same white noise sample. The statistics embedded in the sounds influence the spectral density of the sound excerpts, as depicted in the cochleagram. The bottom panels display the broadband envelopes of the repeated and novel sounds in both experiments for the same three sounds displayed above (Motorcycle Idling and Idling Boat). Despite the spectral distribution

$$\mu_k = \sum_t w(t)x(t) \quad \text{(i)}$$

$$\frac{\sigma^2}{\mu_k^2} = \frac{\sum_t w(t)(x_k(t) - \mu_k)^2}{\mu_k^2} \quad \text{(ii)}$$

$$\eta_k = \frac{\sum_t w(t)(x_k(t) - \mu_k)^3}{\sigma_k^3} \quad \text{(iii)}$$

where μ_k is the mean, σ_k^2 is the variance, η_k is the skewness at each cochlear channel, and w(t) is a windowing function, with the constraint that $\sum_t w(t) = 1$ (McDermott & Simoncelli, 2011). The variance was normalized by the squared mean to make it dimensionless, as the skewness (McDermott & Simoncelli, 2011). These marginal moments captured the sparsity of the time-averaged subband envelopes. Note that, following the original authors recommendations, compared to the previous version of the model (McDermott & Simoncelli, 2011), our implementation omitted the Kurtosis from the marginal moments, because proven to be not very informative.

The model included pairwise correlations between each cochlear channel and the eight nearest neighbors. Cross–band correlations capture broadband events that would activate more cochlear bands at the same time (Nelken et al., 1999).

The cross–band correlation coefficient can be computed as follow:

$$c_{jk} = \frac{\sum_t w(t)\left(x_j(t) - \mu_k\right)}{\sigma_j \sigma_k}, j, k \in [1 \ldots 32]$$

such that $(k - j) \in [1, 2, 3, 5, 8, 11, 16, 21]$.

Note that to capture the envelope power at different modulation rates, the modulation subband variance was normalized by the corresponding total cochlear envelope variance.

Another fundamental measure was the power at each modulation rate, namely the modulation power. The modulation power represents the major statistics of interest concerning the modulation bands and may reflect modulation-tuned properties in the thalamic neurons response (Dau et al., 1997; Miller et al., 2002). To calculate it, the model measured the variance of each modulation subband and normalized it by the total envelope variance as follow:

$$\sigma_{k,n} = \frac{\sum_t w(t)\left(b_{k,n}(t) - \mu_{k,n}\right)^2}{\sigma_k^2}, k \in [1 \ldots 32], n \in [1 \ldots 20]$$

Finally, the model employed octave-spaced modulation filters (McDermott & Simoncelli, 2011) to measure correlations (C1 and C2) between modulation subbands of different cochlear channels. The frequency responses of the filters (seven filters, with center frequencies in octave steps from 1.56 to 100 Hz) were half-cosines on a log-scale and more broadly tuned with $= \sqrt{2}$.

C1 was computed as follow:

$$C1_{jk,n} = \frac{\sum_t w(t)\tilde{b}_{j,n}(t)\tilde{b}_{k,n}(t)}{\sigma_{j,n}\sigma_{k,n}}, j \in [1 \ldots 32], (k - j) \in [1, 2], n \in [2 \ldots 7],$$

and

$$\sigma_{j,n} = \sqrt{\sum_t w(t)\tilde{b}_{j,n}(t)^2}$$

where $\tilde{b}_{k,n}(t)$ are the resulting bands of the correlation filters.

C2 is calculated as follow

$$C2_{k,mn} = \frac{\sum_t w(t)d^*_{k,m}(t)\alpha_{k,n}(t)}{\sigma_{k,m}\sigma_{k,n}}, k \in [1 \ldots 32], m \in [1 \ldots 6], (n - m) = 1$$

where * denote the complex conjugate, α is the analytic signal of the modulation bands comprising the response of the filter and its quadratic twin, and d is the frequency doubling resulting from squaring the analytic signal (complex-valued).

α and d are calculated as follow:

$$\alpha_{k,n}(t) = \tilde{b}_{k,n}(t) + iH(\tilde{b}_{k,n}(t))$$

and

$$d_{k,n} = \frac{\alpha^2_{k,n}(t)}{||\alpha_{k,n}(t)||}$$

where $i = \sqrt{-1}$ and H is the Hilbert transform.

For each sound texture, a set of time-averaged texture statistics were measured and included in a parameter vector, ζ,

between repeated and novel sounds being more similar in the Local Features experiment, the broadband envelope of novel and repeated sounds correlated more in Summary Statistics ($r = .6$) than in the Local Features experiment ($r = -.03$). This reveals that the temporal structure of the sounds was influenced by the original white noise sample used during synthesis, rather than the similarity of the embedded statistics, which in turn influenced the spectral content of sounds. (C) Temporal and Statistics dissimilarity between novel and repeated sound pairs. Two different metrics (correlation coefficients and signal-to-noise ratio, SNR) are used to measure the similarity between specific stimulus features (broadband envelope and time-averaged auditory statistics, respectively) of repeated and novel sounds. In the left panel, boxplots represent the correlation coefficients (r) measured between the broadband envelopes of repeated and novel sounds in each experiment (Local Features, Summary Statistics) and duration (40, 478 ms). In the right panel, we show the average SNRs measured between auditory statistics of repeated and novel sounds. Higher SNRs would indicate higher statistical similarity between sound pairs. Within sound duration, the broadband envelopes between repeated and novel sounds were always statistically more similar in the Summary Statistics experiment than in the Local Features one. By contrast, auditory statistics similarity showed a double dissociation according to sound duration. Boxplots outline the median and quartiles of the distributions. ***P < 00.001.

which was subsequently utilized to generate the synthetic sounds.

3.2. Sound texture synthesis

We synthesized sound textures using the Sound Texture Synthesis Toolbox and following the procedure described in McDermott and Simoncelli (2011). The parameter vector (ζ) of time-averaged statistics was measured for each of the 7s original sound textures ($n = 54$; listed in Table S1, Supplementary Information).

The measured statistics were then imposed on a white noise sample (seed). The algorithm measured the statistics from the seed and iteratively adjusted them to match the ones extracted from the original sound textures. To perform this computation, we used a conjugate gradient descent based on the "minimize" function by Carl Rasmussen (embedded in the toolbox, see above; McDermott & Simoncelli, 2011) which compares the total squared error of statistics measured from the synthetic sound with those from the original signal. Specifically, the procedure monitored the convergence between the statistics of the two signals by calculating the signal-to-noise ratio (SNR). The SNR was determined by comparing the squared error of a statistic class (summed across all statistics within that class) to the sum of the squared statistic values in that class. The iteration procedure was stopped when all statistic classes had an SNR of 30 dB or greater or after reaching 60 iterations. Convergence was achieved when the average SNR of all statistic classes reached 20 dB or higher (McDermott & Simoncelli, 2011). By seeding the synthesis algorithm with different Gaussian noises, we could generate distinct exemplars of the same sound texture, including auditory statistics that converged with sound duration but retained a different fine structure (local features). For each sound texture, the synthesis algorithm was initialized with four different white noise samples resulting in four different synthetic exemplars of the same original texture.

All four synthetic exemplars were cut into excerpts of 40 or 478 ms to which a 20 ms half-Hann window was applied (10 ms at the beginning and 10 ms at the end) to reduce edge artifacts. Excerpts were numbered according to their position along the 5-s signal (i.e., excerpt 1 occurs from 1 to 40 ms, excerpt 2 from 41 to 80 ms, and so on). We will refer to this number as their positional index (see below). All excerpts were equalized to the same root mean squared amplitude (rms) of .1 and had a sampling rate of 20 kHz.

3.3. Data acquisition

The brain activity was recorded using a 306-channel whole head MEG system comprising 204 planar gradiometers and 102 magnetometers (Neuromag TRIUX, Elekta). Signal was recorded with a sampling rate of 1000 Hz and filtered online at 0.1 Hz. Prior to recording the MEG, digital head-shape points (HSP) were measured on the scalp using a Polhemus Fastrak system (Polhemus, Colchester, VT). Approximately 500 points for subject were acquired, including three fiducials (nasion and left and right preauricular points). Three electrodes were applied to measure ocular and cardiac artifacts: two EOGs (vertical and horizontal channels attached above and lateral to the right eye) and one ECG.

Prior to commencing the experiment, we conducted standard clinical audiometry spanning from 125 to 8 kHz utilizing an AS608 Basic audiometer (Interacoustics, Middelfart, Denmark) to evaluate participants' auditory capabilities.

Participants were then accompanied in the magnetically shielded room (AK3B, Vacuumschmelze, Hanau, Germany) where the MEG system was located and sat comfortably in front of a screen located at a ~110 cm distance. The experimental instructions and a fixation cross were back projected on the translucent screen. Stimuli and triggers were produced and delivered through the VPixx system (comprising DATAPixx2 display driver, PROPixx DLP LED projector, and RESPONSEPixx response box by VPixx Technologies Inc., Saint-Bruno, Canada). Sounds were delivered by fMRI-compatible air-tube earphones to avoid magnetic interference in the MEG room. The experiment was programmed in MATLAB using the Psychtoolbox-3 (Brainard, 1997). Additionally, a class-based abstraction layer (https://gitlab.com/thht/o_ptb; Hartmann & Weisz, 2020) programmed on top of Psychtoolbox was used to facilitate the management of audio and triggers delivery in DataPixx.

3.4. Experimental protocol

The experiment consisted of 4 blocks of 5.4min each. Participants listened to streams of sounds but were instructed to ignore them and simply press a button when a target one occurred. The target sound was very infrequent (maximum three sounds per block) and consisted of a 50 ms pure tone with a frequency of 2200 Hz, an amplitude of 50 dB SPL, and a sampling rate of 20 kHz.

Within each block, 648 sounds were presented at a continuous stimulation rate (one sound every 500 ms). The stream in each block contained sounds of one specific length, either short (40 ms) or long (478 ms), so that the duration of the stream was kept constant, but the amount of acoustic information changed according to the size of the single sounds (Fig. 1A). The sound streams consisted of instances of synthetic excerpts, representing snippets cut from exemplars of the same sound texture or a different one. Specifically, stimuli were presented in triplets: two sounds were identical (repeated; $n = 432$), while the third one was different (novel; $n = 216$). All three sound excerpts within a triplet had the same positional index in the synthetic exemplar (i.e., all of them were starting and ending at the same time point along the corresponding 5s synthetic sounds they were extracted from).

Compared to the repeated sounds, in one experimental context, the novel sound systematically varied for its local features, while in the other for its summary statistics (Fig. 1B). That is, in the Local Features experiment, the third sound was drawn from a synthetic exemplar of the same sound texture but derived from a different white noise sample. The original white noise sample used to initialize the synthesis is expected to affect the temporal structure and the statistical values of the sounds measured at the high temporal resolution, that is in its local properties. In the other experiment (Summary Statistics), the third sound originated from the same white noise sample used to generate the two identical repeated

sounds but was constrained by a different set of statistics. The summary statistics of the imposed Sound Texture affect the fine-grained spectral density measured at low temporal resolution (see below; Fig. 1C). Noticeably, the texture pairs in this experimental condition were matched according to their similarity based on previous evaluations (McDermott et al., 2013). Overall, this presentation scheme permits controlling for expectancy effects, as the novel sound always occurs as a third element in the triplet. Moreover, creating regularities and then disrupting them (by presenting two identical sounds followed by a different one) permits the measurement of discriminative responses driven by local and summary computations without an explicit task. To generalize local and summary computations to a vast pool of sounds, the excerpts used in the experiments were drawn randomly among the available ones; thus, the presentation order and stimuli were never the same between participants and blocks (Berto et al., 2022). This allowed us to specifically investigate the computation of interest while generalizing it to many different stationary sounds, disregarding their idiosyncratic properties (i.e., frequency, pitch, timbre, and amplitude of single sound excerpts).

Throughout the experiment, participants kept their eyes open and were instructed to look at a fixation cross appearing at the center of the screen.

No part of the study analyses and procedures was preregistered prior to the research being conducted.

3.5. *Temporal structure and time-averaged statistics similarities between excerpt pairs*

To test whether the similarity in the temporal structure of repeated and novel sound excerpts was influenced by the original white noise sample used to initialize the synthesis (different in Local Features; same in Summary Statistics), we used the Auditory Texture Model (McDermott & Simoncelli, 2011; see above) to obtain the cochleagrams of each excerpt pair presented in the experiments across all participants, for each experiment (Local or Summary) and duration (40 or 478 ms). Note that the same sound excerpt can appear more than once in this distribution, as we selected the precise sound sequence presented in the experiment to participants which included sounds from the same pool of available ones. Furthermore, each participant received idiosyncratic stimulation; that is, stimuli were randomly selected and paired among the available ones sharing the desired properties and sequences were never the same across participants. We then averaged over frequency channels ($n = 32$) at each time-point to obtain the broadband envelopes. Pearson's *rs* transformed to Fisher-z-scores were used as a metric to assess the similarity between the broadband envelopes in the sound pairs presented in the Local Features and the Summary Statistics experiment. This procedure was done for each excerpt pair presented in each experiment and duration ($n = 19{,}008$). The normalized coefficients in Local Features were statistically compared to the ones in Summary Statistics (two-tailed *t*-tests) within each sound duration. Results showed that the envelopes were always significantly more similar when the excerpts originated from the same white noise sample compared to when they did not. That is, the correlation coefficients were higher in Summary Statistics experiment compared to Local Features one, both at short ($P < .001$; Local: mean = 1.45, std = .47; Summary: mean = 2.20, std = .68) and long ($P < .001$; Local: mean = .36, std = .25; Summary: mean = 1.02, std = .46) durations. Correlation coefficients (*r*) are displayed in Fig. 1C, left panel.

On the other hand, we expected statistical values between repeated and novel sounds to be more similar depending on sound duration and experimental context. To address the statistical similarity between the couples of excerpts (repeated and novel) in the study, we selected all the sound pairs (repeated and novel) presented in the study, as done before.

For each sound excerpt, we extracted the set of summary statistics using the functions "*generate_subbands_and_envs.m*" and "*measure_texture_stats_from_envs.m*" in the Sound Texture Synthesis Toolbox (http://mcdermottlab.mit.edu/downloads.html; McDermott & Simoncelli, 2011).

We then calculated the similarity between statistics from repeated and novel sounds by estimating the Signal-to-noise ratio (SNR) as in Berto et al. (2022). Specifically, for each excerpt pairs (repeated and novel) and separately for each statistic class (envelope mean, variance, skew, cross-correlation, modulation power, C1, and C2), we calculated the total squared error at each cochlear channel by subtracting the value of statistics measured from novel sound from the one measured from the repeated one and elevating by a power of 2. When statistics had more than one dimension (i.e., modulation bands), the values in the other dimensions were summed prior to computing the error (Berto et al., 2023; McDermott & Simoncelli, 2011). We then defined the SNR as follows:

$$SNR = 10\ log_{10}\left(\frac{\sum_k StatRep(k)^2}{\sum_k \varepsilon(k)}\right), k \in [1, 2, 3\ldots, 32]$$

where k is a cochlear channel ($n = 32$), *StatRep* represents the value of a statistic class calculated from a repeated sound, and ε is the squared error between the same repeated sound and the corresponding novel one for that statistic class.

SNRs were computed separately for each statistic class in each excerpt pair and then averaged across statistics to obtain one average SNR for each sound pair (Berto et al., 2023). Higher SNRs would indicate that, overall, summary statistics in the repeated sounds were more similar to the ones in the novel sounds. Average SNRs for each experiment (Local Features, Summary Statistics) and duration (Short 40 ms, Long 478 ms) are displayed in Fig. 1C, right. In order to test whether there was a difference in SNRs between experiments and within duration, we performed non-parametric Wilcoxon rank sum tests comparing Local Features and Summary Statistics at either short (40 ms) or long (478 ms) sound duration.

Results showed that when sounds were short, the SNRs were significantly lower in the Local Features experiment ($P < .001$; Local: mean = 8.32, std = 1.25) compared to the Summary Statistics one (mean = 9.91, std = 2.42), meaning that statistical values were significantly more dissimilar in the former. By contrast, when sounds were long, SNRs were lower in the Summary Statistics experiment ($P < .001$; mean = 6.74,

std = 2.09) compared to Local Features one (mean = 9.31, std = 1.22).

Overall, these analyses show that while broadband envelope similarity between repeated and novel sounds is influenced by their original white noise seed, the similarity of auditory statistics changes based on sound duration.

Based on these results, we would expect greater activation in Local Features experiment compared to Summary Statistics one when sounds are short, and the opposite when sounds are long.

4. Data analysis

4.1. *Preprocessing of MEG data*

Right after the recording, a signal space separation method from the Maxfilter software (MEGIN Oy, Espoo, Finland) was used to correct for different head positions across blocks within each participant and to clean channels from external interference (Taulu et al., 2005).

Preprocessing of the max-filtered MEG data was performed using MNE-Python version 1.1.1 (Gramfort et al., 2013) running on Python 3.9.16. For each participant, the signal recorded for each block was concatenated; segments contaminated, by either the beep sound or the button press, were marked for rejection. Independent component analysis (using the fast ICA method; Hyvarinen, 1999) was run to detect stereotypical artifacts (eye movements, blinks, and heartbeat). We filtered the signal below 1 Hz and downsampled it to 250 Hz to reduce computational time and space on storage. To account for the presence of different type of channels (e.g., EOG, ECG, MEG sensors), and therefore different units (volts or tesla), the data from each channel type were pre-whitened, that is, they were scaled by the standard deviation across all channels (z-standardized). ICA was fit to the data using all 306 channels, without any prior dimensionality reduction. A specific number of components was then automatically selected for each subject based on the given variance level and accounting for rank deficiency of the data. To select which components to remove, we used a semi-automatic template-matching procedure (Campos Viola et al., 2009). Specifically, we manually selected a component (template) from one subject which best represented each artifact (one template for the eyes and one for the heart; in this case, we selected the 2 templates from 2 different subjects). Templates were chosen after visual inspection of the topography of the inverse weights and double-checked by looking at the IC activation scroll. To have additional proof that the selected templates represented eye and cardiac artifacts, we also performed Pearson correlation between all components for the selected subjects and their EOG and ECG components. Then we used the corrmap function (implemented in MNE-Python) to detect all the other components in our dataset (across all subjects) which correlated with the template above a certain threshold (.85 for eye; .60 for heart). Thresholds were selected so that corrmap could find at least one IC for each subject that matches the template. Implementing the template-matching method allowed to improve and speed up bad component detection even for those subjects where EOG and ECG channels were compromised. After the components to exclude were selected, they were removed from the original unfiltered, full-resolution (1000 Hz) dataset. On average, 2.68 components per subject were removed (std = .82; range = 2–5). The signal was then lowpass filtered at 40 Hz (one-pass, zero-phase, non-causal FIR filter; windowed time-domain design, window type = hamming; passband ripple = .0194; stopband attenuation = 53 dB; upper passband edge = 40 Hz; upper transition bandwidth = 10 Hz; -6 dB cutoff frequency = 45 Hz; filter length = 331 samples) and highpass filtered at 0.1 Hz (one-pass, zero-phase, non-causal FIR filter; windowed time-domain design, window type = hamming; passband ripple = .0194; stopband attenuation = 53 dB; lower passband edge = .10 Hz; lower transition bandwidth = .10 Hz; -6 dB cutoff frequency = .05 Hz; filter length = 33,001 samples).

4.2. *Epoching*

Data was epoched into segments from −100 to 500 ms from the onset of the novel stimulus and downsampled to 250 Hz. To correct for the physical delay of 16.5 ms in sound delivery due to transmission in the air-conducting earphones, the signal was shifted in time by .0165 s. For each participant, within all the 216 epochs, the ones compromised by the occurrence of the infrequent target (beep sound) and/or corresponding button press (minimum 1, maximum 6 per block) were excluded from the analysis; for each sound duration, the remaining number of epochs was equalized between different experimental conditions (the difference depended on the number of beeps presented in each specific condition block).

All valid epochs (range 208–215) were then averaged to compute event-related fields (ERFs) and baseline correction was applied by subtracting the averaged pre-stimulus period activation from −100 to 0 ms from each time point post-stimulus.

4.3. *Source localization*

A semi-automatic co-registration pipeline was used to compute the head models. This approach has been shown to lead to comparable results as compared to manual co-registration ones (Houck & Claus, 2020). Participants' head shapes were co-registered to a "standard" model created by the combination of 40 MRI scans of real brains (fsaverage; Fischl, 2012). For each subject, co-registration was performed in 4 steps. In step 1, we loaded the max-filtered head-shape points (HSP) and defined the fsaverage template from Freesurfer (Fischl, 2012). In step 2, we estimated three fiducial points from the fsaverage template and aligned them to the fiducials in the digitized HSP. In step 3, the outcome was refined using the Iterative Closest Point (ICP; Besl & McKay, 1992) algorithm with a small number of iterations (n = 6). Any point in the HSP with a distance from the scalp above 5 mm was considered an outlier and omitted. Finally, in step 4, ICP was performed once again, with a higher number of iterations (n = 20) for the final co-registration fit. Head models and sensors were displayed at the end of any stage for visual inspection, to evaluate the quality of the fit and make sure that the head model was comprised within the sensors and did not fall outside. For three participants, co-registration was

successful after step 3 and worsened at step 4, which was then omitted. For one participant, co-registration was effective at step 2, so steps 3 and 4 were both ignored. Finally, for one participant, co-registration failed at all steps, thus the subject was excluded from further data analysis.

The co-registered head models were used to calculate the forward solution.

First, we computed the single-layer boundary element model (BEM (Akalin-Acar & Gençer, 2004); to create a BEM solution for the fsaverage template brain. Second, to define the position and orientation of the sources, we create a decimated dipole grid on the white matter surface using an icosahedron subdivision (ico-4) which included 2562 sources for hemispheres. Assuming a cortical surface of 1000 cm^2, the approximate spacing between the grid points was 6.2 mm, and each voxel occupied 39 mm^2 of the cortical surface.

To control for field patterns related to noisy sources (e.g., human-based or environmental), we weighted each channel by a full noise covariance matrix computed on the nearest empty room recording. That is, prior to any daily data acquisition, 2 min of empty room measurements (data collected with no subject) were recorded in the MEG; depending on the date of testing, the closest measurement was found for each subject and preprocessed in the exact same way (max filtered; high-pass filter: 0.1 Hz; low-pass filter: 40 Hz). The Noise covariance matrix and the true rank were computed from the empty room measurement and used to calculate the inverse solution based on the L2 Minimum Norm Estimates.

Finally, Source Time Courses were estimated by applying the inverse operator (lamba = 2; SNR = 3; method = minimum norm estimates (MNE); Hämäläinen & Ilmoniemi, 1994) on the evoked data and pre-stimulus periods (from −100 to 0 ms) were cropped to reduce computational space and time (as statistics were performed at latencies above 0; see Statistical Analyses below).

4.4. *Statistical Analyses*

Since repeated sounds were always identical in both Local features and Summary Statistics, assessing whether the responses to the novel sounds differed between experiments would reveal if and which brain correlates discriminate based on local features and summary statistics changes. Thus, we measured statistical differences between experiments (Local vs. Summary) and within duration (Short 40 ms and Long 478 ms) for the evoked response to the novel sounds in source space.

Statistics were assessed using eelbrain, a Python toolkit (Brodbeck et al., 2022).

Source estimates were imported in eelbrain from MNE-Python for each subject, experiment, and duration. A cluster-based permutation test for paired samples was used to test for differences in source activation at any time-point (Maris & Oostenveld, 2007). The procedure works as follows. We ran a two-tailed *t*-test, with a critical alpha of .05 at all latencies (from 0 to 500 ms) and all sources. To form spatial clusters, an adjacency matrix is computed based on the fsaverage template brain and the ico-4 source space. To address the multiple comparison problem, we used a threshold-free cluster enhancement procedure (TFCE; Smith & Nichols, 2009). For every permutation of the data (n = 10,000), we computed a parameter map and processed it with the cluster enhancement algorithm in steps of .1. Subsequently, for each permutation, the maximum value of the test statistic was measured across the whole map to obtain the distribution of t-values under the null hypothesis. Based on its position in the distribution, each data point in the parameter map is assigned a *p*-value. To avoid considering spurious results, we assumed a stringent spatiotemporal criterion, and accepted only clusters of activation perduring at least 20 ms and including a minimum of 20 sources.

5. Results

In this study, we addressed whether, depending on sound duration, a change in specific sound properties (local details or summary statistics) engaged different brain correlates. That is, we tested if the auditory cortex is organized in sub-regions for selective computations based on the temporal resolution at which the sound change has occurred. To this aim, we compared the response to novel sounds between Local Features and Summary Statistics experiments within short (40 ms) and long (478 ms) sound duration. Any dissociation according to sound duration would indicate that the brain is endowed with distinct neural substrates for the processing of local details and summary statistics. This was clearly supported by our findings.

Prior to performing analysis in source space, as a control analysis, we run a spatiotemporal cluster-based permutation (Maris & Oostenveld, 2007) comparing experiments (Local vs. Summary) within each sound duration (40 or 478 ms; see Supplementary Information). In Fig. 2, we report the time course of the evoked response -Global Field Power (GFP)- to novel sounds of planar gradiometers-at sensors compatible with auditory cortex activations in the right and left hemispheres. These plots highlight the overall higher contribution of right auditory regions in response to novel sounds. This difference is especially pronounced for Local Features at short sound duration (40 ms) and Summary Statistics at long ones (478 ms; see Fig. 2).

Multiple regions of the right hemisphere were selectively engaged for Local Feature changes occurring with short duration. Indeed, when comparing the brain response to novel sounds between Local Features and Summary Statistics experiments for short sound duration (40 ms), we found a large cluster of higher activation for the Local Features experiment selectively in the right hemisphere ($P < .001$; max t-value = 8.06; Cohen's d = 1.22; Fig. 3A). The cluster started 148 ms after stimulus onset, lasted until 224 ms (duration: 76 ms), and spatially comprised multiple (177) voxels of cortical sources. The cluster involved voxels embedded in several cortical regions in the auditory cortex, specifically: the superior temporal sulcus (STS) and gyrus (STG), the anterior transverse temporal gyrus (Heschl gyrus; HG), the planum temporale (PT), and the posterior segment of the lateral fissure. Other regions outside of the auditory cortex included: the central operculum, the angular and supramarginal gyri (AG and SMG), the postcentral and central sulci, the superior segment of the circular sulcus of the insula, the Jensen sulcus,

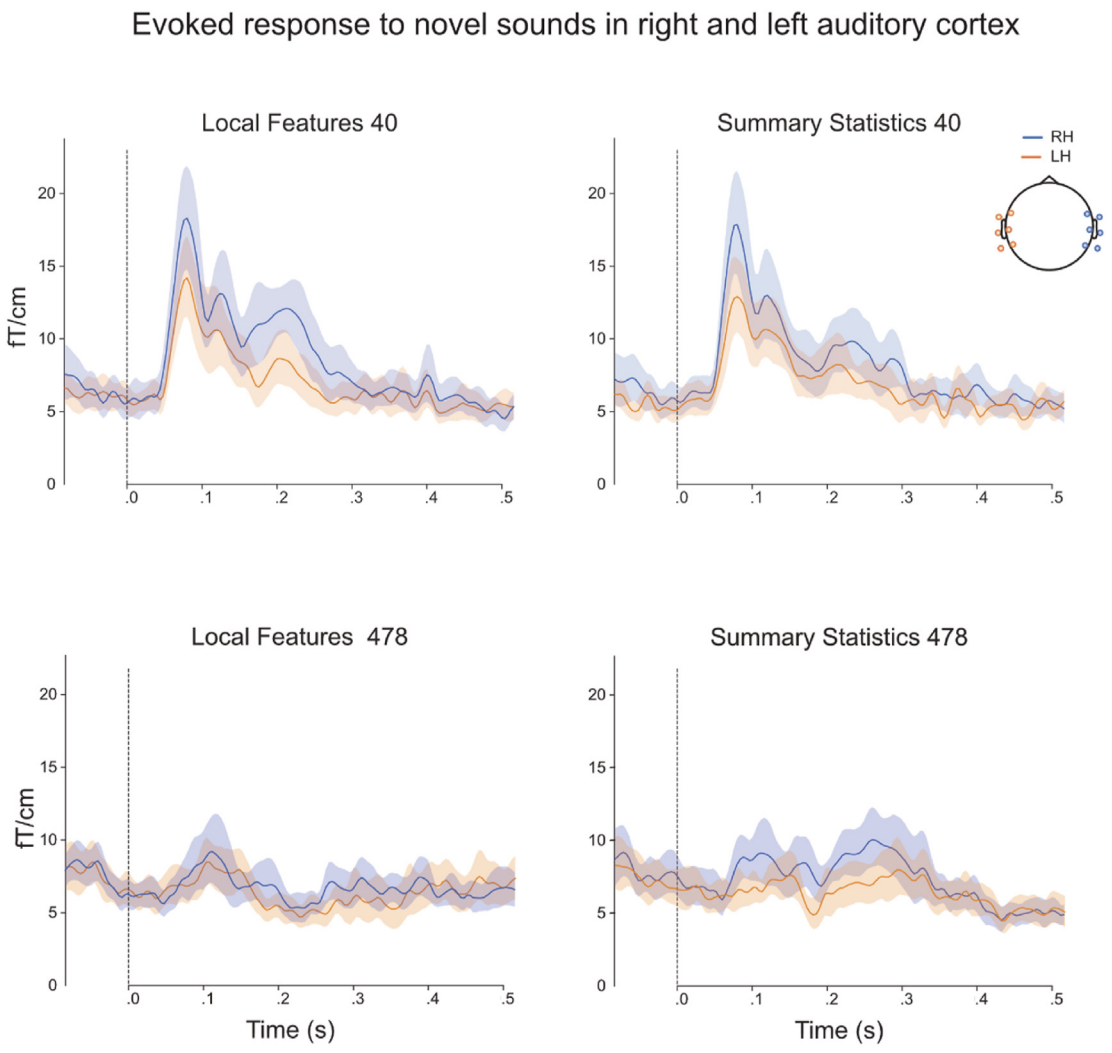

Fig. 2 – Evoked response at the sensor level. Global Field Power (GFP) of the evoked response to novel sounds of planar gradiometers in the left and right auditory cortex are plotted separately for each experimental condition (Local Features and Summary Statistics at 40 and 478 ms). The plots show that the right auditory cortex is more engaged in auditory discrimination compared to the left one, especially for Local Features at short duration and Summary Statistics at long one. Shaded areas represent a confident band of 95% (the threshold for a bootstrap estimation of the confidence interval).

the interparietal sulcus, and transverse parietal sulci (Table 1A).

Conversely, changes associated with Summary Statistics occurring for long sounds (478 ms) engaged both left and right hemispheres. For sounds of this duration, two large regions were significantly more responsive to novel sounds in the Summary Statistics experiment as compared to the Local Features one: (i) one in the right hemisphere and (ii) the other in the left one (Fig. 3B and C).

(i) The cluster found in the right hemisphere ($P < .001$; t-max = 6.62; Cohen's $d = 1.11$) perdured from 180 to 312 ms after stimulus onset (duration = 132 ms) and comprised 589 voxels of cortical sources. Some of the sources overlapped with the cluster observed in response to local features at the short duration, specifically around the primary auditory cortex, Heschl gyrus (HG), some temporal auditory regions such as STG, PT, STS, the lateral fissure, and other areas such as the Jensen sulcus, AG, SMG, the circular sulcus of the insula, the interparietal and transverse parietal sulci, and the central and postcentral sulci. The other non-overlapping regions comprised frontal areas, including portions of the superior, middle, and inferior frontal sulci (S-M-IFS) and gyri (S-M-IFG), and the inferior part of the precentral sulcus and gyrus; some temporal regions such as the transverse temporal sulcus, the middle temporal gyrus (MTG), the lateral fissure, and several parts of the insular cortex and central operculum; finally, some posterior regions in the occipital lobe (Table 1B).

(ii) The other large cluster was found in the left hemisphere (p-value = .004; t-max = 5.77; Cohen's $d = 1.13$; Fig. 3C). The left localized cluster started 168 ms after stimulus onset and lasted until 332 ms (duration = 164 ms). It included 289 voxels of cortical space, located mainly

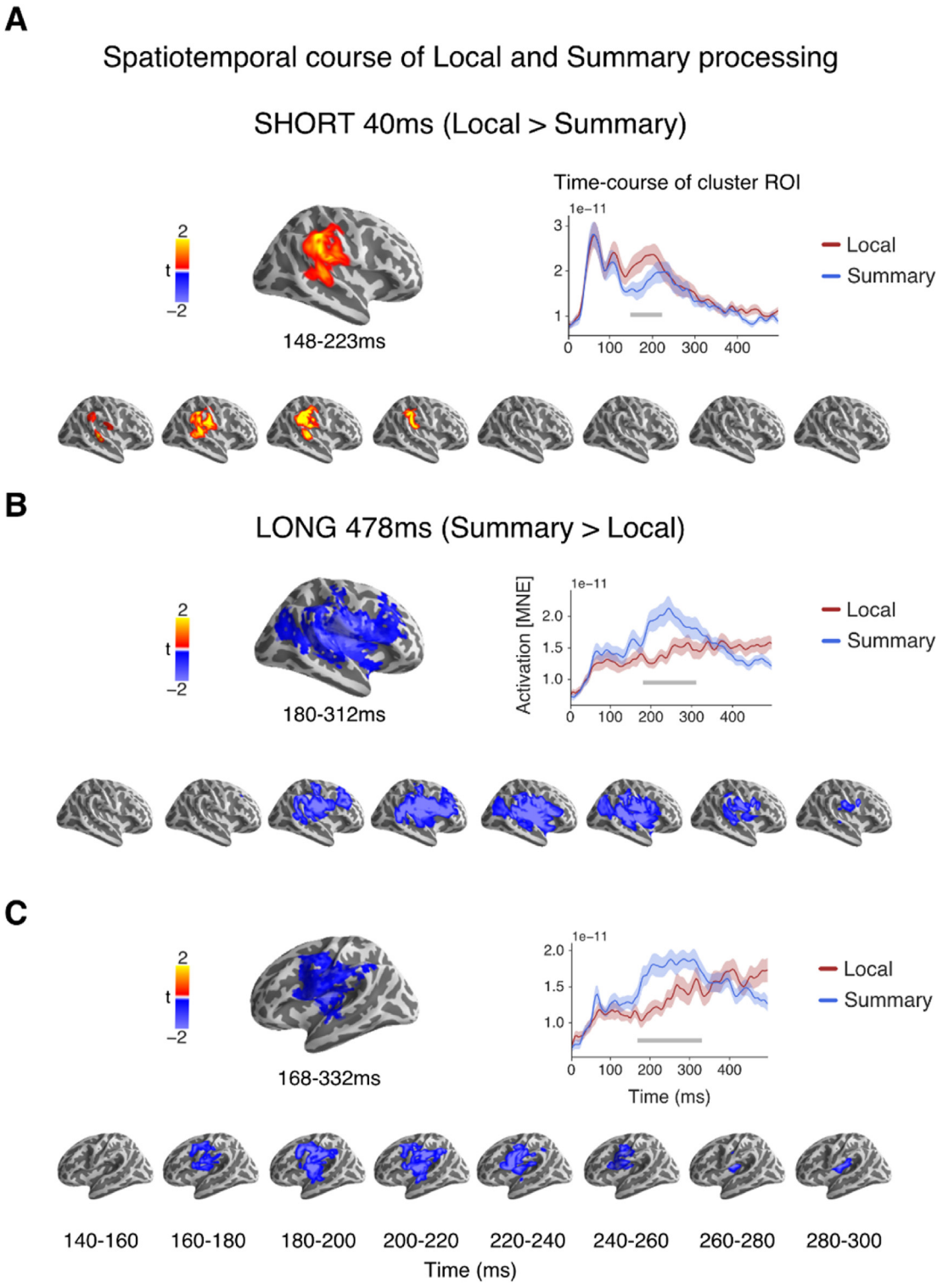

Fig. 3 – Clusters of different activation between experiments at short (40 ms) and long (478 ms) durations. (A) Source localization of regions whose activity was higher for Local Features as compared to Summary Statistics when sounds were short (40 ms). (B) Source localization of regions in the right hemisphere whose activity was higher for Summary Statistics than Local Features when sounds were long (478 ms). (C) Source localization of regions in the left hemisphere in which activity was higher for Summary Statistics than Local Features when sounds were long (478 ms). For each figure (A, B, or C), the top left panel shows the t-values of significant differences between the conditions plotted on the brain surface, at the latencies displayed below. The top right panel represents the time course of activation (computed with MNE) extracted from the selected regions of interest pertaining to the cluster of activation shown on the left for both experiments (Local or Summary). Shaded areas are the standard error of the mean (SEM). Gray lines represent latencies where activations significantly differ between experiments ($P < .05$). The bottom panel displays the spatiotemporal course of cluster t-maps plotted on the brain surface at latencies displayed below.

Table 1 – List of cortical regions in the clusters.

A. Cortical regions in RH Cluster for short (40 ms) sounds	% of voxels
Planum temporale or temporal plane of the superior temporal gyrus *	88.24
Supramarginal gyrus *	74.24
Posterior ramus (or segment) of the lateral sulcus (or fissure) *	62.5
Postcentral sulcus *	39.68
Subcentral gyrus (central operculum) and sulci	37.14
Sulcus intermedius primus (of Jensen) *	36.36
Lateral aspect of the superior temporal gyrus *	30.95
Superior temporal sulcus (parallel sulcus) *	12.61
Postcentral gyrus *	12.5
Anterior transverse temporal gyrus (of Heschl) *	12.5
Central sulcus (Rolando's fissure) *	8
Superior segment of the circular sulcus of the insula *	7.69
Angular gyrus *	1.92
Intraparietal sulcus (interparietal sulcus) and transverse parietal sulci *	1.49
B. Cortical regions in RH cluster for long (478 ms) sounds	**% of voxels**
Anterior transverse temporal gyrus (of Heschl) *	100
Planum temporale or temporal plane of the superior temporal gyrus *	100
Transverse temporal sulcus	100
Posterior ramus (or segment) of the lateral sulcus (or fissure) *	100
Long insular gyrus and central sulcus of the insula	100
Subcentral gyrus (central operculum) and sulci	100
Vertical ramus of the anterior segment of the lateral sulcus (or fissure)	83.33
Supramarginal gyrus *	81.82
Superior segment of the circular sulcus of the insula *	71.79
Inferior part of the precentral sulcus	68.97
Opercular part of the inferior frontal gyrus	68.18
Sulcus intermedius primus (of Jensen) *	63.64
Superior temporal sulcus (parallel sulcus) *	62.18
Inferior frontal sulcus	58.33
Lateral aspect of the superior temporal gyrus *	52.38
Inferior segment of the circular sulcus of the insula	48.57
Central sulcus (Rolando's fissure) *	48
Short insular gyri	45.45
Angular gyrus *	44.23
Postcentral gyrus *	41.67
Postcentral sulcus *	41.27
Superior frontal sulcus	36.17
Middle frontal gyrus	35.38
Precentral gyrus	28.81
Triangular part of the inferior frontal gyrus	26.67
Superior occipital sulcus and transverse occipital sulcus	21.74
Anterior segment of the circular sulcus of the insula	21.43
Intraparietal sulcus (interparietal sulcus) and transverse parietal sulci *	16.42
Middle frontal sulcus	12.9
Middle occipital gyrus (lateral occipital gyrus)	6.06
Middle temporal gyrus	5.77
Superior frontal gyrus	
C. Cortical regions in LH cluster for long (478 ms) sounds	**% of voxels**
Subcentral gyrus (central operculum) and sulci	92.68
Inferior part of the precentral sulcus	87.1
Anterior transverse temporal gyrus (of Heschl)	54.55
Supramarginal gyrus	46.88

Table 1 – *(continued)*

A. Cortical regions in RH Cluster for short (40 ms) sounds	% of voxels
Precentral gyrus	45
Posterior ramus (or segment) of the lateral sulcus (or fissure)	42.42
Central sulcus (Rolando's fissure)	41.77
Superior segment of the circular sulcus of the insula	40.43
Opercular part of the inferior frontal gyrus	40
Long insular gyrus and central sulcus of the insula	33.33
Sulcus intermedius primus (of Jensen)	33.33
Inferior segment of the circular sulcus of the insula	30.77
Superior part of the precentral sulcus	28.57
Transverse temporal sulcus	25
Postcentral gyrus	21.82
Lateral aspect of the superior temporal gyrus	20.93
Postcentral sulcus	20.29
Short insular gyri	14.29
Planum temporale or temporal plane of the superior temporal gyrus	14.29
Middle frontal gyrus	9.23
Inferior frontal sulcus	8.11
Superior temporal sulcus (parallel sulcus)	4.67
Angular gyrus	4.26
Superior frontal sulcus	2.17

Table 1. Extended labels (from Destrieux et al., 2010) of cortical regions associated with the voxels in the three clusters with higher activation in Local Features for short (40 ms) sound duration and Summary Statistics for long (478 ms) sound duration. For each region, we specify the percentage of cortical surface that was included in the cluster (for instance, when the percentage of voxels is 100, the entire surface of the parcellated area is embedded in the cluster; when it is 50, only half of the voxels in that specific region belongs to the cluster, and so on). Stars * indicate the regions in the right hemisphere that are shared between the first and second clusters (A and B; See also Fig. 3A and B). Cortical regions are labeled using the parcellation atlas available in the FreeSurfer package (Freesurfer v7.3.2, aparc.a2009s; Destrieux et al., 2010).

over the frontal lobe. In particular, the activation differed mostly over the SFS, IFS, IFG, MFG, the precentral, central, and postcentral sulci and gyri, while the cluster included fewer voxels in the temporal and parietal lobes, mainly over smaller portions of STS, STG, PT, HG, the transverse temporal sulcus, the Jensen sulcus, the insular cortex, the lateral fissure, the central operculum, and the SMG and AG (Table 1C).

The percentage of significant voxels for each region is displayed in Table 1. Noticeably, the cluster in the right hemisphere observed in response to summary statistics (long sounds) covered 100% of the cortical space in the primary auditory cortex (see Table 1A), suggesting that processing a change in summary statistics compared to one in local features requires a greater engagement of the right auditory cortex. Importantly, the areas in the right hemisphere responding to changes in local features when sounds were short and summary statistics when sounds were long, overlapped over auditory regions (Table 1A, B). That is, to some extent, the same regions in the right auditory cortex seemed to be involved in local features or summary statistics processing depending on sound duration. However, brain

responses to summary statistics (long sounds) were more extended and involved anterior temporal areas and frontal regions as well.

The latencies of the two effects observed in the right hemisphere were also different, with the response to local features changes (short sounds) preceding in time (32 ms before) the response to summary statistics with long sounds (Fig. 1). These findings suggest that the right auditory cortex is differently engaged by one computation or the other based on sound duration; moreover, detecting a change in summary statistics, compared to local features, engages a broader network of frontotemporal regions outside auditory core ones. Conversely, the left hemisphere was involved selectively in the processing of summary statistics at long durations. The response to summary statistics predominantly involved frontotemporal areas and to a less extent primary auditory regions. When considering its temporal dynamic, the left hemisphere response to summary statistic changes (Fig. 3C) started somewhere in between the responses observed for local and summary statistics changes in the right hemispheres (Fig. 3A and B) and lasted longer.

Overall, these findings suggest that strong hemispheric asymmetries exist in processing Local or Summary acoustic information and that these computations rely on partially different brain sources. Specifically, when sound excerpts were short, a change in local features, with unchanged summary statistics, involved only the right hemisphere, at the level of the auditory cortex (Fig. 3A). As more information was provided (e.g., stimulus duration increases), a change in summary statistics (compared to local features only) elicited higher bilateral responses in a broad network of frontotemporal areas (Fig. 3B and C).

6. Discussion

This study investigated whether the human brain is endowed with specialized neural architectures for the computation of local and summary statistics of stationary sounds.

We employed a validated protocol (see Berto et al., 2022) and the MEG to estimate cortical sources of these two auditory modes of functioning. We presented sound streams composed of triplets of sounds, with the first two elements being repeated and the third systematically varying for either its local features or summary statistics. Specifically, we employed synthetic sound excerpts paired according to their temporal or summary statistics similarity. Using the same original white noise sample for the synthesis led to synthetic outcomes with similar temporal amplitude modulations, disregarding the imposed generative statistics (Fig. 1B and C, left panel). As auditory statistics are averaged in time, when sounds were short, their statistics similarity was influenced by temporal local features; as sound duration increased, summary statistics similarity of excerpts including different generative statistics decreased, even when their input white noise was the same (Fig. 1C, right panel). Thus, sound duration represents a crucial factor for measuring auditory discrimination based on local features or summary statistics. Since we expected local and summary processing to rely on different sound properties (i.e., envelope and time-averaged statistics, respectively), we hypothesized they could also rely on different auditory networks.

Our findings confirmed this prediction. We measured greater activation for changes in local features than in summary statistics, in short sounds. The opposite was found for long sounds, with higher activation in response to a change in summary statistics compared to local features. This clear dissociation suggested that discriminative responses to sound changes are, at least partially, guided by stimulus properties measurable at high or low temporal resolutions. Specifically, when little information is presented, the processing of local features subtends to detecting sound changes; as sound duration increases, local features become less relevant, and deviant discrimination relies on summary statistics. Interestingly, these differences in activation do not require an explicit task, are engaged by a systematic, regular change in incoming stimulation, and represent a general mechanism of encoding the selective type of sound change. Given the large variability of sound textures employed to generate triplets, this effect could not be stimulus-specific. The only constant was the change in local features or statistics between novel and repeated sounds which was ensured by using synthetic sounds.

6.1. *Functional cortical specialization for local features and summary statistics*

Source localization highlighted partially different correlates in response to sound changes based on local features or summary statistics.

For short sounds, a change in local features elicited higher activations selectively in the right hemisphere, involving several regions, including the primary auditory cortex (HG), the PT, the middle portion of the STG, and the SMG (Fig. 3A; Table 1A). The higher response in these areas was sustained for several milliseconds (148–224 ms) and suggested an increased sensitivity of the right auditory cortex to changes in the local details of brief sound excerpts. A recent study highlighted the contribution of both HG and STG in the perceived dissimilarity of sounds and the active role of these regions in processing fine-grained acoustic details (Giordano et al., 2023). Interestingly, in the context of complex-sounds perception, the PT has been described as a computational hub devoted to segregating spectrotemporal features from complex auditory patterns (Griffiths & Warren, 2002). This includes computations relying on basic acoustic properties, such as amplitude (Giraud et al., 2000) and frequency (Hall et al., 2002) modulations. Finally, the right SMG was appointed as part of a network involved in auditory memory (Jerde et al., 2011) and rhythmic perception (Schaal et al., 2017). Higher activations in the right PT and SMG may reflect the perception of a pattern influenced by the acoustic dissimilarity between repeated and novel sound excerpts.

Conversely, for long sounds, both hemispheres were more responsive to the processing of summary statistics changes. Such differential activation involved a broad frontotemporal

network comprising several auditory regions of the right hemisphere (Fig. 3B; Table 1B). Compared to the right hemisphere activation favoring local features changes in short sounds, the one favoring summary statistics at long durations involved more anterior and ventral temporal regions. Previous evidence showed that regions located posterior-laterally to HG tend to encode broadband spectral information with high temporal precision; conversely regions located more antero-ventrally preferably process fine-grained spectral information at lower temporal resolutions (Giordano et al., 2023; Kumar et al., 2007; Santoro et al., 2014). As more information is accumulated, spectral content is summarized in time, and thus, the difference in activation between local and summary statistics extends to the anterior and ventral parts of the auditory cortex. The activation also spread towards the posterior parts of STS, compatibly with the emergence of abstraction mechanisms (Warren et al., 2005). The STG is part of the ventral stream ("what" pathway) for sound recognition; the posterior part, especially laterally to HG, is bilaterally activated by environmental and natural sounds (DeWitt & Rauschecker, 2012; Doehrmann et al., 2008). Moreover, different groups of neurons in STG may be responding to selective properties of sounds; such activations are then processed in higher-order areas (possibly ventrolateral prefrontal cortex; Hjortkjær et al., 2018) to resolve the category-specificity of the information (Giordano et al., 2023). Coherently, not only we observed frontal activations in the right hemisphere but also in the left one.

The left hemisphere was more engaged for a change in summary statistics, selectively at long sound duration. Such activation comprised fewer auditory regions and more areas located in the inferior frontal lobe (Fig. 3C). Inferior frontal regions, known as "auditory-related areas," have been associated with sound-object selection and are responsive to both targets and non-targets sounds, if relevant (Steinschneider et al., 2014). Previous studies suggested that local information supports sound differentiation while the emergence of summary statistics may support sound recognition (e.g., Zhai et al., 2020). In the same vein, the automatic tuning to systematic changes in generative statistics would also signify a systematic change in sound category and involve a network of activation spreading towards non-core auditory regions and frontal auditory-related areas.

Overall, these findings showed that the processing of acoustic changes at high or low temporal resolutions relies on overlapping as well as on different brain regions suggesting a functional specialization in the auditory pathway for the processing of local and summary representations.

6.2. *Hemispheric specialization of auditory computations at high and low temporal resolution*

Previous studies investigated hemispheric specialization in auditory processing (i.e., Flinker et al., 2019; Zatorre & Belin, 2001). Auditory regions of one hemisphere can be more sensitive to specific sound properties than their contralateral homolog. The left hemisphere is predisposed to processing temporal modulations, while the right one spectral information (Zatorre & Belin, 2001). Other studies showed that the left hemisphere specialized in processing information at high temporal resolutions while the right hemisphere at slower temporal rates (Flinker et al., 2019). A recent study in mice demonstrated that activations of single neurons in the right auditory cortex persist longer than in the left one (Neophytou et al., 2022). This could underlie a right hemispheric specialization in retaining brief auditory signals in memory, consistent with previous findings in humans showing longer integration windows in the right STG (Arnal et al., 2015). Hemispheric auditory specializations emerged in areas outside the core auditory regions and the temporal lobe (e.g., SMG Schaal et al., 2013, 2017).

By using a systematic synthesis approach, we could tackle specific auditory computations (local and summary processing) while generalizing to a vast category of stimuli, disregarding the idiosyncratic characteristics of each sound (i.e., pitch, acoustic frequency). We showed that the right and left hemispheres are engaged differently depending on which computation is involved in the discrimination. When sounds were short, the processing of local features was favored and involved auditory regions located in the right hemisphere; when sounds were long, higher activations were observed for a change in summary statistics compared to local features in both hemispheres.

When both hemispheres are involved in a function in the same way, it indicates functional symmetry, which would suggest that both hemispheres contribute equally or similarly to that cognitive process. In the context of auditory discrimination, functional symmetry may have several implications. For instance, both hemispheres could be equally involved in sound processing. However, in our specific case, we report a case of functional asymmetry, when one hemisphere of the brain is more heavily involved or specialized than the other as a function of the features underpinning auditory discriminations. That is, regions in the right hemisphere involved in local feature processing of short sounds were also involved in processing summary statistics changes in long sounds, suggesting their computational role adapted to sound duration. Conversely, the left hemisphere selectively processed summary statistics at long duration; its activation preceded the one in the right hemisphere and was more sustained in time, suggesting it may reflect a mechanism of selective attention. Hemispheric asymmetry may have specific roles in auditory processing. Hemispheric lateralization may suggest that each hemisphere has different cognitive functions that it specializes in, while the other hemisphere supports or complements its function. Several studies have previously documented right hemispheric dominance for auditory discrimination. For instance, the right auditory cortex contributes significantly more than the left one in tasks involving fine auditory temporal discriminations (De Sanctis et al., 2009), audio-spatial perception (Dietz et al., 2014), and comparison between stimuli and/or target detection tasks (Gilmore et al., 2009). Additionally, right-hemispheric dominance has been documented for auditory steady-state responses (ASSR) when participants hear periodic amplitude-modulated sequences of tones (Ross et al., 2005).

The findings described here add to the list of auditory functions that are right lateralized and suggest a predominant

role of the right hemisphere in discriminating local features and summary statistics information in stationary sounds. This highlights the specialized function of the right hemisphere, potentially indicating its superior ability to analyze subtle differences in sound frequencies, tones, and patterns. Importantly, we showed that the right hemisphere indeed dominates in processing local features, but as the duration and complexity of information increase (for example, changes in summary statistics), the left hemisphere also supports processing at the level of frontal regions, thereby complementing automatic auditory discrimination functionality. Supporting this hypothesis, left hemispheric specialization in the attentional selection of fine-grained acoustic information has previously been reported (e.g., Bidet-Caulet et al., 2007).

Altogether, these findings suggest that hemispheric asymmetry for local features and summary statistics processing of stationary sounds may reflect complementary roles of the two hemispheres, with each hemisphere contributing uniquely to acoustic processes at different levels of processing. The right hemisphere may be more involved in general operations subtending to both local and summary computations, such as features retention and their subsequent integration into compact representations. On the other hand, the left hemisphere may have specialized at a higher level of processing. For instance, top-down regulations to guide the averaging of statistics, evaluating whether features belong to a different sound-object when a change in statistics is detected, or, in general, selecting which computations to entrain (local or summary) based on monitoring the amount of entering information.

The asymmetry in auditory processing may have significant implications for understanding how the brain processes language, music, and other auditory stimuli. Moreover, it underscores the complexity of hemispheric specialization in perception and cognitive functioning, highlighting the importance of further research to elucidate the underlying neural mechanisms driving this phenomenon.

7. Conclusion

Our results revealed a clear dissociation in the processing of local features and summary statistics according to sound duration. These findings allowed us to uncover the neural regions associated with these two modes of acoustic representation of stationary sounds. We highlighted how the right hemisphere developed a similar architecture to perform sound discrimination based on local features and summary statistics but possesses the ability to adapt its computations based on sound duration. On the other hand, the left hemisphere was selectively involved at a higher level of processing. By combining computational auditory modeling, a systematic approach to synthesizing sounds, and the measurement of brain activity using MEG, we provide evidence concerning the foundations of different sensory computations and their associated cortical networks.

Overall results revealed that the human brain is endowed with partially overlapping and partially specialized neural architectures for the computation of local and summary statistics of stationary sounds.

Open practices section

The study in this article has earned Open Data, Open Materials and Preregistered badges for transparent practices. The data, materials and preregistered studies are available at: http://doi.org/10.60817/Q5A6-H216.

CRediT authorship contribution statement

Martina Berto: Writing – review & editing, Writing – original draft, Visualization, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Patrick Reisinger:** Writing – review & editing, Visualization, Supervision, Methodology, Investigation. **Emiliano Ricciardi:** Writing – review & editing, Resources, Funding acquisition. **Nathan Weisz:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition. **Davide Bottari:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Investigation, Funding acquisition, Conceptualization.

Fundings

D.B. was supported by the Italian Ministry of University and Research (MUR; PRIN 2017 research grant. Prot. 20177894 ZH). This work was advanced by the research stay of M.B., supported by "Talent at work", Erasmus+ mobility consortium (Grants for traineeships 2021/22). P.R. is supported by the Austrian Science Fund (FWF; Doctoral College "Imaging the Mind"; W 1233-B) and the Austrian Research Promotion Agency (FFG; BRIDGE 1 project "SmartCIs"; 871,232).

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgement

We are thankful to Benz Kaja for the support in developing the repository for the dataset and code used in this work.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cortex.2024.09.020.

Scientific transparency statement

DATA: All raw and processed data supporting this research are publicly available: https://bids-datasets.data-pages.anc.plus.ac.at/auditory/mb_auditory_statistics/, https://github.com/martinaberto/Hemispheric_asymmetry_MEG.

CODE: All analysis code supporting this research is publicly available: https://github.com/martinaberto/Hemispheric_asymmetry_MEG.

MATERIALS: All study materials supporting this research are publicly available: https://github.com/martinaberto/Hemispheric_asymmetry_MEG/tree/main/Experimental%20protocol%20%20.

DESIGN: This article reports, for all studies, how the author(s) determined all sample sizes, all data exclusions, all data inclusion and exclusion criteria, and whether inclusion and exclusion criteria were established prior to data analysis.

PRE-REGISTRATION: No part of the study procedures was pre-registered in a time-stamped, institutional registry prior to the research being conducted. No part of the analysis plans was pre-registered in a time-stamped, institutional registry prior to the research being conducted.

For full details, see the *Scientific Transparency Report* in the supplementary data to the online version of this article.

REFERENCES

Akalin-Acar, Z., & Gençer, N. G. (2004). An advanced boundary element method (BEM) implementation for the forward problem of electromagnetic source imaging. *Physics in Medicine and Biology, 49*(21), 5011. https://doi.org/10.1088/0031-9155/49/21/012

Arnal, L. H., Poeppel, D., & Giraud, A. (2015). Chapter 5—temporal coding in the auditory cortex. In M. J. Aminoff, F. Boller, & D. F. Swaab (Eds.), *Handbook of clinical neurology* (Vol. 129, pp. 85–98). Elsevier. https://doi.org/10.1016/B978-0-444-62630-1.00005-6.

Berto, M., Ricciardi, E., Pietrini, P., & Bottari, D. (2021). Interactions between auditory statistics processing and visual experience emerge only in late development. *iScience, 24*(11), Article 103383. https://doi.org/10.1016/j.isci.2021.103383

Berto, M., Ricciardi, E., Pietrini, P., Weisz, N., & Bottari, D. (2022). Distinguishing fine structure and summary representation of sound textures from neural activity. https://doi.org/10.1101/2022.03.17.484757.

Berto, M., Ricciardi, E., Pietrini, P., Weisz, N., & Bottari, D. (2023). Distinguishing fine structure and summary representation of sound textures from neural activity. *eNeuro*. https://doi.org/10.1523/ENEURO.0026-23.2023

Besl, P. J., & McKay, N. D. (1992). Method for registration of 3-D shapes. *Sensor Fusion IV: Control Paradigms and Data Structures, 1611*, 586–606. https://doi.org/10.1117/12.57955

Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P.-E., Giard, M.-H., & Bertrand, O. (2007). Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. *Journal of Neuroscience, 27*(35), 9252–9261. https://doi.org/10.1523/JNEUROSCI.1402-07.2007

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.

Brodbeck, C., Das, P., Gillis, M., Kulasingham, J. P., Bhattasali, S., Gaston, P., Resnik, P., & Simon, J. Z. (2022). Eelbrain: A Python toolkit for time-continuous analysis with temporal response functions, 2021.08.01.454687) https://doi.org/10.1101/2021.08.01.454687.

Campos Viola, F., Thorne, J., Edmonds, B., Schneider, T., Eichele, T., & Debener, S. (2009). Semi-automatic identification of independent components representing EEG artifact. *Clinical Neurophysiology, 120*(5), 868–877. https://doi.org/10.1016/j.clinph.2009.01.015

Dau, T., Kollmeier, B., & Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration. *The Journal of the Acoustical Society of America, 102*(5), 2906–2919. https://doi.org/10.1121/1.420345

De Sanctis, P., Molholm, S., Shpaner, M., Ritter, W., & Foxe, J. J. (2009). Right hemispheric contributions to fine auditory temporal discriminations: High-density electrical mapping of the duration mismatch negativity (MMN). *Frontiers in Integrative Neuroscience, 3*. https://doi.org/10.3389/neuro.07.005.2009

Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage, 53*(1), 1–15. https://doi.org/10.1016/j.neuroimage.2010.06.010

DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences, 109*(8), E505–E514. https://doi.org/10.1073/pnas.1113427109

Dietz, M. J., Friston, K. J., Mattingley, J. B., Roepstorff, A., & Garrido, M. I. (2014). Effective connectivity reveals right-hemisphere dominance in audiospatial perception: Implications for models of spatial neglect. *Journal of Neuroscience, 34*(14), 5003–5011. https://doi.org/10.1523/JNEUROSCI.3765-13.2014

Doehrmann, O., Naumer, M. J., Volz, S., Kaiser, J., & Altmann, C. F. (2008). Probing category selectivity for environmental sounds in the human auditory brain. *Neuropsychologia, 46*(11), 2776–2786. https://doi.org/10.1016/j.neuropsychologia.2008.05.011

Fischl, B. (2012). FreeSurfer. *NeuroImage, 62*(2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Flinker, A., Doyle, W. K., Mehta, A. D., Devinsky, O., & Poeppel, D. (2019). Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. *Nature Human Behaviour, 3*(4). https://doi.org/10.1038/s41562-019-0548-z. Article 4.

Gilmore, C. S., Clementz, B. A., & Berg, P. (2009). Hemispheric differences in auditory oddball responses during monaural versus binaural stimulation. *International Journal of Psychophysiology, 73*(3), 326–333. https://doi.org/10.1016/j.ijpsycho.2009.05.005

Giordano, B. L., Esposito, M., Valente, G., & Formisano, E. (2023). Intermediate acoustic-to-semantic representations link behavioral and neural responses to natural sounds. *Nature Neuroscience, 26*(4). https://doi.org/10.1038/s41593-023-01285-9. Article 4.

Giraud, A.-L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., & Kleinschmidt, A. (2000). Representation of the temporal envelope of sounds in the human brain. *Journal of Neurophysiology, 84*(3), 1588–1598. https://doi.org/10.1152/jn.2000.84.3.1588

Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research, 47*(1), 103–138. https://doi.org/10.1016/0378-5955(90)90170-T

Gramfort, A., Luessi, M., Larson, E., Engemann, D., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience, 7*. https://www.frontiersin.org/articles/10.3389/fnins.2013.00267.

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences, 25*(7), 348–353. https://doi.org/10.1016/S0166-2236(02)02191-4

Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: Minimum norm estimates. *Medical & Biological Engineering & Computing, 32*(1), 35–42. https://doi.org/10.1007/BF02512476

Hall, D. A., Johnsrude, I. S., Haggard, M. P., Palmer, A. R., Akeroyd, M. A., & Summerfield, A. Q. (2002). Spectral and

temporal processing in human auditory cortex. *Cerebral Cortex, 12*(2), 140–149. https://doi.org/10.1093/cercor/12.2.140
Hartmann, T., & Weisz, N. (2020). An introduction to the objective psychophysics toolbox. *Frontiers in Psychology, 11*. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.585437.
Hjortkjær, J., Kassuba, T., Madsen, K. H., Skov, M., & Siebner, H. R. (2018). Task-modulated cortical representations of natural sound source categories. *Cerebral Cortex, 28*(1), 295–306. https://doi.org/10.1093/cercor/bhx263
Houck, J. M., & Claus, E. D. (2020). A comparison of automated and manual co-registration for magnetoencephalography. *Plos One, 15*(4), Article e0232100. https://doi.org/10.1371/journal.pone.0232100
Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks, 10*(3), 626–634. https://doi.org/10.1109/72.761722
Jerde, T. A., Childs, S. K., Handy, S. T., Nagode, J. C., & Pardo, J. V. (2011). Dissociable systems of working memory for rhythm and melody. *NeuroImage, 57*(4), 1572–1579. https://doi.org/10.1016/j.neuroimage.2011.05.061
Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024
McDermott, J. H., Schemitsch, M., & Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nature Neuroscience, 16*(4). https://doi.org/10.1038/nn.3347. Article 4.
McDermott, J. H., & Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis. *Neuron, 71*(5), 926–940. https://doi.org/10.1016/j.neuron.2011.06.032
McWalter, R., & McDermott, J. H. (2019). Illusory sound texture reveals multi-second statistical completion in auditory scene analysis. *Nature Communications, 10*(1). https://doi.org/10.1038/s41467-019-12893-0. Article 1.
Miller, L. M., Escabí, M. A., Read, H. L., & Schreiner, C. E. (2002). Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *Journal of Neurophysiology, 87*(1), 516–527. https://doi.org/10.1152/jn.00395.2001
Nelken, I., Rotman, Y., & Yosef, O. B. (1999). Responses of auditory-cortex neurons to structural features of natural sounds. *Nature, 397*(6715). https://doi.org/10.1038/16456. Article 6715.
Neophytou, D., Arribas, D. M., Arora, T., Levy, R. B., Park, I. M., & Oviedo, H. V. (2022). Differences in temporal processing speeds between the right and left auditory cortex reflect the strength of recurrent synaptic connectivity. *PLOS Biology, 20*(10), Article e3001803. https://doi.org/10.1371/journal.pbio.3001803
Norman-Haignere, S. V., & McDermott, J. H. (2018). Neural responses to natural and model-matched stimuli reveal distinct computations in primary and nonprimary auditory cortex. *PLOS Biology, 16*(12), Article e2005127. https://doi.org/10.1371/journal.pbio.2005127
Ross, B., Herdman, A. T., & Pantev, C. (2005). Right hemispheric laterality of human 40 Hz auditory steady-state responses. *Cerebral Cortex, 15*(12), 2029–2039. https://doi.org/10.1093/cercor/bhi078
Ruggero, M. A. (1992). Responses to sound of the basilar membrane of the mammalian cochlea. *Current Opinion in Neurobiology, 2*(4), 449–456. https://doi.org/10.1016/0959-4388(92)90179-O
Schaal, N. K., Pollok, B., & Banissy, M. J. (2017). Hemispheric differences between left and right supramarginal gyrus for pitch and rhythm memory. *Scientific Reports, 7*(1). https://doi.org/10.1038/srep42456. Article 1.
Schaal, N. K., Williamson, V. J., & Banissy, M. J. (2013). Anodal transcranial direct current stimulation over the supramarginal gyrus facilitates pitch memory. *European Journal of Neuroscience, 38*(10), 3513–3518. https://doi.org/10.1111/ejn.12344
Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage, 44*(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061
Steinschneider, M., Nourski, K. V., Rhone, A. E., Kawasaki, H., Oya, H., & Howard, M. A. (2014). Differential activation of human core, non-core and auditory-related cortex during speech categorization tasks as revealed by intracranial recordings. *Frontiers in Neuroscience, 8*, 240. https://doi.org/10.3389/fnins.2014.00240
Taulu, S., Simola, J., & Kajola, M. (2005). Applications of the signal space separation method. *IEEE Transactions on Signal Processing, 53*(9), 3359–3372. https://doi.org/10.1109/TSP.2005.853302
Warren, J. D., Jennings, A. R., & Griffiths, T. D. (2005). Analysis of the spectral envelope of sounds by the human brain. *NeuroImage, 24*(4), 1052–1057. https://doi.org/10.1016/j.neuroimage.2004.10.031
Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11*(10), 946–953. https://doi.org/10.1093/cercor/11.10.946
Zhai, X., Khatami, F., Sadeghi, M., He, F., Read, H. L., Stevenson, I. H., & Escabí, M. A. (2020). Distinct neural ensemble response statistics are associated with recognition and discrimination of natural sound textures. *Proceedings of the National Academy of Sciences, 117*(49), 31482–31493. https://doi.org/10.1073/pnas.2005644117
Zuk, N. J., Teoh, E. S., & Lalor, E. C. (2020). EEG-based classification of natural sounds reveals specialized responses to speech and music. *NeuroImage, 210*, Article 116558. https://doi.org/10.1016/j.neuroimage.2020.116558