NeuroImage 258 (2022) 119366

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Auditory timing-tuned neural responses in the human auditory cortices

Martijn van Ackooij[a], Jacob M. Paul[a,b], Wietske van der Zwaag[c], Nathan van der Stoep[a], Ben M. Harvey[a,*]

[a] *Experimental Psychology, Helmholtz Institute, Utrecht University, Heidelberglaan 1, Utrecht 3584 CS, the Netherlands*
[b] *Melbourne School of Psychological Sciences, University of Melbourne, Redmond Barry Building, Parkville 3010, Victoria, Australia*
[c] *Spinoza Centre for Neuroimaging, Meibergdreef 75, Amsterdam 1105 BK, Netherlands*

ARTICLE INFO

Keywords:
Timing
Audition
fMRI
Computational
Modelling

ABSTRACT

Perception of sub-second auditory event timing supports multisensory integration, and speech and music perception and production. Neural populations tuned for the timing (duration and rate) of visual events were recently described in several human extrastriate visual areas. Here we ask whether the brain also contains neural populations tuned for auditory event timing, and whether these are shared with visual timing. Using 7T fMRI, we measured responses to white noise bursts of changing duration and rate. We analyzed these responses using neural response models describing different parametric relationships between event timing and neural response amplitude. This revealed auditory timing-tuned responses in the primary auditory cortex, and auditory association areas of the belt, parabelt and premotor cortex. While these areas also showed tonotopic tuning for auditory pitch, pitch and timing preferences were not consistently correlated. Auditory timing-tuned response functions differed between these areas, though without clear hierarchical integration of responses. The similarity of auditory and visual timing tuned responses, together with the lack of overlap between the areas showing these responses for each modality, suggests modality-specific responses to event timing are computed similarly but from different sensory inputs, and then transformed differently to suit the needs of each modality.

1. Introduction

Neural processing of auditory event timing is important to support several behaviors that rely on fine-scale perception of the auditory timing, for example speech perception and production (Assmann and Summerfield, 1990; Ding et al., 2014; Ding and Simon, 2014), music perception and production (Bridwell et al., 2017; Manning and Schutz, 2013), and perception of relationships between the timing of events in different sensory modalities. All of these processes rely on perception and action planning following timing in the sub-second range.

Previous research has demonstrated that neurons respond when sounds occur. Neural oscillatory signals follow (entrain to) the rates at which repetitive sounds occur in music (Bridwell et al., 2017; Nozaradan et al., 2016), speech (Ding et al., 2014; Ding and Simon, 2014; Henry and Obleser, 2012) and many artificial sounds (Elhilali et al., 2009; Lalor et al., 2009; Will and Berg, 2007). Multisensory integration also relies on synchronization of oscillatory signals in different sensory cortices (Kayser et al., 2008; Kayser and Logothetis, 2009; Romei et al., 2012; van Atteveldt et al., 2014). Motor actions can also be accurately synchronized to auditory temporal structure (Madison and Merker, 2004), which is consistent with a synchronization in the responses of auditory and motor cortices (Tierney and Kraus, 2013).

Oscillatory activity following the timing of sounds suggests that average neural response amplitudes would monotonically increase with the rate of sounds. Similarly, if sounds with longer durations produce larger responses, average neural response amplitudes would increase monotonically with sound duration. In the visual (Hendrikx et al., 2022; Stigliani et al., 2017; Zhou et al., 2018), motor (Merchant et al., 2011) and somatosensory (Saal et al., 2015) systems, neural responses increase with both the duration and frequency of events. However, these monotonic responses are accompanied by timing-tuned responses that peak at different event durations and/or frequencies in different neural populations (Harvey et al., 2020; Hendrikx et al., 2022; Merchant et al., 2013; Protopapa et al., 2019; Saal et al., 2015). These timing-tuned responses abstract the representations of an event's temporal properties from neural responses at the moments the events occurred, and this abstraction from low-level sensory activity may allow higher-level analysis of sensory event timing. In the visual system, the preferred durations and frequencies (producing the maximal response) of each neural population gradually change across the cortical surface, forming topographic maps of event timing (Harvey et al., 2020; Protopapa et al., 2019).

[*] Corresponding author.
E-mail address: b.m.harvey@uu.nl (B.M. Harvey).

https://doi.org/10.1016/j.neuroimage.2022.119366.
Received 11 January 2022; Received in revised form 25 May 2022; Accepted 8 June 2022
Available online 9 June 2022.

These timing-tuned responses appear to be straightforwardly derived from separate responses to specific events, particularly when sensory receptors with different temporal dynamics are available (Buonomano and Merzenich, 1995; Hendrikx et al., 2022; Saal et al., 2015), and these processes seem likely to be applicable to inputs from different sensory modalities.

Our sense of time is often considered to be an abstract, supra-modal process. A supra-modal sense of timing would predict closely related responses to the timing of auditory events in these same multisensory regions. Supporting this viewpoint, EEG studies have shown that visual regions entrain to rhythmic auditory stimuli (Escoffier et al., 2015), and visual rhythm perception can be improved through auditory but not visual training (Barakat et al., 2015). Furthermore, the basal ganglia respond to beat perception in combined audiovisual tasks but not unimodal audio or visual tasks (Grahn et al., 2011).

Alternatively, modality-specific responses to event timing would predict distinct networks responding to different modalities. This modality-specific viewpoint is supported by recent results. Analysis of responses to event timing in early visual areas (Hendrikx et al., 2022; Roseboom et al., 2018; Stigliani et al., 2017; Zhou et al., 2018) suggest that the neural representation of event timing may be derived from the activity of early visual areas in a distinct pathway for analyzing and perceiving visual timing. Furthermore, neuropsychological lesion studies show a clear dissociation between timing in the visual and auditory domain, where a lesion in the auditory cortex led to a specific lack of entrainment to auditory rhythm (without deafness), but left visual rhythm perception intact (Fries and Swihart, 1990). This suggests a distinct pathway for analyzing and perceiving auditory timing.

Perceptual adaptation studies also support modality-specific neural responses to event timing. Both auditory and visual duration perception are affected by the durations of preceding auditory stimuli, showing a repulsive duration aftereffect that the perceived duration of the current stimulus is repelled away from the duration of preceding stimuli (Heron et al., 2012; Huppert and Singer, 1967; Walker et al., 1981). Such repulsive duration aftereffects are often seen as effects of duration tuned responses (Heron et al., 2012; Tsouli et al., 2022). Adaptation to auditory stimulus duration does not affect perceived visual duration, or vice versa (Heron et al., 2012; Walker et al., 1981), consistent with modality-specific neural responses to timing. These perceptual findings suggest that auditory duration-tuned neural populations may exist in humans, and that these are distinct from visual duration-tuned neural populations (Heron et al., 2013).

Even if the timing of events in different sensory modalities is initially analyzed in modality-specific networks, responses from different modalities may converge into supra-modal representations at later stages to allow grouping or comparisons of the timing of events in different modalities, or action planning that can follow timing from any modalities. Notably, higher-level responses to event visual timing are largely found in parietal and frontal areas implicated in multisensory integration and action planning (Harvey et al., 2020). Furthermore, auditory stimuli can affect the perceived timing of visual stimuli, and vice versa, but only when they are perceptually grouped (Klink et al., 2011). Therefore, separate derivation of distinct auditory and visual timing tuned responses does not exclude the possibility of interactions between them.

If auditory timing is derived in a distinct pathway, it is likely this pathway consists of, or overlaps with, the classical auditory processing areas from the primary auditory cortices of Heschl's gyri to the belt and parabelt regions of the temporal cortex. Indeed, multiple studies show that auditory stimuli with different temporal structures can be classified from activity patterns in the auditory cortices. This classification can distinguish between a range of natural stimuli (Norman-Haignere et al., 2015) and musical genres (Nakai et al., 2021). Similarly, decoding pattern classification approaches can distinguish responses to stimuli with different broadband distributions of auditory frequency (pitch) and temporal amplitude modulation (changes in periodicity of the envelope) (Sohoglu et al., 2020). Sohoglu and colleagues further show distinct auditory cortical regions that carry information about changes in pitch, amplitude modulation, and both. Another decoding study reveals regions of the auditory cortex that change their response patterns with pitch, timbre, or both (Allen et al., 2017). Therefore, the auditory cortices are a strong candidate to exhibit timing-tuned responses.

Here we ask whether and where auditory timing tuned responses occur within the human brain and how their properties change within and between brain areas, and how these responses relate to auditory frequency (pitch) selective responses. FMRI studies demonstrate that neural populations around the auditory cortex respond selectively to the periodicity (frequency) of auditory stimuli that sinusoidally vary their amplitude over time (Barton et al., 2012). Here we use discrete sounds with rapid onsets and offsets, and constant amplitude for a variable duration. We present these sounds repetitively at variable rates, which we describe in terms of the event's period (1/frequency) so that it is (like event duration) given in seconds. Previous studies have sometimes described this period in terms of the sound's envelope or described it in terms of the fundamental frequency (1/period of the envelope) of a repetitive sound. While this period is a measure of event frequency, it should not be confused with the auditory frequency (pitch) of the sound: auditory frequency describes the rate of the compression and decompression of air, in frequencies of at least 20 Hz; event frequency describes the rate of onset of specific auditory events, in frequencies below 10 Hz in the current study. Our auditory events were white noise bursts, so have the same power at every audible frequency and were amplitude-modulated to produce events with variable timings.

2. Methods

2.1. Ethics statement

Experimental participants gave written informed consent for all experimental procedures and use of the data collected from them. All experimental procedures were approved by the ethics committee of University Medical Center Utrecht.

2.2. Participants

Six participants (aged 26-37, three female, all right-handed) with previous MRI experience were recruited to participate. As in previous similar studies using pRF modelling of 7T fMRI data (Harvey et al., 2013, 2020; Harvey and Dumoulin, 2017a; Klein et al., 2014), we aim for a detailed assessment of functional neural response properties in each participant (Gordon et al., 2017) and therefore focus on internal replication using a large amount of data from each participant rather than large numbers of participants (Baker et al., 2021).

All participants had normal or corrected to normal vision and normal hearing. None were musicians.

2.3. Auditory stimuli

All auditory stimuli were generated using MATLAB along with functions from the Binaural Sound Creation Toolbox (Akeroyd, 2001) and functions built in house. Auditory stimuli were presented binaurally using state of the art MRI-compatible piezoelectric headphones (MR ConFon, Model: HP PI US, Cambridge research systems, Rochester, UK), worn over foam earplugs to reduce the sound pressure from the scanner. All stimuli were presented with a sample rate of 44.1 KHz. Participants were first asked to listen to a practice stimulus and to set the volume to a level so they could hear the sound over the scanner noise without it being uncomfortably loud. This was at approximately 110 dB, and the earplugs were rated at -32 dB.

2.4. Duration mapping stimuli

A range of stimuli changing in duration and period was presented. Stimulus duration was defined as the length of time between the onset

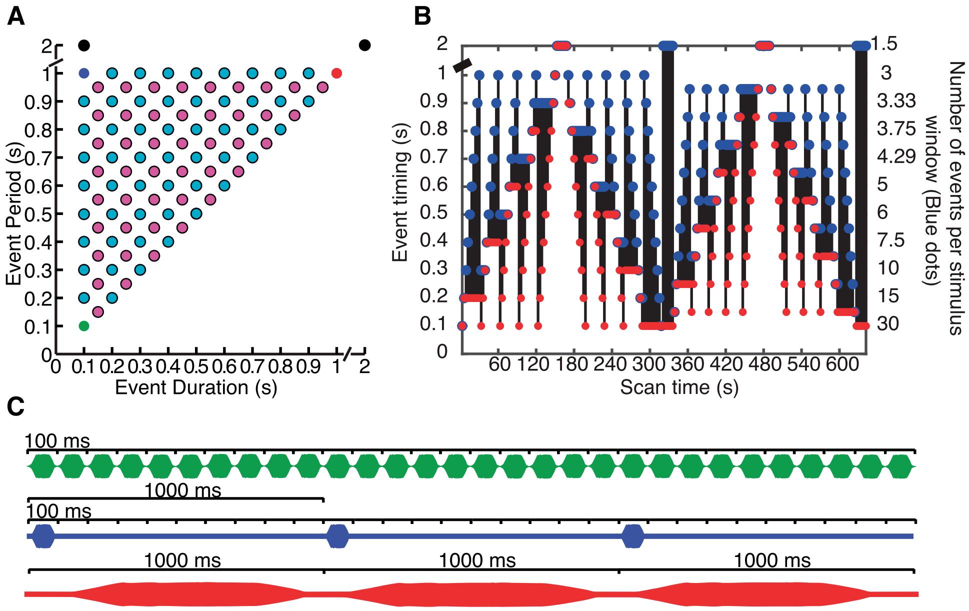

Fig. 1. Timing of presented auditory events. (A) Stimulus space with grid of configurations sampled during the experiment. Cyan and magenta dots denote stimulus configurations present in the first and second half of the scan run, respectively. Black dots denote the control conditions with high periods. Green, blue and red dots denote the duration-period pairs shown in panel C. (B) Sequence of stimulus event timings presented to the participant. Red and blue dots denote the duration and period (respectively) presented at a given time in the scan run, and duration-period pairs are coupled by black lines indicating their presence in the same trial. Both durations and periods are shown on the left axis. The number of times a given event was repeated during a three second window (the stimulus window) before the timing changed is shown on the right axis for the period (blue) dots. (C) Representative sound waveform envelopes for repetitive events of example event timings, repeated for 3 seconds before changing. Detailed sound waveforms are not visible at this scale, but consisted of white noise. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

and offset of a white noise burst (an event). The period was defined as the duration plus the interstimulus interval (i.e., onset to onset). The stimulus sequence systematically traversed the space ranging from 100 milliseconds to 1000 milliseconds in both duration and period dimensions (See Fig. 1a). As the duration can never exceed the period, the resulting space consists only of the upper left half of the square duration-period space. The sequence was constructed in such a way that it was not straightforward for participants to track where in the stimulus sequence they currently were and which timing was being presented.

Each unique stimulus configuration (duration-period pair, denoted by black lines in Fig. 1B) was presented multiple times in a three second interval, depending on the amount of repetitions of the current period that would fit. As three seconds was not an integer multiple of all period values used, the timing of events drifted slightly from synchronization with the 3-second stimulus window around which timing changes were designed. However, this drift never went beyond ±250 ms, and the changes in timing of consecutive events were only 100 ms, so this drift was not perceptible. The presented event timings were used for analysis.

All events were white noise bursts with uniform frequency spectra at auditory frequencies and randomly generated phases of different frequencies. As such, there were no spectral cues and no relationship between auditory frequency and event timing.

No timing judgements were required. To keep participants alert, we used an orthogonal auditory frequency oddball detection task and instructed participants in this task before scanning. To keep the frequency of oddballs and key presses the same for all event timings, only the first event of any three seconds could be an oddball. In these oddballs, the auditory frequency was either high- or low-pass filtered. Ten percent of 3-second stimulus windows began with oddballs, and participants pressed different buttons as quickly as possible to indicate whether an oddball they detected was high or low. These oddballs' positions were pseudo-random, counterbalanced across scan runs by creating twenty scan runs with mutually exclusive oddball positions. As such, summed across all scan runs, each 3-second stimulus window contained exactly one high-pass oddball, one low-pass oddball, and required two different responses.

2.5. Tonotopy stimuli

Tonotopic mapping stimuli were presented binaurally using the same setup as for the auditory duration stimuli. Each sequence consisted of a systematic progression of 27 pure frequency tones ranging from 88 to 7965 Hertz. All 27 frequencies were logarithmically spaced following the sequence generating equation:

$$f_n = f_0 \cdot 2^{\frac{300\cdot(n-1)}{1200}} = f_0 \cdot 2^{\frac{(n-1)}{4}}, \text{ with } f_0 = 88, \text{ and } n \in \{1, 2, \ldots, 26, 27\} \tag{1}$$

During the sequence, tones were presented sequentially for 1.5 seconds each. After all 27 tones were presented there was a period of silence for 3 seconds after which the sequence was repeated in a reversed order (from 7965Hz down to 88Hz). This cycle of increasing followed by decreasing frequency progressions was repeated six times in each run, totaling 522 s. To avoid habituation and temporal expectation of these tones, each 1.5 s presentation contained eight separate tones with the same frequency and intermixed durations of 50 or 200 ms (4 of each, randomly ordered) (Da Costa et al., 2011). The interval between consecutive tones was always 50 milliseconds. The perceived loudness of the same sound pressure varies with auditory frequency (Neuhoff et al., 1999), and it was important for participants to hear all tones over the scanner noise. We corrected for differences in perceived loudness by attenuating sound intensity levels of individual frequencies following the standard equal-loudness curves (ISO 226:2003, see Suzuki and Takeshima, 2004). These equal loudness curves vary sound pressure as a function of frequency such that all frequencies are perceptually equally loud to a 1000Hz tone. This led to a range of sound intensity levels varying from 56dB to 83dB depending on auditory frequency. The resulting frequency response function is for a fixed perceptual loudness rather than a fixed sound pressure.

2.6. MRI data acquisition

We acquired MRI data on a 7T Philips Achieve scanner. Briefly, we acquired T1-weighted anatomical scans, automatically segmented these with Freesurfer, then manually edited labels to minimize segmentation errors using ITK-SNAP. This provided a highly accurate cortical surface model at the grey-white matter border to characterize cortical organization. We acquired T2*-weighted functional images using a 32-channel head coil at a resolution of 1.688×1.688×1.7 mm, with 57 interleaved slices of 128×128 voxels. The resulting field of view was 216×216×96.9 mm. TR was 1500 ms, TE was 22.49 ms, and flip angle was 70 degrees. We used a gradient echo sequence with SENSE acceleration factor 2.4, multiband factor three and anterior-posterior encoding. This fMRI sequence acquired (three simultaneous) slices 19 times per 1500 ms TR,

Table 1

Supplement to Fig. 3, table of all models tested describing different elements of their construction and their average variance explained.

Bar nr. in Fig. 3	Model Name	Number of Free Params	Free params	Predictors	Equation	Shape of RF	Average variance explained
1	Constant (Linear Frequency Only)	0	N/A	N/A	$z(t) \propto c$	Flat line	16 %
2	Linear Duration + Linear Frequency	1	x	- Amplitude Ratio on duration	$z(t) \propto x(t)$	Linear response	12.5 %
3	Linear Duration + Compressive Frequency	2	a p	- Amplitude ratio on duration - compressive exponent	$z(t) \propto a \cdot x(t) + \frac{F(s)^p}{F(s)}$	Subadditive response	16.2 %
4	Compressive Duration + Compressive Frequency	3	a p_x p_f	- Amplitude ratio on duration - compressive exponent on duration - compressive exponent on frequency	$z(t) \propto a \cdot \frac{x(t)^{p_x}}{x(t)} + \frac{F(s)^{p_f}}{F(s)}$	Subadditive response	16.2 %
5	Duration Tuned + Compressive Frequency	3	x_{pref} σ p	- preferred duration - tuning width - compressive exponent on frequency	$z(t) \propto \frac{F(s)^p}{F(s)} \cdot e^{-\frac{1}{2}(\frac{x(t)-x_{pref}}{\sigma})^2}$	1D tuned response with compression	20.2 %
6	Circular Duration and Period	3	x_{pref} y_{pref} σ	- preferred duration - preferred period - tuning width	$z(t) \propto e^{-\frac{1}{2}((\frac{x(t)-x_{pref}}{\sigma})^2+(\frac{y(t)-y_{pref}}{\sigma})^2)}$	2D circular tuned response	20.2 %
7	Circular (log) Occupancy and (log) Period	3	x_{pref} y_{pref} σ	- preferred occupancy - preferred period - tuning width	$z(t) \propto e^{-\frac{1}{2}((\frac{\ln(x(t))-\ln(x_{pref})}{\sigma})^2+(\frac{\ln(y(t))-\ln(y_{pref})}{\sigma})^2)}$	2D circular tuned response	20.8 %
8	Circular Duration and Period tuned + compressive exponent on frequency	4	x_{pref} y_{pref} σ p	- preferred duration - preferred period - tuning width - compressive exponent on frequency	$z(t) \propto \frac{F(s)^p}{F(s)} \cdot e^{-\frac{1}{2}((\frac{x(t)-x_{pref}}{\sigma})^2+(\frac{y(t)-y_{pref}}{\sigma})^2)}$	2D circular tuned response with compression	21.1 %
9	Anisotropic occupancy and period tuned	5	x_{pref} y_{pref} σ_{major} σ_{minor} θ	- preferred occupancy - preferred period - major tuning width - minor tuning width - angulation of RF	$z(t) \propto e^{-\frac{1}{2}((\frac{Y}{\sigma_{major}})^2+(\frac{X}{\sigma_{minor}})^2)}$ $X =$ $\cos(\theta) \cdot (x(t) - x_{pref}) - \sin(\theta) \cdot (y(t) - y_{pref})$ $Y =$ $cos(\theta) \cdot (y(t) - y_{pref}) + sin(\theta) \cdot (x(t) - x_{pref})$	2D anisotropic tuned response	20.4 %
10	Anisotropic (log) occupancy and (log) period tuned	5	x_{pref} y_{pref} σ_{major} σ_{minor} θ	- preferred log of occupancy - preferred log of period - major tuning width - minor tuning width - angulation of RF	$z(t) \propto e^{-\frac{1}{2}((\frac{Y}{\sigma_{major}})^2+(\frac{X}{\sigma_{minor}})^2)}$ $X = \cos(\theta) \cdot (\ln(x(t)) - \ln(x_{pref})) - \sin(\theta) \cdot (\ln(y(t)) - \ln(y_{pref}))$ $Y = cos(\theta) \cdot (\ln(y(t)) - \ln(y_{pref})) + sin(\theta) \cdot (\ln(x(t)) - \ln(x_{pref}))$	2D anisotropic tuned response	20.6 %
11	Anisotropic occupancy and period tuned + compressive exponent	6	x_{pref} y_{pref} σ_{major} σ_{minor} θ p	- preferred occupancy - preferred period - major tuning width - minor tuning width - angulation of RF - compressive exponent on frequency	$z(t) \propto \frac{F(s)^p}{F(s)} \cdot e^{-\frac{1}{2}((\frac{Y}{\sigma_{major}})^2+(\frac{X}{\sigma_{minor}})^2)}$ $X =$ $\cos(\theta) \cdot (x(t) - x_{pref}) - \sin(\theta) \cdot (y(t) - y_{pref})$ $Y =$ $cos(\theta) \cdot (y(t) - y_{pref}) + sin(\theta) \cdot (x(t) - x_{pref})$	2D anisotropic tuned response with compression	20.4 %
12	Anisotropic (log) duration and (log) period tuned + compressive exponent	6	x_{pref} y_{pref} σ_{major} σ_{minor} θ p	- preferred log of duration - preferred log of period - major tuning width - minor tuning width - angulation of RF - compressive exponent on frequency	$z(t) \propto \frac{F(s)^p}{F(s)} \cdot e^{-\frac{1}{2}((\frac{Y}{\sigma_{major}})^2+(\frac{X}{\sigma_{minor}})^2)}$ $X = \cos(\theta) \cdot (\ln(x(t)) - \ln(x_{pref})) - \sin(\theta) \cdot (\ln(y(t)) - \ln(y_{pref}))$ $Y = cos(\theta) \cdot (\ln(y(t)) - \ln(y_{pref})) + sin(\theta) \cdot (\ln(x(t)) - \ln(x_{pref}))$	2D anisotropic tuned response with compression	21.6 %
13	Anisotropic duration and period tuned + compressive exponent	6	x_{pref} y_{pref} σ_{major} σ_{minor} θ p	- preferred duration - preferred period - major tuning width - minor tuning width - angulation of RF - compressive exponent on frequency	$z(t) \propto \frac{F(s)^p}{F(s)} \cdot e^{-\frac{1}{2}((\frac{Y}{\sigma_{major}})^2+(\frac{X}{\sigma_{minor}})^2)}$ $X =$ $\cos(\theta) \cdot (x(t) - x_{pref}) - sin(\theta) \cdot (y(t) - y_{pref})$ $Y =$ $cos() \cdot (y(t) - y_{pref}) + sin(\theta) \cdot (x(t) - x_{pref})$	2D anisotropic tuned response with compression	26.4 %

and so produced audible gradients every 79 ms. We used a 2nd-order image-based B0 shim of the functional scan's field of view (MRCodeTool, MRCode, Zaltbommel, Netherlands). This covered most of the brain but omitted the posterior lobe of the cerebellum and the anterior temporal lobes, where 7T fMRI has low response amplitudes and large spatial distortions.

Each scan lasted for 663 seconds and began with 6 dummy scans to ensure steady state of the signal (which were discarded), followed by 436 functional volumes. Six to eight repeated runs were acquired within the same session. In each session, we acquired a top-up scan recorded with the opposite phase-encoding direction from the main functional runs to correct for image distortion in the gradient encoding direction, (Andersson et al., 2003), and consists of 10 TRs.

2.7. Preprocessing

Co-registration of functional data to the high-resolution anatomical space was performed using a custom pipeline (Paul et al., 2022) in AFNI (afni.nimh.nih.gov)(Cox, 1996). A single transformation matrix was constructed, incorporating all the steps from the raw data to the cortical surface model to reduce the number of interpolation steps to one. No other spatial or temporal smoothing procedures were applied. A T1 image with the same resolution, position and orientation as the functional data was first used to determine the transformation to a higher resolution (1mm isotropic) whole-brain T1 image (3dUnifize, 3dAllineate). For the fMRI data, we first applied motion correction to two series of images that were acquired using opposing gradient encoding directions (3dvolreg). Subsequently, we determined the distortion transformation between the average images of these two series (3dQwarp). We then determined the transformation in brain position between and within functional scans (3dNwarpApply). Then we determined the transformation that co-registers this functional data to the T1 acquired in the same space (3dvolreg). We applied the product of all these transformations to every functional volume to transform our functional data to the whole-brain T1 anatomy. We repeated this for each fMRI session to transform all their data to the same anatomical space.

We then imported these data into Vistasoft's mrVista framework (github.com/vistalab/vistasoft) for analysis and model fitting. We averaged the resulting data together separately for duration mapping and tonotopy. For both conditions, we also made further averages of the data from odd and even scans for cross-validation. For tonotopy only, we then averaged repeated stimulus cycles within the same scan, leaving 58 fMRI responses measurements to which we fit response models.

2.8. fMRI analyses

2.8.1. Timing-selective neural response models

To characterize the aggregate response of the neural populations in each fMRI voxel, we use forward models to predict the observed fMRI time courses that would result from neural response functions with various parametric relationships between event duration and event frequency. This approach lets us first determine the form of the parametric response function that best predicts responses from all responsive voxels, and then determine the parameters of this response function that best predict the responses in each voxel. Following our previous work on responses to visual event timing (Harvey et al., 2020; Hendrikx et al., 2022) we tested the hypothesis that the brain shows responses that vary with event duration and period following an anisotropic Gaussian function. However, neural response functions with various levels of complexity have been described for many stimulus features. These vary from constant (the same response amplitude per event, regardless of the event's properties) to multivariate nonlinear functions (e.g., the classical circular-symmetric Gaussian tuned response function of visual field position, Hubel and Wiesel, 1959). The timing changes in our repetitive stimuli can also be parameterized in terms of event duration, event frequency (the number of events per second in each 3 second stimulus window), event period (1/frequency), occupancy (event period divided by event duration), and interstimulus interval (event period minus event duration). We therefore tested candidate models that predict neural responses amplitudes to each event as either a monotonic function of event duration and/or frequency, or a Gaussian function of various combinations of these parameters.

The simplest model we tested predicts a constant response to every event, regardless of its duration and period. In essence, this response model is no more than a detector of event offsets. We denote this as:

$$z(t) \ \propto \ c \tag{2}$$

where $z(t)$ denotes the amplitude of the observed signal at event time t. We predict response amplitudes on a per-event basis, but these responses accumulate over a few seconds due to fMRI's measurement of slow changes in blood flow and oxygenation. The fMRI response amplitude will then increase linearly with event frequency.

A more complex response to each event would linearly follow the duration of the event. We denote this as:

$$z(t) \ \propto \ x(t) \tag{3}$$

here, $x(t)$ refers to the magnitude of the physical quantity presented at time point t and $t-1$.

The monotonic model is easily extended to a sub-additive function that allows for a sub-additive accumulation of responses with increasing event frequency, giving us

$$z(t) \ \propto \ a \cdot x(t) + \frac{F(s)^p}{F(s)} \tag{4}$$

$F(s)$is the frequency with which a fixed stimulus state appears during a period of time s. Notice that the response amplitude is computed *per event*, but the sub-additive effect of a number of occurrences of a single stimulus state is computed as a constant over the entire stimulus window$s$, of three seconds (2 TR's of the acquisition sequence). $p \leq 1$is the exponent with which the compression occurs and is one of the parameters estimated by the model. The ratio $\frac{F(s)^p}{F(s)}$ replaces the constant response per event with a component that increases sub-additively with event frequency.

More complicated candidate neural response functions could include a Gaussian tuned response to one timing parameter, duration or occupancy, supplanting the monotonic response to duration:

$$z(t) \ \propto \ \frac{F(s)^p}{F(s)} \cdot e^{-\frac{1}{2}\left(\frac{x(t)-x_{pref}}{\sigma}\right)^2} \tag{5}$$

With additional parameters x_{pref} referring to preferred duration or period andσbeing the extent of the tuning function.

Lastly, we hypothesized that responses to events of variable duration and period would be predicted most closely by using a multivariate tuned representation, similar to the two-dimensional receptive fields seen in visual field position encoding and two-dimensional response functions seen in visual timing encoding (Harvey et al., 2020). A tuned Gaussian response function can be extended to two dimensions by adding a preference and a range in the second dimension. This can be computed through the following equation:

$$z(t) \ \propto \ \frac{F(s)^p}{F(s)} \cdot e^{-\frac{1}{2}(\left(\frac{x(t)-x_{pref}}{\sigma_x}\right)^2 + \left(\frac{y(t)-y_{pref}}{\sigma_y}\right)^2)} \tag{6}$$

However, this limits the response function's orientation to the duration or period axis exclusively, while for visual timing-tuned responses we have demonstrated an interaction between the two terms in the form of an angulated Gaussian. Angulation can be introduced by expressing the gaussian in terms of the major and minor axes such that

$$z(t) \ \propto \ \frac{F(s)^p}{F(s)} \cdot e^{-\frac{1}{2}(\left(\frac{Y}{\sigma_{major}}\right)^2 + \left(\frac{X}{\sigma_{minor}}\right)^2)} \tag{7}$$

where X and Y are expressed in terms of the angulation parameter θ

$$X \ = \ cos(\theta) \cdot \left(x(t) - x_{pref}\right) - sin(\theta) \cdot \left(y(t) - y_{pref}\right) \tag{8}$$

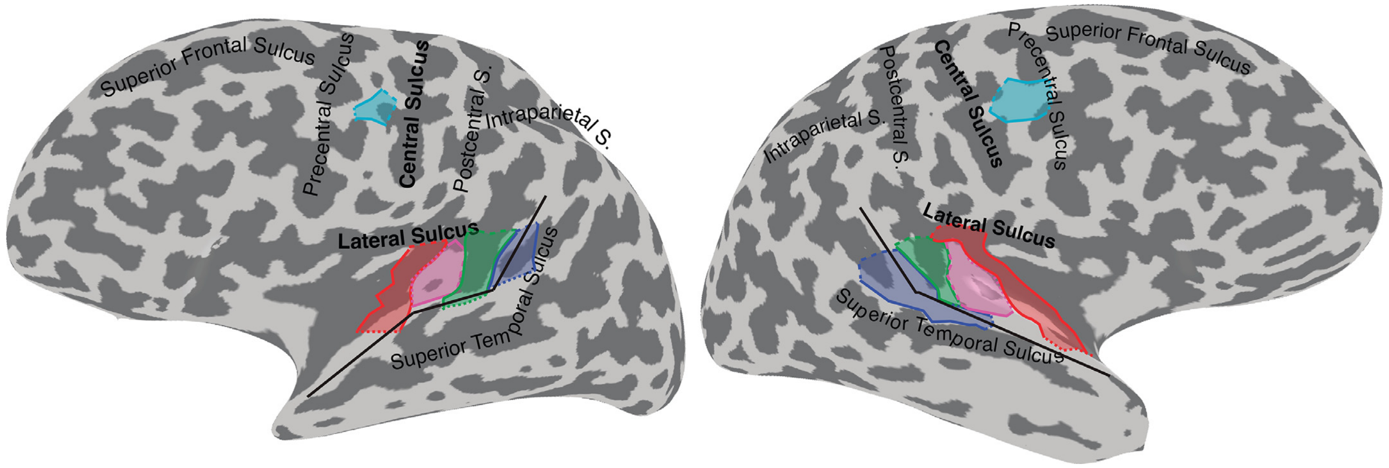

Fig. 2. Inflated cortical surface rendering of one participant showing anatomical landmarks and the ROIs used to analyze the data. The magenta lines surround the Heschl's gyri ROI, ATA is in red, ATM in green, ATP in blue, and ATPM in cyan. Fine and coarse dashed lines represent the ends of each ROI used for analysis of changes in response properties with cortical distance across the ROI. The crest of the Superior Temporal Gyrus (STG) is marked by the black line.

$$Y = cos(\theta) \cdot \left(y(t) - y_{pref}\right) + sin(\theta) \cdot \left(x(t) - x_{pref}\right) \tag{9}$$

In Eqs. (5), (6), (8) and (9), x and y can be duration and period respectively, occupancy and period respectively, duration and inter-event interval respectively, or logarithms of these parameters.

2.9. Auditory frequency selective neural response models

This tuning in a logarithmic domain has been shown in the primary auditory cortex and the surrounding regions (Da Costa et al., 2011), as well as in other modalities (Harvey and Dumoulin, 2017a, 2017b). We use the logarithmic case in testing candidate response functions using the auditory frequency data. This is done in a one-dimensional case as follows

$$z(t) \propto e^{-\frac{1}{2}\left(\frac{ln(x(t))-ln\left(x_{pref}\right)}{\sigma}\right)^2} \tag{10}$$

where ln is the natural logarithm, x is auditory frequency, and all other components are as specified in the section on duration selective response functions.

2.10. Testing candidate models

All of the aforementioned candidate models are tested using the population receptive field modelling paradigm (Dumoulin and Wandell, 2008). This modeling approach can evaluate the predictions any candidate parametric response function (Harvey et al., 2020; Harvey and Dumoulin, 2017a, 2017b) for a large set of candidate combinations of response function parameters. For every stimulus time point, we computed the response function's amplitude at the presented stimulus state, giving a candidate prediction of the neural response time course expected if this candidate response function was responding to this stimulus sequence. This candidate neural response time course was then convolved with the canonical hemodynamic response function (Glover, 1999) to generate a predicted fMRI time course expected if this response function was responding to this stimulus sequence. We determined the correlation between this prediction and the measured fMRI response time course at each recording site. We then fit the free parameters of the model to find the response function parameters that maximize the correlation between the fMRI time course prediction and the measured fMRI response time course.

We used cross-validation to compare the prediction of different candidate response models despite the different numbers of free parameters. We took the best-fitting parameter set from the odd numbered scans and quantified the correlation between its prediction and the measured fMRI time course from the even numbered scans, and vice versa, giving the cross-validated variance explained by each candidate model. All resulting cross-validated model fits for each tested neural response model were averaged within each region of interest in each hemisphere for each cross validation split. These fits were compared between models with pairwise two-sided comparisons to give t-statistics and p-values of differences between model fits.

2.11. Region of interest identification

Using the best fitting model, we then defined regions of interest (ROIs) for further analysis. These were contiguous responsive areas which were consistently located relative to anatomical landmarks in different hemispheres. While these ROIs were anatomically defined regions of the cortical surface, we exclude responses with below 10% cross-validated variance explained in the best performing response model from all further analyses as these do not show clear response modulation by auditory event timing, so not all voxels within each anatomical ROI are analyzed. The first ROI was Heschl's gyri (transverse temporal gyri), the anatomical location of the primary auditory cortices. We subsequently defined three regions surrounding the auditory cortex that include the classical belt and parabelt regions. We named these by their anatomical locations, preceded by "AT" for "auditory timing", following conventions for naming visual field maps (Wandell et al., 2005), numerosity maps (Harvey and Dumoulin, 2017a; Hofstetter et al., 2021) and visual timing maps (Harvey et al., 2020). Auditory timing anterior (ATA) extended from the anterior edge of Heschl's gyri along the lateral bank of the lateral sulcus into the fundus of the lateral sulcus, following the planum polare (medial surface of the superior temporal gyrus (STG), Fig. 2). This resembles the locations of the medial belt areas of the macaque auditory cortex (Kaas and Hackett, 2000). Auditory timing medial (ATM) extended from the posterior edge of Heschl's gyri to the upper bank of the STG including the planum temporale and the inferior part of the supramarginal gyrus. This resembles the locations of the lateral belt areas of the macaque auditory cortex (Kaas and Hackett, 2000). Auditory timing posterior (ATP) ran alongside ATM, extending from the top of the STG along its lateral bank, consistent with the anatomical location of Wernicke's area or the macaque auditory parabelt region (Kaas and Hackett, 2000). A fourth region, auditory timing premotor (ATPM), was located on the precentral gyrus at its junction with the middle frontal gyrus, consistent with the part of the premotor cortex.

2.12. Transformations of ROI locations to standard templates

We transformed each individual participant's brain to the N27 (Talairach) standard template using AFNI's surface-based co-registration tools (3dQwarp, 3dAllineate). We then identified the centre of each ROI on the surface and passed this through the same transformation to determine centre coordinates in Talairach space. Finally, we used the tal2mni tool to transform these centre coordinates into MNI space. After each transformation of the individual participant's ROI centres, we calculated the mean and standard deviation in each template space (Provided in supplementary Table 1).

2.13. Analysis of parameter distribution within ROIs

2.13.1. Changes in preferred duration and period

To characterize the cortical organization of timing selectivity within these ROIs, we analyzed the observed progression of preferred durations and periods over cortical surface between anatomically defined landmarks. Previous work has focused on linear and u-shaped progression of response preferences binned by cortical distance (Arcaro et al., 2011; Harvey et al., 2013, 2020; Harvey and Dumoulin, 2017a), interpreted as one or two bordering topographic maps respectively (e.g., one descending and one ascending progression for a u-shaped profile). Our anatomically-defined ROIs do not necessarily consist of precisely one topographic progression, particularly because the auditory belt and parabelt each include several contiguous areas with distinct functional properties (Moerel et al., 2014), cytoarchitectonic properties (Kaas and Hackett, 2000; Moerel et al., 2014) and connectivity. To allow the possibility of several topographic map progressions within a single ROI, we introduced a cyclical sinusoidal fit (with frequency and phase as free parameters) to the mean preferred durations or frequencies of 2 mm cortical distance bins between the ends of each ROI. If ROIs consistently showed the same number of cycles in a similar phase in different hemispheres this would provide evidence of multiple topographic progressions within the ROI. We used a general linear model to quantify the contributions of both the best fitting cyclical and linear functions to the observed changes in preferred duration and period between the anatomically-defined ends of each ROI. Subsequently, we compute the corresponding t-statistics and probabilities in each ROI and hemispheres, correcting probabilities for the number of ROIs tested by using false discovery rate correction (Benjamini and Hochberg, 1995).

We further tested whether estimates of duration and period preference showed a repeatable spatial structure across measurements (regardless of topographic organization) by testing for a positive correlation between the duration or period preferences estimated from odd and even scan runs. In converting the correlation coefficient to a probability of the observed correlation, we divided the number of surface vertices in the ROI by the upsampling factor between the acquired functional data and the cortical surface model in which response models were fit to account for the transformation from the scanned EPI resolution (1.7 mm^3) to the higher resolution (1.0 mm^3) cortical surface model space. This upsampling factor was 2.89 (i.e. 1.7^2) because our ROIs were restricted to a (folded) 2-dimensional cortical surface.

We also asked how the preferred duration and preferred period parameters of timing models were related. We used the same procedure to test for correlations between preferred event duration and preferred event period estimates, here using the data from all scan runs combined. Finally, we asked whether timing preferences were related to auditory frequency (tonotopic) preferences. We again used the same procedure to test for correlations between preferred event duration and preferred auditory frequency.

2.13.2. Relationships between preferred duration and response function extent

For each ROI we also looked for changes in the response functions' major and the minor extents with preferred duration, in data from the same ROI grouped across participants. To visualize these changes, we binned the recording sites within every 50 ms increase in preferred duration, calculating the mean and standard error of the response function extent. Our previous study on responses to visual event timing (Harvey et al., 2020) revealed that response function extent first increased with preferred duration, and second increased as the preferred duration moved away from the middle of the presented range. To test both progressions independently, it was necessary to include the same number of recording sites with preferred durations on either side of the middle of the presented range. Therefore, we first split the recording sites into those with preferred durations above and below the middle of the presented range. From the larger group, we repeatedly discarded a random selection so both groups had the same count, following a bootstrapping procedure. We then fit a general linear model to this combined data, with three predictors: preferred duration; the absolute difference between the preferred duration and the middle of the presented range; and a constant to capture the intercept of the progressions. We repeated this procedure in 1000 bootstrap permutations, each permutation discarding a different random selection. We took the mean t-statistic for each predictor across permutations, and converted this to a probability taking into account the (fixed) number of recording sites, corrected for upsampling as already described to give a probability for each relationship. We repeated this procedure for the bin means in Supplementary Fig. 4 to give the best fitting lines. We determined 95% confidence intervals by plotting fit lines from all permutations and finding the 2.5% and 97.5% percentiles of their values.

2.13.3. Analysis of differences between ROIs

In previous studies of visual timing-tuned (Harvey et al., 2020; Paul et al., 2022), numerosity-tuned (Harvey and Dumoulin, 2017a) and visual field position tuned responses (Amano et al., 2009; Dumoulin and Wandell, 2008; Harvey and Dumoulin, 2011), the parameters of the response functions change between ROIs, reflecting hierarchical transformations of quantity and spatial representations between brain areas. To test for similar hierarchical transformations between auditory timing ROIs, we compared several properties of the ROIs and their responses between ROIs. For each ROI in each hemisphere of each participant, we quantified the ROI's cortical surface area (Fig. 6A). We then quantified the mean of several properties across the recording sites within the ROI: model variance explained (Fig. 6B), preferred duration and period (Fig. 6C), the extent of the response function along its major and minor axes (Fig. 6F), and the aspect ratio of the response function (i.e. major axis extent / minor axis extent) (Fig. 6G), the orientation of the response function's major axis (Fig. 6F), and the compressive exponent on event frequency (Fig. 6I). Finally, we quantified the interquartile range of preferred durations (Fig. 6D) and preferred periods (Fig. 6E) of recording sites within each ROI.

We used linear mixed effects models to determine how the mean response model parameters differed between ROIs. These models included ROI as a fixed factor and participant as a random factor, because the quality of fMRI data varies between sessions and participants. Marginal tests for the fixed effects were adjusted using Satterthwaite degrees of freedom approximation (Satterthwaite, 1941, 1946). To determine which ROIs differed in response model parameters, we determined corrected post-hoc multiple comparisons using Tukey's honestly significant difference test (Tukey, 1949), which gives the marginal means and confidence intervals shown in Fig. 6.

3. Results

3.1. Regions of interest

We used 7T fMRI to record neural responses to auditory events with variable timing (Fig. 1). These were repetitive white noise bursts that gradually varied in the two timing properties: event duration and event period. Event duration describes the time from the onset to the offset of

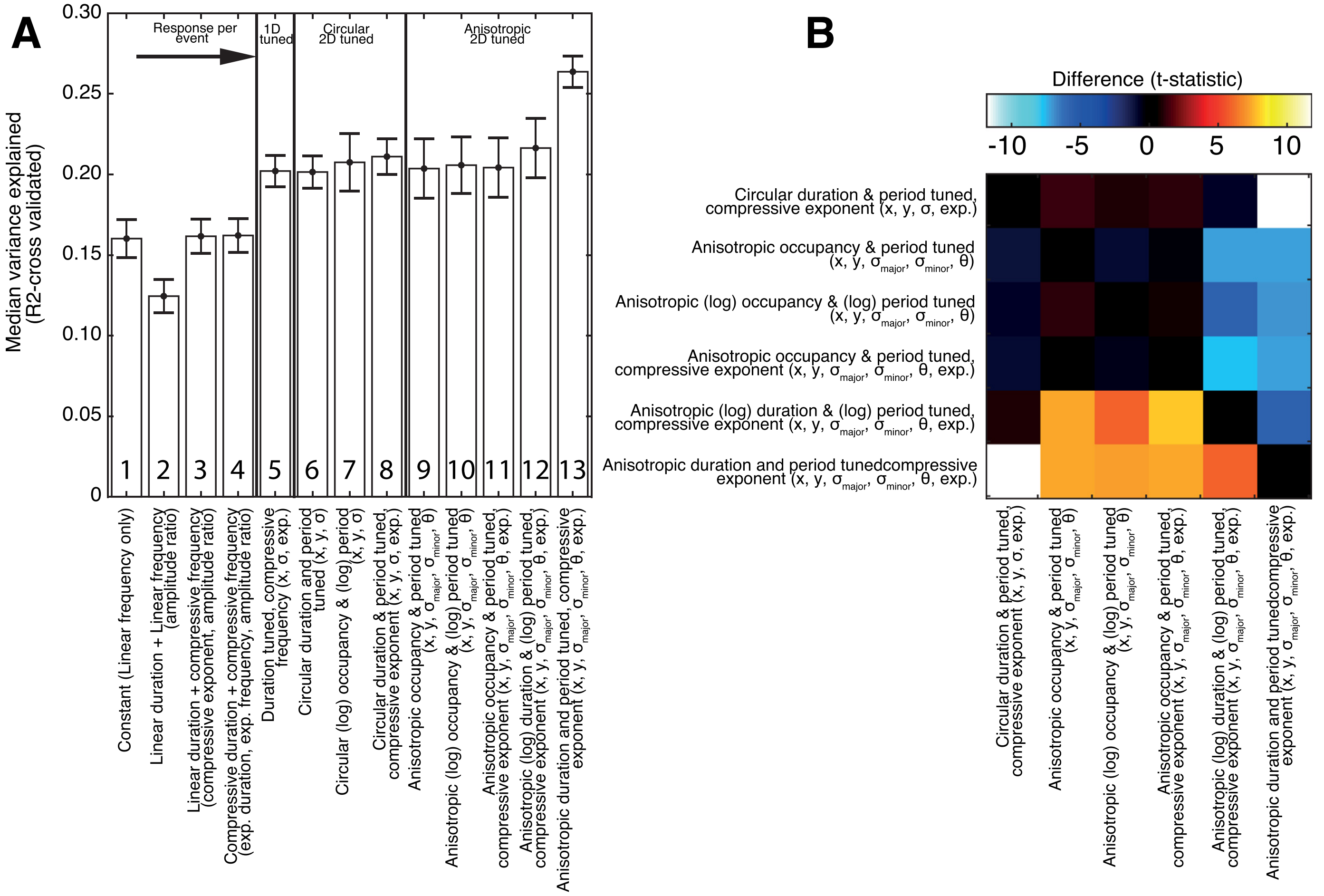

Fig. 3. Candidate timing response models and their fits. (A) We compared the ability of different response models to predict the responses observed in our ROIs. We fit the parameters of each model in one half of our scans and determined the proportion of the variance in the complementary half that the resulting model explained. This revealed that models tuned to event duration and period, with a nonlinear compressive accumulation of response amplitude with event frequency (furthest right) best captured the observed responses. Numbers in the bars can be used to refer to Table 1, which provides full details on each model. (B) T-statistics of paired comparisons between the variance explained in cross-validated data by the best fitting models.

an auditory event, while event period describes the time from the onset of one event to the onset of the next. In each participant, we use timing-dependent neural response models to capture responses in five regions of interest (ROIs) in areas of each hemisphere (see Methods, Fig. 2) that changed their responses with auditory event timing.

3.2. Model comparison

To determine how the neural populations in our ROIs encoded event timing, we fit neural response models of varying complexity and evaluated model fits in cross-validated data. We exclude responses with below 10% cross-validated variance explained in the best performing response model from all further analyses as these do not show clear response modulation by auditory event timing. The simplest monotonic neural response model, where response amplitude increases linearly with event frequency only, explained some response variance (mean over all participants, hemispheres, ROIs and data split halves: 16.0%) suggesting that the neural populations within our voxels give a distinct response to each event (Fig. 3A). The same model's performance on visual timing responses (Harvey et al., 2020) was very similar (17%). Adding a monotonic component where response amplitude also increase linearly with event duration counterintuitively decreases model fits (12.5%, t = -6.4, $p = 3.7 \times 10^{-9}$, n = 120) in a two-sided t-test paired across region of interest measurements), showing that including a response component describing a linear response to duration leads to overfitting. However, we would not necessarily expect response amplitude to increase linearly with either event duration or frequency, as previous studies of monotonic responses to visual event timing in early visual cortex (Hendrikx et al., 2022; Zhou et al., 2018) show that these responses accumulate sub-additively with duration and frequency. Adding a compressive nonlinearity to the responses to duration and frequency produces a response model that predicts responses more closely than any other monotonic response model (16.2%, t = 3.1, p = 0.002).

Beyond these monotonic increases with event duration and frequency, we tested how tuned, timing selective components affected model fits. Adding a Gaussian tuned response to duration and keeping a compressively increasing response to event frequency predicts responses better than the best monotonic model (20.2%, t = 10.0, $p < 10^{-10}$). However, adding a tuned response to two timing dimensions, following a circular-symmetric Gaussian function, predicts responses more closely than this one-dimensional Gaussian response model. Allowing duration and period tuning to have different extents and to interact with each other (i.e., an anisotropic Gaussian function) gives the best model we tested (26.4%, t = 5.74, $p = 8.2 \times 10^{-8}$ against the next best model). The same model's performance on visual timing responses (Harvey et al., 2020) was again similar (31.7%). However, it would also be possible to have a similar function tuned for different timing parameters (period occupancy or inter-stimulus interval) or logarithmic spacing of these timing parameters. These models have the same complexity as the best performing model but predict the observed responses significantly less well (all t >= 5.74, $p < 8.2 \times 10^{-8}$). Overall, neural response model performance here seems to depend on how closely the modelled response function can approximate this anisotropic function of duration and period.

The best performing neural response model captured timing selectivity with an anisotropic Gaussian neural response function (Supplementary Fig. 1). This predicts observed responses significantly more closely than all other models when treating all hemispheres, ROIs, and data splits as independent measures (n=120). However, this difference compared to the next best model (a similar anisotropic Gaussian function of log(duration) and log(period)) does not reach significance with statistical inference at the individual participant level (i.e. biological replication across individual humans), when all measures from one participant are grouped and participants are treated as the independent measure (n = 6, p = 0.129, t = 1.815), but does so on the hemisphere level where each hemisphere of each participant is taken as an individual measure (n = 12, p = 0.022, t = 2.671). We also repeated this analysis for each of our five ROIs separately and found the same neural response model best predicted the responses in every ROI. The best performing model's response function has several parameters: preferred duration, preferred period, tuning widths along major and minor axes, angulation of the major axis, and the compressive exponent on the increase in response amplitude with event frequency. In the following sections we therefore ask how these parameters change within and between ROIs.

3.3. Distribution of timing preferences within ROIs

Within each of our ROIs, we observed a range of timing preferences (Fig. 4 for an example participant, Supplementary Figs. 2 and 3 for all participants). These timing preferences were significantly positively correlated between repeated measures (split halves of odd and even scan runs) in 36 of 60 individual ROI examples for duration preferences, and in 34 of 60 ROI examples for period preferences (Fig. 5A for an example hemisphere, Supplementary Figs. 4 and 5 for all hemispheres). In both cases, the set of correlation coefficients was significantly greater than zero in a one-sample t-test (preferred duration: t = 11.71, $p < 10^{-10}$, n = 60; preferred period: t = 9.63, $p < 10^{-10}$). However, despite this repeatable variation of timing tuning properties between voxels, the spatial organization of timing preferences across the cortical surface did not show a clear structure, at least at the spatial resolution of our data (Supplementary Fig. 6). First, we tested for a linear progression of timing preferences across each ROI. We found no consistent correlation between preferred duration and distance across the ROIs. In some ROIs, there appeared to be changes back and forth from long to short preferred durations, so we also tested whether a sinusoidal function of cortical surface distance, or a combination of linear and sinusoidal functions, was consistently correlated with the ROIs duration preferences. Again, no such correlation was consistently found. As such, while there is a repeatable variation of timing preferences between voxels, the spatial structure of these tuning preferences could not be determined in our data.

We also tested whether preferred duration and preferred period estimates within each ROI were correlated with each other, here using models fit to all data rather than split halves. These response preferences were significantly positively correlated in 37 of 60 individual ROI examples (Fig. 5B for an example hemisphere, Supplementary Fig. 7 for all hemispheres). Again, the set of correlation coefficients was significantly greater than zero in a one-sample t-test (t = 10.56, $p < 10^{-10}$, n = 60).

We also tested for relationships between the extent of the response function and its preferred duration in each ROI, either following a linear function (consistent with the scalar property of duration perception (Gibbon et al., 1997; Harvey et al., 2020) or a v-shaped function with the smallest functions in the middle of the presented range (consistent with the regression towards the mean in duration perception (Harvey et al., 2020; Jazayeri and Shadlen, 2010)). Neither of these relationships consistently reached significance in any ROI (Supplementary Fig. 8).

Finally, we tested whether timing preferences were related to tonotopic auditory frequency preferences (Figs. 4 and 5C for an example participant, Supplementary Figs. 8 and 9 for all participants), and only found significant positive correlations in 12 of 60 individual ROI examples, with significant negative correlations in 1 of 60 individual ROI examples. Here, the set of correlation coefficients was marginally but significantly different from zero in a one-sample t-test (t = 2.11, p=0.04, n=60), perhaps suggesting some relationship between timing and auditory frequency preferences, although certainly a far weaker relationship than between repeated measures of timing preferences or between duration and period parameters.

3.4. Parameter changes between ROIs

We tested for differences in response model parameters between ROIs using linear mixed effects modeling (fixed factor: ROI, random factor: participant). The cortical surface areas of responsive voxel groups differed significantly between ROIs ($p = 4.38 \times 10^{-10}$, F(4, 50) = 14.90)(Fig. 6A). Specifically, post-hoc multiple comparisons demonstrated that Heschl's gyri (the location of the primary auditory cortices), and ATA and ATM (immediately neighboring the primary auditory cortices) were significantly larger than the other ROIs, and ATPM was significantly smaller than ATP. ATPM covered less than 50 mm^2 of the cortical surface on average, whereas all other ROI's on average covered 130 mm^2 or more.

The response variance explained by the best-fitting timing-tuned response model differed significantly between ROIs ($p = 3.92 \times 10^{-5}$, F(4, 50) = 8.13)(Fig. 6B). Specifically, post-hoc multiple comparisons demonstrated that model fits were significantly better in Heschl's gyri, ATA and ATM than in ATP and ATPM. A similar pattern was observed for the response variance explained by auditory frequency-tuned (tonotopic mapping) response models ($p < 10^{-10}$, F(4, 50) = 45.63 in a similar mixed effects model), though auditory frequency-tuned response models fit significantly better in Heschl's gyri than ATA and ATM (Fig. 6J). As such, auditory frequency responses are clearest in Heschl's gyri, while timing-tuned responses are clearest in the auditory cortex belt (ATA and ATM).

Mean preferred duration did not change significantly between ROIs (p=0.269, F(4, 50)=1.34) (Fig. 6C). The interquartile range of preferred durations present in each ROI also did not differ significantly between ROIs (p = 0.117, F(4, 50) = 1.95) (Fig. 6D).

Mean preferred period did not differ significantly between ROIs (Fig. 6C). However, the interquartile range of preferred periods in each ROI did differ significantly between ROIs ($p < 2.77 \times 10^{-6}$, F(4, 50)=10.57) (Fig. 6E). Specifically, ATP's interquartile range was significantly higher than in other ROIs.

Quite unlike previous findings for visual event timing-tuned responses (Harvey et al., 2020) the response function's extent along both its major and minor axes did not change significantly between ROIs (Fig. 6F) and as a result the aspect ratio of this response function did not change significantly (Fig. 6G). The angulation of the response function parameter did not change significantly between ROIs either (Fig. 6H). Furthermore, the response model's compressive exponent parameter, which captures the subadditive accumulation of responses to repeated events, did not change significantly between ROIs (Fig. 6I). Therefore, unlike in the hierarchical processing of visual event timing, there is no evidence of a progressive sharpening of response functions or integration of responses to repeated events in the auditory modality.

4. Discussion

The current study aimed to identify regions where neural response show tuned responses to the timing of auditory events and how the parameters of this response function change within and between these regions. We used a combination of ultra-high field fMRI of responses to

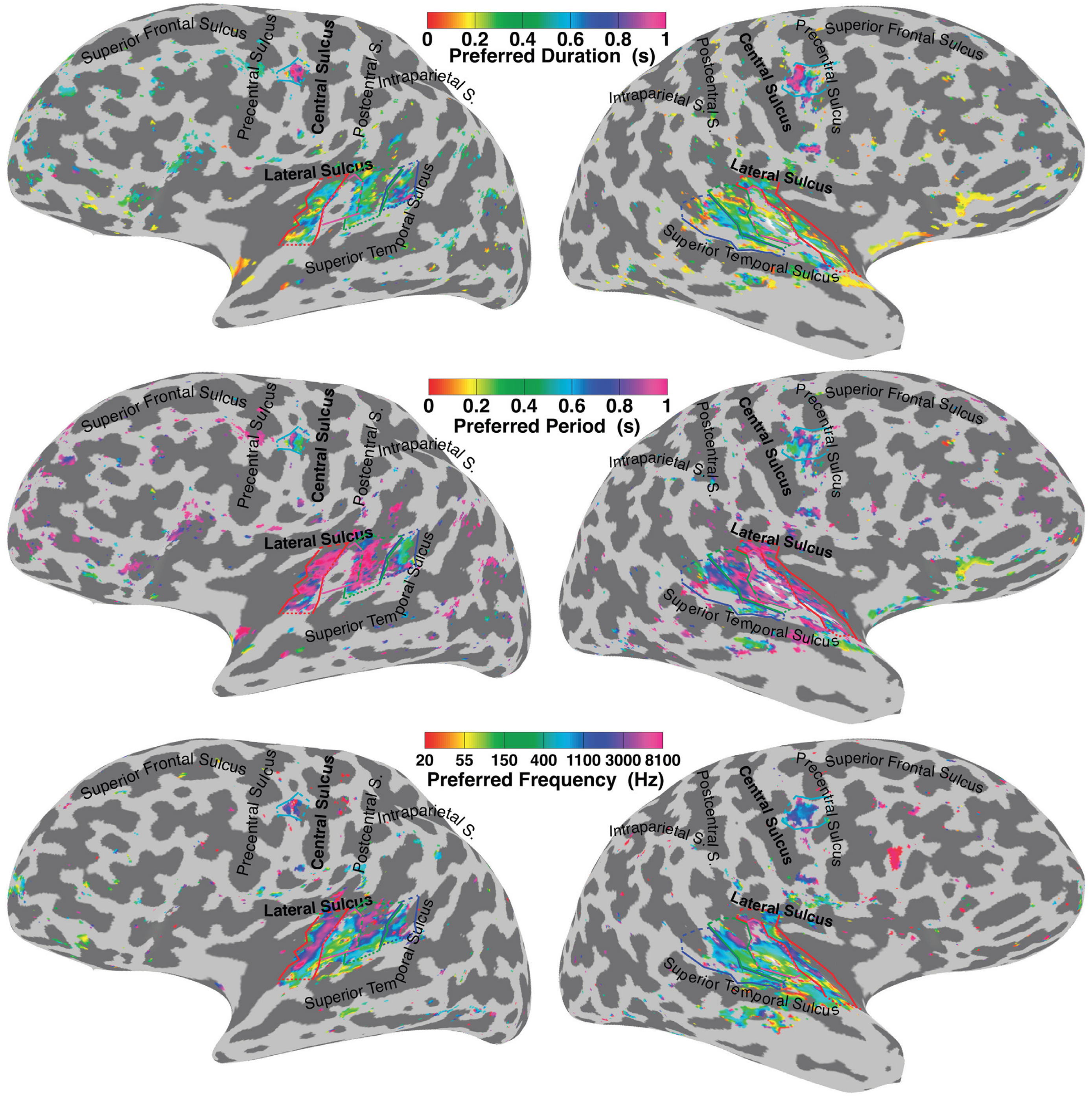

Fig. 4. Best fitting response function parameters projected onto the inflated cortical surface of one representative participant. (A). Preferred auditory event duration from the best fitting two-dimensional anisotropic gaussian response model. (B). Preferred auditory event period. (C). Preferred auditory frequency from the tonotopic mapping stimulus response.

auditory events that changed in timing, and neural model-based analyses of the resulting responses. We found regions showing tuned responses to auditory timing in and around Heschl's gyri (consistent with the locations of the primary auditory cortices and macaque lateral and medial auditory cortex belt) and extending into the superior temporal gyrus (consistent with Wernicke's area and the macaque parabelt), as well as a small responsive region in the premotor cortex. Comparison between candidate tuned and monotonic neural response models demonstrated that the responses of these regions were best captured by a two-dimensional anisotropic gaussian response function, tuned to event duration and period, with response amplitudes accumulating subadditively with increasing event frequency. We found a repeatable variation of timing preferences within each region, indicating a spatial separation on the cortical surface of neural populations with preferences for different event timings. However, there was no clear structure to the cortical organization of the timing preferences within these regions, and no relationship between tuning function extents and preferred durations. We also found a consistent relationship between response preferences for event duration and period, but at most a very weak relationship between timing preferences and tonotopic auditory frequency preferences.

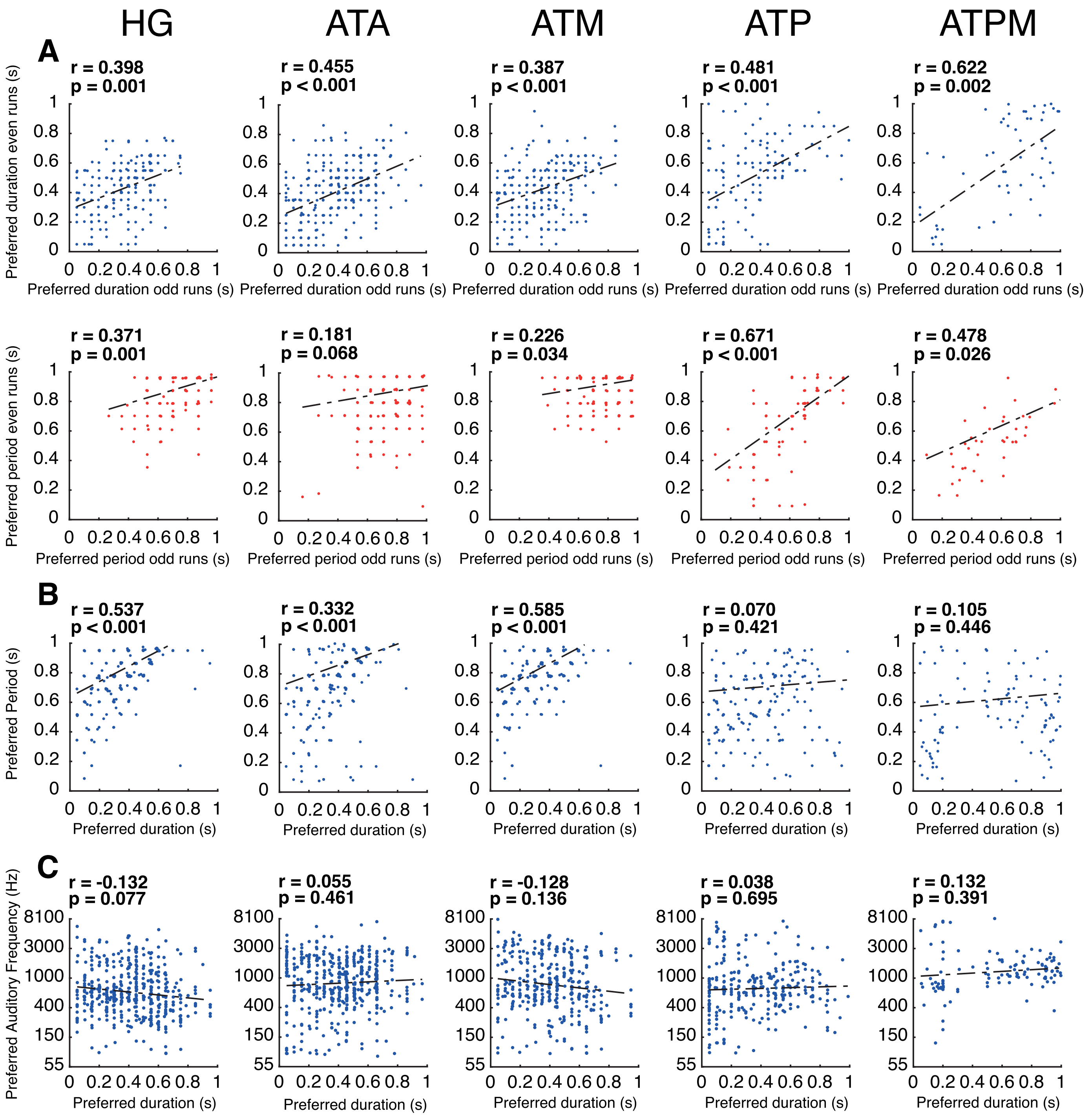

Fig. 5. Correlations between response model parameters in an example right hemisphere. (A) Correlations between estimates of preferred duration or preferred period from independent sets of scan runs, for the set of voxels within each ROI of an example hemisphere. Response model parameters were consistently positively correlated between repeated measures. (B) Correlations between preferred duration and preferred period estimates (from all scan runs), which were also consistently positively correlated. (C) Correlations between preferred duration and preferred auditory frequency estimates (from all scan runs), which were not consistently correlated.

Responses differed between these regions: those immediately adjacent to the primary auditory cortices were larger and gave clearer responses, while tonotopic responses to auditory frequency were clearest in the primary auditory cortices themselves.

Perhaps the most striking feature in the current dataset is that the neural response function whose predictions best fit the observed data is the same neural response function reported in our previous study of responses to visual event duration (Harvey et al., 2020). This is notable given that the encoding of visual information in early visual brain areas and auditory information in the auditory cortex differ in several ways: responding to positions on the retina and the cochlea respectively; encoding position and auditory frequency respectively; and being located in very different brain areas. Despite these differences, it appears that the encoding temporal information is very similar between modalities, suggesting closely related computational mechanisms. As in the visual cortex (Hendrikx et al., 2022), this encoding may be derived from ear-

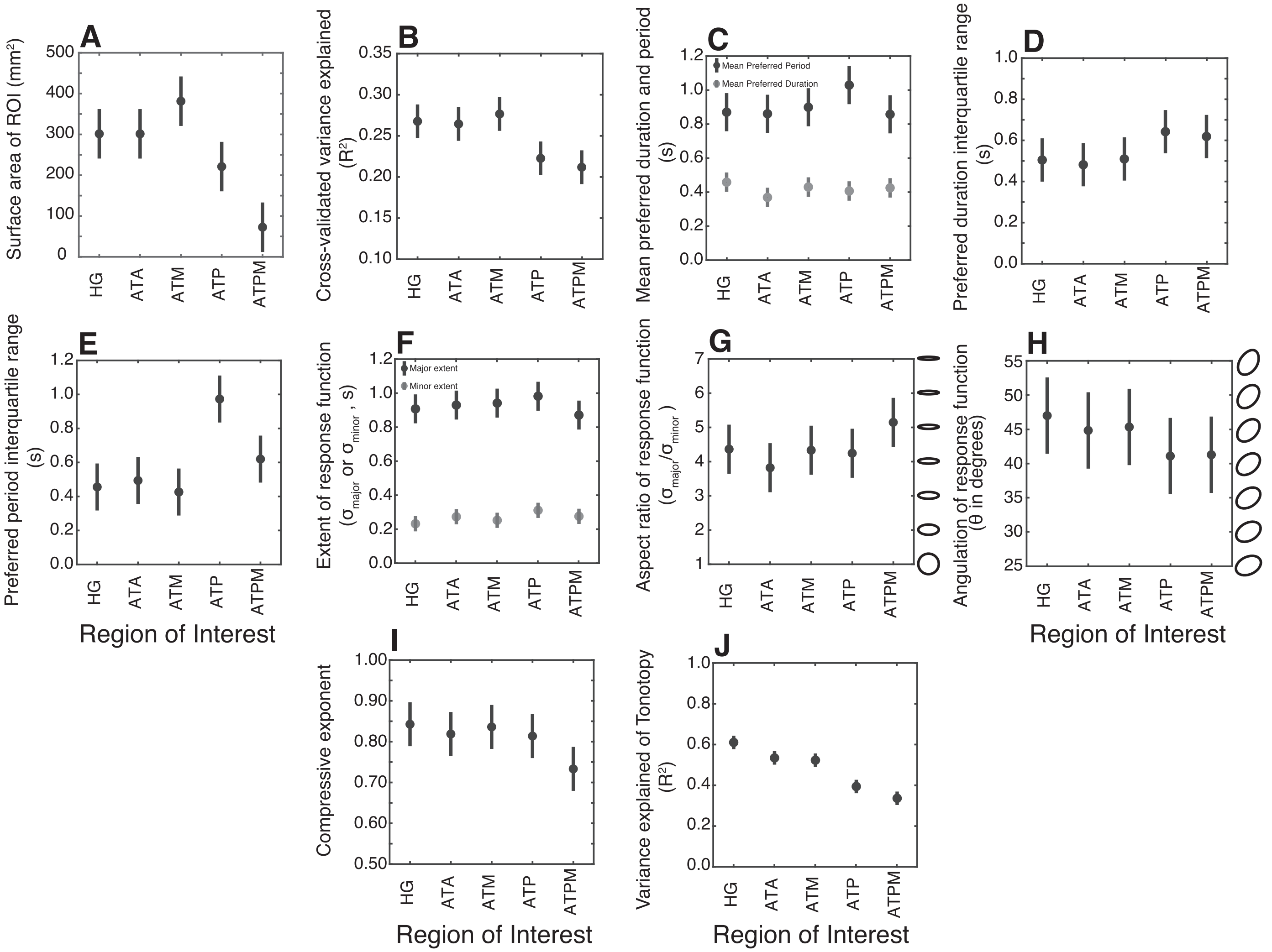

Fig. 6. Differences in response model parameters between ROIs. Points represent the population marginal mean of values across hemipsheres. Many of these parameters show significant differences between ROIs, typically between the ROIs immediately surrounding Heschl's gyri (ATA and ATM) and those further away (ATP and ATPM). Error bars are 95% confidence intervals: separable error bars show significant differences at $p<0.05$. (A) Surface area of ROI. (B) Cross-validated variance explained of the 2D anisotropic gaussian model tuned to duration and period, expressed in R^2. (C) Mean preferred duration (lower bars in grey) and period (upper bars in black). (D) Interquartile range of preferred durations. (E) Interquartile range of preferred periods. (F) Extent of response function. Extent along the minor axis are the lower bars shown in grey, extent along the major axis are the upper bars shown in black. (G) Aspect ratio of the tuning function, measured in extent across major axis over extent across minor axis. Circles on the right of G show the change in aspect ratio going from a round, isotropic response function on the bottom, to a response function that is more sensitive along the minor compared to the major axis on the top. (H) Angulation of response function relative to the x-axis in the duration-period domain. The circles on the right of figure H show differing ranges of response function angulation, going from a nearly horizontal angulation on the bottom to a near vertical angulation at the top. (I) Compressive exponent on frequency. (J) Variance explained of the tonotopy models in the same ROIs as all other figures.

lier responses that monotonically increase with event duration and frequency. In both cases, the earlier monotonic response amplitudes are also likely to increase with stimulus strength (visual contrast or auditory loudness) which must be normalized in the computation of true timing-tuned responses. Visual contrast normalization between primary and extrastriate visual areas is well established (Aqil et al., 2021; Carandini and Heeger, 2011; Kastner and Pinsk, 2004) as is loudness normalization in primary auditory cortices (Behler and Uppenkamp, 2016; Neuner et al., 2014), suggesting that event timing can be straightforwardly separated from other stimulus features that affect global response amplitudes.

How might this timing tuned representation affect how the brain analyses event timing? First, a timing-tuned representation makes explicit in neural responses the timing information that is implicitly present in the response dynamics of early sensory areas (Hendrikx et al., 2022). We speculate that this might first allow timing information to be further analyzed in a hierarchy of timing-driven responses and allow the timing of responses in different sensory modalities to be compared and grouped. An explicit representation of timing may allow the same responses to event timing regardless of auditory frequency, while implicit encoding of timing in neural response dynamics separates timing information into different auditory frequencies. A timing-tuned representation also means different neural populations respond to different event timings. Auditory events with different timings often have different roles in human behavior, such as speech and music, and processing these different event timings with distinct neural populations may allow neural populations to specialize in one type of auditory information (Norman-Haignere et al., 2015). Even within these types of auditory information, timing differs between musical genres and speech emphasis (for example), potentially producing distinct neural responses in each case (Nakai et al., 2021).

Despite the similarities between visual and auditory timing-tuned responses, we found a highly modality-specific representation of temporal information. We found no region that was part of both visual and auditory networks of timing-tuned responses. Unlike early visual areas which show monotonic increases in response amplitude with event duration and rate (Hendrikx et al., 2022; Stigliani et al., 2017; Zhou et al., 2018), primary auditory cortices already showed evidence of timing-tuned responses. We therefore speculate that the earlier monotonic responses to auditory event duration and frequency may be present in subcortical auditory nuclei such as the cochlea nuclei, superior olivary complex, inferior colliculus and medial geniculate nucleus of the thalamus. These regions process incoming auditory information prior to Heschl's gyrus and might therefore be the analogue to the primary visual cortex in terms of monotonic versus tuned processing of incoming event timing information. However, the current field of view and scanning sequence type (Moeller et al., 2010; Moerel et al., 2015) did not allow for us to study the subcortical nuclei in detail and we know of no previous studies of how subcortical responses are affected by event timing. We found auditory timing-tuned responses in the primary auditory cortices and in surrounding areas, which appear to correspond to the auditory cortex belt (ATA and ATM) and parabelt of the macaque, apparently extending into Wernicke's area in the superior temporal gyrus (ATP). Therefore, as for visual event timing, the auditory timing network consisted of several areas that are well established to be involved in auditory processing. The presence of timing-tuned responses throughout the early auditory cortices may reflect the vital role of timing in interpreting auditory inputs. This may underlie co-localized (but independent) responses to both timing and auditory frequency. However, we find the strongest responses to auditory frequency in the primary auditory cortices (Heschl's gyrus) and the strongest responses to timing also extend into the immediately surrounding areas, ATA and ATM. This implies that responses to timing may be more preserved into later auditory processing stages than responses to auditory frequency, though it is too soon to draw strong conclusions on this point. We also found similar responses in a small area of the premotor cortex (ATPM), which also shows auditory frequency-tuned (tonotopic) responses in our data, though we can find no previous reports of auditory-driven responses here.

ATPM is close to a right-hemisphere premotor area that has previously been shown to hold information about visual event duration in group data (Hayashi et al., 2018), though where we have used the same individual participants to map visual timing preferences (Harvey et al., 2020), the nearby visual timing map TFS is consistently anterior to ATPM. Similarly, the visual timing map TLS is consistently posterior to ATP, though again they are nearby. As such, while we find no evidence of generalized responses to timing of events in different modalities, there are tuned responses to visual and auditory event timing in nearby areas. This may allow these distinct responses to interact (Tsouli et al., 2022) particularly in processing multimodal stimuli where the timing of events in different senses is perceptually grouped (Klink et al., 2011). So while the computational mechanisms of timing processing appear similar between modalities, these are implemented in different brain areas with different inputs. This speaks against a single, abstracted, modality independent neural representation of event timing, which is consistent with the modality-specific nature of duration adaptation effects (Heron et al., 2012), for example. But the similar encoding mechanisms within these distinct representations is consistent with the many similarities between auditory and visual timing perception (Barne et al., 2018; Zélanti and Droit-Volet, 2011).

Despite the similarities between vision and audition in the computation and representation of timing, there are also differences between the changes in response function parameters both within and between brain areas. While there is a repeatable spatial structure to both auditory and visual timing preferences within each region we examine, we find no clear evidence of topographic organization of auditory timing preferences. This contrasts with the clear topographic organization of visual event timing preferences, which we showed using very similar data quality and analyses. This may be because the brain areas showing auditory timing-tuned responses are considerably smaller than those showing visual timing-tuned responses: for example, the tonotopic maps of the auditory cortices are far smaller than the visual field maps. If the areas we observe surrounding the primary auditory cortices do indeed correspond to the histologically-defined auditory cortex belt and parabelt, these regions would be very small, around 3×5 mm each (Kaas and Hackett, 2000; Moerel et al., 2014). Our fMRI protocols are not designed to reveal the spatial structure of response preferences at such fine scales: our voxels are 1.6 mm across, allowing only two or three voxels across each subdivision of the belt. Therefore, if topographic organization of timing preferences across a belt subdivision is present, we may be unable to resolve it.

In visual timing maps, we have also found clear relationships between timing preferences and response function extents (i.e. tuning widths), specifically that response functions are smallest near the middle of the presented timing range (Harvey et al., 2020). We interpret this as a potential neural basis of Vierordt's law (von Vierordt, 1866), the attraction of perceived event timing towards the middle of the presented range (Jazayeri and Shadlen, 2010), which is also found (though weaker) in auditory timing perception (Murai and Yotsumoto, 2016). Conversely, we find no relationship between auditory timing preferences and response function extents. This could be because each voxel samples a neural population with a wider range of timing preferences. However, this would predict broader tuning functions for auditory than visual timing-selective populations, while in fact we observe smaller tuning functions for auditory than visual timing. Therefore, we propose that auditory timing-tuned populations may represent all event timings as finely as possible, to give the most accurate possible representation of the fine-scale temporal structure of sounds. This is consistent with a smaller Weber fraction for distinguishing auditory event timing than visual event timing, and also with the weaker attraction of perceived auditory timing towards the middle of the presented range (Murai and Yotsumoto, 2016).

We also found changes in the parameters of timing selective response between visual timing-tuned areas, suggesting a hierarchical processing of visual event timing. The most striking of these changes is a progressive integration of responses to multiple visual events, shown by a decrease in our response model's compressive exponent parameter from about 0.8 (almost independent responses to each event) to about 0.2 (little effect of event frequency). We do not see this at all in auditory event timing-tuned areas: the compressive exponent is around 0.85 in all areas, dropping slightly but not significantly in ATMP. This suggests distinct neural responses to different auditory events, with very little interaction of these responses even at high event frequencies. Again, we propose that this may maintain the most accurate and complete representation possible of the fine-scale temporal structure of sounds, with the possible exception of premotor areas. Therefore, many of these differences between auditory and visual timing-tuned responses are consistent with differences in auditory and visual perception, although our interpretation of the potential roles of these neural response differences in perception remains speculative.

The main spatial structure of auditory responses in the cortex is tonotopic, following auditory frequency and so mapping the spatial structure of the cochlea. Auditory frequency describes the sinusoidal vibration of a sound wave in the range of around 20 Hz to 20,000 Hz, which humans perceive as pitch or tone. Our event timing mapping stimuli change over time, but at a much lower rate from 0.5 to 10 Hz, i.e., with a period from 2 to 0.1 seconds. Furthermore, these stimuli are not sinusoidally modulated tones, but are white noise bursts changing between silence and full intensity (and back to silence) using a raised-cosine gate over 10% of each event's duration. As white noise bursts, these stimuli contain all auditory frequencies at the same intensity. As such, there is no relationship between event timing and auditory frequency in our timing mapping stimuli.

We also used a tonotopic mapping procedure (Da Costa et al., 2011) to examine relationships between neural responses to auditory frequency and event timing in our ROIs. We found at most a very weak relationship between auditory frequency and auditory event timing preferences. Given that repeated measures of auditory event timing preferences are well correlated, there are two potential interpretations of this lack of relationship between these preferences, which are not mutually exclusive. First, the spatial scale of changes in auditory event timing preferences is finer than the spatial scale of auditory frequency preferences. Therefore, for a given auditory frequency there would be voxels with a broad range of timing preferences. Second, the spatial progressions of auditory event frequency and auditory event timing may be in orthogonal directions. The idea of a cortical progression of temporal structure preferences as an orthogonal dimension to tonotopy in the auditory cortex has been proposed before (Barton et al., 2012). The temporal structure in that study differed from ours in that their stimulus amplitude was sinusoidally modulated, rather than changing rapidly between silence and full intensity. Such a sinusoidally modulated stimulus does not have events of meaningful durations, but rather an ongoing stimulus described by its temporal frequency.

These independent responses to auditory frequency and event timing are also in the large neural populations of fMRI voxels. It is possible that different sub-populations of neurons respond to auditory frequency and event timing, or that single neurons are tuned to both features with no relationship between the response preferences in these two dimensions (Hofstetter et al., 2021; Tsouli et al., 2022)

Although perception of sound's durations and rates are often studied separately, our results show correlated and interacting selectivity for event duration and period (a measure of event rate). This suggests that event duration- and rate-selective neural responses may be estimated by a single mechanism (Hartcher-O'Brien et al., 2016). Any mechanism that captures the time between transient neural responses (which occur at both stimulus onset and offset) could give identical responses to duration and period in our stimuli fMRI scanning is very loud, which may affect the interpretation of any auditory fMRI study. Our fMRI sequence produced audible gradients every 79 ms. To dampen this noise and keep the stimulus clearly audible, participants' ears were plugged and covered with the headphones that produced our stimuli. Although the scanner noise remained audible, it should not directly affect our estimates of auditory timing preferences because the scanner noise never changed during a scanning run and the resulting hemodynamic response to the scanner noise was at steady state (ongoing for 22.5 seconds) before acquiring the data we modelled. As such, responses to the scanner noise added a constant, unchanging component to the resulting signal. Nevertheless, neural response adaptation to the timing of the scanner noise may indirectly affect responses to our stimuli. Repeated presentation of sounds with fixed timing repels the perceived timing of subsequent auditory events (Huppert and Singer, 1967; Walker et al., 1981). For other quantities like visual event timing (Hayashi et al., 2015) and numerosity (Piazza et al., 2004), such repetition suppresses responses to similar stimuli and affects neural response functions (Tsouli et al., 2021), likely underlying perceptual adaptation effects (Tsouli et al., 2022). We would therefore expect scanner noise with a period of 79 ms to suppress responses to stimuli with high event frequencies (our shortest event period was 100 ms), and thereby potentially increase the preferred event period that produced the largest responses. Assuming these adaptation effects are limited to the adapted sensory modality (Heron et al., 2012; Walker et al., 1981), adaptation to auditory event period may explain why the mean preferred auditory event periods we observed (874 ms) were far longer than the mean preferred visual event periods previously described (556 ms), while the mean preferred auditory and visual event durations were very similar (auditory: 437 ms; visual 452 ms).

While our experimental design uses a relatively small number of participants, the large amount of data collected in each participant allows a detailed analysis of functional neural response properties with high statistical confidence (Baker et al., 2021; Gordon et al., 2017). Given that each participant shows a five spatially separate responsive regions in each hemisphere, and we use multiple measures of these responses for cross validation, this allows high statistical confidence in the best response model and its parameters. However, our analyses of differences in response model properties between responsive regions include participant as a random factor, which consistently reaches significance. This relatively small sample of young, healthy adults does not allow us to meaningfully characterize these differences between participants. These differences may result from trivial factors that can lead to large individual differences in fMRI data (participant motion, hemodynamic response variations, brain size, position in the coil), but might instead be related to differences in perception (Tsouli et al., 2022), task performance and attention that we have not yet investigated.

Our results suggest interesting directions for future work. First, participant groups including aging participants (Silva et al., 2021), children (Gomez et al., 2018) still learning language or music, or a range of perceptual abilities (Song et al., 2015) may clarify the importance of these responses for perception, development and aging. However, given the location of timing tuned responses in and around the auditory cortex, showing causal involvement in auditory timing perception by disrupting these areas with TMS may be unfeasible: this would likely disrupt auditory processing globally, rather than timing representations specifically.

Furthermore, we measure responses to event timing with broadband auditory frequency (white noise) stimuli, and responses to auditory frequency with a range of event timings, in different experiments. Although the two responses were found together in the large neural populations of fMRI voxels, it therefore remains unclear whether timing-tuned responses depend on specific frequencies and vice-versa. Future experiments could vary timing and frequency in one experiment to determine the spectro-temporal receptive field of each voxel. This would determine whether responses to timing are abstracted across all frequencies or depend on the voxel's preferred frequency being present in the stimuli.

Even using ultra-high field (7T) fMRI and fairly high scan resolution (1.7 mm^2) we find no clear spatial structure of timing preferences. However, the subdivisions of the auditory cortex belt and parabelt (where we see the clearest responses) are very small, around 3×5 mm each (Kaas and Hackett, 2000; Moerel et al., 2014), so their internal spatial structure would not be detectable at these resolutions. Even higher field strengths and scan resolutions, or dense invasive recordings, may reveal such structure.

We also omit subcortical structures for early auditory processing (cochlea nuclei, superior olivary complex, inferior colliculus and medial geniculate nucleus of the thalamus) and temporal processing (basal ganglia and cerebellum). The responses of these structures may provide a more complete picture of how auditory timing-tuned responses are derived (Hendrikx et al., 2022) and whether these are eventually integrated with visual timing-timed responses.

Overall, auditory timing selectivity resembles responses to visual timing and other quantities, but is focused in the auditory cortices. This suggests that the human brain performs similar analyses of the temporal structure of its inputs from different sensory modalities. We propose that this similar processing underlies similarities between timing perception in different modalities, without the need for responses to different modalities to converge onto shared neural populations. In contrast to the hierarchical and integrative processing of visual event timing, auditory timing responses have very similar parameters between brain areas, suggesting it may be beneficial for the brain to keep a specific response to each auditory event. Most of the brain areas involved are tightly grouped in and around Heschl's gyri and may correspond to the primary auditory cortex and its belt and parabelt regions described in macaques. While there is a repeatable spatial structure to timing preferences, we did not find evidence of topographic organization of timing preferences at the spatial resolution of our data. As the observed spatial structure was not strongly related to tonotopic auditory frequency preferences, the fine spatial structure of auditory timing preferences may allow functional subdivisions of these auditory association cortices to be determined. The unexpected finding of both auditory frequency and event timing-selective responses in a small area of the ventral caudal premotor cortex may reflect a route by which responses to auditory events could guide motor responses to auditory stimuli (Graziano et al., 1999).

Data and code availability statement

Ethical constraints prevent us from sharing the medical imaging data sets (MRI scans) generated in the current study to public repositories. These raw data sets are available from the corresponding author upon reasonable request, depending on agreements not to share these data publicly. Model parameters underlying all statistical analyses and response data for all model fitting are publicly available at the following DOIs:

- Timeseries data: https://doi.org/10.6084/m9.figshare.19849879
- Response Model parameters: https://doi.org/10.6084/m9.figshare.19849264

The code that supports the findings of this study is available from the following repositories:

- vistasoft (https://github.com/vistalab/vistasoft)
- vistasoftAddOns (https://github.com/benharvey/vistasoftAddOns)
- fMRI_preproc (https://github.com/MvaOosterhuis/fMRI_preproc)
- AnalysisScript (https://github.com/MvaOosterhuis/AuditoryTiming_Analysis)

Declaration of Competing Interests

The authors declare no competing interests.

Credit authorship contribution statement

Martijn van Ackooij: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Jacob M. Paul:** Methodology, Formal analysis, Investigation, Writing – review & editing. **Wietske van der Zwaag:** Methodology, Writing – review & editing. **Nathan van der Stoep:** Writing – review & editing, Supervision. **Ben M. Harvey:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Acknowledgments

This work was supported by the Netherlands Organization for Scientific Research (452.17.012 to B.M.H. and TTW_VI.Vidi.198.016 to W.vdZ.).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119366.

References

Akeroyd, M.A. (2001). Binaural SOUND creation toolbox for MATLAB. 28.

Allen, E.J., Burton, P.C., Olman, C.A., Oxenham, A.J., 2017. Representations of pitch and timbre variation in human auditory cortex. J. Neurosci. 37 (5), 1284–1293. doi:10.1523/JNEUROSCI.2336-16.2016.

Amano, K., Wandell, B.A., Dumoulin, S.O., 2009. Visual field maps, population receptive field sizes, and visual field coverage in the human MT+ complex. J. Neurophysiol. 102 (5), 2704–2718. doi:10.1152/jn.00102.2009.

Andersson, J.L.R., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage 20 (2), 870–888. doi:10.1016/S1053-8119(03)00336-7.

Aqil, M., Knapen, T., Dumoulin, S.O., 2021. Divisive normalization unifies disparate response signatures throughout the human visual hierarchy. Proc. Nat. Acad. Sci. U.S.A. 118 (46), e2108713118. doi:10.1073/pnas.2108713118.

Arcaro, M.J., Pinsk, M.A., Li, X., Kastner, S., 2011. Visuotopic organization of macaque posterior parietal cortex: a functional magnetic resonance imaging study. J. Neurosci. 31 (6), 2064. doi:10.1523/JNEUROSCI.3334-10.2011.

Assmann, P.F., Summerfield, Q., 1990. Modeling the perception of concurrent vowels: vowels with different fundamental frequencies. J. Acoust. Soc. Am. 88 (2), 680–697. doi:10.1121/1.399772.

Baker, D.H., Vilidaite, G., Lygo, F.A., Smith, A.K., Flack, T.R., Gouws, A.D., Andrews, T.J., 2021. Power contours: optimising sample size and precision in experimental psychology and human neuroscience. Psychol. Methods 26 (3), 295–314. doi:10.1037/met0000337.

Barakat, B., Seitz, A.R., Shams, L., 2015. Visual rhythm perception improves through auditory but not visual training. Curr. Biol. 25 (2), R60–R61. doi:10.1016/j.cub.2014.12.011.

Barne, L.C., Sato, J.R., de Camargo, R.Y., Claessens, P.M.E., Caetano, M.S., Cravo, A.M., 2018. A common representation of time across visual and auditory modalities. Neuropsychologia 119, 223–232. doi:10.1016/j.neuropsychologia.2018.08.014.

Barton, B., Venezia, J.H., Saberi, K., Hickok, G., Brewer, A.A., 2012. Orthogonal acoustic dimensions define auditory field maps in human cortex. Proc. Natl. Acad. Sci. 109 (50), 20738–20743. doi:10.1073/pnas.1213381109.

Behler, O., Uppenkamp, S., 2016. The representation of level and loudness in the central auditory system for unilateral stimulation. Neuroimage doi:10.1016/j.neuroimage.2016.06.025.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57 (1), 289–300.

Bridwell, D.A., Leslie, E., McCoy, D.Q., Plis, S.M., Calhoun, V.D., 2017. Cortical sensitivity to guitar note patterns: EEG entrainment to repetition and key. Front. Hum. Neurosci. 11, 90. doi:10.3389/fnhum.2017.00090.

Buonomano, D.V., Merzenich, M.M., 1995. Temporal information transformed into a spatial code by a neural network with realistic properties. Science 267 (5200), 1028–1030. doi:10.1126/science.7863330.

Carandini, M., Heeger, D.J., 2011. Normalization as a canonical neural computation. Nat. Rev. Neurosci. 13 (1), 51–62. doi:10.1038/nrn3136.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29 (3), 162–173. doi:10.1006/cbmr.1996.0014.

Da Costa, S., van der Zwaag, W., Marques, J.P., Frackowiak, R.S.J., Clarke, S., Saenz, M., 2011. Human primary auditory cortex follows the shape of Heschl's Gyrus. J. Neurosci. 31 (40), 14067–14075. doi:10.1523/JNEUROSCI.2000-11.2011.

Ding, N., Chatterjee, M., Simon, J.Z., 2014. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. Neuroimage 88, 41–46. doi:10.1016/j.neuroimage.2013.10.054.

Ding, N., Simon, J.Z., 2014. Cortical entrainment to continuous speech: functional roles and interpretations. Front. Hum. Neurosci. 8. doi:10.3389/fnhum.2014.00311.

Dumoulin, S.O., Wandell, B.A., 2008. Population receptive field estimates in human visual cortex. Neuroimage 39 (2), 647–660. doi:10.1016/j.neuroimage.2007.09.034.

Elhilali, M., Xiang, J., Shamma, S.A., Simon, J.Z., 2009. Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. PLoS Biol. 7 (6), e1000129. doi:10.1371/journal.pbio.1000129.

Escoffier, N., Herrmann, C.S., Schirmer, A., 2015. Auditory rhythms entrain visual processes in the human brain: evidence from evoked oscillations and event-related potentials. Neuroimage 111, 267–276. doi:10.1016/j.neuroimage.2015.02.024.

Fries, W., Swihart, A.A., 1990. Disturbance of rhythm sense following right hemisphere damage. Neuropsychologia 28 (12), 1317–1323. doi:10.1016/0028-3932(90)90047-R.

Gibbon, J., Malapani, C., Dale, C.L., Gallistel, C.R., 1997. Toward a neurobiology of temporal cognition: advances and challenges. Curr. Opin. Neurobiol. 7 (2), 170–184. doi:10.1016/S0959-4388(97)80005-0.

Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage 9 (4), 416–429. doi:10.1006/nimg.1998.0419.

Gomez, J., Natu, V., Jeska, B., Barnett, M., Grill-Spector, K., 2018. Development differentially sculpts receptive fields across early and high-level human visual cortex. Nat. Commun. 9 (1), 788. doi:10.1038/s41467-018-03166-3.

Gordon, E.M., Laumann, T.O., Gilmore, A.W., Newbold, D.J., Greene, D.J., Berg, J.J., Ortega, M., Hoyt-Drazen, C., Gratton, C., Sun, H., Hampton, J.M., Coalson, R.S., Nguyen, A.L., McDermott, K.B., Shimony, J.S., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., Nelson, S.M., Dosenbach, N.U.F., 2017. Precision functional mapping of individual human brains. Neuron 95 (4), 791–807. doi:10.1016/j.neuron.2017.07.011, e7.

Grahn, J.A., Henry, M.J., McAuley, J.D., 2011. FMRI investigation of cross-modal interactions in beat perception: audition primes vision, but not vice versa. Neuroimage 54 (2), 1231–1243. doi:10.1016/j.neuroimage.2010.09.033.

Graziano, M.S.A., Reiss, L.A.J., Gross, C.G., 1999. A neuronal representation of the location of nearby sounds. Nature 397 (6718), 428–430. doi:10.1038/17115.

Hartcher-O'Brien, J., Brighouse, C., Levitan, C.A., 2016. A single mechanism account of duration and rate processing via the pacemaker–accumulator and beat frequency models. Curr. Opin. Behav. Sci. 8, 268–275. doi:10.1016/j.cobeha.2016.02.026.

Harvey, B.M., Dumoulin, S.O., 2011. The relationship between cortical magnification factor and population receptive field size in human visual cortex: constancies in cortical architecture. J. Neurosci. 31 (38), 13604–13612. doi:10.1523/JNEUROSCI.2572-11.2011.

Harvey, B.M., Dumoulin, S.O., 2017a. A network of topographic numerosity maps in human association cortex. Nat. Hum. Behav. 1 (2), 0036. doi:10.1038/s41562-016-0036.

Harvey, B.M., Dumoulin, S.O., 2017b. Can responses to basic non-numerical visual features explain neural numerosity responses? Neuroimage 149, 200–209. doi:10.1016/j.neuroimage.2017.02.012.

Harvey, B.M., Dumoulin, S.O., Fracasso, A., Paul, J.M., 2020. A network of topographic maps in human association cortex hierarchically transforms visual timing-selective responses. Curr. Biol. 30 (8), 1424–1434. doi:10.1016/j.cub.2020.01.090, e6.

Harvey, B.M., Klein, B.P., Petridou, N., Dumoulin, S.O., 2013. Topographic representation of numerosity in the human parietal cortex. Science 341 (6150), 1123–1126. doi:10.1126/science.1239052.

Hayashi, M.J., Ditye, T., Harada, T., Hashiguchi, M., Sadato, N., Carlson, S., Walsh, V., Kanai, R., 2015. Time adaptation shows duration selectivity in the human parietal cortex. PLoS Biol. 13 (9). doi:10.1371/journal.pbio.1002262, undefined-undefined.

Hayashi, M.J., van der Zwaag, W., Bueti, D., Kanai, R., 2018. Representations of time in human frontoparietal cortex. Commun. Biol. 1 (1), 1–10. doi:10.1038/s42003-018-0243-z.

Hendrikx, E., Paul, J., van Ackooij, M., van der Stoep, N., & Harvey, B. (2022). Derivation of visual timing-tuned neural responses from early visual stimulus representations. doi:10.21203/rs.3.rs-579436/v1

Henry, M.J., Obleser, J., 2012. Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. Proc. Natl. Acad. Sci. 109 (49), 20095–20100. doi:10.1073/pnas.1213390109.

Heron, J., Aaen-Stockdale, C., Hotchkiss, J., Roach, N.W., McGraw, P.V., Whitaker, D., 2012. Duration channels mediate human time perception. Proc. Biol. Sci. 279 (1729), 690–698. doi:10.1098/rspb.2011.1131.

Heron, J., Hotchkiss, J., Aaen-Stockdale, C., Roach, N.W., Whitaker, D., 2013. A neural hierarchy for illusions of time: duration adaptation precedes multisensory integration. J. Vis. 13 (14), 4.

Hofstetter, S., Cai, Y., Harvey, B.M., Dumoulin, S.O., 2021. Topographic maps representing haptic numerosity reveals distinct sensory representations in supramodal networks. Nat. Commun. 12 (1), 221. doi:10.1038/s41467-020-20567-5.

Hubel, D.H., Wiesel, T.N., 1959. Receptive fields of single neurones in the cat's striate cortex. J. Physiol. 148 (3), 574–591. doi:10.1113/jphysiol.1959.sp006308.

Huppert, F., Singer, G., 1967. An aftereffect in judgment of auditory duration. Percept. Psychophys. 2 (11), 544–546. doi:10.3758/BF03210263.

Jazayeri, M., Shadlen, M.N., 2010. Temporal context calibrates interval timing. Nat. Neurosci. 13 (8), 1020–1026. doi:10.1038/nn.2590.

Kaas, J.H., Hackett, T.A., 2000. Subdivisions of auditory cortex and processing streams in primates. Proc. Natl. Acad. Sci. 97 (22), 11793–11799. doi:10.1073/pnas.97.22.11793.

Kastner, S., Pinsk, M.A., 2004. Visual attention as a multilevel selection process. Cogn. Affect. Behav. Neurosci. 4 (4), 483–500. doi:10.3758/cabn.4.4.483.

Kayser, C., Logothetis, N.K., 2009. Directed interactions between auditory and superior temporal cortices and their role in sensory integration. Frontiers in Integrative Neuroscience 3, 7. doi:10.3389/neuro.07.007.2009.

Kayser, C., Petkov, C.I., Logothetis, N.K., 2008. Visual modulation of neurons in auditory cortex. Cereb. Cortex 18 (7), 1560–1574. doi:10.1093/cercor/bhm187.

Klein, B.P., Harvey, B.M., Dumoulin, S.O., 2014. Attraction of position preference by spatial attention throughout human visual cortex. Neuron 84 (1), 227–237. doi:10.1016/j.neuron.2014.08.047.

Klink, P.C., Montijn, J.S., van Wezel, R.J.A., 2011. Crossmodal duration perception involves perceptual grouping, temporal ventriloquism, and variable internal clock rates. Atten. Percept. Psychophys. 73 (1), 219–236. doi:10.3758/s13414-010-0010-9.

Lalor, E.C., Power, A.J., Reilly, R.B., Foxe, J.J., 2009. Resolving precise temporal processing properties of the auditory system using continuous stimuli. J. Neurophysiol. 102 (1), 349–359. doi:10.1152/jn.90896.2008.

Madison, G., Merker, B., 2004. Human sensorimotor tracking of continuous subliminal deviations from isochrony. Neurosci. Lett. 370 (1), 69–73. doi:10.1016/j.neulet.2004.07.094.

Manning, F., Schutz, M., 2013. "Moving to the beat" improves timing perception. Psychon. Bull. Rev. 20, 1133–1139. doi:10.3758/s13423-013-0439-7.

Merchant, H., Pérez, O., Zarco, W., Gámez, J., 2013. Interval tuning in the primate medial premotor cortex as a general timing mechanism. J. Neurosci. 33 (21), 9082–9096. doi:10.1523/JNEUROSCI.5513-12.2013.

Merchant, H., Zarco, W., Pérez, O., Prado, L., Bartolo, R., 2011. Measuring time with different neural chronometers during a synchronization-continuation task. Proc. Nat. Acad. Sci. U.S.A. 108 (49), 19784–19789. doi:10.1073/pnas.1112933108.

Moeller, S., Yacoub, E., Olman, C.A., Auerbach, E., Strupp, J., Harel, N., Uğurbil, K., 2010. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. Magn. Reson. Med. 63 (5), 1144–1153. doi:10.1002/mrm.22361.

Moerel, M., De Martino, F., Formisano, E., 2014. An anatomical and functional topography of human auditory cortical areas. Frontiers in Neuroscience 8, 225. doi:10.3389/fnins.2014.00225.

Moerel, M., De Martino, F., Uğurbil, K., Yacoub, E., Formisano, E., 2015. Processing of frequency and location in human subcortical auditory structures. Sci. Rep. 5 (1), 17048. doi:10.1038/srep17048.

Murai, Y., Yotsumoto, Y., 2016. Timescale- and sensory modality-dependency of the central tendency of time perception. PLoS One 11, e0158921. doi:10.1371/journal.pone.0158921.

Nakai, T., Koide-Majima, N., Nishimoto, S., 2021. Correspondence of categorical and feature-based representations of music in the human brain. Brain Behav. 11 (1), e01936. doi:10.1002/brb3.1936.

Neuhoff, J.G., McBeath, M.K., Wanzie, W.C., 1999. Dynamic frequency change influences loudness perception: a central, analytic process. J. Exp. Psychol. Hum. Percept. Perform. 25 (4), 1050–1059. doi:10.1037//0096-1523.25.4.1050.

Neuner, I., Kawohl, W., Arrubla, J., Warbrick, T., Hitz, K., Wyss, C., Boers, F., Shah, N.J., 2014. Cortical response variation with different sound pressure levels: A combined event-related potentials and FMRI study. PLoS ONE 9 (10), e109216. doi:10.1371/journal.pone.0109216.

Norman-Haignere, S., Kanwisher, N.G., McDermott, J.H., 2015. Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. Neuron 88 (6), 1281–1296. doi:10.1016/j.neuron.2015.11.035.

Nozaradan, S., Peretz, I., Keller, P.E., 2016. Individual differences in rhythmic cortical entrainment correlate with predictive behavior in sensorimotor synchronization. Sci. Rep. 6 (1), 20612. doi:10.1038/srep20612.

Paul, J.M., van Ackooij, M., Ten Cate, T.C., Harvey, B.M., 2022. Numerosity tuning in human association cortices and local image contrast representations in early visual cortex. Nat. Commun. 13 (1), 1340. doi:10.1038/s41467-022-29030-z.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., Dehaene, S., 2004. Tuning curves for approximate numerosity in the human intraparietal sulcus. Neuron 44 (3), 547–555. doi:10.1016/j.neuron.2004.10.014.

Protopapa, F., Hayashi, M.J., Kulashekhar, S., Zwaag, W., van der, Battistella, G., Murray, M.M., Kanai, R., Bueti, D., 2019. Chronotopic maps in human supplementary motor area. PLoS Biol. 17 (3), e3000026. doi:10.1371/journal.pbio.3000026.

Romei, V., Gross, J., Thut, G., 2012. Sounds reset rhythms of visual cortex and corresponding human visual perception. Curr. Biol. 22 (9–2), 807–813. doi:10.1016/j.cub.2012.03.025.

Roseboom, W., Fountas, Z., Nikiforou, K., Bhowmik, D., Shanahan, M., & Seth, A.K. (2018). Time without clocks: human time perception based on perceptual classification (p. 172387). doi:10.1101/172387

Saal, H.P., Harvey, M.A., Bensmaia, S.J., 2015. Rate and timing of cortical responses driven by separate sensory channels. ELife 4, e10450. doi:10.7554/eLife.10450.

Satterthwaite, F.E., 1941. Synthesis of variance. Psychometrika 6 (5), 309–316. doi:10.1007/BF02288586.

Satterthwaite, F.E., 1946. An Approximate distribution of estimates of variance components. Biom. Bull. 2 (6), 110–114. doi:10.2307/3002019.

Silva, M.F., Harvey, B.M., Jorge, L., Canário, N., Machado, F., Soares, M., d'Almeida, O.C., Castelo-Branco, M., 2021. Simultaneous changes in visual acuity, cortical population receptive field size, visual field map size, and retinal thickness in healthy human aging. Brain Struct. Funct. 226 (9), 2839–2853. doi:10.1007/s00429-021-02338-0.

Sohoglu, E., Kumar, S., Chait, M., Griffiths, T.D., 2020. Multivoxel codes for representing and integrating acoustic features in human cortex. Neuroimage 217, 116661. doi:10.1016/j.neuroimage.2020.116661.

Song, C., Schwarzkopf, D.S., Kanai, R., Rees, G., 2015. Neural population tuning links visual cortical anatomy to human visual perception. Neuron 85 (3), 641–656. doi:10.1016/j.neuron.2014.12.041.

Stigliani, A., Jeska, B., Grill-Spector, K., 2017. Encoding model of temporal processing in human visual cortex. Proceedings of the National Academy of Sciences 114 (51), E11047–E11056. doi:10.1073/pnas.1704877114.

Suzuki, Y., Takeshima, H., 2004. Equal-loudness-level contours for pure tones. J. Acoust. Soc. Am. 116 (2), 918–933. doi:10.1121/1.1763601.

Tierney, A., Kraus, N., 2013. The ability to move to a beat is linked to the consistency of neural responses to sound. J. Neurosci. 33 (38), 14981–14988. doi:10.1523/JNEUROSCI.0612-13.2013.

Tsouli, A., Cai, Y., van Ackooij, M., Hofstetter, S., Harvey, B.M., Te Pas, S.F., van der Smagt, M.J., Dumoulin, S.O., 2021. Adaptation to visual numerosity changes neural numerosity selectivity. Neuroimage 229, 117794. doi:10.1016/j.neuroimage.2021.117794.

Tsouli, A., Harvey, B.M., Hofstetter, S., Cai, Y., van der Smagt, M.J., te Pas, S.F., Dumoulin, S.O., 2022. The role of neural tuning in quantity perception. Trends Cogn. Sci. doi:10.1016/j.tics.2021.10.004.

Tukey, J.W., 1949. Comparing individual means in the analysis of variance. Biometrics 5 (2), 99. doi:10.2307/3001913.

van Atteveldt, N., Murray, M.M., Thut, G., Schroeder, C.E., 2014. Multisensory integration: flexible use of general operations. Neuron 81 (6), 1240–1253. doi:10.1016/j.neuron.2014.02.044.

von Vierordt, K. (1866). Der Zeitsinn nach Versuchen—Deutsche Digitale Bibliothek. http://www.deutsche-digitale-bibliothek.de/item/FYWLZJGGXML3NAVQJNQ5ZQJW6QBPM5IH

Walker, J.T., Irion, A.L., Gordon, D.G., 1981. Simple and contingent aftereffects of perceived duration in vision and audition. Percept. Psychophys. 29 (5), 475–486. doi:10.3758/BF03207361.

Wandell, B.A., Brewer, A.A., Dougherty, R.F., 2005. Visual field map clusters in human cortex. Philos. Trans. R. Soc. B 360 (1456), 693–707. doi:10.1098/rstb.2005.1628.

Will, U., Berg, E., 2007. Brain wave synchronization and entrainment to periodic acoustic stimuli. Neurosci. Lett. 424 (1), 55–60. doi:10.1016/j.neulet.2007.07.036.

Zélanti, P.S., Droit-Volet, S., 2011. Cognitive abilities explaining age-related changes in time perception of short and long durations. J. Exp. Child. Psychol. 109 (2), 143–157. doi:10.1016/j.jecp.2011.01.003.

Zhou, J., Benson, N.C., Kay, K.N., Winawer, J., 2018. Compressive temporal summation in human visual cortex. J. Neurosci. 38 (3), 691–709. doi:10.1523/JNEUROSCI.1724-17.2017.