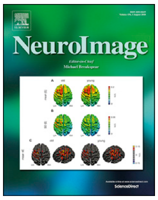
NeuroImage 176 (2018) 29–40

Contents lists available at ScienceDirect

NeuroImage

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



What's what in auditory cortices?

Chrysa Retsa [a,*,1], Pawel J. Matusz [a,b,1], Jan W.H. Schnupp [c,d,2], Micah M. Murray [a,e,f,g,2]

[a] The LINE (Laboratory for Investigative Neurophysiology), Radiology Department and Department of Clinical Neurosciences, University Hospital Center and University of Lausanne, 1011, Lausanne, Switzerland
[b] Information Systems Institute, University of Applied Sciences Western Switzerland, HES-SO Valais, 3960, Sierre, Switzerland
[c] Department of Biomedical Sciences, City University of Hong Kong, To Yuen Street, Kowloon, Hong Kong
[d] Department of Physiology, Anatomy and Genetics, University of Oxford, Parks Road, Oxford, OX1 3PT, UK
[e] The EEG Brain Mapping Core, Center for Biomedical Imaging (CIBM), University Hospital Center and University of Lausanne, 1011, Lausanne, Switzerland
[f] Department of Ophthalmology, University of Lausanne, Fondation Asile des Aveugles, Lausanne, Switzerland
[g] Department of Hearing and Speech Sciences, Vanderbilt University, Nashville, TN, USA

ARTICLE INFO

Keywords:
Auditory
Object
Functional segregation
Auditory evoked potential (AEP)
'What' and 'where' pathways

ABSTRACT

Distinct anatomical and functional pathways are postulated for analysing a sound's object-related ('what') and space-related ('where') information. It remains unresolved to which extent distinct or overlapping neural resources subserve specific object-related dimensions (i.e. who is speaking and what is being said can both be derived from the same acoustic input). To address this issue, we recorded high-density auditory evoked potentials (AEPs) while participants selectively attended and discriminated sounds according to their pitch, speaker identity, uttered syllable ('what' dimensions) or their location ('where'). Sound acoustics were held constant across blocks; the only manipulation involved the sound dimension that participants had to attend to. The task-relevant dimension was varied across blocks. AEPs from healthy participants were analysed within an electrical neuroimaging framework to differentiate modulations in response strength from modulations in response topography; the latter of which forcibly follow from changes in the configuration of underlying sources. There were no behavioural differences in discrimination of sounds across the 4 feature dimensions. As early as 90ms post-stimulus onset, AEP topographies differed across 'what' conditions, supporting a functional sub-segregation within the auditory 'what' pathway. This study characterises the spatio-temporal dynamics of segregated, yet parallel, processing of multiple sound object-related feature dimensions when selective attention is directed to them.

Introduction

The perceived aspects of sounds, such as their particular pitch or timbre or location, are constructs of neural activity, and not a simple and direct reflection of physical sound properties. It is postulated that the perception of auditory information is the result of parallel processing along multiple functional pathways; one pathway is chiefly involved in determining the identity of sound objects, and another is devoted to determining their location (Alain et al., 2001; Rauschecker and Tian, 2000). Findings from animal studies (Romanski et al., 1999; Rauschecker and Tian, 2000; Tian et al., 2001; see also Perrodin et al., 2015 and Hackett, 2015 for reviews of studies in humans and non-human primates) as well as from human neuropsychological, functional imaging, magneto-electrophysiological (M/EEG), lesion as well as virtual lesion studies (Maeder et al., 2001; Morosan et al., 2001; Clarke et al., 2002; Zatorre et al., 2002; Anourova et al., 2001; Lewald et al., 2004; Scott, 2005; Tardif et al., 2008) seem to support this distinction between an anterior/ventral, "what" pathway, and a posterior/dorsal, "where" auditory pathway. Both pathways are thought to include the primary auditory cortices (PAC) and then diverge, such that the ventral pathway involves the rostral superior temporal cortex and ventral subdivisions of frontal and prefrontal cortices, and the dorsal pathway involves the caudal superior temporal cortex, parietal cortex and dorsal subdivisions of frontal and prefrontal cortices (Kaas and Hackett, 1999; Romanski et al., 1999; Lomber and Malhotra, 2008). However, the functional organization of the putative "what" stream understood in humans remains

[*] Corresponding author.
E-mail address: chrysa.retsa@gmail.com (C. Retsa).
[1] equal first-author contribution.
[2] equal senior-author contribution.

https://doi.org/10.1016/j.neuroimage.2018.04.028
Received 16 January 2018; Received in revised form 4 April 2018; Accepted 12 April 2018
Available online 18 April 2018
1053-8119/© 2018 Elsevier Inc. All rights reserved.

poorly characterised.

The neural processing of sounds can be thought of as giving rise to identification of numerous perceptual features and "dimensions", both spatial and non-spatial (i.e., object-identity-related). That is, the same syllable spoken by two different speakers may be considered the "same auditory object" if all that matters is the message and not the messenger, but they are different objects in terms of who is talking. Similarly, a syllable may or may not be perceived as "the same" if its pitch changes substantially. In Indo-European, non-tonal languages (e.g., English, German, French), voice pitch is an important feature in prosody, but changing voice pitch does not change the meaning of words; but this is not the case in tonal languages (e.g., Chinese or Thai), where changing pitch may result in the perception of different words, with very different meanings. Furthermore, the identity of a speaker as someone large/small or older/younger will depend on both the formant frequencies and the pitch of a spoken syllable. Systematic variations in both pitch and/or formant structure of vocalizations, in addition to sound source location, are thus very useful tools to dissect the functional organization of the "what" and "where" streams within the auditory pathway.

These two classes of auditory perceptual dimensions are indeed fine-grained. However, it remains unclear if this is mirrored by comparably fine-grained dissociable spatio-temporal representations and neural circuits, both within and beyond traditional auditory cortices (i.e. core and belt fields). Previous M/EEG studies have shown that dissociable processing along auditory "what" and "where" pathways is observed as early as 100ms post-stimulus (De Santis et al., 2007; Leavitt et al., 2011; Ahveninen et al., 2006; Anourova et al., 2001). Notably, typically, these effects are based on the contrast between responses to a single object-related feature and a single space-related feature (or selective attention to such). Some comparisons involved the neural responses to pitch versus location processing (Alain et al., 2001; Anourova et al., 2001, 2003; De Santis et al., 2007; Warren and Griffiths, 2003; Paltoglou et al., 2011), phonetic versus location processing (Ahveninen et al., 2006; Tian et al., 2001) or object identification versus location processing (Herrmann et al., 2002; Leavitt et al., 2011). Others have focused on sub-divisions within the auditory "where" stream. On the one hand, there is evidence that location and motion processing are subserved by partially distinct neural systems beginning from ~250ms post-stimulus onset (Ducommun et al., 2002). On the other hand, partially segregated and non-linearly interacting responses have been observed across different spatial acoustic cues (i.e. inter-aural intensity and timing differences) (Tardif et al., 2006). However, it is also worth noting that the support for (even partially) selective brain activation for different sound dimensions is not universal (e.g. Bidet-Caulet et al., 2005; Maeder et al., 2001; Rama et al., 2000; Zatorre et al., 1999; Zatorre et al., 1994).

The contradictory results might partly arise also due to differing definitions of what constitutes an auditory "object". Some research has focused on semantic dimensions; demonstrating, e.g., distinctions between sounds of living and man-made environmental sounds (Engel et al., 2009; Murray et al., 2006; Lewis et al., 2005) or vocalizations (De Lucia et al., 2010; Belin et al., 2004; Belin et al., 2000; Bruneau et al., 2013; see also Perrodin et al., 2015 for a review of studies in non-human primates). Such comparisons involved different acoustic inputs that engendered also distinct semantic perceptions, leaving unresolved whether different object features of the same acoustic inputs are differentially processes. Others have compared more rudimentary object-related dimensions. For example, in a positron emission tomography (PET) study (Zatorre et al., 1992) study, participants were required to perform either a phoneme or a pitch discrimination task on speech sounds. While phonetic discrimination resulted in increased activity in parts of the left hemisphere Broca's area and in the left superior parietal cortex, pitch discrimination elicited increased activity in right prefrontal cortex. At present, it remains contentious as to whether there is a dedicated pitch-related centre and to what extent, if any, pitch is processes in a segregated manner from features such as timbre (reviewed in Griffiths and Hall, 2012). Similarly, an fMRI study comparing speaker identity versus vowel processing showed distinct neural activation patterns for the same stimuli depending on the task-relevant dimension, with speaker discrimination relying on distinct right middle superior temporal gyrus/sulcus (STG/STS) activation and vowel discrimination relying on right posterior temporal cortex activation (pSTS; Bonte et al., 2014). Another fMRI study investigating voice identity versus verbal processing showed selective right anterior STS (aSTS) activation related to the speaker task and activation in the bilateral fusiform/lingual region related to the verbal task (von Kriegstein et al., 2007). In addition, a MEG study also investigating phonological versus speaker processing showed more posterior and superior sources active during speaker categorization compared to vowel categorization (Obleser et al., 2004).

Notably, in terms of sub-segregation of networks responsive to object dimensions, specifically, a recent fMRI study (Allen et al., 2017) and an MEG study (Gutschalk and Uppenkamp, 2011) investigating differences between pitch and timbre processing and pitch and vowel processing, respectively, found no evidence for anatomical distinctions between regions dedicated to the processing of the above dimensions. In the former study, traditional univariate analyses did not reveal distinct pitch and timbre processing within auditory cortices. That the two tasks may engage distinct sub-circuits within the same regions was instead revealed only by multi-voxel pattern analysis (MVPA) reported in the same study (see also Griffiths and Hall (2012) for a similar argument regarding the benefits of MVPA for identifying functionally specialized circuits for pitch). Differences across these interleaved representations of different "what" dimensions could be more readily gauged by studying the spatio-temporal dynamics underlying their activation.

Evidence for the selective activation for different sound object dimensions has been indeed most consistently provided by methods sensitive to spatio-temporal brain response dynamics. In one EEG study exploring vowel and speaker related processing using one-back tasks (Bonte et al., 2009), distinct task-specific processing for the two dimensions was observed relatively late (after 300ms post-stimulus onset). Intertrial-phase-coherence (ITC) analysis of the EEG data indicated a left-hemisphere bias for vowel processing and a right-hemisphere bias for speaker processing. A MEG study investigating discriminations along similar stimulus dimensions (speaker versus speech recognition) identified an earlier point of divergence between their neural processing (~200ms), with right pSTS and right aSTS both related to the processing of the speaker's voice, whereas left STS found to be specifically related to the processing of speech information (Schall et al., 2015). Evidence for the dissociable brain activation in response to different sound object dimensions has also been provided in EEG studies targeting the mismatch negativity (MMN) component (Giard et al., 1995; Deouell et al., 1998). Giard et al. (1995) observed differences in the topographic distribution of MMNs elicited by deviance across multiple but "lower-level" perceptual dimensions, i.e., frequency, intensity and duration. Moreover, often MMN to double deviants - where two stimulus dimensions within a given stimulus synchronously deviate - is as large as the sum of the MMNs to each of the individual feature deviants (Paavilainen, 2013). This additivity has been partially attributed to the involvement of distinct neural populations in the processing of the different stimulus features (Paavilainen et al., 2001). Notwithstanding, not all studies using double deviants have observed this additive effect (e.g. Hay et al., 2015). In addition, Schairer et al. (2001) did not obtain evidence of separate source locations for frequency, intensity and duration MMN. Thus, there remains a need for multi-dimensional investigations of sound processing within the auditory "what" pathway, and perhaps with other experimental paradigms. Only one previous study, conducted in ferrets, examined how sound stimuli varying in more than two "higher-level" perceptual dimensions are encoded by neurons in the auditory cortex (Bizley et al., 2009). Using vowels systemically varying in pitch, timbre and spatial location, Bizley et al. demonstrated that the three dimensions are encoded in a *de facto* interdependent manner in primary auditory cortex and anterior auditory fields. However, similar studies in humans have not been conducted.

To summarise, we believe that the reason behind the discrepant results with respect to the degree of distinct activations in response to different sound object dimensions in humans are three-fold: 1) using passive tasks that encourage the conflation of the highly plastic dimensions of auditory perception, 2) using a very limited number of dimensions fails to emulate ethologically-relevant, real-world situations, and 3) using insensitive analytical techniques ignoring either the temporal or spatial information. Consequently, here we investigated the presence versus absence of functional specialisation of the "what" pathway in humans by focusing on the spatio-temporal brain dynamics underlying the processing of sound defined across multiple object-related dimensions. We have addressed the shortcomings of the existing studies in several important ways.

First, we used speech sounds varying across four perceptual dimensions: three object-related dimensions (pitch, syllable type, speaker identity) and one spatial dimension (left/right location). Participants were always engaged in a two-alternative-forced-choice task, either discriminating the syllable type ("ta" or "ti"), syllable pitch (high or low), speaker (man or boy), or the sound location (left or right). Such multi-feature focus better reflects the fact that in everyday life multiple aspects of sounds are processed concurrently. Second, identical sounds were used across all four tasks. The only manipulation was the perceptual dimension that the participants had to discriminate on a given block of trials. Previous fMRI, PET, EEG and MEG work has shown that selective attention enhances the activity within task-relevant areas of the human auditory cortex and modulates activity within the auditory "where" and "what" pathways in a feature-specific fashion (Ahveninen et al., 2006; Petkov et al., 2004; Alho et al., 2003; Woldorff et al., 1993; Hillyard et al., 1973). The latter has been suggested to be the result of increased selectivity of neural populations based on task requirements (Ahveninen et al., 2006). The earlier described studies (phonetic vs. speaker processing, Bonte et al., 2014; Obleser et al., 2004), demonstrating different areas involved in the processing of distinct sound dimensions, had employed active tasks where the dimension to be attended was manipulated. In contrast, studies using passive listening tasks (Allen et al., 2017; Gutschalk and Uppenkamp, 2011) did not provide evidence for selective activations – at least when univariate analyses of the data were used. This suggests that the division of labour within the "what" pathway in the auditory cortex is enhanced and may become apparent only when the participants listen to and isolate the different attributes of the sounds. Thus, in the present study, we investigated how speech sound representations are modulated in a task-dependent manner and, specifically, if selective attention can modulate the brain activity induced by different sounds as a function of the different task-relevant "what" dimensions independently. Third, we recorded AEPs and analysed them within an electrical neuroimaging framework in order to investigate differences both in the response strength and the topography of the electric field at the scalp, with the latter reflecting changes in the configuration of brain generators.

As this enabled us to capitalise on the added value of investigating differences in both spatial (which brain regions are involved) as well as temporal (when are they involved) sound-elicited brain activity, we predicted we would be able to identify differences in the timing and topography of neural responses that accompany a participants' shift in attention across four different "perceptual dimensions" of the sound stimuli, namely the syllable type, pitch, speaker or location respectively. As cross-dimension differences in AEPs have been previously demonstrated with active-task paradigms, we expected that directing attention to such different attributes of exactly the same sounds could offer a robust way of identifying differences in the functional organisation within the "what" pathway.

Materials and methods

Participants

Nineteen healthy unpaid volunteers (11 female; aged 24–49 years; mean ± SD = 27.5 ± 5 years) provided informed consent to participate in the experiment. All procedures were approved by the Cantonal Ethics Committee. The data of 3 participants were excluded from further analyses due to excessive artefacts during recording. Fifteen of the remaining participants were right-handed and one was left-handed, as assessed with the Edinburgh questionnaire (Oldfield, 1971). None of the participants reported having current or prior neurological or psychiatric illnesses. All participants reported normal hearing and had normal or corrected-to-normal vision.

Apparatus and stimuli

The participants were seated at the center of a sound-attenuated chamber (whisper room model 102126 E) and acoustic stimuli were delivered over insert earphones (Etymotic model ER-4P; www.etymotic.com) at a sample rate of 48 kHz. Stimulus intensity was approximately 75 dB SPL at the ear.

We chose to systematically change a sample of vocalizations from a syllable set recorded by the Cambridge Centre for the Neural Basis of Hearing (CNBC) to generate our stimulus set. The original recordings were kindly provided by Prof. Roy Patterson (see Ives et al., 2005 for details). We chose the syllables/ta/and/ti/, from the original syllable set, which had been spoken by a single male adult speaker in a quiet room recorded with a Shure SM58-LCE microphone and digitized at 48kHz. From these recordings we generated natural sounding morphs of the/-ta/and/ti/syllables that were identical in sound intensity and duration and differed only in systematic shifts of their harmonic and formant frequencies to create high or low pitched versions in the voice of a man or a boy respectively. The recordings were decomposed into fundamental frequency (pitch, F0), aspiration and formant signals using Kawahara and Irino's STRAIGHT speech analysis software (Kawahara and Irino, 2005) for subsequent resynthesis, with the appropriate adjustments to the signals' duration, pitch or formant structure. To generate the "baritone" voice typical of an adult man, the formant frequencies of the syllable were scaled down by a factor of 0.9, and to create low and high pitched syllables in that voice, the fundamental frequency (F0) was scaled to either 77.8 or 155.6 Hz, respectively (half an octave above and below A2, 110Hz). In contrast, to create the "alto" voice of a primary school age boy the formants were scaled up by a factor of 1.4, and the F0s were set to 311 and 622 Hz (±0.5 octaves around A4) respectively to generate low and high pitched syllables in that voice. The scale factors used here were chosen after informal experimentation so as to obtain resynthesized syllables which, despite the tight control over their formant and fundamental frequencies nevertheless sound like natural exemplars of male adult or infant speech sounds. WAV files of the sound stimuli are available as supplementary materials."

ce of a primary school age boy the formants were scaled up by a factor of 1.4, and the F0s were set to 311 and 622 Hz (+/-0.5 octaves around A4) respectively to generate low and high pitched syllables in that voice. The scale factors used here were chosen after informal experimentation so as to obtain resynthesized syllables which, despite the tight control over their formant and fundamental frequencies nevertheless sound like natural exemplars of male adult or infant speech sounds. WAV files of the sound stimuli are available as supplementary materials."

Supplementary video related to this article can be found at https://doi.org/10.1016/j.neuroimage.2018.04.028.

To vary the perceived spatial location of the syllables, these morphed syllables were presented in a "virtual acoustic space" at a distance of 1m, 60° to the left or right off the midline in front of the person's head, at eye level. To add realism, a small amount of reverberation was added to the sound by adding "specular reflections", that is, each wall floor and ceiling of the room were treated as "sound mirrors" which will reflect sound essentially without frequency filtering but a flat reduction in amplitude. We chose an absorption coefficient of 0.6 for the virtual walls of this simple room model, an appropriate value to approximate the quite highly absorbent walls of the actual recording room. The original virtual sound source location was mirrored along each of the walls, floor and ceiling of to create virtual reflected mirror sources. Mirrored sources were further reflected across the opposite walls, ceiling or floor to generate second order reflections. The sound presented over the headphones to each ear was then computed as the superposition of sounds from the original virtual source as well as each of the first and second order reflections. Each of these virtual direct and indirect sources was filtered using the most appropriate head related transfer function (HRTF) impulse response given the angle of that sound source to the centre of the listener's head, delayed to reflect the distance of the source from the head assuming a speed of sound of 343 m/s, and attenuated to model the distance (inverse square law) as well as, for reflected sources, the corresponding absorption coefficients. HRTF impulse responses were obtained from head number 6 of the Sydney-York SYMARE database (http://sydney.edu.au/engineering/electrical/carlab/symare.htm; Jin et al., 2014). The E-prime software controlled stimulus delivery and recorded the participants' behavioural performance (www.pstnet.com/eprime). Spectrograms of the sound stimuli are shown in Fig. 1s

Procedure

Participants performed a discrimination task comprising 4 different types of blocks of trials varying on the dimension-to-be-discriminated (Table 1). Each block was presented twice, resulting in a total of 8 blocks, and the order of blocks was counter-balanced across participants. Sound acoustics were held constant across blocks; only the specific instructions differed between blocks. Sounds of two spatial locations (left or right), two relative pitch levels (high or low), two speakers (man or boy) and two syllables ("ta" or "ti") were used, resulting in 16 different stimuli, each presented 10 times in each block. All sounds had duration of 550ms. Therefore, each block contained 160 trials, which took approximately 5 min. A short instruction was presented before the start of each block to indicate to the participants which feature they would have to discriminate, for example: "Press 1 if the syllable is 'ti', press 2 if it is 'ta'". A fixation cross in the centre of the screen was presented at the beginning of the trial and remained visible during the presentation of the sound; participants were asked to fixate on it and not to move their eyes. The participant's task was to indicate as quickly and as accurately as possible whether the presented sound was i) presented on the left or right side - in the "spatial" block ("where" condition), or ii) low or high pitch - in the "pitch" block (first "what" condition), or iii) a child or a man - in the "speaker" block (second "what" condition), or iv) a 'ti' or a 'ta' - in the "syllable" block (third "what" condition). The same response buttons (placed below the participant's right index and right major finger) were used in all tasks. After each response, the next trial started after a randomized interval of 300–600ms at steps of 100ms.

Table 1
Attended block type and features of stimuli that had to be discriminated.

Block type	Features to-be-discriminated	
Location	left side (50%)	right side (50%)
Pitch	low (50%)	high (50%)
Speaker	school age boy (50%)	adult man (50%)
Syllable	ti (50%)	ta (50%)

EEG recording and pre-processing

Continuous EEG was acquired at 1024 Hz through a 128-channel Biosemi ActiveTwo AD-box (www.biosemi.com), referenced to the common mode sense (CMS; active electrode) and grounded to the driven right leg (DRL; passive electrode), which functions as a feedback loop driving the average potential across the electrode montage to the amplifier zero. Prior to epoching, the EEG was filtered (low-pass 40Hz; high-pass 0.1Hz; removed DC; 50Hz notch; using a second-order Butterworth filter with −12 db/octave roll-off that was computed linearly in both forward and backward directions to eliminate phase shifts). Peri-stimulus epochs spanning 100ms pre-stimulus to 500ms post-stimulus were averaged for each of the 4 conditions (attend to location, pitch, speaker or duration, respectively) and for each participant, to calculate the auditory evoked potentials (AEPs). Epochs were rejected based on an automated artefact rejection criterion of ±80 μV as well as visual inspection for eye blinks, eye movements or other sources of transient noise. The average number (±SEM) of accepted EEG epochs for each participant and each of the above four conditions was 281 ± 6.7, 285 ± 5.8, 286 ± 5.9 and 288 ± 5.7, respectively. These values did not significantly differ ($F(3,45) = 0.34$, $p > 0.75$). Bad channels were identified before averaging and excluded from the artefact rejection. These data at artefact electrodes from each participant were interpolated using 3-D splines prior to group averaging (Perrin et al., 1987). The average number of interpolated channels was five. In addition, data were baseline corrected using the 100ms pre-stimulus period and recalculated against the average reference.

ERP analyses

Differences in the processing of the sounds as a function of which of the four sound dimensions was attended to were examined using a multistep analysis procedure, referred to as electrical neuroimaging, which involves both local and global measures of the electric field on the scalp, and which has been described in detail previously (Koenig et al., 2014; Michel and Murray, 2012; Tzovara et al., 2012; Murray et al., 2008; Michel et al., 2004). This analytical approach focuses on reference-independent measures of the electric field at the scalp and

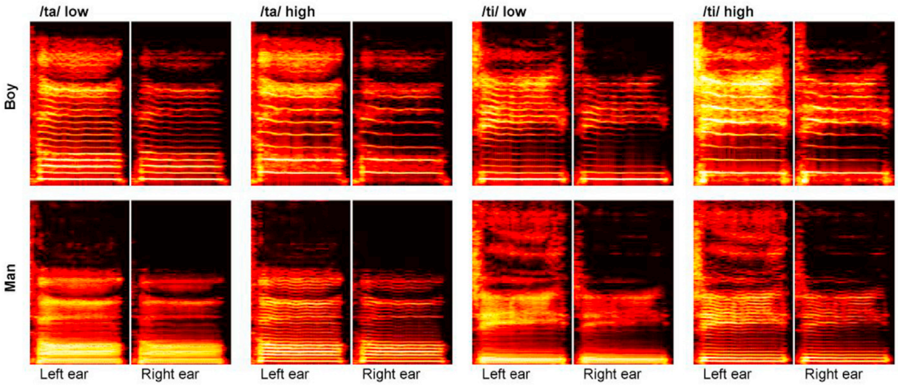


Fig. 1. Spectrograms of the 16 syllable stimuli used in this experiment. Each pair of panels shows the left and right ear stimuli respectively for one syllable (either/ta/or/ti/at either low of high voice pitch, as indicated above each panel). They are plotted using a linear frequency axis (vertical) from 0 to 8kHz and a time axis (horizontal) covering 0–0.55 s. The color scale saturates over a 120 dB range. The top row of panels shows the syllables for the synthesized "boy" voice, achieved by systematically shifting the formant frequencies. The bottom row shows the corresponding sounds for the "older man'"s voice. Only the sounds for the virtual acoustic space location at 45 deg. to the participant's left are shown - the stimuli to the participant's right are essentially mirror symmetric.

allows for differentiation between effects that arise from modulation in the strength of responses of statistically indistinguishable brain generators and alterations in the configuration of the active generators.

We first analysed the AEP voltage waveform data from each scalp electrode as a function of time using a one-way ANOVA with the within-subject factor of condition (4 levels: attend location, pitch, speaker or syllable, respectively). For this analysis we used an average reference as well as a temporal criterion for the detection of statistically significant effects (>10ms continuously at 1024 Hz sampling rate) in order to correct for temporal auto-correlation at individual electrodes (Guthrie and Buchwald, 1991). Similarly, a spatial criterion (effects were considered statistically significant only if they entailed >10% of the electrodes of the 128-channel montage at a given latency) was applied in order to address spatial correlation. Analyses of the AEP voltage waveform data were included here to provide readers with a sense of the general waveform shape and to link the canonical ERP analysis approaches and the electrical neuroimaging framework. Note, however, that the shapes of AEP voltage waveforms are reference-dependent and thus do not serve here as a primary basis for interpretation (Murray et al., 2008).

The electrical neuroimaging analyses comprised the following steps. First, changes in the strength of the electric field at the scalp as a function of which sound feature was attended to were assessed using global field power (GFP) for each participant and stimulus condition (Lehmann, 1987). GFP is calculated as the root mean square across the electrode montage (vs. the average reference) – i.e. $\mathrm{RMS}_{(\text{average reference})}$ as described in Lehmann and Skrandies (1980). A stronger GFP value is therefore indicative of greater and/or more synchronised brain activity, though the root cause cannot be readily asserted based on this measure alone (e.g. increased neural firing rate, increased numbers of active neurons, etc.). However, a modulation of GFP in the absence of reliable evidence for topographic modulations can most parsimoniously be interpreted as a modulation in the strength of responses originating from statistically indistinguishable sources. GFP was analysed as a function of time using a one-way ANOVA for the within-subject factor of condition.

Second, AEP topographic differences between the four attention conditions were identified using global map dissimilarity (GMD), which is another reference-free measure and equals the square root of the mean of squared differences between the potentials measured at each electrode for different conditions, normalised by their GFP (Lehmann and Skrandies, 1980). GMD is an index of topographic differences between two electric fields, and its values range from 0 to 2, with 0 indicating no topographic differences and 2 indicating topographic inversion. Analysis of GMD was performed using the Randomisation Graphical User interface (RAGU; Koenig et al., 2011). RAGU performs a non-parametric randomisation test on the GMD values (colloquially termed “TANOVA”), comparing the observed value at each time point to an empirical distribution based on permutations of the data from all participants and conditions. As topographic differences between conditions follow from changes in the configuration of the underlying neural generators (Lehmann, 1987), this analysis reveals whether and when the four different conditions activate distinct brain networks.

Next, a topographic pattern analysis based on a hierarchical clustering algorithm was performed on the post-stimulus group-average AEPs across experimental conditions (Murray et al., 2008). This clustering identifies stable electric field topographies (“template maps”). The clustering is insensitive to pure amplitude modulations across conditions as the data are first normalised by their instantaneous GFP. The optimal number of template maps that explained the whole group-averaged data set was determined using a modified Krzanowski-Lai criterion (Murray et al., 2008). The clustering makes no assumption regarding the orthogonality of the derived template maps (Pourtois et al., 2008; De Lucia et al., 2010; Koenig et al., 2014). The pattern of maps that was identified in the group-average AEPs were then submitted to a fitting procedure wherein each time point of each individual participant's ERP is labelled according to the template map with which it best correlated spatially (Murray et al., 2008). This yielded a measure of relative map presence (in milliseconds) that was then submitted to a repeated-measure ANOVA with factors of map and condition. This fitting procedure revealed whether a given experimental condition was more often described by one map versus another, and if the observed pattern varied across experimental conditions.

Finally, we estimated the underlying intracranial sources of the AEPs in response to the four different conditions using a distributed linear inverse solution (minimum norm) combined with the LAURA (local autoregressive average) regularisation approach (Grave de Peralta Menendez et al., 2001, 2004; see also Michel et al., 2004 for a review). LAURA selects the source configuration that best mimics the biophysical behaviour of electric vector fields (i.e., activity at one point depends on the activity at neighbouring points according to electromagnetic laws). In our study, as part of the regularization strategy, homogenous regression coefficients in all directions and within the whole solution space were used. The solution space was calculated on a realistic head model that included 3005 nodes, selected from a grid equally distributed within the gray matter of the Montreal Neurological Institute's average brain (courtesy of Grave de Peralta Menendez and Gonzalez Andino; http://www.electrical-neuroimaging.ch/). The head model and lead field matrix were generated within the Spherical Model with Anatomical Constraints (SMAC; Spinelli et al., 2000 as implemented in Cartool (Brunet et al., 2011)). As an output, LAURA provides current density measures; their scalar values were evaluated at each node. Prior basic and clinical research from members of our group and others have documented and discussed in detail the spatial accuracy of the inverse solution model used here (e.g., Grave de Peralta Menendez et al., 2004; Michel et al., 2004; Gonzalez Andino et al., 2005; Martuzzi et al., 2009). The source estimations were calculated after first averaging across time for each participant and condition. The results of the above topographic pattern analysis defined time-periods for which intracranial sources were estimated and statistically compared across conditions. Statistical analysis of source estimations was performed by first taking the average of the AEP across the time-periods demonstrating statistically reliable topographic differences (1) 89–170ms and 2) 262–370ms as detailed below). Then, the mean source estimations for this averaged time-period were performed for each of the 3005 nodes prior to conducting an ANOVA with the within-subjects factor of condition. The statistical significance criterion at an individual solution point was set at $p < 0.05$. Only clusters with at least 10 contiguous significant nodes were considered reliable in an effort to correct for multiple comparisons and was based on randomization thresholds determined with Alphasim software (see also Toepel et al., 2009; De Lucia et al., 2010; Knebel and Murray, 2012; Matusz et al., 2015; Matusz et al., 2016 for similar implementations).

Results

Behavioural results

The differences in sound features between the stimuli in the stimulus set were deliberately chosen to be large and easily distinguishable, and indeed all participants discriminated the sounds in each of the four conditions with high accuracy near-ceiling levels. The mean accuracy rate (percentage of correct responses± standard deviation) in each block was 94.2 ± 5.6% for the location condition, 90.8 ± 5.9% for the pitch condition and 95.3 ± 4.6% and 94.2 ± 10.1% for the speaker and syllable condition respectively. A one-way ANOVA with the within-subject factor of condition showed no statistically significant differences ($F(3,45) = 0.8$, $p = .49$, $\eta_p^2 = .06$, $\omega^2 = .23$) between the different types of blocks, suggesting that the pitch, speaker, syllable or location changes were all approximately equally easily distinguished. Likewise, a one-way ANOVA performed on the mean reaction times showed no significant differences ($F(3,45) = 2$, $p = .08$, $\eta_p^2 = .19$, $\omega^2 = .74$) between the four conditions (751 ± 75ms for the location, 825 ± 76ms for the pitch, 819 ± 109ms for the speaker and 775 ± 93ms for the syllable condition). Because of the high level of accuracy, both correct and incorrect responses were included in the subsequent ERP analyses.

ERP results

Group-averaged AEPs for the Location, Pitch, Speaker and Syllable conditions from an exemplar frontal midline electrode (FCz) are displayed in Fig. 2. Fig. 2 also shows the results of the univariate ANOVA across the full electrode montage, displayed as the percentage of electrodes exhibiting a main effect of type of Condition as a function of time post-stimulus and including a threshold of 10% of the electrode montage. From 75ms post-stimulus onwards, a main effect of type of attended block was observed continuously. Post-hoc pair-wise comparisons via t-tests showed significant differences between Pitch and all the other dimensions in this time-period, as well as differences between location and speaker, and location and syllable. Differences between speaker and syllable were not found to be significant. While this analysis provides a general sense of the timing of ERP modulations, the remainder of the Results section focuses on the findings using the electrical neuroimaging framework described in the Methods. Consistently, analysis of the GFP across the four conditions confirmed what was observed at the waveform level, with statistically significant effects of Condition starting from 120ms post-stimulus onwards.

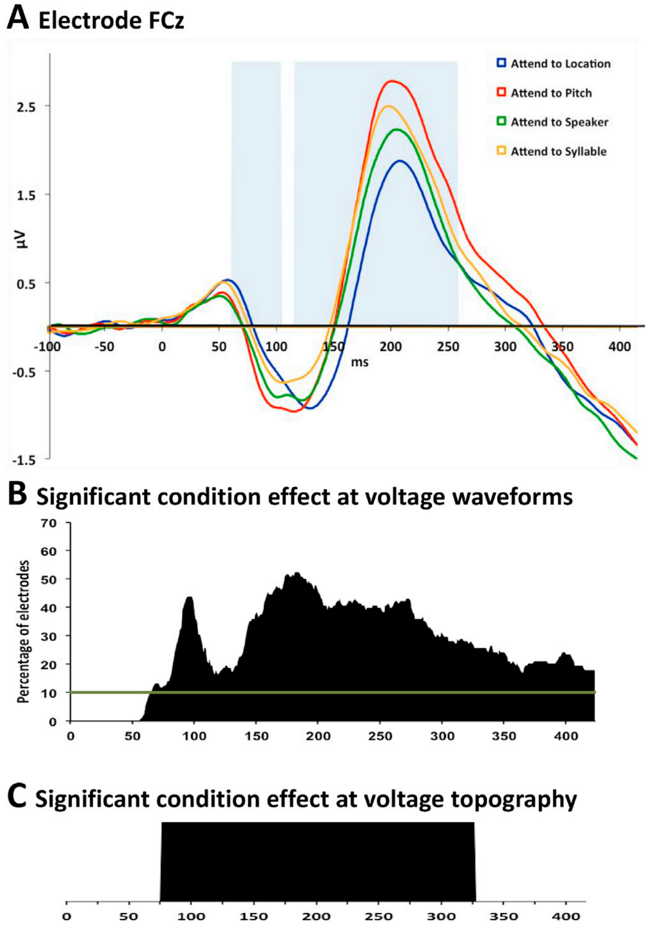


Fig. 2. A. Group-averaged AEPs at an exemplar frontal midline electrode, shown separately for each of the four conditions. The pale blue areas indicate time-periods of statistically significant differences across the 4 conditions. **B.** The results of the millisecond-by-millisecond one-way ANOVA across the electrode montage displaying the percentage of electrodes showing a main effect of condition and meeting the $p < 0.05$ criterion for at least 10ms continuously and across at least 10% of the electrode montage (the area above the green line). **C.** The results of the millisecond-by-millisecond one-way ANOVA on the strength-normalised electric field topography, showing a time-period where the main effect of condition was observed, meeting the $p < 0.05$ criterion for at least 10ms continuously.

Topography analyses across the four conditions, which were based on GMD, revealed significant differences over the 76–330ms post-stimulus period, indicating the activation of distinct configurations of intracranial brain generators, across the 4 different blocks (Fig. 2). Post-hoc t-tests showed significant differences between pitch and position, pitch and speaker as well as between pitch and syllable over two time-periods, first over 75–105ms and then over 130–300ms post-stimulus time-windows. Differences between position and speaker and between position and syllable were observed in the 109–270ms time-window. No significant differences between speaker and syllable were identified. A topographic pattern analysis was then conducted over the full 500ms post-stimulus time-period in order to identify time intervals of stable electric activity at the scalp and determine whether response differences between conditions followed from single or multiple electric configuration changes. This analysis provides a set of so-called “template maps”. Note that a template map refers to a stable ERP topography observed in the group-averaged data that is then used for spatial correlation analyses at the single-subject level across all experimental conditions (the above-mentioned fitting procedure; see Materials and methods for details).

In the present dataset eight different template maps (shown in Fig. 3) accounted for the collective group-averaged dataset. AEPs in four of the six time intervals shown could be captured with just a single template map each, but the 89–170ms and 262–370ms post-stimulus time-periods required two distinct template maps. During both time-periods (89–170ms and 262–370ms) the two identified template maps were observed across all stimulus conditions, albeit appearing for differing relative durations. This pattern observed in the group-averaged ERPs was statistically assessed in the single-subject data using the fitting procedure. This was done separately for these two different time-periods and their respective template maps. The values of the fitting procedure were submitted to a repeated measure ANOVA (one for each time-period) using stimulus condition and template maps as within-subject factors (see bar graphs in Fig. 3). For the first time-period (89–170ms), there was a two-way interaction between condition and map ($F(3,45) = 3.36$, $p < .05$, $\eta_p^2 = 0.19$, $\omega^2 = .72$), which was further tested with post-hoc comparisons, showing that the relative contribution of each map was varying depending on the condition. Also over the second time-period (262–370ms) there was a two-way interaction between condition and map ($F(3,45) = 2.85$, $p < .05$, $\eta_p^2 = 0.17$, $\omega^2 = .64$), showing that the relative contribution of each map was varying depending on the condition (see bar graph on the right in Fig. 3B). Specifically, one template map found to predominate the pitch condition versus any of the other conditions.

Source estimations

Finally, in order to identify the likely brain regions contributing to the differential effects identified during the topographic analyses (89–170ms and 262–370ms), source estimations were calculated over the two time-periods. For this purpose, AEPs for each participant and experimental condition were averaged over i) 89–170ms and ii) 262–370ms periods. The statistical contrasts carried out on these source estimations during the first time-period identified five clusters exhibiting a significant main effect of condition (Fig. 4A). These clusters were located within the right superior parietal lobule (local maximum at 34, -58, 47 mm), the right precuneus (local maximum at 9, -47, 36 mm), the left intraparietal sulcus (IPS) (local maximum at −27, −70, 35 mm), the left superior temporal gyrus (local maximum at −51, −43, 10 mm), and the left inferior temporal lobe (local maximum at −43, −34, −14 mm using the Tailarach and Tournoux (1988) atlas).

Fig. 4B shows axial slices with the results of post-hoc contrasts across pairs of conditions, which allow us to attribute the observed differences

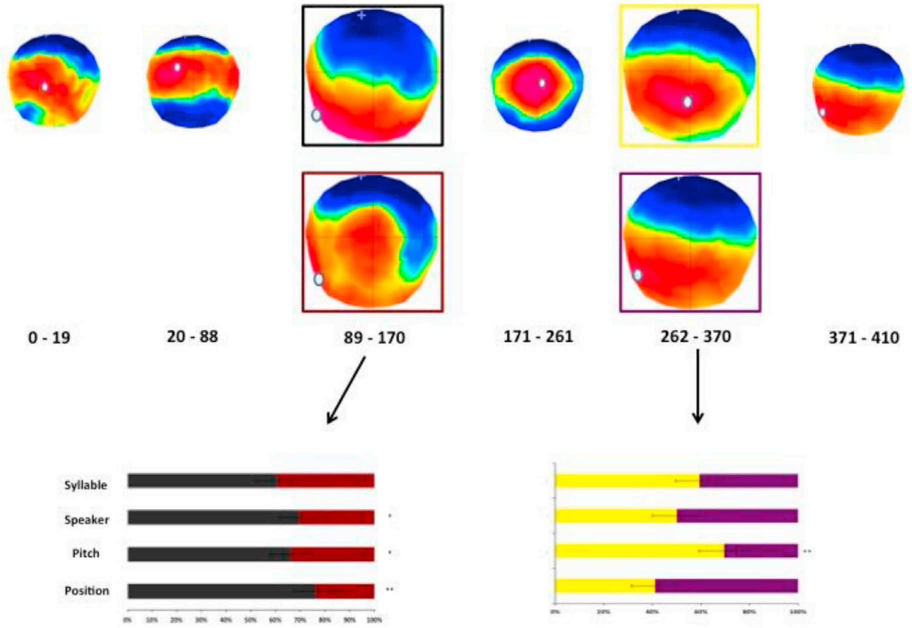


Fig. 3. The topographic pattern analysis identified eight stable topographies (template maps) for all the conditions around the 400ms post-stimulus period. The time-period when each map was observed is indicated. Note that because data are normalized to the instantaneous GFP, the maps are scale-independent and are shown here to illustrate their topographic distribution. One template map was observed in response to all the four conditions in the following time-periods: over the initial 20–88ms post-stimulus period, between 171 and 261ms and during the period after 371ms. At the group-average level, two different template maps were identified over i) the 89–170ms (framed in black and red) and ii) the 262–370ms (framed in yellow and purple) post-stimulus period. The stacked bar graphs show the results of the single-subject fitting procedure, indicating the group-averaged duration each template map was ascribed to in each condition (SEM indicated). Both 2 × 4 repeated measure ANOVAs (one for each time-period) on these duration values revealed an interaction between template map and condition.

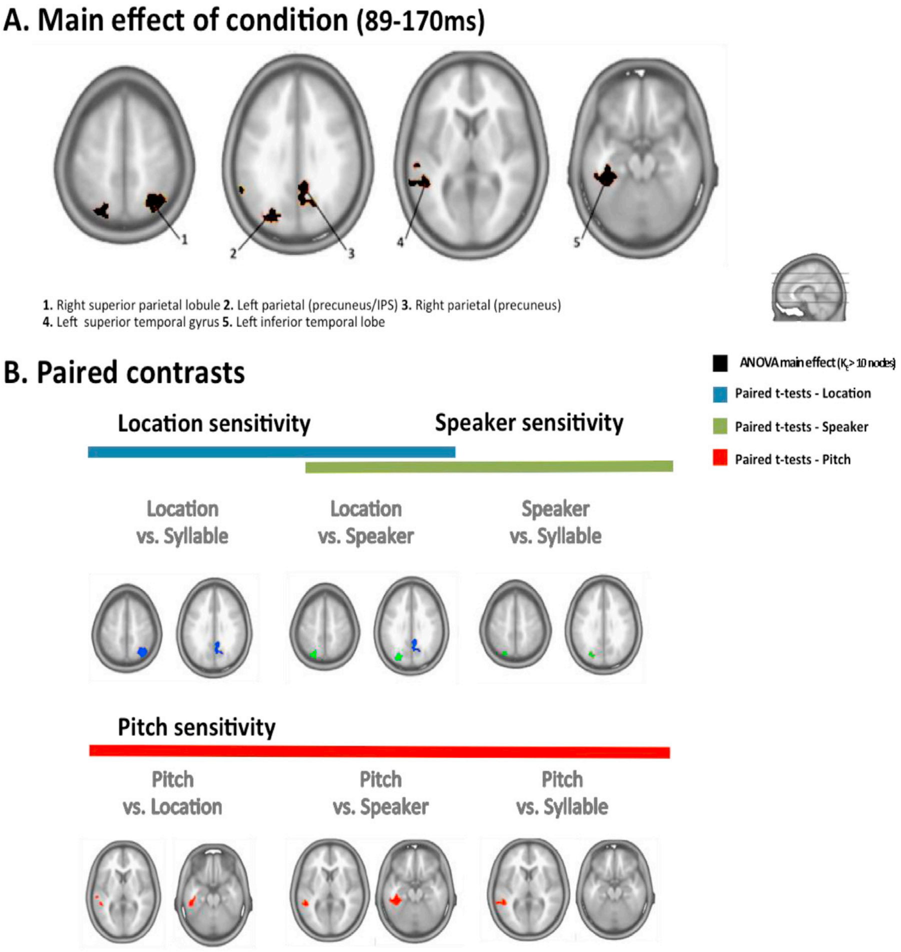

A. Main effect of condition (89-170ms)
1
2
3
4
5
1. Right superior parietal lobule 2. Left parietal (precuneus/IPS) 3. Right parietal (precuneus)
4. Left superior temporal gyrus 5. Left inferior temporal lobe
B. Paired contrasts
ANOVA main effect (K_E > 10 nodes)
Paired t-tests - Location
Paired t-tests - Speaker
Paired t-tests - Pitch
Location sensitivity
Speaker sensitivity
Location vs. Syllable
Location vs. Speaker
Speaker vs. Syllable
Pitch sensitivity
Pitch vs. Location
Pitch vs. Speaker
Pitch vs. Syllable

Fig. 4. A. The results of the one-way ANOVA on source estimates calculated over the 89–170ms post-stimulus time-interval showed a significant main effect of condition within a distributed set of brain regions, indicated here by numbers. Data are shown on axial slices, with the left hemisphere on the left and the nasion upwards. **B.** The results of the post-hoc comparisons (t-values) for the different pairs of conditions. Blue coloured nodes indicate stronger activation for the location condition (within the right superior parietal lobule and the right precuneus), green indicates stronger activation for the speaker condition (within the left IPS) and red indicates stronger activation for the pitch condition (within the left superior temporal gyrus and inferior temporal lobe).

in activation to the various stimulus conditions. Overall, responses were significantly stronger for the Location condition within the right superior parietal lobule and the right precuneus, compared to the other conditions. In addition, responses were significantly stronger for the pitch condition within the left superior temporal gyrus and inferior temporal lobe, whereas stronger activity within the left IPI was observed in response to sounds in the speaker condition. No regions were significantly more active for the syllable condition compared to any of the other conditions.

During the second time-period (262–370ms) six clusters of activation were identified that showed a significant main effect of condition (Fig. 5A). These clusters were located within the right inferior parietal lobule (local maximum at 51, -45, 50 mm), the right middle occipital gyrus (local maximum at 27, -76, −8 mm [or local minimum at 21, -70, −3]), the right middle frontal gyrus (local maximum at 47, 16, 40 mm), the left precuneus/cuneus (local maximum at −15, −82, 35 mm), the left superior temporal gyrus and parietal postcentral gyrus (local maximum at −41, −27, 21 mm) and the left superior and middle temporal gyrus (local maximum at −47, −49, 16 mm). Fig. 5B illustrates the results of post-hoc tests that link these significant effects with comparisons across individual conditions. Stronger activity within the right inferior parietal lobule and the right middle frontal gyrus was observed in response to sounds in the location condition (Fig. 5B). In addition, responses were significantly stronger for the Pitch condition within the two clusters on the left temporal lobe, (Fig. 5B), whereas for the speaker condition stronger activation was observed within the left occipito-parietal area (Fig. 5B). Finally, consistently weaker activation was observed within the right occipital gyrus for the syllable condition compared with the remaining 3 conditions.

Discussion

To verify the presence of sub-segregation within the auditory “what” neural pathway, we compared the spatiotemporal dynamics of brain responses during the discrimination of speech sounds across three distinct object-related dimensions (pitch, syllable type and speaker identity). A fourth, spatial (“where”) dimension was also considered. We chose the spatial, ”where” condition and three “what” dimensions based on the fact that these feature dimensions have been most frequently utilised in extant research on functional segregation of dimensions of sound features. Our results demonstrated robust differences in task-dependent processing of physically identical stimuli, as a function of whether pitch, syllable type, or speaker identity was attended to. We identified two distinct time-periods during which electric field topographies differed between the four attention conditions. The first time-period (89–170ms) included *enhanced* activity in regions selective for attention to i) Pitch: left STG and inferior temporal lobe, ii) Speaker: the left IPS, iii) Location: the right superior parietal lobule and precuneus. The second time-period (262–370ms) demonstrated regions selective for attention to i) Pitch: the left temporal lobe, ii) Speaker: the left occipito-parietal area, iii) Location: the right inferior parietal lobule and right middle frontal gyrus were, iv) Syllable: the right occipital gyrus. Interestingly, this latter activity was *weaker* compared to that in all three remaining conditions.

Task-dependent segregation of “what” and “where” pathways in auditory processing

This sub-segregation of auditory object-related processing began at a similar time-period to the start of the segregation between spatial and non-spatial object dimension processing. During the first time-period (beginning ~90ms post-stimulus), in addition to the topographical differences across the three “what” dimensions, differences were overall observed between the “where” (location) condition and the “what” conditions. The timing of these general “where”/”what” differential effects was similar to the timing of the effects observed in previous ERP and MEG studies (Herrmann et al., 2002; Ahveninen et al., 2006; De Santis et al., 2007; Leavitt et al., 2011). All those studies (including the present study) showed differential processing of object-related versus spatial information starting ~100ms post-stimulus, despite the use of different types of paradigms - passive (De Santis et al., 2007) versus active tasks (including present study; Herrmann et al., 2002; Ahveninen et al., 2006; Leavitt et al., 2011), and types of sounds - band-pass filtered noises, environmental sounds, animal calls, and vowels. Our study provides an

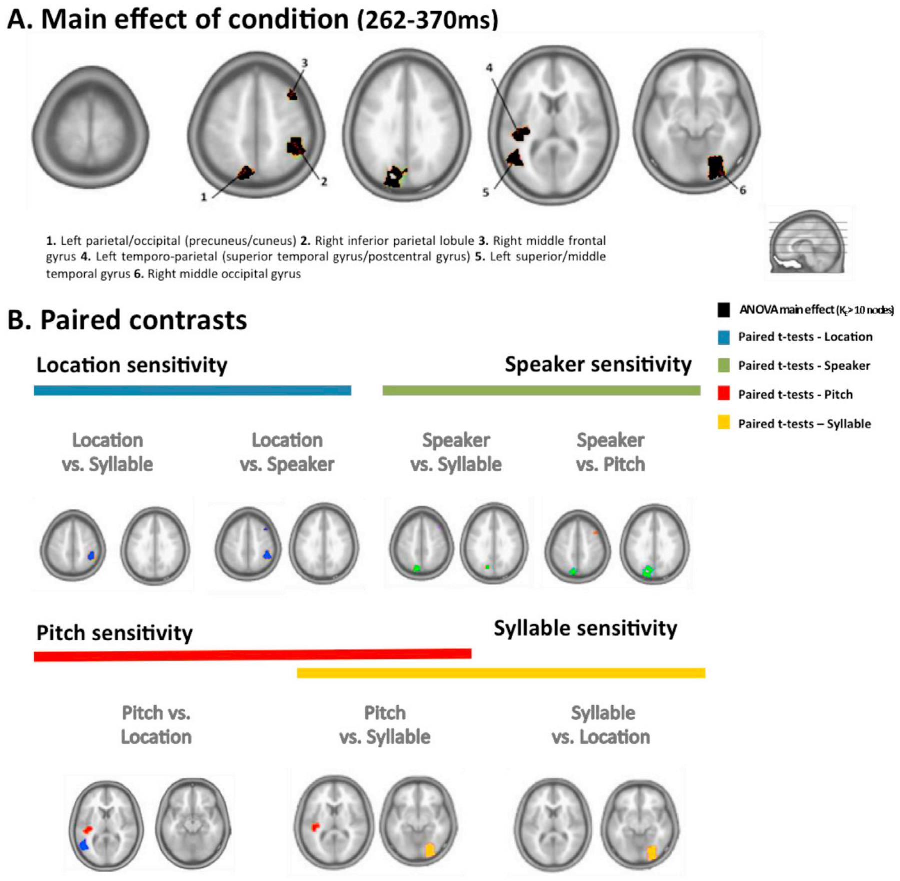

A. Main effect of condition (262-370ms)
3
4
1
2
5
6
1. Left parietal/occipital (precuneus/cuneus) 2. Right inferior parietal lobule 3. Right middle frontal gyrus 4. Left temporo-parietal (superior temporal gyrus/postcentral gyrus) 5. Left superior/middle temporal gyrus 6. Right middle occipital gyrus
B. Paired contrasts
Location sensitivity
Speaker sensitivity
Location vs. Syllable
Location vs. Speaker
Speaker vs. Syllable
Speaker vs. Pitch
Pitch sensitivity
Syllable sensitivity
Pitch vs. Location
Pitch vs. Syllable
Syllable vs. Location
■ ANOVA main effect (K_E>10 nodes)
■ Paired t-tests - Location
■ Paired t-tests - Speaker
■ Paired t-tests - Pitch
■ Paired t-tests – Syllable

Fig. 5. A. The results of the one-way ANOVA on source estimates calculated over the 262–370ms post-stimulus interval showed a significant main effect of condition within a distributed set of brain regions, indicated here by numbers. Data are shown on axial slices with the left hemisphere on the left and the nasion upwards. **B.** The results of the post-hoc comparisons (t-values) for the different pairs of conditions. Blue coloured nodes indicate stronger activation for the location condition (within the right inferior parietal lobule and the right middle frontal gyrus), green indicates stronger activation for the speaker condition (within the left occipito-parietal area) and red indicates stronger activation for the pitch condition (two clusters within the left temporal lobe). Yellow indicates weaker activation for the syllable condition (within the right occipital gyrus).

important confirmation of thesse previous findings, as it utilised a robust, reference-independent analytical methods, which were able to identify both the timing of the differential "what"/"where" processing as well as its brain substrates (see also De Santis et al., 2007 as well as Bidet-Caulet and Bertrand, 2005 for similar examples).

Specifically, our source estimations showed that across both time-periods with topographic differences regions within the right parietal cortex were selectively involved in spatial auditory processing. In the second time-window, the right middle frontal cortex was identified as an additional source selective for spatial processing. Our findings are consistent with previous studies reporting the involvement of these regions in spatial features of auditory stimuli (De Santis et al., 2007; Ducommun et al., 2002, 2004; Alain et al., 2001; Leavitt et al., 2011; Tervaniemi and Hugdahl, 2003 for review). The right lateralization of these effects is also consistent with several previous studies that suggest greater involvement of the right hemisphere in sound localization (Hermann et al., 2002; Anourova et al., 2001; Mathiak et al., 2007).

Task-dependent subsegregation of the "what" pathway in auditory processing

This study constitutes the first demonstration in humans detailing the differences in the spatiotemporal patterns of brain activation during the processing of sounds across multiple object-related feature dimensions. In contrast to the right-lateralized nature of brain areas involved in space-related sound processing, processing of sounds along the three "what" dimensions more strongly activated mainly brain regions within the left hemisphere. The left lateralization of the different "what" dimensions could be attributed to the use of speech sounds in the present study, which likely evoked stronger responses in regions specialized for verbal processing. Previous studies using vocal speech sounds have shown activations within left auditory cortices when contrasted with non-vocal sounds of similar acoustic spectra (Griffiths and Warren, 2002). Similarly, accurate vowel discrimination has been linked to left superior temporal cortex activations (Bonte et al., 2014). All previous studies in humans in this domain have focused on only two dimensions at once, e.g., pitch versus phonetic discrimination. However, outside of the laboratory, sounds typically vary across multiple acoustic dimensions at the same time (Walker et al., 2011). Thus, to ensure ecological validity of findings on functional segregation of sound dimensions, it is essential to establish how the auditory cortex represents sounds across these feature dimensions when the dimensions vary simultaneously. Only one study in non-human animals has manipulated more than two sound dimensions at once, and reported for the majority of neurons across the ferret auditory cortex to be sensitive to variations in sound pitch, timbre as well as location (Bizley et al., 2009). Our findings are the first to verify the extent to which these previous findings in the ferret generalise to humans. We discuss them now in more detail.

For pitch, the discrimination was associated with selective activation within the left STG and MTG across both time-periods and additionally with activity within the left inferior temporal lobe only during the first time-period. These results are consistent with Griffiths and Warren (2002) study showing STG activity related to pitch processing. A few previous studies that investigated the processing of sounds varied across different "what" dimensions have also observed involvement of the STG (Formisano et al., 2008; Kilian-Hutten et al., 2011; Bonte et al., 2014).

For speaker identity, the processing was selectively associated with activity in sources within the left IPS. The human IPS has previously been shown to be involved in abstract representation of size. One fMRI study has reported left IPS activity in response to changes in the acoustic scale of resonant sounds from different object categories – including human voices (von Kriegstein et al., 2007). Our study employed stimuli representing voice of two male speakers, an adult and a child. The size of a sound source often affects the sound it produces, and the acoustic scales of voices of our two speakers differed considerably. Therefore, it seems plausible to assume that the activation of the left IPS reflects the representation of speaker size. This is consistent with the previously formed hypothesis that the IPS in humans deals with a supramodal representation of size (Vogel et al., 2017).

For syllable, the discrimination involved a pattern distinct to the one found for the other two "what" feature dimensions. We observed the main effect of condition, including distinct effects in the processing of sound across the "what" feature dimensions, across two distinct time-periods. These results suggest two stages of differential processing and, by extension, of functional subsegregation. While distinct topographies of AEPs to sound across pitch and speaker identity conditions were observed across both time-periods, syllable processing elicited differential processing by consistently weaker brain activations, within the right occipital cortex, but this activation was observed exclusively in the second time-period. The weaker activation within the occipital cortex would suggest that syllable discrimination involved visual suppression. In support of this idea, previous studies have shown that passive presentation of auditory stimuli can induce de-activation of regions of the visual cortex (Laurienti et al., 2002; Amaral and Langers, 2013). More recently, Amaral and Langers (2015) reported suppression of activity in the primary visual cortex during an active auditory task involving discrimination of consonant-vowel syllables – similarly to our experiment. It will be important for future research to verify the importance of this occipital activity suppression for auditory perception. More generally, the link between the network of regions identified in our study and auditory perception remains to be determined and will require parametric, but equated, variation of task difficulty of discriminations of attended sounds across the feature dimensions. Previous studies investigating task-dependent speaker and vowel - or syllable - processing (Bonte et al., 2009, 2014; Schall et al., 2015; Obleser et al., 2004; von Kriegstein et al., 2007) have observed differential activations related to speaker and syllable discrimination, often within the temporal cortex. However, the timing of these effects have been somewhat inconsistent (i.e. >300ms in Bonte et al., 2009, vs. ~200msmspost-stimulus in Schall et al., 2015). By contrast, the localization of those effects seems to be reliably linked to the STS and its subdivisions. One possibility as to why similar loci were less forthcoming here is that our post-hoc analyses were masked by the loci exhibiting a main effect of condition across the four stimulus dimensions (whereas prior studies involved contrasts between two conditions). When no such masking is applied, we do in fact observe results similar to the published literature. Within the anterior STS bilaterally, source estimations were stronger in response to the speaker than syllable condition over the second time-period (262–370ms). Likewise, during the first time-period (69–170ms) source estimations were stronger in response to the speaker than syllable dimension within the right STG. The overall present pattern of results indicates the brain networks that are selectively activated for different sound dimensions when multiple dimensions vary simultaneously and when attention is parametrically directed to each of them. The current methods have not only used robust, reference-independent measures of brain activity, but are also more temporally resolved than methods used previously. Again, this underscores the added benefits of electrical neuroimaging in studying functional segregation of auditory processing, especially when combined with paradigms emulating the information processing demands characteristic of naturalistic environments (see, e.g., Krakauer et al., 2017, Matusz et al., under review, for similar ideas).

The importance of selective attention in uncovering functional segregation in the "what" auditory pathway

Some recent studies comparing the processing of different auditory object-related dimensions - such as pitch versus timbre, or speaker versus vowel - have provided evidence against the anatomical sub-segregation of neural processing across the "what" object dimensions. Specifically, Allen et al. (2017) used fMRI to investigate whether spatially distinct regions are selectively responsive to changes in pitch versus timbre. The results of mass univariate analyses revealed no clear differences between auditory cortical regions dedicated to processing of these two dimensions. However, results of a multivoxel pattern analysis (MVPA) of the same data reported in the same study have suggested that indeed distinct sub-circuits within the same cortical regions might be engaged by information processing across the two dimensions. A combined MEG/EEG study by Gutschalk and Uppenkamp (2011), led to conclusions similar to those found with a mass univariate analyses by Allen et al. Overlapping responses in the antero-lateral Heschl's gyrus were observed when pitch versus vowel of the stimuli were manipulated, suggesting again a lack of spatial segregation across the "what" auditory pathway. Also, the single-unit study by Bizley et al. (2009) that employed stimuli that varied in pitch, spectral envelope peak (corresponding to timbre) and spatial location, came to similar conclusions. They demonstrated that most neurons across the ferret's auditory cortex are sensitive to sound features across multiple perceptual dimensions, with over 2/3 of the recorded units responding to at least two dimensions *(most often, pitch and timbre).

However, it should be noted that all these studies employed passive listening tasks that potentially have made it more difficult for participants to perceptually separate the examined dimensions, compared to studies that employed active tasks. In contrast, the few relatively recent studies that have employed active tasks, directing participants' attention towards one sound feature dimension versus another, have indeed reported evidence for sub-segregation within the "what" auditory pathway (Bonte et al., 2009, 2014; Obleser et al., 2004; Schall et al., 2015, von Kriegstein et al., 2007; see also Ahveninen et al., 2006 for similar findings with auditory "what"/"where" network differences). The overall evidence supports the potential importance of selective attention as means for revealing distinct networks governing the processing of specific auditory object feature dimensions (see also Bidet-Caulet and Bertrand, 2005).

This contribution of selective attention is what we observed in our study; distinct activation patterns were identified as being involved in the processing of pitch, speaker identity and syllable type, when attention of the participants was directed to each of these features in separate blocks. Stimuli in each block of trials were identical, albeit varying across all four dimensions (pitch, speaker, syllable and location). However, participants were required only to attend to one of these dimensions during each block. It remains to be seen whether functional segregation of brain regions responsible for the processing of features from the distinct "what" dimensions is dependent on attention, or whether similar effects would occur regardless of task demands. In order to do so, a more stimulus-driven analysis of the current data could be performed in the future, looking also at the processing of task-irrelevant features and the interactions between attended and unattended dimensions.

Conclusions

This is the first study to demonstrate how the selective activation for different sound object dimensions can be readily identified by combining parametric manipulations of top-down attention towards specific object dimensions with their spatiotemporal brain response patterns revealed by EEG/ERP electrical neuroimaging. Specifically, we have revealed that distinct cortical networks are preferentially involved as early as 90ms post-stimulus onset depending on whether a listener attends to the identity of the speaker, the pitch of their voice or the spatial location where the speech is coming from. In summary, conceptual models of auditory processing must take fuller consideration of the multiple sound object-related feature dimensions that receive segregated, albeit parallel, treatment.

Role of the funding sources

Financial support was provided by the Swiss National Science Foundation (grants: 320030_149982 and 320030_169206 to M.M.M., PZ00P1_174150 to P.J.M., and the National Centre of Competence in Research project "SYNAPSY, The Synaptic Bases of Mental Disease" [project 51NF40–158776]). P.J.M. receives support from the Pierre Mercier Foundation. P.J.M. and M.M.M. are both supported by Fondation Asile des Aveugles. Also, this work has been supported by a "Royal Society International Exchange Project support grant (to J.S. and M.M.M.). These sources had no further role in this study design, in the data collection and analysis, in the writing of the report, and in the decision to submit the paper for publication.

Conflicts of interest

Authors report no conflict of interest.

References

Ahveninen, J., Jääskeläinen, L.P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., Levänen, S., Lin, F.H., Sams, M., Shinn-Cunningham, B.G., Witzel, T., Belliveau, J.W., 2006. Task-modulated "what" and "where" pathways in human auditory cortex. PNAS 103 (39), 14608–14613.

Alain, C., Arnott, S.R., Hevenor, S., Graham, S., Grady, C.L., 2001. "What" and "where" in the human auditory system. PNAS 98 (21), 12301–12306.

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.

Alho, K., Vorobyev, V.A., Medvedev, S.V., Pakhomov, S.V., Roudas, M.S., Tervaniemi, M., van Zuijen, T., Näätänen, R., 2003. Hemispheric lateralization of cerebral blood-flow changes during selective listening to dichotically presented continuous speech. Cognitive Brain Res. 17 (2), 201–211.

Amaral, A.A., Langers, D.R.M., 2015. Tinnitus-related abnormalities in visual and salience networks during a one-back task with distractors. Hear. Res. 326, 15–29.

Amaral, A.A., Langers, D.R.M., 2013. The relevance of task-irrelevant sounds: hemispheric lateralization and interactions with task-relevant streams. Front. Neurosci. 7, 264.

Anourova, I., Nikouline, V.V., Ilmoniemi, R.J., Hotta, J., Aronen, H.J., Carlson, S., 2001. Evidence for dissociation of spatial and nonspatial auditory information processing. Neuroimage 14, 1268–1277.

Anourova, I., Artchakov, D., Korvenoja, A., Ilmoniemi, R.J., Aronen, H.J., Carlson, S., 2003. Differences between auditory evoked responses recorded during spatial and nonspatial working memory tasks. Neuroimage 20, 1181–1192.

Belin, P., Fecteau, S., Bédard, C., 2004. Thinking the voice: neural correlates of voice perception. Trends Cognitive Sci. 8 (3), 129–135.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309–312.

Bidet-Caulet, A., Bertrand, O., 2005. Dynamics of a temporo-fronto-parietal network during sustained spatial or spectral auditory processing. J. Cong. Neurosci. 17 (11), 1691–1703.

Bidet-Caulet, A., Voisin, J., Bertrand, O., Fonlupt, P., 2005. Listening to a walking human activates the temporal biological motor area. Neuroimage 28, 132–139.

Bizley, J.K., Walker, K.M.M., Silverman, B.W., King, A.J., Schnupp, J.W.H., 2009. Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. J. Neurosci. 29 (7), 2064–2075.

Bonte, M., Hausfeld, L., Scharke, W., Valente, G., Formisano, E., 2014. Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. J. Neurosci. 34 (13), 4548–4557.

Bonte, M., Valente, G., Formisano, E., 2009. Dynamic and task-dependent encoding of speech and voice by phase reorganization of cortical oscillations. J. Neurosci. 29 (6), 1699–1706.

Bruneau, N., Roux, S., Cléry, H., Rogier, O., Bidet-Caulet, A., Barthélémy, C., 2013. Early neurophysiological correlates of vocal versus non-vocal sound processing in adults. Brain Res. 1528, 20–27. https://doi.org/10.1016/j.brainres.2013.06.008.

Brunet, D., Murray, M.M., Michel, C.M., 2011. Spatiotemporal analysis of multichannel EEG: CARTOOL. Comput. Intell. Neurosci. https://doi.org/10.1155/2011/813870.

Clarke, S., Bellmann Thiran, A., Maeder, P., Adriani, M., Vernet, O., Regli, L., Cuisenaire, O., Thiran, J.P., 2002. What and where in Human Audition : Selective Deficits Following Focal Hemispheric Lesions.

De Lucia, M., Clarke, S., Murray, M.M., 2010. A temporal hierarchy for conspecific vocalization discrimination in humans. J. Neurosci. 30 (33), 11210–11221.

Deouell, L.Y., Bentin, S., Giard, M.H., 1998. Mismatch negativity in dichotic listening: evidence for interhemispheric differences and multiple generators. Psychophysiology 35, 355–365.

De Santis, L., Clarke, S., Murray, M.M., 2007. Automatic and intrinsic auditory "what" and "where" processing in humans revealed by electrical neuroimaging. Cereb. Cortex 17, 9–17.

Ducommun, C.Y., Michel, C.M., Clarke, S., Adriani, M., Seeck, M., Landis, T., Blanke, O., 2004. Cortical motion deafness. Neuron 43, 765–777.

Ducommun, C.Y., Murray, M.M., Thut, G., Bellmann, A., Viaud-Delmon, I., Clarke, S., Michel, C.M., 2002. Segregated processing of auditory motion and auditory location : an ERP mapping study. Neuroimage 16, 76–88.

Engel, L.R., Frum, C., Puce, A., Walker, N.A., Lewis, J.W., 2009. Different categories of living and non-living sound-sources activate distinct cortical networks. Neuroimage 47 (4), 1778–1791.

Formisano, E., De Martino, F., Bonte, M., Goebel, R., 2008. "Who" is saying "what"? Brain-based decoding of human voice and speech. Science 322, 970–973.

Giard, M.H., Lavikainen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J., Näätänen, R., 1995. Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory. An event-related-potential and dipole-model analysis. J. Cogn. Neurosci. 7, 133–143.

Gonzalez Andino, S., Murray, M.M., Foxe, J.F., Grave de Peralta Menendez, R., 2005. How single trial neuroimaging contributes to multisensory research. Exp. Brain Res. 166, 298–304.

Grave de Peralta Menendez, R., Gonzalez Andino, S., Lantz, G., Michel, C.M., Landis, T., 2001. Noninvasive localization of electromagnetic epileptic activity. I. Method descriptions and simulations. Brain Topogr. 14, 131–137.

Grave de Peralta Menendez, R., Murray, M.M., Michel, C.M., Martuzzi, R., Gonzalez Andino, S.L., 2004. Electrical neuroimaging based on biophysical constraints. NeuroImage 21, 527–539. https://doi.org/10.1016/j.neuroimage.2003.09.051.

Griffiths, T.D., Hall, D.A., 2012. Mapping pitch representation in neural ensembles with fMRI. J. Neurosci. 32, 13343–13347.

Griffiths, T.D., Warren, J.D., 2002. The planum temporale as a computational hub. Trends Neurosci. 25 (7), 348–353.

Guthrie, D., Buchwald, J.S., 1991. Significance testing of difference potentials. Psychophysiology 28, 240–244.

Gutschalk, A., Uppenkamp, S., 2011. Sustained responses for pitch and vowels map to similar sites in human auditory cortex. Neuroimage 56, 1578–1587.

Hackett, T.A., 2015. Anatomic organization of the auditory cortex. Hanb. Clin. Neurol. 129, 27–53.

Hay, R.A., Roach, B.J., Srihari, V.H., Woods, S.W., Ford, J.M., Mathalon, D.H., 2015. Equivalent mismatch negativity deficits across deviant types in early illness schizophrenia-spectrum patients. Biol. Psychol. 105, 130–137.

Herrmann, C.S., Senkowski, D., Maess, B., Friederici, A.D., 2002. Spatial versus object feature processing in human auditory cortex: a magnetoencephalographic study. Neurosci. Lett. 334, 37–40.

Hillyard, S.A., Hink, R.F., Schwent, V.L., Picton, T.W., 1973. Electrical signs of selective attention in the human brain. Science 182, 177–180.

Ives, D.T., Smith, D.R., Patterson, R.D., 2005. Discrimination of speaker size from syllable phrases. J. Acoust. Soc. Am. 118, 3816.

Jin, C.T., Guillon, P., Epain, N., Zolfaghari, R., van Schaik, A., Tew, A.I., Hetherington, C., Thorpe, J., 2014. Creating the Sydney York morphological and acoustic recordings of ears database, Multimedia. IEEE Trans. 16, 37–46.

Kaas, J.H., Hackett, T.A., 1999. "What" and "where" processing in auditory cortex. Nat. Neurosci. 2 (12), 1045–1047.

Kawahara, H., Irino, T., 2005. Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In: Speech Separation by Humans and Machines. Springer.

Kilian-Hütten, N., Valente, G., Vroomen, J., Formisano, E., 2011. Auditory cortex encodes the perceptual interpretation of ambiguous sound. J. Neurosci. 31 (5), 1715–1720.

Knebel, J.F., Murray, M.M., 2012. Towards a resolution of conflicting models of illusory contour processing in humans. Neuroimage 59, 2808–2817.

Koenig, T., Stein, M., Grieder, M., Kottlow, M., 2014. A tutorial on data-driven methods for statistically assessing ERP topographies. Brain Topogr. 27, 72–83.

Koenig, T., Kottlow, M., Stein, M., Melie García, L., 2011. Ragu: a free tool for the analysis of EEG and MEG event-related scalp field data using global randomization statistics. Comput. Intell. Neurosci., 938925

Krakauer, J.W., Ghazanfar, A.A., Gomez-Marin, A., MacIver, M.A., Poeppel, D., 2017. Neuroscience needs behaviour: correcting a reductionist bias. Neuron 93, 480–490.

Laurienti, P.J., Burdette, J.H., Wallace, M.T., Yen, Y.F., Field, A.S., Stein, B.E., 2002. Deactivation of sensory-specific cortex by cross-modal stimuli. J. Cogn. Neurosci. 14, 420–429.

Leavitt, V.M., Molholm, S., Gomez-Ramirez, M., Foxe, J.J., 2011. "What" and "where" in auditory sensory processing: a high-density electrical mapping study of distinct neural processes underlying sound object recognition and sound localization. Front. Integr. Neurosci. 5, 23.

Lehmann, D., 1987. Principles of spatial analysis. In: Gevins, A., Remond, A. (Eds.), Methods of Analysis of Brain Electrical and Magnetic Signals: Handbook of Electroencephalography and Clinical Neurophysiology, vol. 1. Elsevier, Amsterdam, pp. 309–354.

Lehmann, D., Skrandies, W., 1980. Reference-free identification of components of checkerboard-evoked multichannel potential fields. Electroencephalogr. Clin. Neurophysiol. 48, 609–621.

Lewald, J., Wienemann, M., Boroojerdi, B., 2004. Shift in sound localization induced by rTMS oft he posterior parietal lobe. Neuropsychologia 42, 1598–1607.

Lewis, J.W., Brefczynski, J.A., Phinney, R.E., Janik, J.J., DeYoe, E.A., 2005. Distinct cortical pathways for processing tool versus animal sounds. J. Neurosci. 25 (21), 5148–5158.

Lomber, S.G., Malhotra, S., 2008. Double dissociation of 'what' and 'where' processing in auditory cortex. Nat. Neurosci. 11 (5), 609–616.

Maeder, P., Meuli, R., Bellmann, A., Fornari, E., Thiran, J.P., Pittet, A., Clarke, S., 2001. Distinct pathways involved in sound recognition and localization: a human fMRI study. Neuroimage 14, 802–816.

Mathiak, K., Menning, H., Hertrich, I., Mathiak, K.A., Zvyagintsev, M., Ackermann, H., 2007. Who is telling what from where? A functional magnetic resonance imaging study. Neuroreport 18, 405–409.

Martuzzi, R., Murray, M.M., Reto, A.M., Thiran, J.P., Maeder, P.P., Michel, C.M., Grave de Peralta Menendez, R., Gonzalez Andino, S.L., 2009. Methods for determining frequency- and region- dependent relationships between estimated LFPs and BOLD responses in Humans. J. Neurophysiol. 101, 491–502.

Matusz, P.J., Dikker, S., Huth, A.G., Perrodin, C., (under review). Are we ready for real-world neuroscience? J. Cogn. Neurosci..

Matusz, P.J., Retsa, C., Murray, M.M., 2016. The context -contingent nature of cross-modal activations of the visual cortex. Neuroimage 125, 996–1004.

Matusz, P.J., Thelen, A., Amrein, S., Geiser, E., Anken, J., Murray, M.M., 2015. The role of auditory cortices in the retrieval of single-trial auditory-visual object memories. Eur. J. Neurosci. 41, 699–708.

Michel, C.M., Murray, M.M., 2012. Towards the utilization of EEG as a brain imaging tool. NeuroImage 61, 371–385. https://doi.org/10.1016/j.neuroimage.2011.12.039.

Michel, C.M., Murray, M.M., Lantz, G., Gonzalez, S., Spinelli, L., Grave de Peralta, R., 2004. EEG source imaging. Clin. Neurophysiol. 115, 2195–2222. https://doi.org/10.1016/j.clinph.2004.06.001.

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., Zilles, K., 2001. Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. Neuroimage 13, 684e701.

Murray, M.M., Brunet, D., Michel, C.M., 2008. Topographic ERP analyses: a step-by-step tutorial review. Brain Topogr. 20, 249–264. https://doi.org/10.1007/s10548-008-0054-5.

Murray, M.M., Camen, C., Gonzalez Andino, S.L., Bovet, P., Clarke, S., 2006. Rapid brain discrimination of sounds of objects. J. Neurosci. 26 (4), 1293–1302.

Obleser, J., Elbert, T., Eulitz, C., 2004. Attentional influences on functional mapping of speech sounds in human auditory cortex. BMC Neurosci. 5, 24.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9 (1), 97–113.

Paavilainen, P., 2013. The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: a review. Int. J. Psychophysiol. 88 (2), 109–123.

Paavilainen, P., Simola, J., Jaramillo, M., Näätänen, R., Winkler, I., 2001. Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN). Psychophysiology 38, 359–365.

Paltoglou, A.E., Sumner, C.J., Hall, D.A., 2011. Mapping feature-sensitivity and attentional modulation in human auditory cortex with functional magnetic resonance imaging. Eur. J. Neurosci. 33 (9), 1733–1741.

Petkov, C., Kang, X., Alho, K., Bertrand, O., Yund, E., Woods, D., 2004. Attentional modulation of human auditory cortex. Nat. Neurosci. 7, 658–663.

Perrin, F., Pernier, J., Bertrand, O., Giard, M.H., Echallier, J.F., 1987. Mapping of scalp potentials by surface spline interpolation. Electorencephalogr. Clin. Neurophysiol. 66, 75–81.

Perrodin, C., Kayser, C., Abel, T.J., Logothetis, N.K., Petkov, C.I., 2015. Who is that ? Brain networks and mechanisms for identifying individuals. Trends Cogn. Sci. 19 (12), 783–796.

Pourtois, G., Delplanque, S., Michel, C., Vuilleumier, P., 2008. Beyond conventional event-related brain potential (ERP): exploring the time-course of visual emotion processing using topographic and principal component analyses. Brain Topogr. 20 (4), 265–277.

Rama, P., Paavilainen, L., Anourova, I., Alho, K., Reinikainen, K., Sipila, S., Carlson, S., 2000. Modulation of slow brain potentials by working memory load in spatial and nonspatial auditory tasks. Neuropsychologia 38 (7), 913–922.

Rauschecker, J.P., Tian, B., 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. PNAS 97 (22), 11800–11806.

Romanski, L.M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P.S., Rauscheker, J.P., 1999. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat. Neurosci. 2 (12), 1131–1136.

Schairer, K.S., Gould, H.J., Pousson, M.A., 2001. Source generators of mismatch negativity to multiple deviant stimulus types. Brain Topogr. 14 (2), 117–130.

Schall, S., Kiebel, S.J., Maess, B., von Kriegstein, K., 2015. Voice identity recognition: functional division of the right STS and its behavioral relevance. J. Cogn. Neurosci. 27 (2), 280–291.

Scott, S.K., 2005. Auditory processing – speech, space and auditory objects. Curr. Opin. Neurobiol. 15 (2), 197–201.

Spinelli, L., Andino, S.G., Lantz, G., Seeck, M., Michel, C.M., 2000. Electromagnetic inverse solutions in anatomically constrained spherical head models. Brain Topogr. 13, 115–125.

Tailarach, J., Tournoux, P., 1988. Co-planar Stereotaxic Atlas of the Human Brain. Thieme, New York.

Tardif, E., Spierer, L., Clarke, S., Murray, M.M., 2008. Interactions between auditory 'what' and 'where' pathways revealed by enhanced near-threshold discrimination of frequency and position. Neuropsychologia 46, 958–966.

Tardif, E., Murray, M.M., Meylan, R., Spierer, L., Clarke, S., 2006. The spatio-temporal brain dynamics of processing and integrating sound localization cues in humans. Brain Res. 1092, 161–176.

Tervaniemi, M., Hugdahl, K., 2003. Lateralization of auditory-cortex functions. Brain Res. Rev. 43, 231–246.

Tian, B., Reser, D., Durham, A., Kustov, A., Rauschecker, J.P., 2001. Functional specialization in rhesus monkey auditory cortex. Science 292, 290–293.

Toepel, U., Knebel, J.F., Hudry, J., le Coutre, J., Murray, M.M., 2009. The brain tracks the energetic value in food images. Neuroimage 44, 967–974.

Tzovara, A., Murray, M.M., Bourdaud, N., Chavarriaga, R., Millán, J.D.R., Lucia, M.D., 2012. The timing of exploratory decision-making revealed by single-trial topographic EEG analyses. Neuroimage 60, 1959–1969.

Vogel, S.E., Goffin, C., Bohnenberger, J., Koschutnig, K., Reishofer, G., Grabner, R.H., Ansari, D., 2017. The left intraparietal sulcus adapts to symbolic number in both the visual and auditory modalities: evidence from fMRI. Neuroimage 153, 16–27.

von Kriegstein, K., Smith, D.R.R., Patterson, R.D., Ives, D.T., Griffiths, T.D., 2007. Neural representation of auditory size in the human voice and in sounds from other resonant sources. Curr. Biol. 17, 1123–1128.

Walker, K.M.M., Bizley, J.K., King, A.J., Schnupp, J.W.H., 2011. Multiplexed and robust representations of sound features in auditory cortex. J. Neurosci. 31 (41), 14565–14576.

Warren, J.D., Griffiths, T.D., 2003. Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. J. Neurosci. 23 (13), 5799–5804.

Woldorff, M.G., Gallen, C.C., Hampson, S.A., Hillyard, S.A., Pantev, C., Sobel, D., Bloom, F.E., 1993. Modulation of early sensory processing in human auditory cortex during auditory selective attention. PNAS U. S. A. 90, 8722–8726.

Zatorre, R.J., Belin, P., Penhune, V.B., 2002. Structure and function of auditory cortex: music and speech. Trends Cogn. Sci. 6 (1), 37–46.

Zatorre, R.J., Mondor, T.A., Evans, A.C., 1999. Auditory attention to space and frequency activates similar cerebral systems. Neuroimage 10 (5), 544–554.

Zatorre, R.J., Evans, A.C., Meyer, E., 1994. Neural mechanisms underlying melodic perception and memory for pitch. J. Neurosci. 14 (4), 1908–1919.

Zatorre, R.J., Evans, A.C., Meyer, E., Gjedde, A., 1992. Lateralization of phonetic and pitch discrimination in speech processing. Science 256 (5058), 846–849.