# Audiovisual adaptation is expressed in spatial and decisional codes

Máté Aller[1,2]*, Agoston Mihalik[1]*, Uta Noppeney[1,3]

[1] Computational Neuroscience and Cognitive Robotics Centre, University of Birmingham, Birmingham, UK

[2] MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK

[3]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands

*These authors contributed equally to this work.

Corresponding author:
Máté Aller
mate.aller@mrc-cbu.cam.ac.uk
MRC Cognition and Brain Sciences Unit
University of Cambridge
Cambridge
UK

## Abstract

The brain adapts dynamically to the changing sensory statistics of its environment. The neural circuitries and representations that support this cross-sensory plasticity remain controversial. We combined psychophysics and model-based representational fMRI and EEG to characterize how the adult human brain adapts to misaligned audiovisual signals. We show that audiovisual adaptation moulds regional BOLD-responses and fine-scale neural representations in a widespread network from Heschl's gyrus to dorsolateral prefrontal cortices. Crucially, audiovisual recalibration relies on distinct spatial and decisional codes that are expressed with opposite gradients and timecourses across the auditory processing hierarchy. Early activity patterns in auditory cortices encode sounds in a continuous space that flexibly adapts to misaligned visual inputs. Later activity patterns in frontoparietal cortices code decisional uncertainty consistent with these spatial transformations. Our findings demonstrate that regions throughout the auditory processing hierarchy multiplex spatial and decisional codes to adapt flexibly to the changing sensory statistics in the environment.

## Keywords

audiovisual recalibration, hemifield code, decisional uncertainty, auditory processing pathway, fMRI, EEG, multivariate decoding, pattern component modelling

## Introduction

Throughout life the brain needs to adapt dynamically to changes in the environment and the sensorium. Changes in the sensory statistics evolve across multiple timescales ranging from milliseconds to years and bring cues from different sensory modalities into conflict. Most notably, physical growth, ageing or entering a room with reverberant acoustics can profoundly alter the sensory cues that guide the brain's construction of spatial representations. To maintain auditory and visual spatial maps in co-registration the brain needs to constantly recalibrate the senses (Chen and Vroomen, 2013).

Recalibrating auditory and visual spatial maps is particularly challenging, because the two sensory systems encode space not only in different reference frames (i.e., eye vs. head-centred), but also in different representational formats (Grothe et al., 2010; Kopčo et al., 2009; Maier and Groh, 2009). In vision, spatial location is encoded directly in the sensory epithelium and retinotopic maps of visual cortices ('place code') (Wandell et al., 2007). In audition, azimuth location is computed mainly from interaural time and level differences in the brain stem (Grothe et al., 2010). In primate auditory cortices, sound location is thought to be encoded by activity differences between two neuronal populations, broadly tuned to ipsi- or contralateral hemifields (i.e., 'hemifield code') (McAlpine et al., 2001; Ortiz-Rios et al., 2017; Salminen et al., 2009; Stecker et al., 2005). Less is known about the coding principles at successive processing stages in parietal or prefrontal cortices, in which the hemifield code may be converted into a place code with narrow spatial tuning functions comparable to vision (e.g. ventral intraparietal area (Schlack et al., 2005)).

Mounting behavioural research illustrates the brain's extraordinary ability for cross-sensory plasticity. Most prominently, exposure to synchronous, yet spatially misaligned audiovisual signals biases observer's perceived sound location towards a previously presented visual stimulus – a phenomenon coined ventriloquist aftereffect (Bertelson et al., 2006; Bosen et al., 2018; Bruns et al., 2011; Frissen et al., 2005; Kopčo et al., 2009; Radeau and Bertelson, 1974; Recanzone, 1998; Woods and Recanzone, 2004; Wozny and Shams, 2011; Zierul et al., 2017). This cross-sensory adaptation to intersensory disparities emerges at multiple timescales ranging from milliseconds (Bruns et al., 2011; Park and Kayser, 2019; Wozny and Shams, 2011) to minutes (Kopčo et al., 2009; Recanzone, 1998; Woods and Recanzone, 2004; Zierul et al., 2017) or even days (Zwiers et al., 2003).

Yet, the underlying neural circuitries, mechanisms and representations that support cross-sensory plasticity remain controversial. The frequency selectivity of audiovisual recalibration has initially pointed towards early stages in tonotopically organized auditory cortices (see (Recanzone, 1998) but (Frissen et al., 2005)). By

contrast, the involvement of hybrid spatial reference frames that are neither eye- nor head-centred suggested a pivotal role for parietal cortices or inferior colliculus (Kopčo et al., 2009; Mullette-Gillman et al., 2005).

Importantly, recalibration may arise at multiple levels affecting spatial and decisional computations. While the former alters neural encoding of sound location irrespective of task, the latter affects the read out of decisional choices from those neural representations. The dissociation of the two is challenging for behavioural research that is forced to estimate recalibration from behavioural responses. Likewise, previous neuroimaging studies were not able to disentangle the two, because they employed a spatial localization task that maps each sound location onto one particular response choice, thereby conflating spatial and response selection processes (Bruns et al., 2011; Park and Kayser, 2019; Zierul et al., 2017). To dissociate changes in spatial and decisional representations, we need neuroimaging experiments that map different sound locations onto identical decisional choices (Woods and Recanzone, 2004).

This study combined psychophysics, fMRI, and EEG to investigate how changes in audiovisual statistics mould auditory spatial and decisional coding. We presented participants with auditory stimuli from seven locations along the azimuth in a spatial classification task before and after left and right audiovisual recalibration. Combining model-based analyses of regional mean BOLD-responses and fine-grained EEG and fMRI activity patterns, we temporally and spatially resolved how the brain flexibly adapts spatial and decisional coding across the cortical hierarchy to changes in the sensory statistics.

## Results

To dissociate spatial and decisional representations involved in adaptive coding of auditory space, this study performed extensive psychophysics, functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) experiments on 13 days in each participant (Figure 1A). Each experiment included unisensory auditory pre-adaptation, left and right audiovisual adaptation and unisensory auditory post-adaptation phases.

During pre- and post-adaptation phases participants were presented with auditory stimuli at 7 spatial locations (±12°, ±5°, ±2° and 0°) along the azimuth and performed a left-right classification task (Figure 1B). Participants responded only on a small fraction of trials (22%, so called 'response trials') that were randomly interspersed in the non-response trials. Critically, the fMRI analysis was limited to the 'non-response' trials to minimize the impact of motor activity on neural representations. The response trials enabled us to control observers' attention and assess the impact of audiovisual recalibration on observers' perceived sound location in terms of psychometric functions fitted to behavioural responses.

During the audiovisual adaptation phases, participants were presented on each trial with a sound in synchrony with a visual stimulus that was displaced by 15° to the left (i.e., VA) or right (i.e., AV) of the sound location. This allowed us to assess recalibration effects unconfounded by time and learning effects by comparing the two adaptation directions directly (rather than pre-adaptation relative to post-adaptation). To focus on implicit perceptual recalibration observers were engaged in a non-spatial task during the adaptation phases: they detected small changes in contrast of the visual stimulus that occurred on 10% of the trials (see Figure 1B and Experimental design and procedure section).

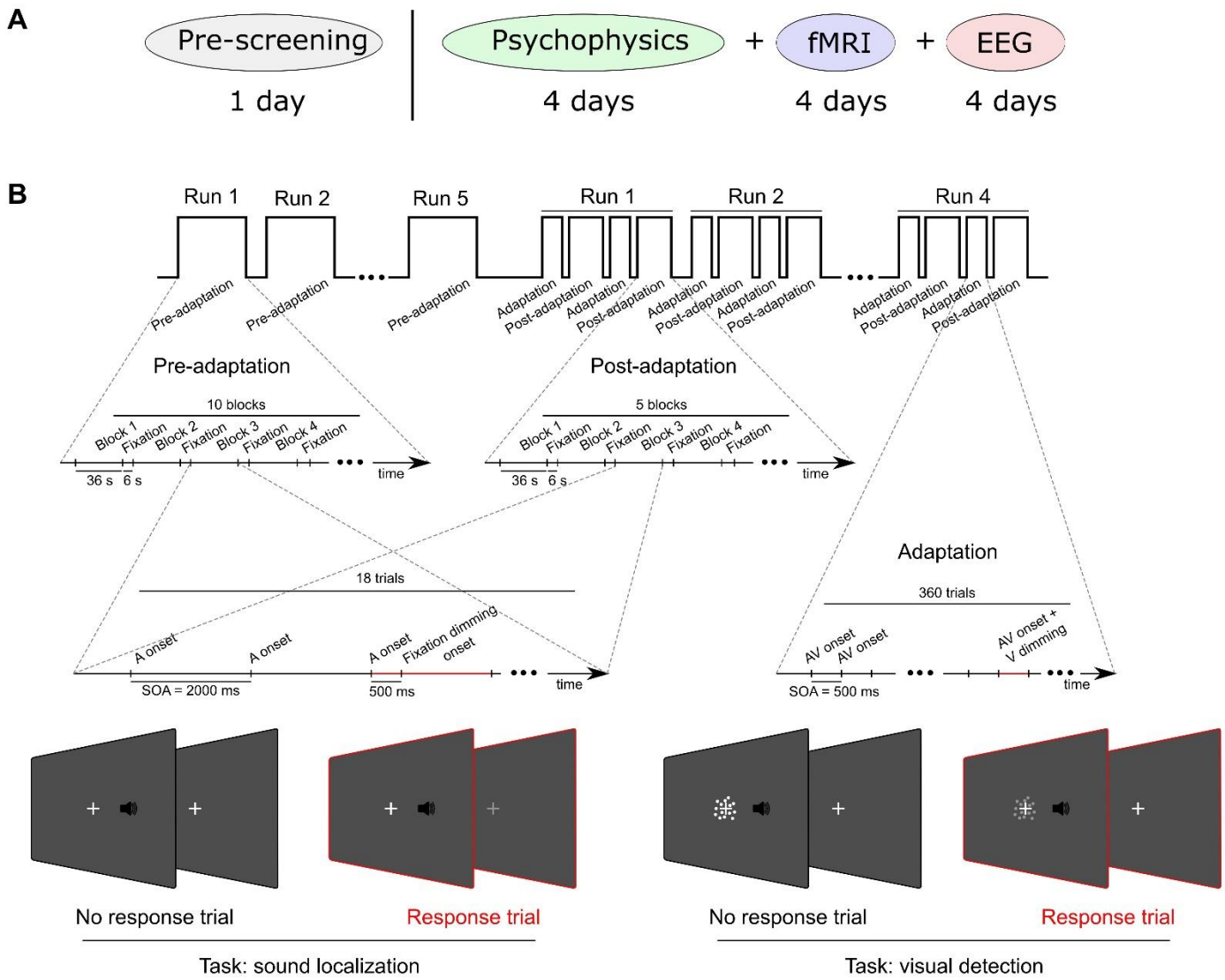


**Figure 1. Study and experimental design.** (**A)** The study included one day of pre-screening, four days of psychophysics testing, four days of fMRI and four days of EEG. (**B)** Each day included i. auditory pre-adaptation, ii. left (i.e., VA) or right (AV) audiovisual adaptation and iii. auditory post-adaptation phase. In the pre-adaptation and post-adaptation phases, observers were presented with unisensory auditory stimuli sampled from seven (±12°, ±5°, ±2° and 0° visual angle) locations along the azimuth. They performed a spatial (left vs. right) classification task on 22% of the trials ('response trial') that were indicated by a brief dimming of the fixation cross 500 ms after sound onset. In the audiovisual adaptation phases, observers were presented with audiovisual stimuli with a spatial disparity of ±15°: the visual signal was sampled from three locations along the horizontal plane (i.e., -5°, 0°, 5°) and the auditory signal was spatially shifted by 15° either to the right (i.e., VA adaptation) or to the left (i.e., AV adaptation) with respect to the visual stimulus. Observers were engaged in a non-spatial visual detection task on 10% of the trials ('response-trial') indicated by the lower contrast of the visual signals. Each day started with the pre-adaptation phase. Next, VA (or AV)-adaptation phases alternated with postVA (or postAV)-adaptation phases. This figure shows the experimental details for the fMRI experiments, which were slightly modified for EEG and psychophysics experiments (for details see Experimental design and procedure section).

## Behavioural results

First, we assessed whether observers maintained attention during the pre- and post-adaptation as well as the adaptation phases. Throughout all experiments observers responded to >88% of the 'response trials' (i.e., 'hits') in the pre-/post-adaptation and the adaptation phases (Table S1). Likewise, they made only < 0.47% false alarms (i.e., responses to non-response trials) resulting in a sensitivity index (d') of > 4 across all

experimental phases (Table S1). This high d' indicates that observers were attentive and followed task instructions. They responded selectively on the 22% (pre/post adaptation phase) or 10% (adaptation phase) of the response trials but refrained from making a response on the non-response trials.

Likewise, eye movement analyses during the psychophysics experiment showed that fixation was well maintained on 94.02% ± 1.00% (mean ± SEM) of the trials throughout the entire experiment. Further, directly comparing post-adaptation phases after left (postVA) and right (postAV) adaptation did not reveal any significant differences in eye movement indices (i.e., % saccades, % eye blinks, and post-stimulus mean horizontal fixation position, for further details see Supplemental Results). This ensures that potential recalibration effects are unlikely to be explained by eye movement confounds.

To assess whether exposure to audiovisual spatially disparate signals in the adaptation phases recalibrates observers' spatial sound representations, we fitted cumulative Gaussians as psychometric functions to the percentage 'perceived right' responses as a function of sound location separately for the pre-, postVA- and postAV-adaptation phases individually for each participant. Figure 2A shows the fitted individual and the across-subjects' mean psychometric functions separately for the psychophysics, fMRI and EEG experiments. Consistent with previous research (Radeau and Bertelson, 1974; Recanzone, 1998; Woods and Recanzone, 2004) observers' perceived sound location in the purely auditory post-adaptation phases shifted towards the displaced visual stimulus from the VA- and AV-adaptation phases. Consequently, the psychometric functions shifted rightwards in the postVA-adaptation phase and leftwards in the postAV-adaptation phase. Bayesian model comparison confirmed that audiovisual recalibration shifts the point of subjective equality (PSE): the full model that allowed for changes in PSE values across pre-, postVA- and postAV-adaptation phases was exceedingly more likely than the static model in which the PSE values were constrained to be equal across conditions (protected exceedance probabilities for the full recalibration model was >0.88 in each of the psychophysics, fMRI and EEG experiments; Table S2). Moreover, paired t-tests of the PSE values (from the full model) between postVA- and postAV-adaptation phases confirmed that the PSEs shifted towards the right after VA-adaption and towards the left after AV-adaptation phases resulting in a significant difference in PSE between postVA- and postAV-adaptation phases (psychophysics: $t(14) = 11.4$, $p < 0.0001$ ; fMRI: $t(4) = 11.3$, $p < 0.001$; EEG: $t(4) = 10.4$, $p < 0.001$) (Figure 2B). Collectively, our behavioural results confirmed that exposure to disparate audiovisual signals recalibrates observers' auditory space as indexed by their left/right sound classification responses.

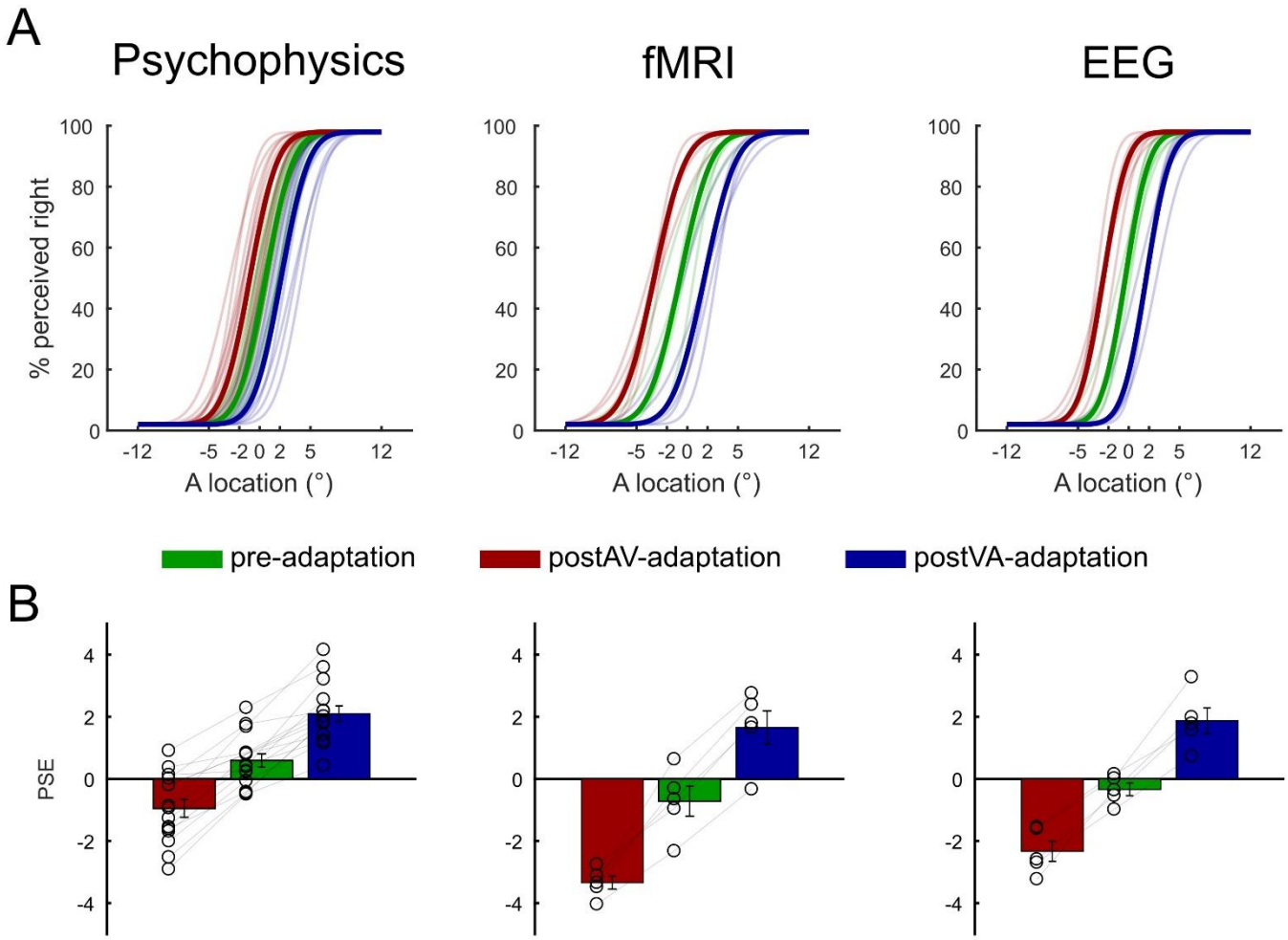


**Figure 2. Behavioural results. (A)** Psychometric functions fitted to the behavioural data from the psychophysics, fMRI and EEG experiments for pre-adaptation (green), postAV- (red) and postVA- (blue) adaptation phases (thin lines for subject-specific psychometric functions, thick line for across subjects' mean). **(B)** Across subjects' mean (± SEM) of the PSE for pre-adaptation (green), postVA- (red) and postAV- (blue) adaptation phases. The subject-specific PSEs are overlaid as circles. The PSE was shifted towards the left for postAV- (red) relative to postVA- (blue) adaptation phases in every single participant. PSE: point of subjective equality, SEM: standard error of the mean.

## fMRI results

### Spatial encoding and recalibration indices

Using multivariate decoding we identified brain areas that (i) encoded sound location and (ii) recalibrated the auditory spatial representations after the adaptation phases. Guided by previous research (Michalka et al., 2016; Ortiz-Rios et al., 2017; Rauschecker and Tian, 2000; van der Heijden et al., 2019; Zatorre et al., 2002; Zierul et al., 2017) we focused on five bilateral regions of interest (ROI) along the auditory spatial processing hierarchy: Heschl's gyrus (HG), higher auditory cortex (hA, mainly planum temporale), intraparietal sulcus (IPS), inferior parietal lobule (IPL) and frontal eye field (FEF). In each of those ROIs we trained a linear support vector regression model (SVR, LIBSVM(Chang and Lin, 2011)) using a 4-fold cross-validation to learn the mapping from BOLD-response activation patterns to external auditory locations based on the pre-adaptation phase data alone. We used this learnt mapping to decode the spatial location from the activation patterns of the pre-adaptation, postVA- and postAV-adaptation activation patterns. We quantified spatial decoding accuracy in terms of the Pearson correlation coefficient between the true and the decoded auditory locations. We observed significantly better than chance decoding accuracies across all

ROIs along the auditory spatial processing hierarchy with a maximal decoding accuracy in planum temporale (see Figure 3A for individuals' and across-subjects' mean Fisher z-transformed Pearson correlation coefficients; right-tailed p-values for the second level bootstrap-based one sample t-tests: HG: $p = 0.028$; hA: $p = 0.012$; IPS: $p = 0.005$; IPL: $p = 0.010$; FEF: $p = 0.019$).

Having established that sound location can be decoded successfully from all ROIs we assessed to what extent these spatial representations are recalibrated by displaced visual stimuli during the adaptation phase. We classified the sound locations that were decoded from the postVA- and postAV-adaptation phases into 'decoded left' and 'decoded right' responses. For each ROI, we computed the recalibration index (RI) as the difference between the fraction of 'decoded right' in the postAV-adaptation phase minus the fraction of 'decoded right' in the postVA-adaptation phase. If the decoded sound location shifts towards the disparate visual signal during the AV- and VA-adaptation phase, we would expect a positive recalibration index. Consistent with this conjecture we observed a RI that was significantly greater than zero in nearly all ROIs (right-tailed p-value of second level bootstrap-based one sample t-test, HG: $p = 0.025$, hA: $p = 0.002$, IPL: $p = 0.010$, FEF: $p = 0.037$, and a trend in IPS: $p = 0.086$).

The results provide initial evidence that regions along the dorsal auditory processing hierarchy encode sound locations that shift towards a spatially disparate visual signal. Yet, the analysis discussed so far only tested for an overall effect of recalibration averaged across all sound locations. In the following, we will characterize the neural representations evoked during the pre-, postVA- and postAV-adaptation phases in greater depth using neurometric functions and representational similarity analyses.

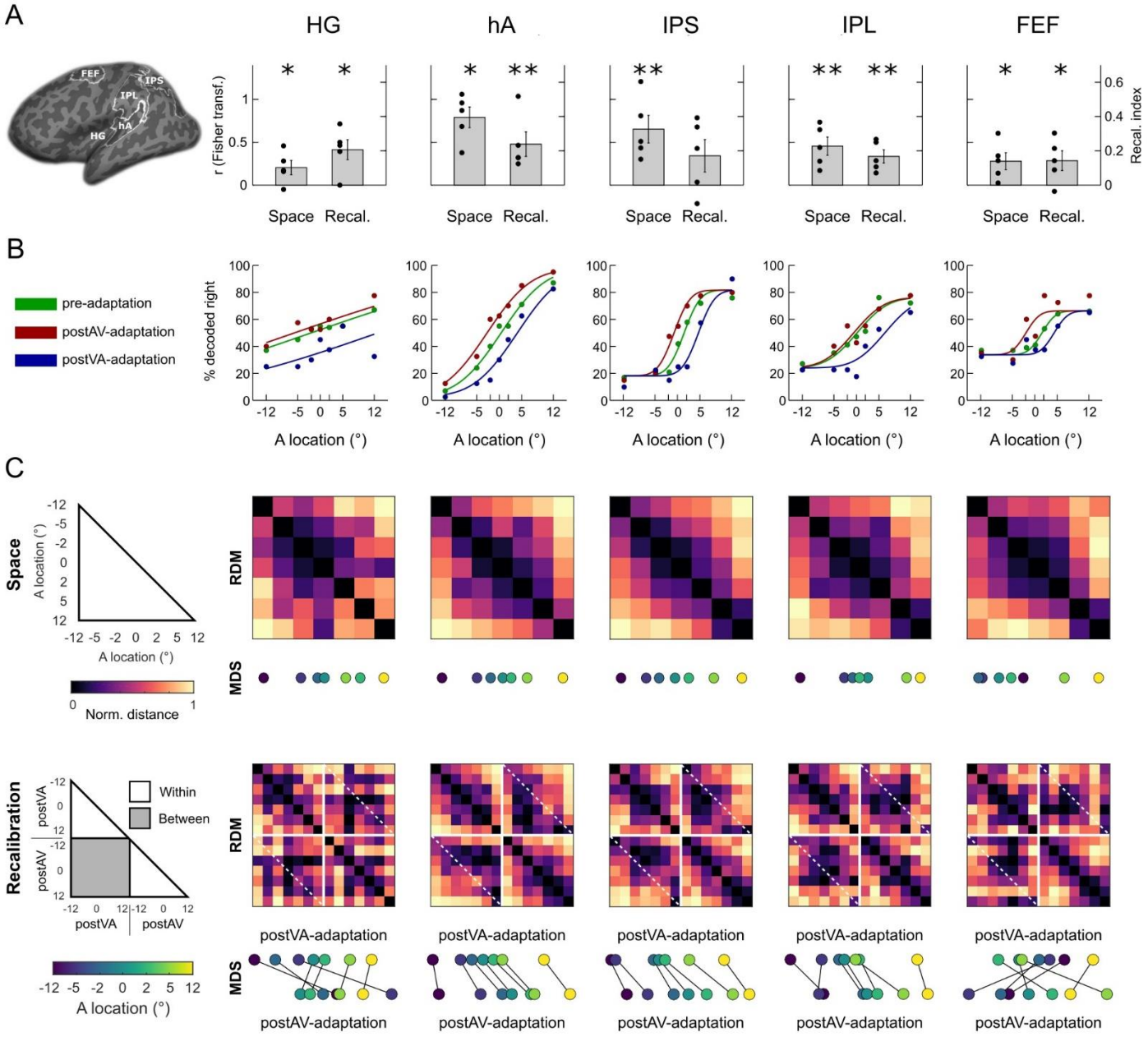


**Figure 3. fMRI multivariate decoding and representational dissimilarity matrices (A)** Across-subjects' mean (±SEM) of the spatial encoding (left bars, Fisher z-transformed correlation coefficient) and recalibration (right bars, difference in fraction 'decoded right' for postAV-adaptation vs. postVA-adaptation) indices in the ROIs as indicated in the figure. *: $p \leqq 0.05$; **: $p \leqq 0.01$. The subject-specific values are overlaid as black dots. **(B)** Neurometric function fitted to the fMRI decoded sound locations across ROIs for pre-adaptation, postVA- and postAV-adaptation phases. **(C)** RDMs (across-subjects' mean) showing the Mahalanobis distances for the fMRI decoded sound locations across ROIs. Top row: 7 x 7 RDM for pre-adaptation phase. Bottom row: 14 x 14 matrix for postVA- and postAV-adaptation. The drawings on the left indicate the organization of the RDMs of spatial locations for pre (top) as well as for postAV vs. postVA-adaptation (bottom) phases. Using MDS we projected each RDM onto a single dimension (i.e., representing space along the azimuth). For illustration purposes, we vertically offset the MDS projections for postVA- and postAV adaptation phases and connected the projections corresponding to the same physical stimulus locations. The MDS results show that the projected locations of the postVA-adaptation phase are shifted towards the left in relation to those of the postAV-adaptation phase particularly in hA and IPS, while a more complex relationship arises in IPL and FEF. HG: Heschl's gyrus; hA: higher auditory cortex; IPL: inferior parietal lobule; IPS: intraparietal sulcus; FEF: frontal eye-field; RDM: representational dissimilarity matrix; MDS: multidimensional scaling.

**Neurometric functions**

As in our behavioural analysis we fitted cumulative Gaussians as neurometric functions to the percentage 'decoded right' (rather than 'perceived right' as in our behavioural analysis) across the different levels of sound location separately for the pre-, postVA- and postAV-adaptation phases at the group level. Figure 3B shows the fitted neurometric functions separately for each of the five ROIs. Consistent with our behavioural findings the neurometric functions were shifted rightwards after VA-adaptation and leftwards after AV-adaptation. Using the Akaike Information Criterion as an approximation to the model evidence, Bayesian model comparison (Akaike, 1998) provided strong evidence that audiovisual recalibration shifts the point of subjective equality (PSE): the difference in ln-model evidence between the full model (allowing for changes in PSE values across pre, postVA- and postAV-adaptation phases) and the static model (in which the PSE values were constrained to be equal across conditions) was > 4.6 ('decisive evidence' (Jeffreys, 1961)) in all ROIs (HG: 12.9, hA: 20.2, IPS: 17.0, IPL: 10.3, FEF: 5.7; Table S3). Bayesian model comparison thus provided decisive evidence that the full model allowing for recalibration outperformed the more constrained static model in all regions.

**Representational similarity analysis**

Linear decoding results indicated that fMRI patterns encode information about sound location and that neural representations are altered by recalibration. Critically, these results are agnostic about the representational and neural coding principles underlying these plastic changes. To characterize the geometry of the neural representations across the seven sound locations we applied representational similarity analysis (RSA (Nili et al., 2014)). For visualization we projected the group level representational dissimilarity matrices (RDMs, averaged across participants) with non-classical multidimensional scaling (MDS) onto a one-dimensional space consistent with our a priori hypothesis that spatial representations are organized along the azimuth (i.e., as physical space). Figure 3C (top row) shows the representational dissimilarity matrices and corresponding MDS results for the pre-adaptation phase. Particularly, the higher auditory cortex (hA, including planum temporale) and IPS encodes the seven sound locations largely consistent with the physical distances of the sound locations. For instance, the MDS distance is greater for the sound locations 12° and 5° than for the locations 5° and 0° in hA and IPS. By contrast, the representational geometry in IPL and FEF does not fully correspond to the distances of the physical sound locations. A similar distance profile can be observed for the postVA- and postAV-adaptation phases (Figure 3C bottom row). Please note that we have projected the entire 14 x 14 matrix onto one single dimension with MDS. Post-hoc we then sorted the neural representations for the postVA- and postAV-adaptation phases into separate rows to visualize the relationship of spatial representations for identical physical sound stimuli during postVA- and postAV-adaptation phases in Figure 3C (bottom row). Across all regions, the MDS shows spatial shifts to the left (resp. right) after AV- (resp. VA)-adaptation. Again, the similarity between neural representations in hA and IPS obeyed the physical order of the sound locations, while this is less the case for IPL or FEF. For instance, in FEF the neural representations for the '-12° sound' is shifted towards the centre (see Figure 3C, FEF in bottom row).

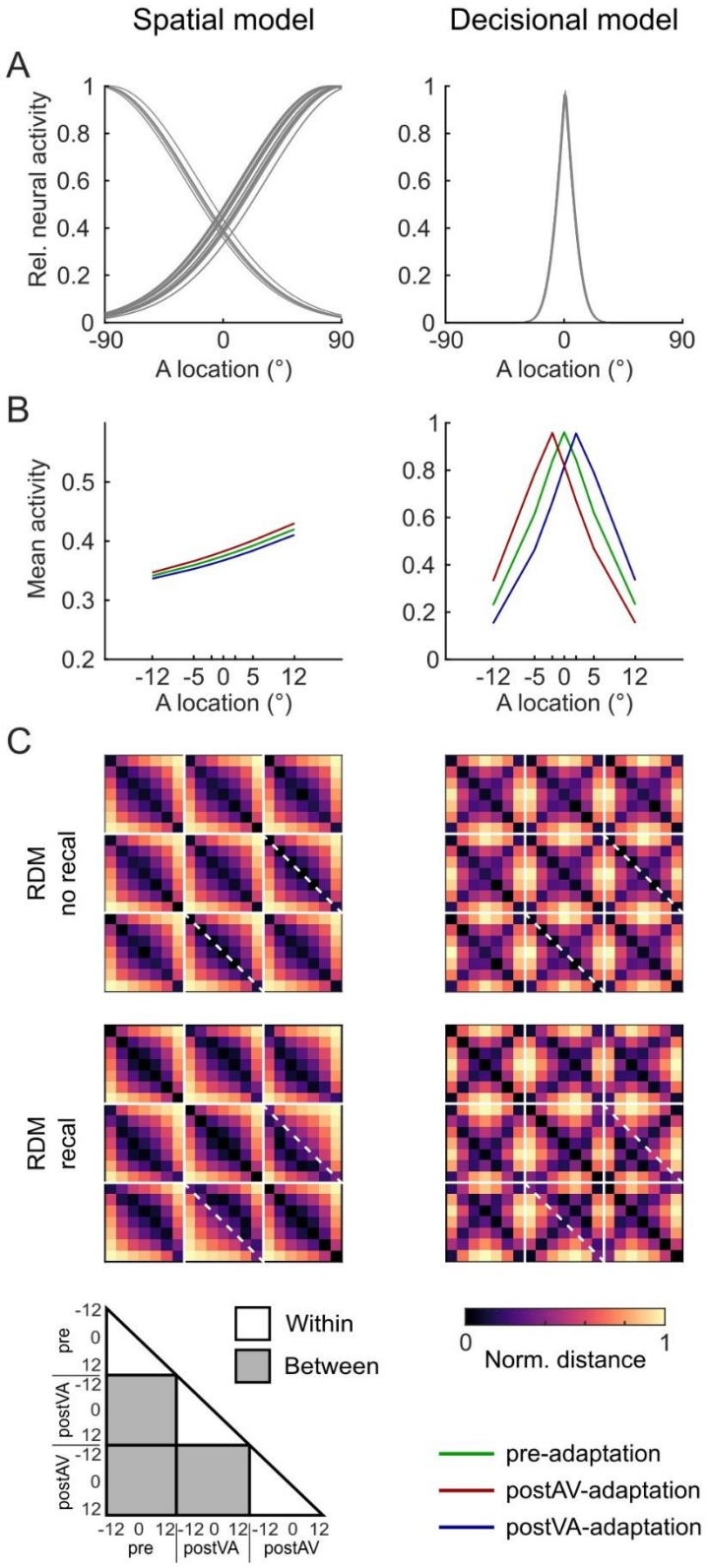


**Figure 4. Spatial and decisional models. (A)** Neural models: The spatial (hemifield) model encodes spatial location in the relative activity of two subpopulations of neurons each broadly tuned either to the ipsi- or contra-lateral hemifield. The ratio of the ipsi- and contralaterally tuned neurons was set to 30%/70% consistent with prior research (Salminen et al., 2009). The decisional model encodes observers' decisional uncertainty as a non-linear function of the distance between the spatial estimates and the spatial classification boundary. **(B)** Predicted mean BOLD-response as a function of sound location along the azimuth in a left hemisphere region for pre-adaptation (green), postVA (blue) and postAV (red)-adaptation. The spatial model predicts a BOLD-response that increases linearly for sound locations along the azimuth. The decisional model predicts BOLD-response that decays with the distance from the decision boundary in

an inverted U-shaped function. Further, this inverted U-shaped function shifts along the azimuth when spatial estimates are recalibration. The model predictions were obtained by averaging the simulated neural activities across 360 neurons in a left hemisphere region. **(C)** Predicted representational dissimilarity matrices (RDM) based on the individual model neural activity profiles across spatial locations (-12° to 12°) and experimental phases (pre, postVA and postAV). We simulated RDMs from the spatial (left) and the decisional (right) model for i. top row: no recalibration, i.e., without a representational shift and ii. bottom row: with recalibration, i.e., with a representational shift. Solid white lines delineate the sub-RDM matrices that show the representational dissimilarities for different stimulus locations within and between different experimental phases. Diagonal dashed white lines highlight the RDM dissimilarity values for two identical physical locations of the postAV-adaptation and the postVA-adaptation phases. Comparing the RDMs with and without recalibration along those dashed white lines shows how the shift in spatial representations towards the previously presented visual stimulus alters the representational dissimilarity of corresponding stimulus locations in postAV- and postVA phases, while the off-diagonals show the dissimilarity values for neighbouring spatial locations.

**Model-based fMRI analysis: dissociating perceptual and decisional codes**

These RSA results raise the critical question whether the differences in representational geometry that the MDS revealed across regions simply reflect noise or arise from mixing of multiple representational components. Most notably, as we discussed in the introduction, recalibration may affect spatial and decisional representations. In the following, we first introduce the neural models for spatial and decisional coding. Using Bayesian model comparison we then assess whether the spatial, the decisional or a combined spatial + decisional model provides a better explanation for the regional mean BOLD-response and/or the BOLD-response patterns across regions.

To model spatial/perceptual representations we used the hemifield model that encodes sound location in the relative activity of two subpopulations of neurons, each broadly tuned either to the ipsi- or the contra-lateral hemifield (McAlpine et al., 2001; Middlebrooks et al., 1994). Following previous research (Salminen et al., 2009) we assumed a preponderance of the neuronal population responding to the contralateral hemifield. Accumulating neurophysiology and neuroimaging research has shown that auditory cortices code sound location consistent with the hemifield model (Ortiz-Rios et al., 2017; Salminen et al., 2009; Stecker et al., 2005). By contrast, it is not yet clear whether frontoparietal cortices represent sound location in a hemifield or place code. For the current study this distinction is less relevant, because the hemifield and place code models make near-indistinguishable predictions for the pattern similarity structure over the selected central locations. Therefore, we applied the hemifield model as a generic spatial model to all regions (see Methods, Discussion and Supplemental Methods for further details).

The decisional model accounts for the fact that the brain needs to map continuous spatial estimates onto binary decisional choices (i.e., left and right spatial classification responses). The decisional model codes observers' decisional uncertainty as a non-linear function of the distance between the spatial estimate from the left/right classification boundary (see (Grinband et al., 2006) and Methods for further details).

Because a region may combine spatial and decisional coding, we first assessed whether (i) the spatial model (S), (ii) the decisional model (D) or (iii) a combined spatial + decisional model (S+D) were the best explanations for our data. In a second step, we used the combined spatial + decisional model to assess the contributions of the spatial and decisional components to recalibration by comparing: (i) spatial + decisional with no recalibration in either model component (S+D), (ii) spatial with recalibration + decisional ($S_R$+D), (iii) spatial + decisional with recalibration (S+$D_R$) and (iv) spatial with recalibration + decisional with recalibration ($S_R$+$D_R$). Because we applied recalibration widely across both hemifields, each model component accounted for recalibration by shifting the 'encoded sound locations' by a constant ±2.3˚ (i.e., across-subjects mean in behavioural PSE shift) to the right (postVA-adapation) vs. left (postAV-adaptation).

As shown in Figure 4, the spatial and decisional models make distinct predictions for the regional mean BOLD response (Figure 4B) and the similarity structure of activity patterns over sound locations (Figure 4C).

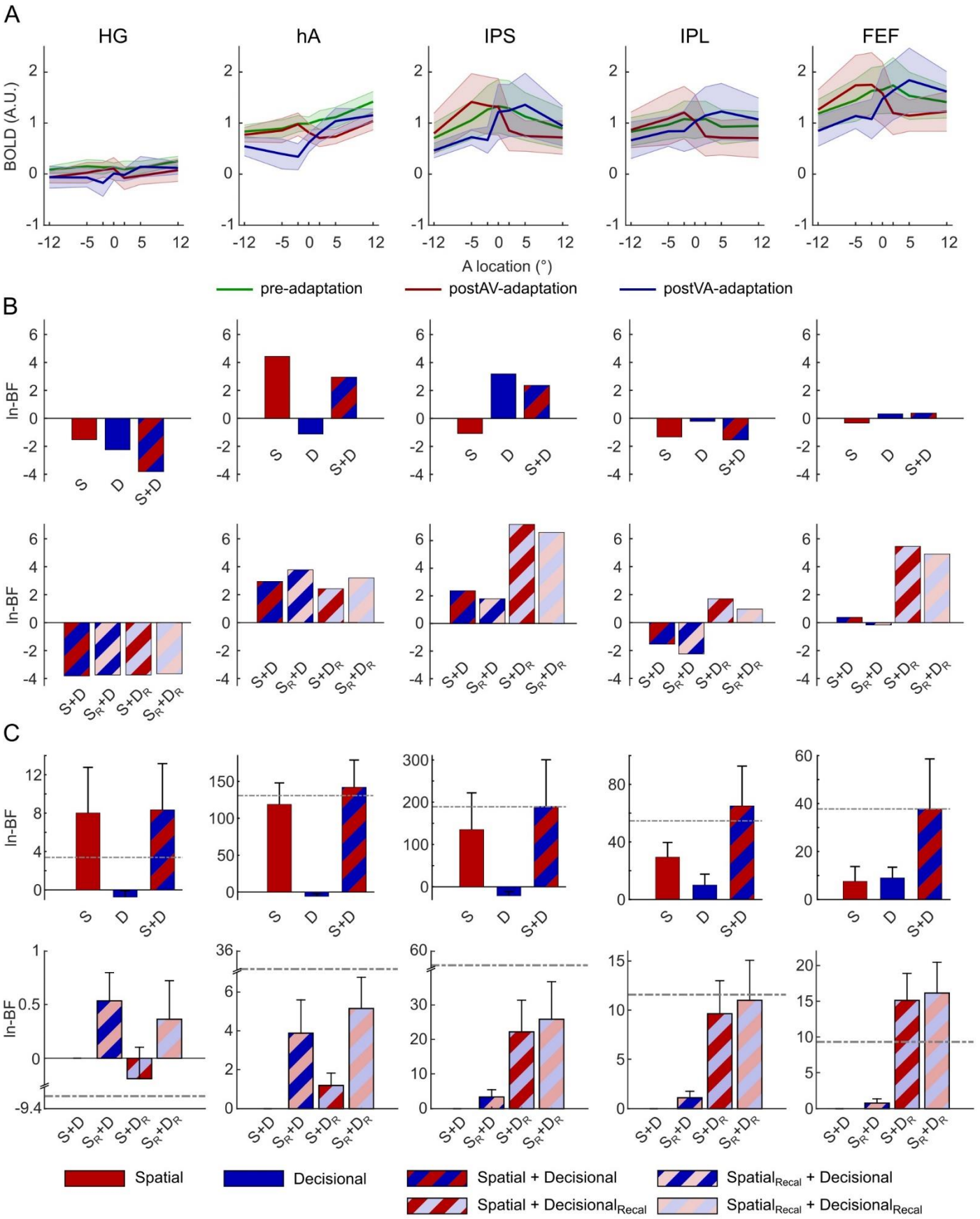


**Figure 5. fMRI results: regional BOLD response and pattern component modelling (PCM). (A)** Across-subjects' mean positive BOLD responses shown as a function of spatial location in pre-adaptation, postVA- and postAV-adaptation phases across ROIs. The shaded areas indicate SEM. **(B)** Results of the linear mixed effects analysis of regional mean BOLD-response across ROIs: ln-Bayes factors (ln-BF) for a specific target model as indicated by the capital letter relative to the null model, which includes only the constant term (see

Methods). Top row: The spatial, decisional, and spatial + decisional models without recalibration. Bottom row: The models factorially manipulate whether the spatial and/or decisional component include recalibration. **(C)** PCM results - Spatial and/or decisional models as predictors for fine grained BOLD-response patterns across ROIs. Across-subjects mean ln-BFs (±SEM) for each target model as indicated by the capital letter. i. Top row: The spatial, decisional and spatial + decisional models without recalibration relative to a null model that allows for no similarity between activity patterns. ii. Bottom row: The models factorially manipulate whether the spatial and/or decisional component account for recalibration. Ln-Bayes factors are relative to the spatial + decisional model (without recalibration) as the null model. Dash-dotted grey lines indicate the relative ln-BF for the fully flexible models as baseline estimated separately for the top and bottom row model comparisons (for details, see Methods). S = spatial model without recalibration, D = decisional model without recalibration, $S_R$ = spatial model with recalibration, $D_R$ = decisional model with recalibration. (Diedrichsen et al., 2018). Lln-Bayes factors of > 1.1 are considered substantial evidence and > 3 as strong evidence for one model over the other (Jeffreys, 1961).

*Regional mean BOLD-response*

The spatial model predicts a linear increase in the regional mean BOLD-response towards stimuli in the opposite hemifield and a small overall change in activation for all sound locations as a result of recalibration towards the contralateral side. By contrast, the decisional model predicts a mean BOLD-response that peaks at the decisional boundary and tapers off with greater distance from the boundary. As a result of recalibration, this peak shifts towards the left or right, because the brain's spatial estimates have been recalibrated leading to a change in the relative distance between the spatial estimates and the decision boundary (see Figure 4B). For instance, the encoded spatial estimate for a sound stimulus at physical location of -5° may be shifted towards the centre after AV recalibration and thereby be associated with a greater decisional uncertainty.

Figure 5A shows the regional mean BOLD-response for our five ROIs (n.b. here shown as for the left hemisphere, see Methods). In hA the mean BOLD-response increased progressively for stimuli along the azimuth as expected under the spatial model. By contrast, in IPS, IPL and FEF the BOLD-response peaked for physical 0° sound location in the pre-adaptation phase, left (i.e., < 0°) sound locations in the postAV-adaptation and right (i.e., > 0°) sound locations in the postVA-adaptation phase as expected under the decisional model. This visual impression was confirmed by Bayesian model comparison of the three (i.e., spatial, decisional, spatial + decisional) linear mixed effects (LME) models. Each LME model used the simulated activity of the spatial and/or decisional models as fixed effects predictors (in addition to a constant regressor) for the 7 (sound locations) x 3 (pre-adaptation, postAV-adaptation, postVA-adaptation) conditions. Figure 5B (top row) shows the ln-Bayes factors for each of the three models relative to a null model that included only the constant as a predictor. In HG the most parsimonious baseline model outperformed the other models suggesting that the regional mean BOLD-response in HG did not provide reliable information about sound location. The Bayes factors provided strong evidence for the spatial model in hA and for the decisional model in IPS. Turning to the recalibration results we observed a small trend towards spatial coding in hA (i.e., $S_R$+D and $S_R$+$D_R$ > S+D and S+$D_R$), but substantial evidence towards decisional coding in IPS, IPL and FEF (i.e., $S_R$+$D_R$ and S+$D_R$ > S+D and $S_R$+D, Figure 5B bottom row, Table S4). This suggests that the mean regional BOLD-response follows the predictions of the hemifield model only in hA. By contrast in IPS, IPL and FEF the regional mean BOLD-response mainly reflects the mapping of the recalibrated spatial estimates onto observers' decisional boundary.

*Pattern component modelling*

Using pattern component modelling (PCM (Diedrichsen et al., 2018)) we investigated whether spatial and decisional codes contributed to the distribution of the fine-grained voxel-specific activity profiles over the 7 (sound locations) x 3 (pre-adaptation, postAV-adaptation, postVA-adaptation) conditions. Importantly, PCM does not model the voxel-specific activity profiles per se, which are unlikely to be consistent across participants, but their similarity structure over conditions as described by the covariance (or more precisely,

second moment) matrix of the activity profile distributions. Moreover, PCM allows us to model the representational structure of activity patterns by a linear combination of components (e.g. spatial and/or decisional). Consistent with our analysis of the regional mean BOLD-response we fitted and compared a baseline model that assumes independence of activation patterns across conditions with the spatial, decisional and combined spatial + decisional PCM models. In a second step we used the combined spatial + decisional model to assess whether recalibration was expressed in the spatial, decisional, or spatial + decisional codes (see Figure 5C). We fitted all target models (e.g., spatial and/or decisional etc.) and a fully flexible model using a leave-one-subject-out cross-validation scheme. The fully flexible model accommodates any possible covariance structure of activity profiles and enables us to assess whether a particular target model accounts for the key representational structure in the observations given the measurement noise and inter-subject variability (see Figure 5C).
As shown in Figure 5C top row, the ln-Bayes factors provided strong evidence for the spatial relative to the decisional model in HG and hA (see Table S5 for exact values). In HG, the spatial component alone was sufficient to reach the threshold set by the fully flexible model. In all other regions, i.e., hA, IPS, IPL and FEF, the combined spatial + decisional model outperformed the single component models suggesting that in those regions spatial and decisional representations together contribute to adaptive coding of auditory space. Moreover, the decisional component was progressively more dominant along the dorsal processing hierarchy. In FEF the decisional model outperformed the spatial model.
The recalibration results further emphasized these opposite gradients for spatial and decisional coding along the auditory processing stream. As shown in Figure 5C bottom row, in HG and hA the models expressing recalibration in spatial coding ($S_R$+D and $S_R$+$D_R$) outperformed those without recalibration in spatial coding (S+D and S+$D_R$). The evidence was substantial in hA, but not conclusive in HG. Conversely in IPS, IPL and FEF we observed strong evidence for the models that expressed recalibration in the decisional code (S+$D_R$ and $S_R$+$D_R$) relative to those that did not (S+D and $S_R$+D, see Table S5 for exact values).

In summary, the initial linear decoding analyses of fMRI BOLD responses suggest that all ROIs recalibrate and adapt spatial representations in response to spatially disparate audiovisual stimuli. Crucially, advanced model-based representational analyses demonstrated that a shift in the mean of decoded locations can result from two distinct, i.e., spatial and decisional, neural codes. While all regions (apart from HG) multiplexed spatial and decisional coding, their contributions arose with opposite representational gradients along the dorsal auditory processing hierarchy. The spatial model provided a better explanation for activations in HG and hA, the decisional model dominated in IPS, IPL and FEF. Collectively, our results demonstrate the critical importance to move beyond simple linear decoding analyses. Only characterizing the representational geometry of the neural representations enabled us to show that adaptive behaviour relies on plastic changes in spatial and decisional codes.

## EEG results

Using EEG, we characterized how spatial and decisional coding evolved across post-stimulus time within a trial. In parallel to our fMRI analysis, we first computed the spatial encoding and the recalibration indices for EEG activity patterns across post-stimulus time to determine when the sound encoding adapts flexibly to misaligned visual signals. Next, we used PCM to assess whether spatial and decisional codes arise at different latencies post-stimulus.

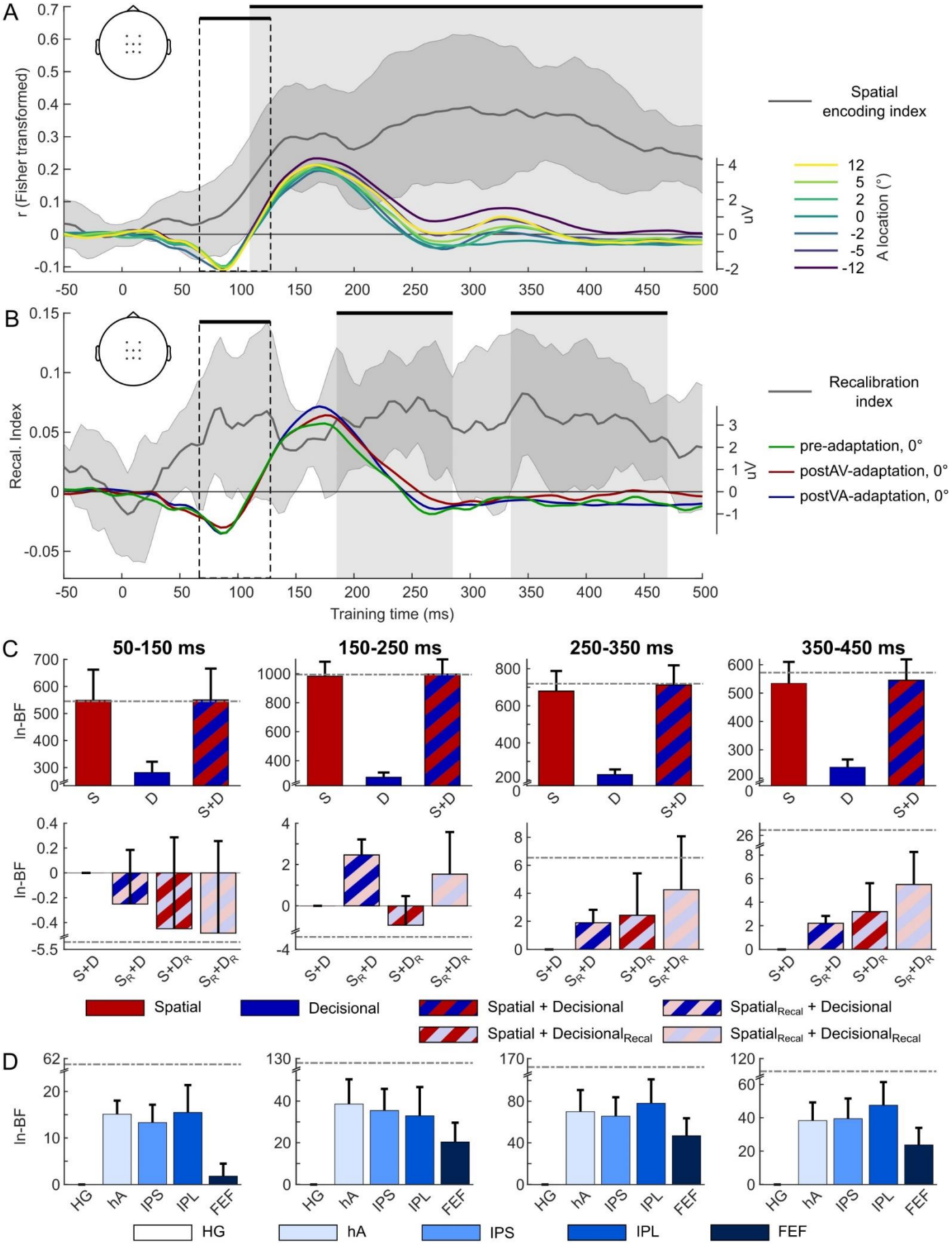


**Figure 6. EEG results: multivariate decoding and pattern component modelling.**

**(A)** Time course of the spatial encoding index (i.e., across-subjects' mean, dark grey line) and of EEG evoked potentials (across-subjects' mean, averaged over central channels, see inset) for the 7 spatial locations in pre-adaptation phases (±12°, ±5°, ±2° and 0° azimuth).
**(B)** Time course of the recalibration index (across-subjects' mean, dark grey line) and the EEG evoked potentials (across-subjects' mean, phases, averaged over central channels, see inset) for sound stimuli presented at 0° azimuth for pre-adaptation, postVA- and postAV-adaptation. The shaded area around the spatial encoding index (A) and the recalibration index (B) indicates SEM. The shaded grey boxes and thick black horizontal lines (on top) indicate time windows in which the decoding accuracy (A) or the recalibration index (B) were significantly greater than zero at $p < 0.05$ corrected for multiple comparisons across the entire [0, 500] ms time window. Areas within the dashed boxes indicate the a priori defined time window [70, 130] ms to focus on the early recalibration effects (Bruns et al., 2011).
**(C)** PCM results – Spatial and/or decisional models as predictors for EEG activity patterns across four time windows. Across subjects' mean ln-BFs (±SEM) for each target model as indicated by the capital letter. i. Top row: The spatial, decisional and spatial + decisional models without recalibration relative to a null model that allows for no similarity between activity patterns. ii. Bottom row: The models factorially manipulate whether the spatial and/or decisional component account for recalibration. Ln-Bayes factors are relative to the spatial + decisional model (without recalibration) as the null model. Dash-dotted grey lines indicate the relative ln-BF for the fully flexible models as baseline. S = spatial model, D= decisional model, $S_R$ = spatial model with recalibration, $D_R$ = decisional model with recalibration.
**(D)** PCM results – BOLD-response patterns from the five ROIs as predictors for EEG activity patterns across four time windows. Across subjects' mean ln-BFs (±SEM) for each target model relative to the model using BOLD-response patterns in HG as predictors. HG: Heschl's gyrus; hA: higher auditory cortex; IPS: intraparietal sulcus; IPL: inferior parietal lobule; FEF: frontal eye-field. ln-Bayes factors of > 1.1 are considered substantial evidence and > 3 as strong evidence for one model over the other (Jeffreys, 1961).

***Spatial encoding and recalibration indices***

To resolve spatial encoding across post stimulus time we trained linear support vector regression models (SVR, LIBSVM (Chang and Lin, 2011)) on the EEG activity patterns of the pre-adaptation trials in overlapping 50 ms time windows sliding from -100 to 500 ms post-stimulus (EEG) and generalized them to trials from pre- and post-adaptation trials in a 4-fold cross-validation scheme (consistent with our fMRI analysis). The cluster based bootstrap test on the Fisher z-transformed Pearson correlation coefficient between the true and the decoded sound locations revealed a significant cluster extending from 110 ms to 500 ms post-stimulus ($p = 0.01$ corrected for multiple comparisons within the entire [-100, 500] ms window). The decoding accuracy was significantly better than chance from about 110 ms post-stimulus, rose steadily and peaked at about 355 ms (Figure 6A). Likewise, we assessed the effect of recalibration in terms of the recalibration index (RI) within the time window in which we observed a significant effect of spatial encoding, i.e., 110 – 500 ms post-stimulus. In the cluster based bootstrap test the RI was significantly positive in two clusters from 185 – 285 ms ($p = 0.019$) and 335 – 470 ms ($p = 0.005$) post stimulus (Figure 6B). Moreover, because previous ERP analyses suggested that the N100 potential is affected by recalibration (Bruns et al., 2011), we performed a temporal ROI analysis selectively on the EEG activity pattern averaged within the N100 time window (i.e., 70 – 130 ms). This temporal ROI analysis showed that the decoding accuracy was significantly better than chance (mean ± SEM: 0.125 ± 0.059; $p = 0.0406$) and the RI was significantly greater than zero (mean ± SEM: 5.963 ± 1.743; $p = 0.0019$).

***Model-based EEG analysis: dissociating perceptual and decisional representations***

*Pattern component modelling*

The fMRI results revealed a mixture of spatial and decisional codes in several ROIs across the dorsal processing hierarchy. Combining EEG and PCM we investigated whether spatial and decisional

representations evolved with different time courses. For this, we compared the null, spatial, decisional, and combined spatial + decisional models as explanations for EEG activity patterns over spatial locations in four consecutive time windows: 50 – 150, 150 – 250, 250 – 350 and 350 – 450 ms post stimulus (Figure 6C). From 50 – 150 ms the spatial model alone provided a sufficient explanation of the EEG data. From 150 – 250 ms and beyond the spatial + decisional model performed decisively better than the spatial model with ln-BFs of at least 10.9. These results suggest that the decisional code becomes progressively more important at later processing stages. This representational gradient across post-stimulus time was also apparent in our recalibration results (Figure 6C, bottom row). Spatial recalibration was expressed mainly via spatial coding (i.e., $S_R$+D and $S_R$+$D_R$ better than S+D and S+$D_R$ by ln-BFs of at least 2.4) from 150 to 250 ms and via decisional coding (i.e., S+$D_R$ and $S_R$+$D_R$ better than S+D and $S_R$+D by ln-BFs of at least 2.4) from 250 ms onwards. Nevertheless, as in our fMRI results recalibration relied jointly on spatial and decisional codes in later stages. Collectively, our EEG results revealed that spatial codes are more prominent in earlier processing stages and decisional coding in later ones. Critically, at later stages recalibration was expressed jointly in spatial and decisional codes.

*Fusing EEG and fMRI in pattern component modelling*

To link fMRI and EEG results more directly, we fused them in PCMs in which the second moment matrices of the BOLD-response activation patterns in one of our five ROIs (i.e., HG, hA, IPS, IPL or FEF) formed the predictor for the distribution of EEG activity patterns over sound locations. Bayesian model comparison across these five fMRI-EEG fusion PCMs confirmed a direct relationship between our fMRI and EEG results (Figure 6D). The PCM with hA (i.e., including planum temporale) explained EEG activity patterns best by ln-BFs of at least 3.2 from 150 – 250 ms and the PCM with IPL from 250 ms onwards (by ln-BFs of at least 8.0). We suspect that FEF has generally less explanatory power for explaining EEG activity patterns because of its smaller size and hence contribution to the EEG potentials measured on the scalp. Collectively, EEG and fMRI spatiotemporally resolved how recalibration alters encoding of sound location and decisional uncertainty that is computed when mapping these spatial estimates onto decisional choices.

## Discussion

Combining psychophysics, fMRI and EEG we demonstrate that audiovisual recalibration moulds fine-grained neural representations in a widespread neural system from Heschl's gyrus to frontoparietal cortices. Critically, this crossmodal plasticity arose in spatial and decisional codes that were expressed with opposite gradients along the auditory processing hierarchy and across post-stimulus time. Early neural activity patterns in planum temporale encoded sound locations in a continuous space that adapted flexibly to misaligned visual inputs. Later activity patterns in frontoparietal cortices coded observers' decisional uncertainty in accordance with this transformed auditory space.

Our behavioural results provide robust evidence that misaligned visual signals recalibrate auditory space (Radeau and Bertelson, 1974; Recanzone, 1998; Woods and Recanzone, 2004). VA- relative to AV-recalibration shifted the point of subjective equality towards the right in every single participant outside and inside the MRI scanner (Figure 2). Because observers were engaged in a non-spatial visual detection task during the recalibration phase, this crossmodal plasticity relied mainly on implicit perceptual rather than choice-related mechanisms (Zaidel et al., 2013) that may have mediated recalibration effects in previous neuroimaging and behavioural research (Bruns and Röder, 2015; Kopčo et al., 2009; Mendonça et al., 2015; Park and Kayser, 2019; Wozny and Shams, 2011).

Consistent with previous research in human and non-human primates our multivariate fMRI analyses show that sound location can be decoded from activity patterns in a widespread network including primary auditory regions, planum temporale and frontoparietal cortices (Ortiz-Rios et al., 2017; Rauschecker and Tian, 2000; Tian et al., 2001; van der Heijden et al., 2019; Zatorre et al., 2002; Zierul et al., 2017). Further,

the recalibration indices and the neurometric functions show that neural representations in all of these regions adapt in response to misaligned visual signals (Figure 3A). These widespread recalibration effects converge with our EEG results showing persistent spatial encoding and recalibration starting with the early auditory N1 component, thought to be generated in primary and secondary auditory cortices (Winkler et al., 2013), and extending until 500 ms post-stimulus (Figure 6B).

Critically, multidimensional scaling revealed not only that VA and AV recalibration shifted the encoded sound locations on average into opposite directions, but also unravelled subtle representational differences across regions (see Figure 3C). While a constant representational shift for VA relative to AV recalibration was observed in planum temporale, a more complex similarity structure emerged in IPL and FEF. To investigate whether this complex representational geometry arose from multiplexing of spatial and decisional codes (Raposo et al., 2014), we assessed whether fMRI and EEG activity can be better captured by the spatial hemifield model, a decisional model or a model that combined spatial and decisional representations (Figure 4). Consistent with previous research, the hemifield model assumed a contralateral predominance (Salminen et al., 2009), such that the predicted regional BOLD-response increased progressively along the azimuth towards the contralateral hemifield. Because the pattern similarity structure is (near) indistinguishable for the hemifield and the place code model over the limited (i.e., [-12° to 12°]) sound locations of the current study (see Figure S1), we applied the hemifield model as a generic spatial model to auditory as well as frontoparietal areas (Ortiz-Rios et al., 2017; van der Heijden et al., 2019). The decisional model computes choice-related uncertainty from the distance between the spatial estimates and the classification boundary and thus makes predictions for the regional BOLD-response and the activity patterns that are very distinct from those of the hemifield model (see Figure 4, (Grinband et al., 2006)). We assessed the role of spatial and decisional coding by comparing a range of models that factorially manipulated whether they included these spatial and decisional components with and without recalibration.

Bayesian model comparison revealed a double dissociation, with the spatial model providing a better explanation for the regional BOLD-response and fine-grained activity patterns in planum temporale and the decisional model in frontoparietal cortices. The regional BOLD response increased linearly for stimuli along the azimuth in planum temporale (Figure 5A), but followed an inverted U-shaped function that flexibly shifted along the azimuth in response to misaligned visual signals in frontoparietal cortices (Figure 5B top and bottom row) (Ortiz-Rios et al., 2017; Petkov et al., 2006; Rauschecker and Tian, 2000; van der Heijden et al., 2019; Zatorre et al., 2002; Zierul et al., 2017; Zimmer and Macaluso, 2005) . Likewise, the recalibration effects in the fine-grained activity patterns were better captured by the spatial model in planum temporale, but by the decisional model in frontoparietal cortices. Nevertheless, consistent with a strategy of multiplexing, the spatial model contributed significantly to the fine-grained activity patterns – though not the regional BOLD-response – in IPS, IPL and even FEF (Figure 5C top row) (Figure 5B top row). Potentially, the hemifield code may be converted into a place code in parietal cortices as suggested by a previous neurophysiology in non-human primates showing narrow auditory tuning functions in ventral intraparietal area (Schlack et al., 2005).

The hierarchical organization with spatial coding in auditory cortices and mainly decisional coding in frontoparietal cortices was mirrored in their temporal evolution post-stimulus. Recalibration of spatial coding was reflected in early EEG activity patterns (150 – 250 ms), while adaptation of decisional uncertainty was more prominent in later activity (after 250 ms, Figure 6C bottom row). We could even directly link the double dissociations that we observed across regions with fMRI and across post-stimulus time with EEG by fusing fMRI and EEG within the same pattern component models: the BOLD-response patterns in hA were the better predictor for early (150 – 250 ms) EEG activity, while IPL accounted more for EEG activity from 250 ms onwards (Figure 6D).

Collectively, our fMRI and EEG results demonstrate that the brain adapts flexibly to the changing statistics of the environment. They also highlight the importance to move beyond simple linear decoders (e.g., support vector regression, linear discriminant analysis) and characterize the fine-grained representational

geometry with advanced model-based analyses. Only the latter was able to show that later recalibration effects in frontoparietal cortices reflect mainly decisional read out rather than genuine spatial coding mechanisms. The expression of recalibration in distinct spatial and decisional codes opens the intriguing possibility that recalibration may not always impact both codes together. Instead, the representations and neural circuitries involved in recalibration may depend on the particular task and/or the duration of recalibration. For instance, while our study focused on implicit recalibration using a non-spatial visual detection task, other studies have shown rapid recalibration effects when observers performed spatial localization tasks both on the adapting audiovisual stimuli and the auditory test stimuli (Bruns and Röder, 2015; Mendonça et al., 2015; Park and Kayser, 2019; Wozny and Shams, 2011). It remains to be shown whether these short-term recalibration effects in explicit spatial tasks similarly involve plastic changes in the spatial representations or selectively alter the mapping from spatial estimates onto decisional choices leaving the representation of continuous auditory space largely preserved. In line with this suggestion, accumulating research suggests that recalibration relies on multiple mechanisms that operate over different timescales (Bosen et al., 2018; Watson et al., 2019). Future studies are therefore needed that disentangle the contributions of spatial and decisional codes depending on recalibration duration and task context in designs that map different spatial locations onto identical decisional choices.

So far, we have emphasized that spatial and decisional codes are expressed with opposite functional gradients across the cortical hierarchy and post-stimulus time. Crucially, Bayesian model comparison also indicated that the combined spatial + decisional model significantly outperformed the more parsimonious decisional or spatial models throughout the cortical hierarchy. Likewise, a small benefit was observed for the pattern component models in which recalibration was expressed in both spatial and decisional codes. These findings suggest that all higher order cortical regions (i.e., hA, IPS, IPL and FEF) multiplexed perceptual and decisional codes as they dynamically adapt to misaligned visual inputs (Raposo et al., 2014). Moreover, because multiplexing of spatial and decisional codes arose mainly later from 250 ms post-stimulus, the coding of decisional uncertainty in auditory cortices (hA) is likely to reflect top-down influences from frontoparietal areas (Mihalik and Noppeney, 2020; Werner and Noppeney, 2010).

In conclusion, combining psychophysics, model-based representational fMRI and EEG we showed that all regions throughout the auditory processing hierarchy multiplex spatial and decisional codes that flexibly adapt to changes in the input statistics across the senses. Crucially, the expression of spatial and decisional codes evolves with different time courses and opposite gradients along the auditory processing stream. Early neural activity in planum temporale encoded sound location within a continuous auditory space. Later frontoparietal activity mainly encoded observers' decisional uncertainty that flexibly adapts in accordance with these transformations of auditory space.

# Methods

## Participants

Nineteen right-handed participants took part in the pre-screening session. 15 of those participants (10 females, mean age = 22.1; SD = 4.1) returned for the psychophysics part of the study. Participants were included after the pre-screening session if (i) they provided reliable spatial discrimination responses in the auditory pre-adaptation test, i.e., the just noticeable difference (JND) of the psychometric function fitted to the spatial discrimination responses was < 4°; (ii) reliably discriminated between response and non-response trials, i.e., sensitivity index (d') > 2.5 and (iii) reliably detected the target luminance change during the audiovisual adaptation phase, i.e., d' > 2.5 (see Experiment design and procedure for details). Six of the psychophysics participants that had the highest recalibration effect size were selected to participate in the fMRI and EEG experiments (see (Smith and Little, 2018)). Five of those participants (4 females, mean age = 22.2; SD = 5.1) completed the whole study (one was an author of the study, A.M.). One participant was excluded after 3 fMRI days because their sensitivity index d' during the adaptation phase was ≤ 2.5 (see

above iii). All participants had no history of neurological or psychiatric illnesses, normal or corrected-to-normal vision and normal hearing. All participants gave informed written consent to participate in the study, which was approved by the research ethics committee of the University of Birmingham (approval number: ERN_11_0470AP4) and was conducted in accordance with the principles outlined in the Declaration of Helsinki.

## Stimuli

The auditory stimulus consisted of a burst of white noise with a duration of 50 ms and 5 ms on/off ramp delivered at a 75 dB sound pressure level. To create a virtual auditory spatial signal, the noise was convolved with spatially specific head-related transfer functions thereby providing both binaural and monaural cues for sound location ((Gardner and Martin, 1995)). Head-related transfer functions from the available locations in the MIT database (http://sound.media.mit.edu/resources/KEMAR.html) were interpolated to the desired spatial locations. For the EEG experiment, scanner background noise was superimposed on the spatial sound stimuli to match the task environment of the fMRI experiment. The visual stimulus was a cloud of 15 white dots (diameter = 0.4° visual angle) sampled from a bivariate Gaussian distribution with a vertical and horizontal standard deviation of 1.5° and a duration of 50 ms presented on a dark grey background (90% contrast) in synchrony with the auditory stimuli.

## Experimental design and procedure

The psychophysics, fMRI, EEG experiments conformed to the same design. Each included three phases: (i) unisensory auditory pre-adaptation, (ii) audiovisual adaptation (AV- or VA-adaptation) and (iii) unisensory auditory post-adaptation (postAV-adaptation or postVA-adaptation) (Figure 1B). All experiments (i.e., psychophysics, fMRI, EEG) included two days for left (i.e., VA) and two days for right (i.e., AV) adaptation (except for one participant who completed the whole EEG experiment over two days) resulting in a total of 12 days of experimental testing + 1 day pre-screening for each participant. The order of left and right adaptation days was counterbalanced across participants. Throughout all experiments participants were instructed to fixate a central fixation cross (0.5˚ diameter).

### *Auditory pre- and post-adaptation*

In the unisensory auditory pre- and post-adaptation phases (Figure 1B), participants were presented with auditory stimuli that were sampled uniformly from 7 spatial locations (±12°, ±5°, ±2° and 0° visual angle) along the azimuth (stimulus onset asynchrony (SOA) = 2000 ms ± 200 ms jitter). Across all experiments (including psychophysics), participants performed a two-alternative forced-choice left-right spatial classification task explicitly only on a fraction of trials (22%), the so-called 'response trials'. The response trials were indicated by a brief (i.e., 200 ms duration) dimming of the fixation cross to 55% of the initial contrast 500 ms after sound onset. They were randomly interspersed amongst non-response trials to maintain attention.

Participants indicated their left-right spatial classification response by pressing one of two buttons with their index or middle fingers of their left or right hand. The response hand was alternated over runs within a day to further control for potential motor confounds in our neuroimaging analysis (see fMRI multivariate decoding section below). The order of left and right response hands was counter-balanced across days.

The fMRI decoding was based only on the non-response trials to minimize the effects of motor confounds. The inclusion of response trials enabled us to estimate psychometric functions. Moreover, task performance and maintenance of attention were monitored using the signal sensitivity index *d'* that was computed by comparing the response rates on response trials with those on non-response trials (see Signal detection measures section below).

To increase design efficiency (in the fMRI study), stimuli were presented from one of the 7 possible locations in a pseudorandomized order in ~36 s blocks of 18 stimuli interleaved with 6 s fixation. In the pre-adaptation

phase one run consisted of 10 blocks (i.e., 18 stimuli x 10 blocks = 180 stimuli) and 9 fixation intervals and lasted for ~7 minutes. After five blocks we inserted a longer fixation interval of 15 s. In the post-adaptation phase, each run started with an audiovisual adaptation block (see next section) followed by 5 post-adaptation blocks. In total, a run included 2 audiovisual adaptation and 10 post-adaptation blocks (~14 minutes). As in the pre-adaptation phase, the post-adaptation blocks were interleaved with 6 s fixation.
The EEG experiment followed the same overall structure as the fMRI experiment, except that the fixation intervals were not included to reduce the duration of the experiment. In the psychophysics experiment, runs in the pre-adaptation phase consisted of 5 stimulus blocks (i.e., 18 stimuli x 5 blocks = 90 stimuli) and 4 fixation intervals (~3.5 minutes).
To maintain participants' engagement with the task, they received visual feedback after every 5 pre- or post-adaptation blocks (i.e., 90 trials) about their detection performance (i.e., the dimming of the fixation cross, n.b. not their localization performance) based on the sensitivity index *d'* (i.e., d' > 3.5: smiling face, 2.5 < d' < 3.5: neutral face, 2.5 < d': sad face).
The pre-adaptation phase of the fMRI (resp. EEG) experiment included 3600 auditory trials (2800 no response + 800 response trials, i.e., 18 stimuli per block x 10 blocks per run x 5 runs per day x 4 days). The postVA (resp. postAV)-adaptation phase included 1440 auditory trials (1120 no response + 320 response trials) for postAV-adaptation and the same amount of trials for postVA-adaptation phase (i.e., 18 stimuli per block x 10 blocks per run x 4 runs per day x 2 days). In the psychophysics experiment, the pre-adaptation phase included 1800 auditory trials (1400 no response + 400 response trials, i.e., 18 stimuli per block x 5 blocks per run x 5 runs per day x 4 days). The postVA (resp. postAV)-adaptation phase included 1800 auditory trials (1400 no response + 400 response trials) for postAV-adaptation and the same number of trials for postVA-adaptation phase (i.e., 18 trials per block x 10 blocks per run x 5 runs per day x 2 days)

***Audiovisual adaptation***

In the audiovisual adaptation phase, participants were presented with spatially disparate (±15° visual angle) audiovisual stimuli (SOA = 500 ms): the visual stimulus component was uniformly sampled from three locations along the azimuth (i.e., -5°, 0°, 5°) to ensure that the visual stimulus did not introduce spatial biases. The auditory stimulus was spatially shifted on separate days by 15° either to the right (i.e., VA-adaptation) or to the left (i.e., AV-adaptation) with respect to the visual stimulus. Hence, we included the following audiovisual stimulus location pairs: [A = -20°, V = -5°], [A = -15°, V = 0°], [A = -10°, V = 5°] in (right) AV-adaptation phases and [A = 10°, V = -5°], [A = 15°, V = 0°], [A = 20°, V = 5°] in (left) VA-adaptation phases. The locations of the audiovisual stimulus pairs were fixed within mini blocks of 5 (psychophysics, i.e., duration of 2.5 s) or 20 (fMRI, EEG, i.e., duration of 10 s) consecutive trials. We did not introduce fixation intervals between the mini blocks within the audovisual adaptation phase.
Participants were instructed to detect slightly dimmer visual stimuli (80% of normal contrast), pseudorandomly interspersed on 10% of the trials (i.e., so-called 'response trials'). The non-spatial task ensures maintenance of participants' attention and introduces audiovisual recalibration mainly at the perceptual rather than decisional or motor response level. To allow sufficient time for responding (given the short SOA of 500 ms), we ensured that each response trial was followed by at least 3 consecutive non-response trials.
One audiovisual adaptation block included either 360 trials (for the first adaptation block in psychophysics and typically in fMRI and EEG, i.e., 3 minutes) or 120 trials (for the remaining adaptation blocks in psychophysics, i.e., 1 minute). To maintain participants' engagement with the task, they received visual feedback after every adaptation block (i.e., 360 or 120 trials) about their detection performance (i.e. of the dimmer visual stimuli) based on the sensitivity index *d'* (i.e., d' > 3.5: smiling face, 2.5 < d' < 3.5: neutral face, 2.5 < d': sad face).
In total, both fMRI and EEG studies included 5760 audiovisual adaptation trials (5184 no response + 576 response trials) for the (left) VA-adaptation phase (i.e., 360 trials per block x 2 blocks per run x 4 runs per day x 2 days). Likewise, there were 5760 audiovisual trials for the (right) VA-adaptation phase.

### *Organization of each testing day*

Each day started with 5 pre-adaptation runs amounting to 900 trials per day (i.e. 18 trials per block x 10 stimulation blocks per run x 5 runs). The subsequent 4 runs (resp. 5 runs in psychophysics) included 2 VA (or AV)-adaptation blocks and 10 postVA (or postAV)-adaptation blocks, such that each VA (or AV)-adaptation block was followed by 5 postVA (or postAV)-adaptation blocks. Hence, each day included 2880 trials AV adaptation trials (i.e, 360 per block x 2 blocks per run x 4 runs per day in fMRI and EEG and 2400 AV adaptation trials in psychophysics). Further, each day included 720 post-adaptation trials in fMRI and EEG (i.e. 18 trials per block x 10 blocks per run x 4 runs per day) and 900 trials in psychophysics (because of 5 runs per day, see Figure 1B). The AV adaptation and post-adaptation blocks were interleaved with 15 s of fixation.

## Experimental setup

In all experiments, visual and auditory stimuli were presented using Psychtoolbox version 3.0.11 (Brainard, 1997; Kleiner et al., 2007) under MATLAB R2011b (MathWorks Inc.) on a MacBook Pro running Mac OSX 10.6.8 (Apple Inc.). In the psychophysics and EEG experiments, participants were seated at a desk with their head rested on a chinrest. Two accessory rods were mounted on the chin rest serving as forehead rest and allowing stable and reliable head positioning. Visual stimuli were presented at a viewing distance of 60 cm via a gamma-corrected 24'' LCD monitor (ProLite B2483HS, iiyama Corp.) with a resolution of 1920 x 1080 pixels at a frame rate of 60 Hz. Auditory stimuli were delivered via circumaural headphones (HD 280 Pro, Sennheiser electronic GmbH & Co. KG) in the psychophysics experiment and via in-ear earphones (E-A-RTONE GOLD, 3M Company Auditory Systems) in the EEG experiment. Participants used a standard USB keyboard for responding. In the fMRI experiment, visual stimuli were back projected to a plexiglass screen using a D-ILA projector (DLA-SX21, JVC, JVCKENWOOD UK Ltd.) with a resolution of 1400 x 1050 pixels at a frame rate of 60 Hz. The screen was visible to the subject through a mirror mounted on the magnetic resonance (MR) head coil and the eye-to-screen distance was 68 cm. Auditory stimuli were delivered via a pair of MR compatible headphones (MR Confon HP-VS03, Cambridge Research Systems Ltd). Participants responded using a two-button MR-compatible keypads (LXPAD 1x5-10M, NATA Technologies). Exact audiovisual onset timing in adaptation trials was confirmed by recording visual and auditory signals concurrently with a photodiode and a microphone.

## Eye movement recording

Eye movement recordings were calibrated in the recommended field of view (32° horizontally and 24° vertically) for the EyeLink 1000 Plus system (SR Research Ltd.) with the desktop mount at a sampling rate of 2000 Hz. Eye position data were on-line parsed into events (saccade, fixation, eye blink) using the EyeLink 1000 Plus software. The 'cognitive configuration' was used for saccade detection (velocity threshold = 30°/sec, acceleration threshold = 8000°/sec$^2$, motion threshold = 0.15°) with an additional criterion of radial amplitude > 1°. Fixation position was post-hoc offset corrected. In the fMRI experiment, precise positioning of participants' heads inside the scanner bore was critical for the sensitive measurement of spatial recalibration, so that high quality eye movement recordings were not possible.

## Behavioural analysis for psychophysics, fMRI and EEG experiments

### *Signal detection measures*

In the unisensory auditory pre- and post-adaptation phases as well as the AV- and VA-adaptation phases participants were instructed to respond only on a fraction of trials (22% auditory pre- and post-adaptation,

10% AV- and VA-adaptation), the so-called 'response trials'. This enabled us to assess their performance in terms of the signal sensitivity measure d':

$$d' = Z(p_{hit}) - Z(p_{false\ alarm})$$

where $p_{hit}$ is the hit rate (i.e., fraction of 'response trials' with responses). Hits are 'response trials', on which observers gave a response. False alarms are 'non-response trials', on which observers gave a response. Hence, this analysis considered only whether observers gave a response irrespective of whether the spatial response (in the pre/post adaptation phases) was correct. 100% hit rate and 0% false alarm rate were approximated by 99.999% and 0.001%, respectively, to enable the calculation of Z-scores.

### *Psychometric functions*

For the pre-, postVA- and postAV-adaptation phases we estimated psychometric functions to the unisensory auditory spatial classification responses obtained from the response trials. We fitted cumulative Gaussians as psychometric functions (PF) to the percentage 'perceived right' responses as a function of stimulus location (±12°, ±5°, ±2° and 0°) (Figure 2). The models were fitted individually to the behavioural data of each participant with maximum-likelihood estimation as implemented in the Palamedes toolbox ((Prins and Kingdom, 2018), www.palamedestoolbox.org). To enable reliable parameter estimates for each participant, we employed a multi-condition fitting using the following constraints: (i) the just noticeable differences (JND or slope parameter) were set equal across all conditions (i.e., pre-, postVA- and postAV-adaptation phases); (ii) guess and lapse rates were set equal to each other and (iii) equal across all conditions. Furthermore, (iv) we constrained the fitted guess and lapse rate parameters to be within 0 and 0.1.

For statistical inference, we first assessed the goodness of fit of the cumulative Gaussian for each condition and participant using a likelihood ratio test. This likelihood ratio test compares the likelihood of participants' responses given the model that is constrained by a cumulative Gaussian function (i.e., our 'target model') to the likelihood given by a so-called 'saturated' model that models observers' responses with one parameter for each stimulus location in each condition. The resulting likelihood ratio for the original data set is then compared with a null distribution of likelihood ratios generated by parametrically bootstrapping the data (5000x) from the 'target model' fitted to the original data set and refitting the 'target' and 'saturated models'. Since the likelihood ratio for the original data set was not smaller in any of the participants than 5% of the parametrically bootstrapped likelihood ratios (i.e., $p > 0.05$), we inferred sufficient goodness of fit for all participants. Second, we assessed whether AV- and VA-adaptation induced a shift in participants' perceived location of unisensory auditory stimuli by comparing a 'static or no adaptation' model with an 'adaptation' model. The static model assumes no effect of audiovisual spatial adaptation on unisensory auditory spatial representations by constraining the point of subjective equality (PSE) for the pre-adaptation, postVA-adaptation, and postAV-adaptation PFs to be equal. The adaptation model includes three PSE values for the pre-adaptation, postVA-adaptation, and postAV-adaptation PFs thereby accommodating potential shifts in the PSE values as a result of audiovisual recalibration. For each participant, we calculated the log-likelihood for each of the two models and computed the Akaike Information Criterion (AIC (Akaike, 1998)) as an approximation to the model evidence. The AIC depends on both model complexity and model fit. Using these AICs we performed Bayesian model comparison at the random effects level as implemented in SPM12 to obtain the protected exceedance probability (the probability that a given model is more likely than any other model, beyond differences due to chance) for the candidate models (Rigoux et al., 2014).

## fMRI data acquisition and analysis

### *fMRI data acquisition*

We used a 3T Philips Achieva scanner to acquire both T1-weighted anatomical images (TR/TE/TI, 7.4/3.5/min. 989 ms; 176 slices; image matrix, 256 x 256; spatial resolution, 1 x 1 x 1 $mm^3$ voxels) and T2*-weighted echo-planar images (EPI) with blood oxygenation level-dependent (BOLD) contrast (fast field echo; TR/TE, 2800/40 ms; 38 axial slices acquired in ascending direction; image matrix, 76 x 75; slice thickness, 2.5

mm; interslice gap, 0.5 mm; spatial resolution, 3 x 3 x 3 $mm^3$ voxels). On each of the four days we acquired five auditory pre-adaptation runs and 4 audiovisual adaptation runs. Each fMRI run started and ended with 10 s fixation and the first 4 volumes of each run were discarded to allow for T1 equilibration effects.

***Pre-processing and general linear model***

The data were analysed with Statistical Parametric Mapping (SPM12; Wellcome Trust Centre for Neuroimaging, London, UK; http://www.fil.ion.ucl.ac.uk/spm/; (Friston et al., 1994)). Scans from each participant were realigned using the first as a reference, unwarped and corrected for slice timing. The time series in each voxel was high pass filtered to 1/128 Hz. The EPI images were spatially smoothed with a Gaussian kernel of 3 mm FWHM and analysed in native space. The data were modelled in a mixed event/block related fashion with regressors entered into the design matrix after convolving the unit impulse (representing the onset of a single trial) or the block with a canonical hemodynamic response function and its first temporal derivative. In the unisensory auditory pre- and post-adaptation phases, unisensory sound stimuli were modelled as events separately for each of our 7 (sound location) x 2 (response vs. no-response) x 3 (pre, post-VA, post-AV) conditions. In the AV- and VA-adaptation phases, AV stimulus presentations were modelled as blocks separately for the 3 (visual locations) x 2 (VA vs. AV adaptation) conditions. In addition, we modelled all response trials during the adaptation phase with a single regressor irrespective of stimulus position to account for motor responses. Realignment parameters were included as nuisance covariates to account for residual motion artefacts. Condition-specific effects for each subject were estimated according to the general linear model (GLM) separately for each run. To minimise confounds of motor response, we limited all subsequent fMRI analyses to the parameter estimates pertaining to the 'no response' trials alone. For the BOLD response analysis, we computed contrast images comparing auditory activation with fixation baseline (auditory stimulus at a particular location > fixation) for each condition in each subject (averaged over runs) resulting in 21 contrast images (i.e., 7 (sound location) x 3 (pre, postVA, postAV)). Moreover, we computed a contrast and associated t-image that compared all 21 sound conditions relative to fixation baseline. This contrast would be later used to identify sound-responsive voxels.

For the multivariate decoding and representational similarity analyses, we applied multivariate spatial noise normalization to the parameter estimates using the noise covariance matrix obtained from the residuals of the GLM and the optimal shrinkage method described in (Ledoit and Wolf, 2004). Finally, we normalized each image by dividing the values of each voxel by the Euclidean norm of this particular image.

***Regions of interest for fMRI analysis***

We defined five regions of interest (ROI) that have previously been implicated in auditory spatial processing based on neurophysiology and neuroimaging research (Rauschecker and Tian, 2000; Schlack et al., 2005; van der Heijden et al., 2019). All regions were combined from the left and right hemispheres, unless otherwise stated.

Heschl's gyrus (HG), higher auditory cortex (hA) and (IPL) were defined using the following parcellations of the Destrieux atlas of Freesurfer 5.3.0 (Destrieux et al., 2010): (i) HG: Heschl's gyrus and anterior transverse temporal gyrus; (ii) hA: higher auditory cortex, i.e., transverse temporal sulcus, planum temporale and posterior ramus of the lateral sulcus; (iii) IPL: inferior parietal lobule, i.e., supramarginal gyrus and inferior part of the postcentral sulcus. The intraparietal sulcus (IPS) and frontal eye field (FEF) were defined using the following group-level retinotopic probabilistic maps (Wang et al., 2015): (iv) IPS: IPS0, IPS1, IPS2, IPS3, IPS4, IPS5 and SPL1; (v) FEF: hFEF. All probabilistic maps were thresholded to a probability of 0.1 (i.e., probability that a vertex belongs to a particular ROI) and inverse normalized into each participant's native space.

***fMRI multivariate decoding – Spatial encoding and recalibration indices***

We trained linear support vector regression (SVR) models as implemented in LIBSVM 3.17 (Chang and Lin, 2011) to accommodate the continuous nature of the auditory locations. More specifically, we extracted the voxel response patterns in a particular region of interest (e.g., HG) from the pre-whitened and normalized parameter estimate images pertaining to the magnitude of the BOLD response for each auditory location

condition and run. To minimize effects of motor processing on the decoded response, we only used the parameter estimate images pertaining to the 'no response trials' (i.e., on which participants did not make a response). In a 4-fold stratified cross-validation scheme, SVR models were trained with default SVR hyper-parameters (C = 1, ν = 0.5) to learn the mapping from the condition-specific fMRI response patterns (i.e., examples) to the condition specific spatial locations (i.e., labels) using examples selectively from the unisensory auditory pre-adaptation runs of all but one fold(Rohe and Noppeney, 2016, 2015). This learnt mapping was then used to decode the spatial locations from the voxel response patterns in the remaining fold (i.e., separate runs) of the pre-adaptation examples and all postVA- and postAV-adaptation examples. This procedure was repeated across all folds. As a result of the 4-fold cross-validation scheme, we obtained one predicted spatial location for each example of the pre-adaptation test runs and four predicted spatial locations for each example of the post-adaptation test runs. In the latter case, we calculated one final predicted location for each example by averaging across the four predicted locations.

To determine whether a region of interest encodes auditory spatial representations, we computed the Pearson correlation coefficients between the true and the decoded auditory locations for the pre-adaptation runs for each participant as a 'spatial encoding index'.

To determine whether auditory spatial representations in a region of interest are recalibrated by concurrent visual signals, we binarized the predicted auditory locations into left vs. right predictions and computed the difference in the fraction of 'decoded right responses' (i.e., conditional probability) between auditory postVA- and postAV-adaptation phases as a 'recalibration index' (RI).

$$RI = p('decoded\ right'|postAV\ examples) - p('decoded\ right'|postVA\ examples)$$

Importantly, postVA- and postAV-adaptation phases were matched in terms of time, exposure, training, and other non-specific effects.

To allow for generalization at the population level, we entered the subject-specific Fisher z-transformed spatial encoding and recalibration indices into separate bootstrap-based one sample t-tests against zero at the group level (Efron and Tibshirani, 1993). Briefly, an empirical null distribution of t-values was generated from the data by resampling the subject-specific values with replacement, subtracting the across-subjects mean and computing the one-sample t-statistic for each bootstrap sample. A right-tailed p-value was computed as the proportion of t-values in the bootstrapped t-distribution greater than the observed t-value.

***fMRI multivariate decoding – neurometric functions***

Consistent with our behavioural analysis we assessed spatial encoding and recalibration in our ROIs using neurometric functions (NF, (Rohe and Noppeney, 2018)). For this, we first binarized the decoded sound locations into 'decoded left' and 'decoded right' and computed the percentage of 'decoded right' for each sound location and participant. Because of the lower signal-to-noise ratio of fMRI data, we fitted cumulative Gaussian functions to the percentage of 'decoded right' averaged across participants as a function of stimulus location (±12°, ±5°, ±2° and 0°). To enable reliable parameter estimates, we employed multi-condition fitting using the following constraints: (i) the JNDs (i.e., slope parameter) were constrained to be equal across all conditions; (ii) guess and lapse rates were set to be equal to each other and (iii) constrained to be equal across all conditions; (iv) lapse rates were set to be within 0 and 0.45 (to account for the lower signal-to-noise ratio of fMRI data). Likelihood ratio test as described in the behavioural analysis section confirmed adequate goodness of fit for the neurometric functions in each ROI (i.e., $p>0.05$ across all ROIs, Table S3).

We assessed whether AV- and VA-adaptation induced a shift in the decoded location of the unisensory auditory stimuli. Consistent with our behavioural analysis, we compared a 'static or no adaptation' model with an 'adaptation' model. The static model assumes no effect of audiovisual spatial adaptation on unisensory auditory spatial representations by constraining the point of subjective equality (PSE) for the pre-adaptation, postVA-adaptation, and postAV-adaptation test PFs to be equal. The adaptation model includes three PSE values for the pre-adaptation, postVA-adaptation and post-AV-adaptation test PFs thereby

accommodating potential shifts in the PSE values as a result of audiovisual recalibration. Using Bayesian model comparison we assessed the two models using the Akaike Information Criterion (AIC (Akaike, 1998)) as an approximation to the model evidence.

### *fMRI multivariate pattern analysis – representational dissimilarity analyses and multidimensional scaling*

We characterized the spatial representations using representational similarity analysis (RSA, as implemented in the rsatoolbox (Nili et al., 2014). To this end, we generated 21 condition specific contrast images for the 7 auditory spatial locations x 3 (pre-adaptation, postVA-adaptation, postAV-adaptation) by averaging parameter estimate images across fMRI runs for each participant. We then characterized the geometry of spatial representations using representational dissimilarity matrices (RDMs) based on the Mahalanobis distances for each participant and each ROI (i.e., HG, hA, IPS, IPL, FEF) separately for pre-adaptation, postVA- and postAV-adaptation phases. For pre-adaptation, we computed a rank-transformed (and scaled between 0 and 1) 7 x 7 RDM based on the 7 auditory contrast images from the pre-adaptation phase. To characterize and visualize the effect of AV- and VA-adaptation on the representational geometry of spatial representations we computed a 14 x 14 matrix (i.e., 7 spatial locations for postVA-adaptation and 7 spatial locations for postAV-adaptation) where we rank-transformed (and scaled between 0 and 1) the 7 x 7 sub-components of the post-adaptation RDMs (i.e., separately for the postAV- and postVA-adaptation). For visualization (see Figure 3C) we averaged the rank transformed matrices across participants. To provide a more intuitive visualization of the representational structure, we projected these group level RDMs (i.e., averaged across participants) onto a one-dimensional space using non-classical multidimensional scaling (MDS) with non-metric scaling. We constrained the MDS to one single dimension given our a priori hypothesis that representations of spatial locations are organized along one spatial dimension along the azimuth.

## EEG data acquisition and analysis

### *EEG data acquisition*

Continuous EEG signals were recorded from 64 channels using Ag/AgCl active electrodes arranged in 10-20 layout (ActiCap, Brain Products GmbH, Gilching, Germany) at a sampling rate of 1000 Hz with FCz as reference. Channel impedances were kept below 10kΩ.

### *EEG pre-processing*

Pre-processing was performed with the FieldTrip toolbox ((Oostenveld et al., 2011), http://www.fieldtriptoolbox.org/). Raw data were high pass filtered at 0.1 Hz, inspected for bad channels, re-referenced to an average reference, and low pass filtered at 45 Hz. Bad channels were identified by visual inspection, rejected and interpolated using the neighbouring channels. Trial epochs for the unisensory auditory pre-adaptation and the auditory post-adaptation conditions were extracted from 100 ms pre-stimulus to 500 ms post-stimulus (i.e., the onset of the response cue on the response trials), baseline corrected and down-sampled to 200 Hz. Epochs containing artefacts within the time window of interest (i.e., between 0 – 500 ms post-stimulus) were identified based on visual inspection and rejected. Furthermore, based on eyetracking data, trials were rejected if they (i) contained eye blinks or (ii) saccades or (iii) the eye gaze was away from the fixation cross for more than 2 degrees (% rejected trials across subjects mean ± SEM: 8.2 ± 1.0%). Grand average ERPs were computed by averaging all trials for each condition first within each participant and then across participants.

For the multivariate analysis, we applied spatial multivariate noise normalization to the individual trials using a noise covariance matrix estimated separately for each time point and the optimal shrinkage method described in (Ledoit and Wolf, 2004). Furthermore, the EEG activity patterns were divided by their Euclidean norm separately for each time point and trial for normalization. Since the EEG responses on trials with and without behavioural response were identical until 500 ms post-stimulus, we pooled over response and no response trials in all EEG analyses.

***EEG multivariate decoding – spatial encoding and recalibration indices***

We trained linear support vector regression models (SVR, LIBSVM (Chang and Lin, 2011)) on evoked potentials (16 trial-averages) separately for 50 ms moving time windows from 0 ms to 500 ms post-stimulus in a 4-fold (i.e., including ~50 evoked potentials per condition per fold) stratified cross-validation procedure with default values for hyperparameters ($C = 1$, $\nu = 0.5$). First, data from all conditions were randomly split into four folds under the constraint that all conditions and auditory locations were equally frequent in each fold. Second, to maximize the signal to noise ratio we computed evoked potentials averaged across sets of 16 randomly sampled individual trials from the same condition within each fold (Grootswagers et al., 2017). Third, we then trained the SVR model to learn the mapping from activation patterns (16 trial averages) to external auditory space using examples selectively from the pre-adaptation conditions from all but one fold (Aller and Noppeney, 2019). The learnt mapping was then used to decode the spatial location from the EEG activity patterns of the pre-adaptation test examples in the remaining fold and all post-adaptation examples. As a result, we obtained one predicted spatial location for each example of the pre-adaptation test runs and four predicted spatial locations for each example of the post-adaptation test runs. In the latter case, we calculated one predicted location for each example by averaging across those four predicted locations.

To resolve spatial encoding across time we trained SVR models on evoked potential patterns concatenated over 50 ms time windows, which were shifted from -100 ms to 500 ms post-stimulus in increments of 5 ms (Grootswagers et al., 2017). The onset of the windows was registered to the final time point, so that no future information was included in the time window. To increase the reliability of the decoding we repeated the stratified cross-validated SVR decoding procedure 50 times; each time we formed new evoked potentials by averaging across randomly sampled 16 trials. We report results averaged across those 50 cross-validated SVR decodings.

As in the fMRI analysis, we assessed spatial encoding in terms of the Pearson correlation coefficient between the true and decoded auditory locations, i.e., the 'spatial encoding index' computed separately for each 50 ms time window from 0 ms to 500 ms. Likewise, we computed the 'recalibration index' for each time window to examine whether spatial representations are altered as a result of audiovisual adaptation. The spatial encoding index and the recalibration index were entered into a random effects (i.e., between subjects) bootstrap-based t-test against zero (Efron and Tibshirani, 1993). We report results corrected for multiple comparisons across the entire 0 to 500 ms time window using a cluster based correction with an auxiliary cluster defining height threshold of $p < 0.05$ uncorrected (Maris and Oostenveld, 2007).

Based on our a priori hypothesis that spatial location is encoded in the N100 potential (see (Bruns et al., 2011)), we also performed between subjects bootstrap-based t-tests on the spatial encoding index and the recalibration index by decoding the evoked potentials averaged within a 70 ms – 130 ms time window.

## Spatial hemifield and decisional models for fMRI and EEG

In the current study observers performed a binary spatial classification task on the sound stimuli from seven different locations enabling us to dissociate spatial and decisional coding in the brain. In the following we will describe the spatial and decisional encoding models that we used to model the regional mean BOLD-response and the fMRI/EEG pattern similarity structure over spatial locations.

*Spatial hemifield model*

The hemifield model (McAlpine et al., 2001; Ortiz-Rios et al., 2017; Salminen et al., 2009; Stecker et al., 2005) encodes sound location in the relative activity of two sub-populations of neurons each broadly tuned either to the ipsi- or contra-lateral hemifield (Figure 4A). We simulated 360 neurons with broad Gaussian tuning functions. The standard deviation was set to 64˚. The means of the tuning functions were sampled uniformly from 80˚ to 100˚ for the neuronal population tuned to the contralateral hemifield and from -80˚ to -100˚ azimuth for the neuronal population tuned to the ipsilateral hemifield. Consistent with previous research

(Salminen et al., 2009) the ratio of the ipsi- and contralaterally tuned neurons was set to 30%/70% (see Figure 4A, S1A).
For the pre-adaptation conditions we sampled neural responses from the seven sound locations (i.e., -12˚, -5˚, -2˚, 0˚, 2˚, 5˚, 12˚) in our paradigm. For the post-adaptation conditions, we sampled again from these seven locations for the non-recalibration model. For the recalibration model we sampled the neural responses from the above locations shifted by 2.3˚ to the right (postVA-adaptation) or left (postAV-adaptation). The shift by 2.3˚ was calculated as the difference between the across-subjects mean PSE values in postVA- and postAV-adaptation phases from the psychometric functions.
While accumulating research has shown that sound location is coded by broadly tuned neural populations in auditory cortices in mammals (Werner-Reiss and Groh, 2008; Wood et al., 2019), less is known about auditory spatial coding in parietal or frontal cortices from single cell neurophysiology. For instance, one previous neurophysiology study in non-human primates has suggested that neurons in parietal area VIP have narrow auditory receptive fields that are congruent with visual receptive fields (Schlack et al., 2005). Critically, the current study selected only near-central sound locations to optimize the characterization of recalibration rather than to dissociate between different spatial encoding principles. For those near-central locations, the hemifield model and a place code model with narrow tuning functions make indistinguishable predictions for the pattern similarity structure (see Supplemental Methods and Figure S1 for further information). We have therefore used the hemifield model as a generic perceptual spatial encoding model and not that comparable predictions could be made with the place code model.

*Decisional model*
In the decisional model the activity of a neuron encodes observers' decisional uncertainty that depends non-linearly on the distance between observers' spatial estimates and their left-right spatial classification boundary (Grinband et al., 2006) according to

$$decisional\ uncertainty = \frac{\big||F(x|\,\mu,\sigma) - 0.5| - 0.5\big|}{0.5}$$

where $F(x|\,\mu,\sigma)$ is the cumulative normal distribution function with mean $\mu$ and standard deviation $\sigma$ evaluated at spatial location $x$:

$$F(x|\,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-(k-\mu)^2}{2\sigma^2}}\,dk$$

The standard deviations of the cumulative normal distribution underlying the neural uncertainty functions were set to 10˚ and their means were uniformly sampled between -1˚ and +1˚ (see Figure 4A). We simulated responses from 360 neurons.
Exactly as for the spatial model, we sampled neural responses from the seven sound locations (i.e., -12˚, -5˚, -2˚, 0˚, 2˚, 5˚, 12˚) for the pre-adaptation conditions. For the post-adaptation conditions, we sampled again from these seven locations for the non-recalibration model. For the recalibration model we sampled the neural responses from the above locations shifted by 2.3˚ to the right (postVA-adaptation) or left (postAV-adaptation) as we did for the hemifield model.
As shown in Figure 4B and C, the spatial and decisional models make distinct predictions for the regional mean BOLD-response and the pattern similarity structure over the 21 conditions = 7 spatial locations x 3 phases (pre, postVA, postAV).

***Regional mean BOLD-response: spatial and decisional linear mixed effects models***
*Regional mean BOLD-response:* For each of the 2 (hemisphere: left, right) x 5 (ROI: HG, hA, IPS, IPL, FEF) regions we selected the 20 most reliably responsive voxels, i.e., with the greatest t-values for all unisensory sound conditions relative to fixation. For each of those 10 regions we extracted the BOLD-response magnitude for each of the 7 locations x 3 phases (pre-adaptation, postAV, postVA) and formed the regional mean.

*Linear mixed effects modelling*: To account for lateralization effects (Ortiz-Rios et al., 2017), we performed separate analyses for each region and hemisphere. Separately for each hemisphere we used the spatial (resp. decisional) model to generate predicted regional mean BOLD-responses for each of the 21 conditions = 7 locations x 3 phases (pre-adaptation, postAV, postVA) by averaging activations of 360 simulated neurons. We generated seven linear mixed effects (LME) models that varied in their fixed effects predictors: (i) The null LME included only a single intercept term. (ii) The spatial LME model (S) included the predictor from the spatial encoding model without recalibration and an intercept term. (iii) The decisional LME model (D) included the predictor from the decisional model without recalibration and an intercept term. The remaining LME models included spatial, decisional and intercept terms (i.e., three fixed effects regressors) and factorially manipulated whether the spatial and/or the decisional predictor modelled recalibration: (iv) spatial without recalibration + decisional without recalibration (S+D), (v) spatial with recalibration + decisional without recalibration ($S_R$+D), (vi) spatial with recalibration + decisional without recalibration (S+$D_R$) and (vii) spatial with recalibration + decisional with recalibration ($S_R$+$D_R$). Subject level effects were included as random effects.
For each of the 2 (hemisphere: left, right) x 5 (ROI: HG, hA, IPS, IPL, FEF) regions we fitted the seven LME models using maximum likelihood estimation and computed the Bayesian information criterion (BIC) as an approximation to the model evidence. The figures show the natural logarithm of the Bayes factors (ln-Bayes factors) averaged across hemispheres for LME models (ii)-(vii) relative to the null LME.
For visualization purposes only (i.e., Figure 5A), we averaged the BOLD-responses across hemispheres whilst controlling for the fact that the effects of spatial location and recalibration differ between the hemispheres as indicated by our model comparison results (Ortiz-Rios et al., 2017). For the pre-adaptation conditions, we averaged the BOLD-responses for [-12°, -5°, -2°, 0°, 2°, 5°, 12°] in the left hemisphere with [12°, 5°, 2°, 0°, -2°, -5°, -12°] in the right hemisphere. For the post-adaptation conditions, we averaged the BOLD-responses for [-12°, -5°, -2°, 0°, 2°, 5°, 12°] of the postVA-adaptation conditions in the left hemisphere with [12°, 5°, 2°, 0°, -2°, -5°, -12°] of the postAV-adaptation conditions in the right hemisphere. In this way, we obtained estimates of the BOLD-response magnitude for each of the 21 conditions for a representative left hemisphere region across all five ROIs (Figure 5A).

***Multivariate pattern: pattern component modelling of fMRI and EEG data***

To assess whether spatial and/or decisional models can explain the fine-scale fMRI or EEG activity patterns across the 7 (sound location) x 3 (pre, post-VA, post-AV) = 21 conditions we combined Pattern Component Modelling (PCM (Diedrichsen et al., 2018), https://github.com/jdiedrichsen/pcm_toolbox) and Bayesian model comparison. Like the more widely used representational similarity analyses, pattern component modelling allows one to investigate whether specific representational structures – as implied for instance by the spatial or decisional model – are expressed in fMRI or EEG activity patterns. Critically, PCM goes beyond standard representational similarity analyses by allowing us to assess arbitrary mixtures of representational components such as a combined spatial + decisional model (i.e., with unknown mixture weights).
PCM estimates the likelihood of the observed neural activity under a multivariate linear model:

$$\boldsymbol{Y} = \boldsymbol{ZU} + \boldsymbol{XB} + \mathbf{E}$$
$$\boldsymbol{u_p} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{G})$$
$$\boldsymbol{\varepsilon_p} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{I\sigma^2})$$

$\boldsymbol{Y}$ is an (N x P) data matrix. N is equal to the number of conditions (i.e., 21) times the number of runs; P is equal to the number of measurement channels (fMRI voxels or EEG channels). Each row in $\boldsymbol{Y}$ includes the pre-whitened activity pattern, i.e., across-voxel parameter estimate patterns from the first-level GLM in fMRI and across-electrode evoked potentials in EEG. $\boldsymbol{Z}$ is a (N x 21 conditions) design matrix specifying the experimental conditions with 21 columns (i.e., 21 conditions = 7 sound locations x [pre, postVA, postAV]). $\boldsymbol{U}$ is a matrix of true activity profiles (21 conditions x number of channels). $\boldsymbol{X}$ is the design matrix for the patterns of no interest $\boldsymbol{B}$ (e.g., condition-independent means of each voxel-specific activity profile). $\mathbf{E}$ is the N x P matrix with the error terms. PCM treats the true activity profiles $\boldsymbol{u_p}$ of each measurement channel (the columns of $\boldsymbol{U}$) as random variables sampled from a multivariate normal distribution $\boldsymbol{u_p} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{G})$.

Here, $\boldsymbol{G}$ is the second moment matrix of activity profiles, which describes the similarity structure over the 21 conditions in the experiment.
$\boldsymbol{G}$ can be expressed as a linear combination of second moment matrices (e.g., generated from the decisional and/or spatial model) with $\exp(\theta_h)$ determining the weight of each second moment component to the second moment matrix $\boldsymbol{G}$ estimated from the measured neural activity profiles over conditions.

$$G = \sum_h \exp(\theta_h) G_h$$

PCM estimates the marginal likelihood of the observed data given the model and the model parameters $\boldsymbol{\theta}$ (i.e., the weights for the different second moment component matrices, signal strength, noise variance) by integrating over the actual activity profiles:

$$p(\boldsymbol{Y}|\boldsymbol{\theta}) = \int p(\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{\theta}) p(\boldsymbol{U}|\boldsymbol{\theta}) d\boldsymbol{U}$$

Hence, the estimation of the individual activity profiles is not necessary and the second moment matrix $\boldsymbol{G}$ (i.e., in our study a 21 x 21 matrix) is a sufficient statistic to compute the marginal likelihoods (Diedrichsen et al., 2018).

Consistent with our LME analysis, we generated second moment component matrices ('pattern components') as predictors for PCM based on the activations of 360 simulated neurons from the spatial, respectively decisional models. We defined, estimated and compared the following PCM models for fMRI and EEG data:
*Null PCM* assumes that all conditions are independent, i.e., there is no similarity between any of the activity patterns of our 21 conditions (i.e., the 2nd moment matrix is the identity matrix).
*Spatial PCM* (S) assumes that the activity patterns are generated by the spatial model without recalibration.
*Decisional PCM* (D) assumes that the activity patterns are generated by the decisional model without recalibration.
*Combined (spatial + decisional) PCMs* assume that the activity patterns are a weighted linear combination of the patterns generated by the spatial and the decisional model. We factorially manipulated whether the spatial and/or decisional model accommodate audiovisual recalibration:

- (S+D) spatial component without recalibration and decisional component without recalibration
- ($S_R$+D) spatial component with recalibration and decisional component without recalibration
- (S+$D_R$) spatial component without recalibration and decisional component with recalibration
- ($S_R$+$D_R$) spatial component with recalibration and decisional component with recalibration

*Free (i.e., fully flexible) PCM* imposes no constraints on the second moment matrix and can therefore accommodate any pattern similarity structure. This model provides an upper benchmark for comparison with the hypothesis-driven constrained models. If a model performs at least as good as the cross-validated free PCM, it is sufficiently complex and captures all consistent variation in the data (Diedrichsen et al., 2018).
*fMRI and EEG fusion PCMs* (only for EEG PCM analysis)*:* We integrated fMRI and EEG analyses by fusing them via PCM. To assess directly whether the regional fMRI activity patterns evolve with different time courses we computed five PCMs that included pattern components based on fMRI activation patterns from one particular ROI. Each PCM included only one pattern component generated from the BOLD-response patterns of one of the five ROIs (combined over both hemispheres): HG, hA, IPS, IPL and FEF.
*Model estimation and comparison.* In our fMRI analysis these PCM models were applied to the pre-whitened parameter estimates from the first level GLM analysis separately for each of the 5 fMRI ROIs (i.e., HG, hA, IPS, IPL, FEF pooled over both hemispheres). In our EEG analysis they were applied to pre-whitened evoked EEG potentials averaged within individual runs (20 trials) and within each of the 4 EEG time windows (i.e., 50 – 150 ms, 150 – 250 ms, 250 – 350 ms, 350 – 450 ms).
We estimated the parameters of the PCM models in a leave-one-subject-out cross-validation scheme at the group level (Diedrichsen et al., 2018). To account for potential correlations of parameter estimates from the same imaging run (n.b. they were all computed relative to a common baseline) we modelled the run mean

as a fixed effect in fMRI when comparing the spatial, decisional, and spatial + decisional models without recalibration. However, modelling the run mean as a fixed effect captures the recalibration effects that arise across runs. We have therefore estimated the PCM with run effects as random when comparing the four spatial + decisional models that factorially manipulate whether the spatial and/or decisional model incorporate recalibration. In EEG, we modelled the run effect as a random effect across all model comparisons, because here the EEG activity patterns were not estimated relative to one common baseline. The marginal likelihood for each model and subject from the leave-one-subject-out cross-validation scheme was used as an approximation to the model evidence. We compared the models using the natural logarithm of the Bayes factors averaged across participants, which is invariant to the number of participants, hence it provides a more stringent criterion for model selection than the ln-Bayes factors summed across participants (Diedrichsen et al., 2018). Ln-Bayes factors of > 1.1 are considered substantial evidence and > 3 as strong evidence for one model over another one (Jeffreys, 1961).

## Acknowledgements

This research was funded by the European Research Council (https://erc.europa.eu/; grant number: ERC-2012-StG_20111109 multsens).

## Author Contributions

Conceptualization: M.A., A.M., and U.N.; Methodology: M.A. and A.M.; Software: M.A. and A.M.; Formal Analysis: M.A., A.M., and U.N.; Investigation: M.A. and A.M.; Resources: U.N.; Data Curation: M.A. and A.M.; Writing - Original Draft: M.A., A.M., and U.N.; Writing - Review & Editing: M.A., A.M., and U.N.; Visualization: M.A. and A.M.; Supervision: U.N.; Funding Acquisition: U.N.

## Declaration of Interests

The authors have declared that no competing interests exist.

## Supplemental Information

**Supplemental Results**

**Supplemental Methods**

**Table S1. Behavioural results: hit rate, d' and PSE in the sound localization and adaptation tasks.**

**Table S2. Psychometric functions for psychophysics, fMRI and EEG experiments: Bayesian model comparison of the 'static model' and the 'adaptation model'**

**Table S3. Neurometric functions for fMRI: Bayesian model comparison of the 'static model' and the 'adaptation model'**

**Table S4. fMRI regional BOLD response: Spatial and decisional Linear mixed effects models**

**Table S5. fMRI activation pattern: Pattern component modelling**

**Table S6. EEG PCM analysis ln-Bayes factors**

**Figure S1. Comparison of spatial models based on hemifield and place codes.**

## References


Akaike, H., 1998. A New Look at the Statistical Model Identification, in: Parzen, E., Tanabe, K., Kitagawa, G. (Eds.), Selected Papers of Hirotugu Akaike, Springer Series in Statistics. Springer New York, New York, NY, pp. 215–222. https://doi.org/10.1007/978-1-4612-1694-0_16

Aller, M., Noppeney, U., 2019. To integrate or not to integrate: Temporal dynamics of hierarchical Bayesian causal inference. PLOS Biol. 17, e3000210. https://doi.org/10.1371/journal.pbio.3000210

Bertelson, P., Frissen, I., Vroomen, J., de Gelder, B., 2006. The aftereffects of ventriloquism: patterns of spatial generalization. Percept. Psychophys. 68, 428–436. https://doi.org/10.3758/BF03193687

Bosen, A.K., Fleming, J.T., Allen, P.D., O'Neill, W.E., Paige, G.D., 2018. Multiple time scales of the ventriloquism aftereffect. PLOS ONE 13, e0200930. https://doi.org/10.1371/journal.pone.0200930

Brainard, D.H., 1997. The Psychophysics Toolbox. Spat. Vis. 10, 433–436.

Bruns, P., Liebnau, R., Röder, B., 2011. Cross-modal training induces changes in spatial representations early in the auditory processing pathway. Psychol. Sci. 22, 1120–6. https://doi.org/10.1177/0956797611416254

Bruns, P., Röder, B., 2015. Sensory recalibration integrates information from the immediate and the cumulative past. Sci. Rep. 5, 12739. https://doi.org/10.1038/srep12739

Chang, C.-C., Lin, C.-J., 2011. LIBSVM : a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, 1–27. https://doi.org/10.1145/1961189.1961199

Chen, L., Vroomen, J., 2013. Intersensory binding across space and time: a tutorial review. Atten. Percept. Psychophys. 75, 790–811. https://doi.org/10.3758/s13414-013-0475-4

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

Diedrichsen, J., Yokoi, A., Arbuckle, S.A., 2018. Pattern component modeling: A flexible approach for understanding the representational structure of brain activity patterns. NeuroImage, New advances in encoding and decoding of brain signals 180, 119–133. https://doi.org/10.1016/j.neuroimage.2017.08.051

Efron, B., Tibshirani, R.J., 1993. An Introduction to the bootstrap. Springer US, Boston, MA. https://doi.org/10.1007/978-1-4899-4541-9

Frissen, I., Vroomen, J., de Gelder, B., Bertelson, P., 2005. The aftereffects of ventriloquism: generalization across sound-frequencies. Acta Psychol. (Amst.) 118, 93–100. https://doi.org/10.1016/j.actpsy.2004.10.004

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-P., Frith, C.D., Frackowiak, R.S.J., 1994. Statistical parametric maps in functional imaging: A general linear approach. Hum. Brain Mapp. 2, 189–210. https://doi.org/10.1002/hbm.460020402

Gardner, W.G., Martin, K.D., 1995. HRTF measurements of a KEMAR. J. Acoust. Soc. Am. 97, 3907–3908. https://doi.org/10.1121/1.412407

Grinband, J., Hirsch, J., Ferrera, V.P., 2006. A Neural Representation of Categorization Uncertainty in the Human Brain. Neuron 49, 757–763. https://doi.org/10.1016/j.neuron.2006.01.032

Grootswagers, T., Wardle, S.G., Carlson, T.A., 2017. Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. J. Cogn. Neurosci. 29, 677–697. https://doi.org/10.1162/jocn_a_01068

Grothe, B., Pecka, M., McAlpine, D., 2010. Mechanisms of sound localization in mammals. Physiol. Rev. 90, 983–1012. https://doi.org/10.1152/physrev.00026.2009

Jeffreys, H., 1961. Theory of probability, 3rd ed. Oxford University Press, Oxford, UK.

Kleiner, M., Brainard, D., Pelli, D., 2007. What's new in Psychtoolbox-3?, in: Perception, 36 (EVCP Abstract Supplement). Alezzo.

Kopčo, N., Lin, I.-F., Shinn-Cunningham, B.G., Groh, J.M., 2009. Reference frame of the ventriloquism aftereffect. J. Neurosci. 29, 13809–13814. https://doi.org/10.1523/JNEUROSCI.2783-09.2009

Ledoit, O., Wolf, M., 2004. A well-conditioned estimator for large-dimensional covariance matrices. J. Multivar. Anal. 88, 365–411. https://doi.org/10.1016/S0047-259X(03)00096-4

Maier, J.X., Groh, J.M., 2009. Multisensory guidance of orienting behavior. Hear. Res. 258, 106–12. https://doi.org/10.1016/j.heares.2009.05.008

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

McAlpine, D., Jiang, D., Palmer, A.R., 2001. A neural code for low-frequency sound localization in mammals. Nat. Neurosci. 4, 396–401. https://doi.org/10.1038/86049

Mendonça, C., Escher, A., van de Par, S., Colonius, H., 2015. Predicting auditory space calibration from recent multisensory experience. Exp. Brain Res. 233, 1983–1991. https://doi.org/10.1007/s00221-015-4259-z

Michalka, S.W., Rosen, M.L., Kong, L., Shinn-Cunningham, B.G., Somers, D.C., 2016. Auditory Spatial Coding Flexibly Recruits Anterior, but Not Posterior, Visuotopic Parietal Cortex. Cereb. Cortex 26, 1302–1308. https://doi.org/10.1093/cercor/bhv303

Middlebrooks, J.C., Clock, A.E., Xu, L., Green, D.M., 1994. A panoramic code for sound location by cortical neurons. Science 264, 842–844. https://doi.org/10.1126/science.8171339

Mihalik, A., Noppeney, U., 2020. Causal inference in audiovisual perception. J. Neurosci. https://doi.org/10.1523/JNEUROSCI.0051-20.2020

Mullette-Gillman, O.A., Cohen, Y.E., Groh, J.M., 2005. Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus. J. Neurophysiol. 94, 2331–2352. https://doi.org/10.1152/jn.00021.2005

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., Kriegeskorte, N., 2014. A toolbox for representational similarity analysis. PLOS Comput. Biol. 10, e1003553. https://doi.org/10.1371/journal.pcbi.1003553

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2011. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011, 156869. https://doi.org/10.1155/2011/156869

Ortiz-Rios, M., Azevedo, F.A.C., Kuśmierek, P., Balla, D.Z., Munk, M.H., Keliris, G.A., Logothetis, N.K., Rauschecker, J.P., 2017. Widespread and Opponent fMRI Signals Represent Sound Location in Macaque Auditory Cortex. Neuron 93, 971–983. https://doi.org/10.1016/j.neuron.2017.01.013

Park, H., Kayser, C., 2019. Shared neural underpinnings of multisensory integration and trial-by-trial perceptual recalibration in humans. eLife 8. https://doi.org/10.7554/eLife.47001

Petkov, C.I., Kayser, C., Augath, M., Logothetis, N.K., 2006. Functional Imaging Reveals Numerous Fields in the Monkey Auditory Cortex. PLOS Biol. 4, e215. https://doi.org/10.1371/journal.pbio.0040215

Prins, N., Kingdom, F.A.A., 2018. Applying the Model-Comparison Approach to Test Specific Research Hypotheses in Psychophysical Research Using the Palamedes Toolbox. Front. Psychol. 9. https://doi.org/10.3389/fpsyg.2018.01250

Radeau, M., Bertelson, P., 1974. The after-effects of ventriloquism. Q. J. Exp. Psychol. 26, 63–71. https://doi.org/10.1080/14640747408400388

Raposo, D., Kaufman, M.T., Churchland, A.K., 2014. A category-free neural population supports evolving demands during decision-making. Nat. Neurosci. 17, 1784–1792. https://doi.org/10.1038/nn.3865

Rauschecker, J.P., Tian, B., 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc. Natl. Acad. Sci. U. S. A. 97, 11800–11806. https://doi.org/10.1073/pnas.97.22.11800

Recanzone, G.H., 1998. Rapidly induced auditory plasticity: the ventriloquism aftereffect. Proc. Natl. Acad. Sci. U. S. A. 95, 869–875. https://doi.org/10.1073/pnas.95.3.869

Rigoux, L., Stephan, K.E., Friston, K.J., Daunizeau, J., 2014. Bayesian model selection for group studies - Revisited. NeuroImage 84, 971–985. https://doi.org/10.1016/j.neuroimage.2013.08.065

Rohe, T., Noppeney, U., 2018. Reliability-Weighted Integration of Audiovisual Signals Can Be Modulated by Top-down Attention. eNeuro 5, ENEURO.0315-17.2018. https://doi.org/10.1523/ENEURO.0315-17.2018

Rohe, T., Noppeney, U., 2016. Distinct computational principles govern multisensory integration in primary sensory and association cortices. Curr. Biol. 1, 1–6. https://doi.org/10.1016/j.cub.2015.12.056

Rohe, T., Noppeney, U., 2015. Cortical hierarchies perform Bayesian causal inference in multisensory perception. PLOS Biol. 13, e1002073. https://doi.org/10.1371/journal.pbio.1002073

Salminen, N.H., May, P.J.C., Alku, P., Tiitinen, H., 2009. A population rate code of auditory space in the human cortex. PLOS ONE 4, e7600. https://doi.org/10.1371/journal.pone.0007600

Schlack, A., Sterbing-D'Angelo, S.J., Hartung, K., Hoffmann, K.-P., Bremmer, F., 2005. Multisensory Space Representations in the Macaque Ventral Intraparietal Area. J. Neurosci. 25, 4616–4625. https://doi.org/10.1523/JNEUROSCI.0455-05.2005

Smith, P.L., Little, D.R., 2018. Small is beautiful: In defense of the small-N design. Psychon. Bull. Rev. 1–19. https://doi.org/10.3758/s13423-018-1451-8

Stecker, G.C., Harrington, I.A., Middlebrooks, J.C., 2005. Location coding by opponent neural populations in the auditory cortex. PLoS Biol. 3, 0520–0528. https://doi.org/10.1371/journal.pbio.0030078

Tian, B., Reser, D., Durham, A., Kustov, A., Rauschecker, J.P., 2001. Functional specialization in rhesus monkey auditory cortex. Science 292, 290–293. https://doi.org/10.1126/science.1058911

van der Heijden, K., Rauschecker, J.P., de Gelder, B., Formisano, E., 2019. Cortical mechanisms of spatial hearing. Nat. Rev. Neurosci. 20, 609–623. https://doi.org/10.1038/s41583-019-0206-5

Wandell, B.A., Dumoulin, S.O., Brewer, A.A., 2007. Visual field maps in human cortex. Neuron 56, 366–383. https://doi.org/10.1016/j.neuron.2007.10.012

Wang, L., Mruczek, R.E.B., Arcaro, M.J., Kastner, S., 2015. Probabilistic Maps of Visual Topography in Human Cortex. Cereb. Cortex N. Y. N 1991 25, 3911–3931. https://doi.org/10.1093/cercor/bhu277

Watson, D.M., Akeroyd, M.A., Roach, N.W., Webb, B.S., 2019. Distinct mechanisms govern recalibration to audio-visual discrepancies in remote and recent history. Sci. Rep. 9, 8513. https://doi.org/10.1038/s41598-019-44984-9

Werner, S., Noppeney, U., 2010. Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. J. Neurosci. 30, 2662–2675. https://doi.org/10.1523/JNEUROSCI.5091-09.2010

Werner-Reiss, U., Groh, J.M., 2008. A rate code for sound azimuth in monkey auditory cortex: implications for human neuroimaging studies. J. Neurosci. 28, 3747–3758. https://doi.org/10.1523/JNEUROSCI.5044-07.2008

Winkler, I., Denham, S., Escera, C., 2013. Auditory Event-related Potentials, in: Jaeger, D., Jung, R. (Eds.), Encyclopedia of Computational Neuroscience. Springer, New York, NY, pp. 1–29. https://doi.org/10.1007/978-1-4614-7320-6_99-1

Wood, K.C., Town, S.M., Bizley, J.K., 2019. Neurons in primary auditory cortex represent sound source location in a cue-invariant manner. Nat. Commun. 10, 1–15. https://doi.org/10.1038/s41467-019-10868-9

Woods, T.M., Recanzone, G.H., 2004. Visually induced plasticity of auditory spatial perception in macaques. Curr. Biol. 14, 1559–64. https://doi.org/10.1016/j.cub.2004.08.059

Wozny, D.R., Shams, L., 2011. Recalibration of auditory space following milliseconds of cross-modal discrepancy. J. Neurosci. 31, 4607–12. https://doi.org/10.1523/JNEUROSCI.6079-10.2011

Zaidel, A., Ma, W.J., Angelaki, D.E., 2013. Supervised calibration relies on the multisensory percept. Neuron 80, 1544–1557. https://doi.org/10.1016/j.neuron.2013.09.026

Zatorre, R.J., Bouffard, M., Ahad, P., Belin, P., 2002. Where is "where" in the human auditory cortex? Nat. Neurosci. 5, 905–9. https://doi.org/10.1038/nn904

Zierul, B., Röder, B., Tempelmann, C., Bruns, P., Noesselt, T., 2017. The role of auditory cortex in the spatial ventriloquism aftereffect. NeuroImage 162, 257–268. https://doi.org/10.1016/j.neuroimage.2017.09.002

Zimmer, U., Macaluso, E., 2005. High Binaural Coherence Determines Successful Sound Localization and Increased Activity in Posterior Auditory Areas. Neuron 47, 893–905. https://doi.org/10.1016/j.neuron.2005.07.019

Zwiers, M.P., Van Opstal, A.J., Paige, G.D., 2003. Plasticity in human sound localization induced by compressed spatial vision. Nat. Neurosci. 6, 175–181. https://doi.org/10.1038/nn999