JNeurosci
THE JOURNAL OF NEUROSCIENCE

**Research Articles | Behavioral/Cognitive**

# Late cross-modal biases in neural spatial representations revealed by EEG decoding

https://doi.org/10.1523/JNEUROSCI.0242-26.2026

Received: 6 February 2026
Revised: 20 August 2026
Accepted: 1 September 2026

# Late cross-modal biases in neural spatial representations revealed by EEG decoding

Zak Buhmann[1], Amanda K Robinson[2], Jason B Mattingley[2,3] & Reuben Rideaux[1]

[1]School of Psychology, The University of Sydney, Camperdown, Australia

[2]School of Psychology, The University of Queensland, St Lucia, Australia

[3]Queensland Brain Institute, The University of Queensland, St Lucia, Australia

Corresponding author: reuben.rideaux@sydney.edu.au

Acknowledgements: RR was supported by a National Health and Medical Research Council (NHMRC; Australia) Investigator Grant (2026318). JBM was supported by an NHMRC Investigator Grant (2010141).

## ABSTRACT

During audiovisual perception, spatial information from vision and audition is combined, often producing biases such as the ventriloquist effect. While these interactions are well documented behaviourally, it remains unclear when cross-modal information begins to alter modality-specific spatial representations in the brain. Here we used cross-generalised inverted encoding modelling of electroencephalography (EEG) data to track the temporal evolution of spatial representations during a spatial ventriloquist task. Human participants (both sexes) localised audiovisual stimuli with horizontally offset auditory and visual components. Decoders trained on unisensory EEG responses were applied to audiovisual trials to estimate whether and when spatial representations of one modality were biased toward the other. Behaviourally, participants integrated cues but showed slight visual over-weighting. Neural decoding revealed robust spatial representations for both modalities, with early unisensory encoding remaining unaffected by cross-modal input. Cross-modal biases emerged only later (from ~200 ms onwards), within a generalised representational window, and occurred sequentially: auditory representations were biased earlier than visual. These findings indicate that multisensory spatial integration arises via recurrent feedback during later processing stages, while early sensory-specific representations remain independent, providing a neural basis for the temporal dynamics underlying the ventriloquist effect.

## SIGNIFICANCE STATEMENT

Understanding how the brain integrates spatial information across the senses is central to theories of perception, yet the timing of cross-modal influences on modality-specific neural representations remains unclear. Using EEG and cross-generalised inverted encoding models, we tracked auditory and visual spatial representations with millisecond precision. Early spatial representations remained unisensory despite synchronous audiovisual stimulation. A shared, generalised spatial representation emerged at ~200 ms, after which cross-modal biases appeared-first in auditory and later in visual codes. These results show that spatial integration is implemented through late, feedback-driven processes operating on initially independent sensory estimates, resolving a long-standing discrepancy between early multisensory interactions and late spatial ventriloquism.

## INTRODUCTION

Perception depends on transforming physical signals from the world into neural representations. Because sensory signals often share a common external source, the brain integrates information across modalities, resolving ambiguities and enhancing perceptual precision. A canonical example is the spatial ventriloquist effect, in which perceived sound location is biased toward visual input (Bertelson & Radeau, 1981).

A central challenge in multisensory perception is determining whether auditory and visual cues arise from a common source and should be integrated, or from separate sources and processed independently (Rideaux et al., 2021; Rideaux & Welchman, 2018). The brain is thought to resolve this via Bayesian causal inference, using spatial location to judge whether cues share a cause (Rohe & Noppeney, 2015); this decision depends on spatial and temporal correspondence and prior expectations (Van Wanrooij et al., 2010). When disparities are small, signals undergo forced fusion and are integrated automatically, each modality weighted by its relative reliability (Alais & Burr, 2004). This raises the question of where and when spatial biases emerge in neural representations.

Neuroimaging has begun to identify cortical loci of audiovisual causal inference. Mihalik and Noppeney (2020) applied multivariate decoding to fMRI data while participants judged whether synchronous auditory and visual signals at matched or mismatched locations shared a source. These decisions were primarily decodable from dorsolateral prefrontal cortex at the top of the hierarchy, where signals are weighted by bottom-up reliabilities and top-down task relevance. Physical spatial congruency, meanwhile, was decoded from frontoparietal cortices, primarily the frontal eye field and intraparietal sulcus, where integrated spatial representations emerge from bottom-up signals and top-down influences, downstream of early sensory processing (Aller & Noppeney, 2019; Ferrari & Noppeney, 2021). Together, these findings suggest the intraparietal sulcus acts as a hub for audiovisual spatial integration, mediating exchange between sensory cortices and conveying feedback from higher-order regions (Scheliga et al., 2023). This raises a key question: are spatial biases implemented only at

higher association stages following causal inference, or do they directly alter modality-specific representations?

It seems reasonable to expect integration to follow causal inference, preventing fusion of signals from different sources, consistent with traditional hierarchical models, in which low-level areas encode each modality separately before integration at higher association areas (Calvert, 2001; Felleman & Van Essen, 1991; Mesulam, 1998). More recent perspectives, however, suggest audiovisual processing is distributed across cortical levels, with dynamic interactions across timescales rather than at one stage (Scheliga et al., 2023; Senkowski & Engel, 2024). Early auditory and visual spatial information is handled separately by planum temporale and primary visual cortex (Bonath et al., 2007; Callan et al., 2015; Mihalik & Noppeney, 2020) , yet these primary areas still receive multimodal input despite a 'preferred sense' (Hsieh et al., 2012; Liang et al., 2013; Vetter et al., 2014). Audiovisual interactions often appear within 100 ms of onset (Molholm et al., 2002; Senkowski et al., 2011).

Such early interactions challenge the intuition that signals stay independent until a causal decision is made (Mihalik & Noppeney, 2020). Yet amplitude modulations in primary sensory cortices need not imply integration of spatial representations, and may instead reflect nonspecific temporal facilitation rather than genuine spatial fusion. Synchronous processing of a non-preferred stimulus (e.g. auditory input in primary visual cortex) may boost the preferred-stimulus response without shifting its spatial representation (Mihalik & Noppeney, 2020; Rohe & Noppeney, 2016). Thus, it remains unclear when cross-modal information begins to bias modality-specific representations.

We tested the hypothesis that cross-modal spatial biases emerge only after early unisensory encoding, implemented later by feedback from higher-order areas. Using cross-generalised inverted encoding modelling of EEG responses during a modified spatial ventriloquist paradigm, we tracked when audiovisual stimuli alter modality-specific spatial codes, training decoders on unisensory data and testing them on audiovisual responses. Early auditory and visual spatial representations remained unisensory, with perceptual biases

89 emerging later via sequential cross-modal influences — supporting models in which spatial
90 integration arises through recurrent feedback rather than early sensory fusion.

## METHOD

***Participants.*** Forty-eight human adults were recruited in return for payment (*M* = 24.48 years; min = 18 years; max = 54 years; 36 females; 46 right-handed). The study was approved by The University of Queensland Human Research Ethics Committee, and informed consent was obtained from all participants. Participants self-reported no neurological or psychiatric disorders, and had normal visual acuity (assessed using a standard Snellen eye chart).

***Materials and procedure.*** The experiment consisted of three conditions, in which the presented stimuli were either visual, auditory, or combined (audiovisual). The order in which conditions were presented was counterbalanced across participants. Before the main task, participants completed two practice rounds each of the auditory and visual conditions. During the practice, participants were directly shown the correct location after each behavioural response. This practice was omitted for the audiovisual condition, because providing specific location feedback might have primed participants to attend to one modality over the other. For the audiovisual task, participants were simply instructed to judge the location of each 'stimulus'. The modality was intentionally left ambiguous to avoid a systematic difference in attentional focus between auditory and visual stimuli for the audiovisual condition.The experiment was conducted in a dark, acoustically and electromagnetically shielded room. Stimuli were presented on a 24-inch ViewPixx monitor (VPixx Technologies Inc., Saint-Bruno, QC) with 1920x1080-pixel resolution and a refresh rate of 120 Hz. Viewing distance was maintained at 54 cm using a chinrest. Stimuli were generated in MATLAB v2021b (Inc., 2021) using the Psychophysics Toolbox (Brainard, 1997). Auditory stimuli were played through two loudspeakers (Q Acoustics, 3010i) placed either side of the display (25-75 W, 6 Ω). An EyeLink 1000 infrared eye tracker recording gaze direction at a sampling rate of 1000 Hz (SR Research Ltd., 2009).

***Stimuli.*** Visual stimuli were Gaussian blobs (0.2 contrast, 16º diameter) presented for 16ms on a mid-grey background. Auditory stimuli were 100 ms clicks with a flat 850 Hz tone embedded within a decay envelope (sample rate = 44, 100 Hz; volume = 60 dBA SPL, as

measured at the ears). Audiovisual stimuli were temporally synchronous combinations of the auditory and visual stimuli, but were spatially offset from each other by ±6º with no other changes to stimulus properties. As audiovisual cues are integrated according to their relative reliability (Alais & Burr, 2004), the visual stimuli were intentionally diffuse to equate the reliability of visual and auditory localisation. In doing so, we were able to investigate how visual cues bias auditory representations, and vice versa. To manipulate spatial location, stimuli were presented from 31 locations along the horizontal meridian of the display. Stimuli were centred on linearly spaced locations from 15º visual degrees to the left and right of the display centre. Auditory stimuli were played through two speakers placed equidistantly either side of the display (58 cm apart). The perceived source location of auditory stimuli was manipulated via changes to interaural intensity and timing (Whitworth & Jeffress, 1961; Wightman & Kistler, 1992). Specifically, auditory stimuli toward the edges of the display had marginally higher volume and were played slightly earlier from the nearest speaker, whereas those toward the centre had more uniform volume with less delay between speakers, as in our previous work (Buhmann et al., 2026).

***Experimental design*.** Participants viewed and/or listened to a random sequence of one to 20 stimuli from horizontal locations along the display. Stimulus presentations were separated by a jittered inter-stimulus interval of 500-700 ms. At the end of each sequence, participants were tasked with localising the horizontal position of the last presented stimulus by clicking on its estimated location with the mouse cursor. Having participants respond only to the last stimulus provided an increased number of stimulus presentations available for EEG analyses while still recording behavioural responses within the limited time frame of the experiment. To minimize eye movements, participants were asked to fixate on a black dot presented 8º above the display centre (see **Fig. 1** for an example trial). Stimulus modality was either auditory, visual, or audiovisual, and trials were blocked with short (~2 min) breaks between conditions. Each condition consisted of 12 blocks of one modality, and each block consisted of 10 trials with a pseudorandomised number of 1-20 stimuli presentations, such that each condition

yielded a total of 1250 stimulus presentations. Each block was followed by a feedback display indicating the average accuracy of participants' responses. In the audiovisual condition, accuracy was calculated relative to the visual cue.

***EEG data pre-processing.*** EEG data were recorded using a 64-channel BioSemi system at a sampling rate of 1024 Hz, which was down-sampled to 512 Hz during preprocessing. Signals were recorded with reference to the common mode sense/driven mode sense-reference level correction, with bipolar electrodes placed above and below the eye, at the temples, and at each mastoid to monitor for eye-movements and muscle artifacts. EEG preprocessing was undertaken in MATLAB using custom scripts and the EEGLAB toolbox (Delorme & Makeig, 2004). Data were high- and low-pass filtered between 0.1 Hz and 45 Hz to remove baseline drifts and high frequency noise, and were baseline-corrected from the average amplitude over the first 100 ms before stimulus onset. Data were then re-referenced according to the average of all 64 channels. Analyses were stimulus locked, with ERP responses segmented into 1100 ms epochs from 100 ms before stimulus presentation to 1000 ms after. No other preprocessing steps or artifact rejection were applied, as these are more likely to reduce statistical power (Delorme, 2023).

***Neural decoding.*** To characterize the neural representations of stimuli, we used an inverted modelling approach to reconstruct the spatial location of audio and visual cues based on EEG recordings (Brouwer & Heeger, 2011; Harrison et al., 2023). Analyses were performed separately for visual, auditory, and audiovisual stimuli. We first created a forward model that characterised the patterns of activity in the EEG electrodes given the 31 locations of the presented stimuli. The forward model was used to obtain the inverse model that described the transformation from electrode activity to stimulus location. For the unisensory conditions, we used a 10-fold cross-validation approach where 90% of the data were used to obtain the inverse model on which the remaining 10% of the data were decoded. Cross-validation was repeated 10 times such that all the data were decoded. For the audiovisual condition, we trained the inverted model on the location of stimuli from the unisensory (auditory or visual) data before using it to decode the location of audiovisual stimuli. In doing so, we investigated whether the decoded audiovisual locations were biased by cross-modal influence (e.g.

whether a visual-only trained decoder would localise audiovisual stimuli towards the offset auditory stimulus). By training the decoder on auditory or visual stimuli alone, any bias in the decoded audiovisual location should indicate a change in the unisensory representation of stimulus location. To investigate the extent to which the neural representation of location from auditory and visual cues is shared, we performed a cross-generalisation decoding analysis between both unisensory conditions (i.e. trained the decoder on unisensory auditory and tested on unisensory visual, and vice-versa). This allowed us to identify when the representation of spatial location transitioned from a sensory-specific pattern into a generalised form, common to both visual and auditory processing streams.

For the purposes of these analyses, we assume that EEG electrode noise is isotropic across stimulus locations and additive with the signal, and as such should not systematically interact with the coding of stimulus locations. Electrodes used in the analysis were obtained from the analysis approach of Buhmann et al. (2026), which used similar stimuli and inverted decoding. Electrodes were selected to maximise the signal-to-noise ratio of location information, which was primarily represented in posterior electrodes for visual and audiovisual stimuli, and in central electrodes for auditory stimuli (see inset of **Fig. 3A**). The encoding model contained five hypothetical channels, with evenly distributed idealised location preferences between -15º to +15º viewing angles to the left and right of the display centre. Each channel consisted of a half-wave rectified sinusoid raised to the second power. The channels were arranged such that an idealised tuning curve of each location preference could be expressed as a weighted sum of the five channels. The observed activity for each presentation can be described by the following linear model:

$$B = WC + E$$

where *B* indicates the EEG recordings (*m* electrodes x *n* presentations), *W* is a weight matrix (*m* electrodes x five channels) that describes the transformation from EEG activity to stimulus location, *C* denotes the hypothesized channel activities (five channels x *n* presentations), and *E* indicates the residual errors. To compute the inverse model we estimated the weights that,

when applied to the data, would reconstruct the channel activities with the least error. Due to the correlation between neighbouring electrodes, we took noise covariance into account when computing the model to optimize it for EEG data (Harrison et al., 2023; Kok et al., 2017; Mostert et al., 2015; Rideaux et al., 2023). Specifically, **B** and **C** were demeaned such that their average over presentations equalled zero for each sensor and channel, respectively. The inverse model was then estimated using a subset of the data, selected through cross-fold validation (10 folds). The hypothetical responses of each of the six channels were calculated from the training data in each fold, resulting in the response row vector $\mathbf{c}_{train,i}$ of length $n_{\text{train}}$ presentations for each channel *i*. The weights on the sensors $\mathbf{w}_i$ were then obtained through least squares estimation for each channel:

$$\mathbf{w}_i = B_{train}\mathbf{c}_{train,i}^{\mathsf{T}}\left(\mathbf{c}_{\text{train,i}}\mathbf{c}_{train,i}^{\mathsf{T}}\right)^{-1}$$

where $\boldsymbol{B}_{train}$ indicates the (*m* sensors × $n_{\text{train}}$ presentations) training EEG data. Subsequently, the optimal spatial filter $\mathbf{v}_i$ to recover the activity of the *i*th channel was obtained as follows (Mostert et al., 2015):

$$\mathbf{v}_i = \frac{\tilde{\Sigma}_i^{-1}\mathbf{w}_i}{\mathbf{w}_i^{\mathsf{T}}\tilde{\Sigma}_i^{-1}\mathbf{w}_i}$$

where $\tilde{\Sigma}_i$is the regularized covariance matrix for channel *i*. Incorporating the noise covariance in the filter estimation leads to the suppression of noise that arises from correlations between sensors. The noise covariance was estimated as follows:

$$\hat{\underset{i}{\Sigma}} = \frac{1}{n_{\text{train}} - 1}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{\mathsf{T}}$$

$$\boldsymbol{\varepsilon}_i = \mathbf{B}_{train} - \mathbf{w}_i\mathbf{c}_{train,i}$$

where $n_{\text{train}}$ is the number of training presentations. For optimal noise suppression, we improved this estimation by means of regularization by shrinkage using the analytically determined optimal shrinkage parameter (Mostert et al., 2015), yielding the regularized

covariance matrix $\tilde{\Sigma}_i$. We then used the inverse model to reconstruct the stimulus location from the EEG recordings. The residual error term *E*, and its associated noise covariance, was estimated separately within the training data of each analysis. For analyses trained on auditory-only or visual-only data, the error term was therefore estimated independently for each model and was not pooled across modalities.

To assess how well the forward model captured location information in the neural signal per modality, decoding accuracy was represented in arbitrary units as the similarity of the decoded location to the presented location. To calculate decoding accuracy, a channel response curve was created for each trial and time point, by calculating the dot product between the channel responses and the forward model. The channel response curve was then normalized to between ±1. The predicted location was given by the peak of the curve, while the decoding accuracy was indicated by the value (between ±1) at the location of the true stimulus location. In the audiovisual condition, accuracy was calculated separately relative to the auditory and visual locations. To assess if the decoder showed a bias towards the offset stimuli in the audiovisual condition, we calculated the average difference between the predicted and actual stimulus locations (separately for auditory and visual cues). ***Statistical analyses.*** Statistical analyses were performed in MATLAB v2025a. For the behavioural analysis, a one-sample Kolmogorov-Smirnov test for each condition revealed all conditions violated assumptions of normality. A non-parametric two-sided Wilcoxon signed-rank test was therefore used to test for significant differences in behavioural precision between all conditions. For the neural data, as stimuli were offset by ±6º visual degrees, those that appeared at the edges of the display (i.e. positions -15º to -10º and +10º to +15º) did not have a matched number of left or right offsets. These positions were removed from EEG decoding analyses for the audiovisual condition, averaging 216 trials out of 1250. To reduce the potential influence of decision- and motor-related activity, we also removed the final presentation from of each trial, where participants responded to the stimulus location, totalling 120 trials from each condition. Any participants with average decoding performance below 1.5 standard

deviations from the group average over the epoch were excluded, though no participants exceeded this threshold.

For decoding performance, a one-sample *t*-test was used to compare decoding accuracy to chance (i.e., zero). Next, the summed value of computed *t* statistics associated with each comparison (separately for positive and negative values) was calculated within contiguous temporal clusters of significant values. We then simulated the null distribution of the maximum summed cluster values using permutations ($n = 20{,}000$) over the sign of the data, from which we derived the 95% percentile threshold value. Clusters identified in the data with a summed effect-size value less than the threshold were considered spurious and removed. To compare the onset of significant decoding accuracy or bias between auditory and visual conditions, we used a leave-two-out jack-knifing approach (1128 unique permutations for 48 participants) to create subsampled distributions (Robinson et al., 2025). Significant clusters identified in the subsampled data were compared to the 95% percentile threshold value computed from the respective original sample. This yielded a distribution of onset times for each condition. Paired-sample *t*-tests were used to test for significant differences.

***Data availability.*** The behavioural and EEG data, and the scripts used for analysis and figure creation, are available at https://doi.org/10.17605/OSF.IO/EGH76

## RESULTS

***Behavioural performance.*** Participants performed well in localising stimulus location across all conditions. **Figure 2A** illustrates response distributions for a representative participant across conditions. To test for the presence of a systematic bias in participants' responses in the audiovisual condition, we compared the difference between the location of the visual stimulus and the participant's response (**Fig. 2B**). Bias values were sign-flipped for leftward auditory offsets so that positive values consistently indicated bias toward the auditory stimulus. To test for successful integration of stimuli in the audiovisual condition, we calculated the bias predicted by maximum likelihood estimation using the unisensory auditory and visual results (Alais & Burr, 2004). The average bias trended towards being slightly smaller than the maximum likelihood estimation prediction ($z$=-1.96, $p$=.050), indicating participants localised the audiovisual stimulus slightly closer to the visual component relative to the auditory. We found that precision (standard deviation of localisation errors) differed across conditions (**Fig. 2C**); precision for auditory stimuli was significantly poorer than precision for both audiovisual ($z$=3.76, $p$<.001) and visual stimuli alone ($z$=2.11, $p$=.035), while precision for audiovisual stimuli was not significantly different than precision for visual stimuli ($z$=1.47, $p$=.142).

***Inverted encoding.*** We used inverted encoding to calculate the spatial decoding accuracy for auditory and visual stimuli in the unisensory conditions (**Fig. 3A**). Decoding emerged at 95 ms for visual stimuli and 145 ms for auditory stimuli, and was consistently above chance for both conditions for ~850 ms. **Figure 3B** shows decoding accuracy for the cross-trained analysis, where the model was trained on unisensory neural activity and tested on audiovisual activity. Spatial location was reliably decoded from 100 ms after stimulus onset for visual stimuli and 170 ms for auditory stimuli, closely matching the results from the unisensory decoding analyses. To test whether the spatial representations of unisensory components within audiovisual stimuli were systematically influenced by the cross-modal component, we trained a decoder on the unisensory visual responses and tested it on the unisensory auditory, and vice-versa. This cross-generalisation analysis can determine when the representation of

298 spatial location transitioned from a sensory-specific pattern to a generalised form not unique
299 to either modality (**Fig. 3C**). We found reliable cross-generalisation from ~205 to 690 ms
300 following stimulus onset.

301 ***Neural bias*.** Having established the temporal profile of spatial decoding, we next tested
302 whether these representations were systematically biased by cross-modal information. To
303 investigate whether the decoder was biased in localising stimuli towards the offset component,
304 bias was computed as decoder-predicted minus true stimulus location. Biases for left-offset
305 trials were sign-flipped so that positive values reflect bias toward the cross-modal cue (**Fig.**
306 **3D**). Auditory-trained decoding showed significant bias beginning at 195 ms and recurring
307 through 560 ms. Visual-trained decoding showed a single significant bias period (~375–455
308 ms). Notably, these biases occurred within the period of generalised spatial location identified
309 from the cross-decoding analysis in **Figure 3C**. Paired-sample *t*-tests on subsampled data
310 revealed that the onset of cross-generalised decoding emerged significantly earlier than both
311 the auditory bias ($t$=-21.29, $p$<.001) and the visual bias ($t$=-100.98, $p$<.001). The onset of the
312 auditory bias also arose significantly earlier than the visual bias ($t$=-55.28, $p$<.001).

313 Behavioural bias averaged ~3°, whereas neural bias estimates were ~0.25°. The
314 smaller neural bias is likely due to a low signal-to-noise ratio within the recorded neural activity,
315 in that the inverted encoding analyses are less sensitive than behavioural measures to
316 condition-specific information. The decoder's trial-to-trial predictions of stimulus location are
317 subject to considerable noise. As the bias is averaged across many trials, this results in an
318 attraction to the mean (zero). Thus, the decoded bias likely underestimates the neural bias.

319 The neural processes that localise events in space are closely tied to the saccadic eye
320 movements that allow rapid, accurate orienting toward novel stimuli (Sparks, 2002). To
321 examine the potential influence of eye-movements toward presented stimuli, we re-analysed
322 the data using the four electrooculography (EOG) electrodes, which were positioned above
323 and below the left eye and at the temples (**Fig. S1**). After training on audiovisual trials, the

decoder was able to discern auditory and visual stimulus locations significantly above chance. Decoding was also biased for both auditory and visual stimuli at approximately 400 ms after stimulus presentation, suggesting that participants saccades were influenced by both visual and auditory cue location. Critically, however, the early auditory bias observed at 200 ms in the main analyses was not present. To help disentangle neural activity relating to preparatory eye-movements from activity associated with stimulus location, we re-analysed the results using a more stringent preprocessing method with artifact rejection via independent component analysis (SASICA; Chaumon et al., 2015). We found the same pattern of results as in our original analyses (**Fig. S2**). Given this, and because the main analyses used parieto-occipital electrodes to minimise ocular muscle artifacts, we believe eye movements are unlikely to underlie the key decoding result.

## DISCUSSION

The current study aimed to examine how spatial representations of stimuli are influenced by cross-modal input throughout the sensory processing stream. We asked when cross-modal information biases modality-specific spatial representations and whether this provides a neural basis for the ventriloquist effect. Participants integrated audiovisual cues, localising stimuli near the midpoint between auditory and visual components. Critically, multivariate decoding analyses revealed a neural bias, in which the spatial representations of audiovisual components were biased towards the location of the cross-modal component. The bias occurred within a generalised representational window, suggesting hierarchical processing with distinct modality-specific and multimodal spatial representations. The bias also occurred sequentially rather than synchronously, with auditory representations being biased before visual representations. That the bias occurs sequentially despite appearing within a generalised period suggests a complex and hierarchical processing network for spatial information, perhaps with multiple representations of a stimulus.

Participants perceived the source of the event to be roughly midway between the offset auditory and visual stimuli, suggesting successful integration of the multimodal spatial information (Alais & Burr, 2004; Bertelson & Radeau, 1981; Bruns, 2019). However, the perceived location was not consistent with maximum-likelihood estimations and showed a marginal deviation towards the visual component. Response precision for audiovisual stimuli exceeded that for auditory but not visual, which may indicate more frequent attention to the visual modality. It is noteworthy, however, that this pattern of results is consistent with previous work suggesting the primacy of vision in human observers, such that visual information is over-weighted relative information in other modalities, despite optimal integration (Aller et al., 2015; Battaglia et al., 2003; Meijer et al., 2019). As such, the slight visual preference observed in participants' behavioural responses in the current study may reflect greater reliability or salience of the visual stimuli in audiovisual trials.

Decoding results for both the multisensory and unisensory conditions followed a similar overall profile, suggesting that much of the location information within the neural response to unisensory components is carried over into the multisensory response. Visual stimuli were decoded from ~100 ms post-stimulus onset, whereas auditory stimuli were decoded later, at ~140 ms. The later emergence of auditory decoding likely reflects the need for cortical computation of spatial cues, whereas visual space is encoded retinotopically in early cortex. Indeed, audiovisual stimuli are more likely to be judged as synchronous if the sound occurs *after* the visual cue (Vroomen & Keetels, 2010, 2020). Because auditory spatial cues (i.e., interaural time and level differences) must be computed cortically (Whitworth & Jeffress, 1961; Wightman & Kistler, 1992), early A1 responses may not contain spatially separable population patterns and therefore cannot be decoded.

The cross-trained unisensory decoding analyses revealed common activity between the auditory-only and visual-only conditions, which became significant at ~200 ms post-stimulus and peaked at ~400 ms. This reflects a generalised multisensory spatial representation, and is consistent with prior work showing that spatial audiovisual information peaks at ~200 ms (Bednar & Lalor, 2020; Buhmann et al., 2026; Rideaux, 2024). Within the ventriloquist effect, feedback for spatial localisation (i.e. physical cue congruence and causal decisions) has been detected after 200 ms post-stimulus onset (Bonath et al., 2007; Rohe & Noppeney, 2016). Early visual decoding (~100 ms) likely reflects retinotopic activity in V1, whereas the later peak (~200 ms) reflects distributed spatial representations influenced by feedback (Robinson et al., 2021). The auditory spatial information peaked at ~400 ms post-stimulus in the auditory-only condition, while in the audiovisual condition it plateaued after 200 ms post-stimulus. The plateau in the audiovisual condition may reflect a transition from modality-specific to generalised spatial coding, which dilutes sensory-specific patterns. Since the visual information had already peaked prior to the generalised representation, it remained relatively unaffected.

Our key analysis investigated whether a neural bias was present within the processing of audiovisual stimuli. Both auditory-trained and visual-trained decoders showed systematic shifts towards the cross-modal cue. As the bias emerged only after the onset of the generalised representation, early modality-specific spatial codes appear to remain unaffected by cross-modal input. Though evidence for audiovisual integration has been found much earlier (i.e. pre-150 ms post stimulus onset; Senkowski et al., 2011; Zhao et al., 2018), these early effects primarily reflect amplitude modulations rather than changes in feature-selective spatial codes (Aller et al., 2015; Mihalik & Noppeney, 2020; Talsma et al., 2010). The integration of stimulus feature information may occur relatively late in perceptual processing, with the localisation of an audiovisual stimulus being finalised during a stage in which the spatial information is common to both audition and vision.

A late bias in the neural representation of spatial information is consistent with theories of a hierarchical processing network, where the integration of spatial information occurs within higher-order areas influenced by feedback (Friston, 2005; Mihalik & Noppeney, 2020). Causal inference signals from prefrontal regions may feedback to parietal areas, where spatial estimates are integrated (Kayser & Shams, 2015; Mihalik & Noppeney, 2020). Visual and auditory representations remain separate in early sensory cortices before converging in higher-order regions such as the intraparietal cortex (Callan et al., 2015; Scheliga et al., 2023). Maintaining separate sensory representations before integration preserves accuracy and provides a mechanism to detect inconsistencies that would be lost if unisensory information were discarded prematurely. As the bias in the auditory spatial information also occurred significantly earlier than the bias in the visual input, the integration of auditory and visual information may be sequential rather than synchronised. One possibility is that the earlier auditory bias reflects integration occurring as soon as reliable spatial estimates become available. The maintenance of auditory information in working memory seems to involve both visual and auditory cortical pathways, whilst visual information is maintained exclusively within visual areas (Wolff et al., 2020). Though our results concern immediate perceptual processing

rather than working memory, they may reflect a similar pattern. The bias to auditory representations occurs first, which may be due to an early integration with the visual signal. The later visual bias, meanwhile, only occurs once spatial information has entered a general format, as visual information may be maintained in a sensory-specific network isolated from auditory processing areas. Future manipulations of audiovisual timing will be needed to test whether aligning peak representational times produces synchronous rather than sequential biases. We also did not measure which modality participants prioritised when localising audiovisual stimuli, assuming that individual preferences would not systematically bias group-level results. Inter-individual variability in the audiovisual condition may nonetheless reflect individual differences in strategy. Future work should therefore examine how attentional focus shapes the perception and integration of audiovisual stimuli (Talsma et al., 2010).

Localisation and gaze orientation are tightly coupled processes (Sparks, 2002), and saccadic eye movements are therefore difficult to suppress entirely in paradigms such as ours. Decoding analyses undertaken on data recorded from EOG electrodes indicated that on at least some trials participants probably made small saccades toward the auditory and visual stimuli, despite instructions to fixate. These saccades appeared to be slightly biased in the direction of the offset stimulus, supporting the finding that participants perceived audiovisual stimuli as being located between the veridical auditory and visual locations. Similar to results from parietooccipital electrodes, the decoding bias for both auditory and visual conditions was present at 400 ms post-stimulus onset and likely reflected similar processes. Here, the biased representations of stimulus location may influence gaze orientation behaviours, or vice-versa. Critically, however, the early sequential bias found for auditory stimuli was not present when decoding from EOG electrodes. Visual localisation is rapid, whereas auditory location must be computed from interaural time and level differences (Martin et al., 2015; Xiong et al., 2022). This earlier bias may reflect a window in which the neural representation of auditory location is already susceptible to visual input, before auditory information has been integrated in turn to influence the visual representation. The bias observed in the EOG decoding may therefore

441 reflect saccades toward the integrated percept alone, rather than toward sensory-specific
442 representations. Despite existing evidence for early multisensory interactions (Molholm
443 et al., 2002; Senkowski et al., 2011; Zhao et al., 2018), here we found spatial representations
444 remained unisensory until later stages. Together, these findings indicate that multisensory
445 spatial integration is implemented via recurrent feedback during a later processing stage,
446 consistent with Bayesian causal inference frameworks in which independent sensory
447 estimates are retained before being merged.

## REFERENCES


Alais, D., & Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, *14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

Aller, M., Giani, A., Conrad, V., Watanabe, M., & Noppeney, U. (2015). A spatially collocated sound thrusts a flash into awareness. *Frontiers in Integrative Neuroscience*, *9*. https://doi.org/10.3389/fnint.2015.00016

Aller, M., & Noppeney, U. (2019). To integrate or not to integrate: Temporal dynamics of hierarchical Bayesian causal inference. *PLOS Biology*, *17*(4), e3000210. https://doi.org/10.1371/journal.pbio.3000210

Bednar, A., & Lalor, E. C. (2020). Where is the cocktail party? Decoding locations of attended and unattended moving sound sources using EEG. *NeuroImage*, *205*, 116283. https://doi.org/10.1016/j.neuroimage.2019.116283

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, *29*(6), 578–584. https://doi.org/10.3758/BF03207374

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H.-J., & Hillyard, S. A. (2007). Neural Basis of the Ventriloquist Illusion. *Current Biology*, *17*(19), 1697–1703. https://doi.org/10.1016/j.cub.2007.08.050

Bruns, P. (2019). The Ventriloquist Illusion as a Tool to Study Multisensory Processing: An Update. *Frontiers in Integrative Neuroscience*, *13*. https://doi.org/10.3389/fnint.2019.00051

Buhmann, Z., Robinson, A. K., Mattingley, J. B., & Rideaux, R. (2026). Inverted encoding of neural responses to audiovisual stimuli reveals super-additive multisensory enhancement. *eLife*, *13*, RP97230. https://doi.org/10.7554/eLife.97230

473 Callan, A., Callan, D., & Ando, H. (2015). An fMRI Study of the Ventriloquism Effect.
474 *Cerebral Cortex*, *25*(11), 4248–4258. https://doi.org/10.1093/cercor/bhu306

475 Calvert, G. A. (2001). Crossmodal Processing in the Human Brain: Insights from Functional
476 Neuroimaging Studies. *Cerebral Cortex*, *11*(12), 1110–1123.
477 https://doi.org/10.1093/cercor/11.12.1110

478 Delorme, A. (2023). EEG is better left alone. *Scientific Reports*, *13*(1), 2372.
479 https://doi.org/10.1038/s41598-023-27528-0

480 Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the
481 primate cerebral cortex. *Cerebral Cortex (New York, N.Y.*, *1*(1), 1–47.
482 https://doi.org/10.1093/cercor/1.1.1-a

483 Ferrari, A., & Noppeney, U. (2021). Attention controls multisensory perception via two
484 distinct mechanisms at different levels of the cortical hierarchy. *PLOS Biology*,
485 *19*(11), e3001465. https://doi.org/10.1371/journal.pbio.3001465

486 Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal*
487 *Society B: Biological Sciences*, *360*(1456), 815–836.
488 https://doi.org/10.1098/rstb.2005.1622

489 Harrison, W. J., Bays, P. M., & Rideaux, R. (2023). Neural tuning instantiates prior
490 expectations in the human visual system. *Nature Communications*, *14*(1), 5320.
491 https://doi.org/10.1038/s41467-023-41027-w

492 Hsieh, P.-J., Colas, J. T., & Kanwisher, N. (2012). Spatial pattern of BOLD fMRI activation
493 reveals cross-modal information in auditory cortex. *Journal of Neurophysiology*,
494 *107*(12), 3428–3432. https://doi.org/10.1152/jn.01094.2010

495 Kayser, C., & Shams, L. (2015). Multisensory Causal Inference in the Brain. *PLOS Biology*,
496 *13*(2), e1002075. https://doi.org/10.1371/journal.pbio.1002075

Kok, P., Mostert, P., & de Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences*, *114*(39), 10473–10478. https://doi.org/10.1073/pnas.1705652114

Liang, M., Mouraux, A., Hu, L., & Iannetti, G. D. (2013). Primary sensory cortices contain distinguishable spatial patterns of activity for each sense. *Nature Communications*, *4*(1), 1979. https://doi.org/10.1038/ncomms2979

Mesulam, M. M. (1998). From sensation to cognition. *Brain*, *121*(6), 1013–1052. https://doi.org/10.1093/brain/121.6.1013

Mihalik, A., & Noppeney, U. (2020). Causal Inference in Audiovisual Perception. *The Journal of Neuroscience*, *40*(34), 6600–6612. https://doi.org/10.1523/JNEUROSCI.0051-20.2020

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Cognitive Brain Research, Multisensory Proceedings*, *14*(1), 115–128. https://doi.org/10.1016/S0926-6410(02)00066-6

Mostert, P., Kok, P., & De Lange, F. P. (2015). Dissociating sensory from decision processes in human perceptual decision making. *Scientific Reports*, *5*(1), 18253. https://doi.org/10.1038/srep18253

Rideaux, R. (2024). Task-related modulation of event-related potentials does not reflect changes to sensory representations. *Imaging Neuroscience*, *2*, 1–13. https://doi.org/10.1162/imag_a_00266

Rideaux, R., Storrs, K. R., Maiello, G., & Welchman, A. E. (2021). How multisensory neurons solve causal inference. *Proceedings of the National Academy of Sciences*, *118*(32). https://doi.org/10.1073/pnas.2106235118

Rideaux, R., & Welchman, A. E. (2018). Proscription supports robust perceptual integration by suppression in human visual cortex. *Nature Communications*, *9*(1), 1502. https://doi.org/10.1038/s41467-018-03400-y

Rideaux, R., West, R. K., Rangelov, D., & Mattingley, J. B. (2023). Distinct early and late neural mechanisms regulate feature-specific sensory adaptation in the human visual system. *Proceedings of the National Academy of Sciences*, *120*(6), e2216192120. https://doi.org/10.1073/pnas.2216192120

Robinson, A. K., Grootswagers, T., Shatek, S. M., Behrmann, M., & Carlson, T. A. (2025). Dynamics of visual object coding within and across the hemispheres: Objects in the periphery. *Science Advances*, *11*(1), eadq0889. https://doi.org/10.1126/sciadv.adq0889

Robinson, A. K., Grootswagers, T., Shatek, S. M., Gerboni, J., Holcombe, A., & Carlson, T. A. (2021). Overlapping neural representations for the position of visible and imagined objects. *Neurons, Behavior, Data Analysis, and Theory*, *4*(1). https://doi.org/10.51628/001c.19129

Rohe, T., & Noppeney, U. (2015). Sensory reliability shapes perceptual inference via two mechanisms. *Journal of Vision*, *15*(5), 22. https://doi.org/10.1167/15.5.22

Rohe, T., & Noppeney, U. (2016). Distinct Computational Principles Govern Multisensory Integration in Primary Sensory and Association Cortices. *Current Biology*, *26*(4), 509–514. https://doi.org/10.1016/j.cub.2015.12.056

Scheliga, S., Kellermann, T., Lampert, A., Rolke, R., Spehr, M., & Habel, U. (2023). Neural correlates of multisensory integration in the human brain: An ALE meta-analysis. *Reviews in the Neurosciences*, *34*(2), 223–245. https://doi.org/10.1515/revneuro-2022-0065

Senkowski, D., & Engel, A. K. (2024). Multi-timescale neural dynamics for multisensory integration. *Nature Reviews Neuroscience*, *25*(9), 625–642. https://doi.org/10.1038/s41583-024-00845-7

Senkowski, D., Saint-Amour, D., Höfle, M., & Foxe, J. J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *NeuroImage*, *56*(4), 2200–2208. https://doi.org/10.1016/j.neuroimage.2011.03.075

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400–410. https://doi.org/10.1016/j.tics.2010.06.008

Van Wanrooij, M. M., Bremen, P., & John Van Opstal, A. (2010). Acquired prior knowledge modulates audiovisual integration. *European Journal of Neuroscience*, *31*(10), 1763–1771. https://doi.org/10.1111/j.1460-9568.2010.07198.x

Vetter, P., Smith, F. W., & Muckli, L. (2014). Decoding Sound and Imagery Content in Early Visual Cortex. *Current Biology*, *24*(11), 1256–1262. https://doi.org/10.1016/j.cub.2014.04.020

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics*, *72*(4), 871–884. https://doi.org/10.3758/APP.72.4.871

Vroomen, J., & Keetels, M. (2020). Perception of causality and synchrony dissociate in the audiovisual bounce-inducing effect (ABE). *Cognition*, *204*, 104340. https://doi.org/10.1016/j.cognition.2020.104340

Whitworth, R. H., & Jeffress, L. A. (1961). Time vs Intensity in the Localization of Tones. *The Journal of the Acoustical Society of America*, *33*(7), 925–929. https://doi.org/10.1121/1.1908849

Wightman, F. L., & Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America*, *91*(3), 1648–1661. https://doi.org/10.1121/1.402445

Wolff, M. J., Kandemir, G., Stokes, M. G., & Akyürek, E. G. (2020). Unimodal and Bimodal Access to Sensory Working Memories by Auditory and Visual Impulses. *Journal of Neuroscience*, *40*(3), 671–681. https://doi.org/10.1523/JNEUROSCI.1194-19.2019

Zhao, S., Wang, Y., Xu, H., Feng, C., & Feng, W. (2018). Early cross-modal interactions underlie the audiovisual bounce-inducing effect. *NeuroImage*, *174*, 208–218. https://doi.org/10.1016/j.neuroimage.2018.03.036

## FIGURE CAPTIONS

**Figure 1. An example trial in the audiovisual condition.** On each trial, participants were presented with a random stream of 1-20 auditory and visual stimuli that were spatially localised at one of 31 possible locations (±15°) along the horizontal meridian of the display, while they fixated a dot 8° above the centre of the display. The visual and auditory components were always offset by ±6°. The top row displays some possible locations of stimuli. The participants' task was to localise the last presented stimulus as accurately as possible using the mouse cursor. **Note:** The example visual stimuli above have been made more obvious to better communicate the experimental procedure. In the actual experiment, the Gaussian profile was shallower, and each stimulus was more diffuse and harder to localise.

**Figure 2. Behavioural performance is significantly influenced by cross-modal offsets. A)** Trial-by-trial error between target stimulus location and behavioural response for all conditions aggregated across participants. Error in the multisensory condition is presented relative to the visual stimulus. **B)** Precision across participants, calculated as the standard deviation of errors for each stimulus condition, relative to the visual stimulus for the audiovisual condition. **C)** Average bias of behavioural responses across participants in identifying stimulus location for the audiovisual condition. The error bar indicates ±1 SEM. The red cross indicates estimated bias assuming optimal (MLE) integration of unisensory cues. **D)** Histogram of average bias of behavioural responses across participants for the audiovisual location.

**Figure 3. Neural representations of stimulus position are systematically biased by early representations.** Accuracy of decoded locations from neural responses to **A)** unisensory and **B)** multisensory stimuli. Shaded error bars indicate ±1 SEM. Coloured horizontal bars indicate cluster corrected periods that showed a significant difference from chance (zero). Right insets show topographic decoding performance across significant clusters. **C)** Cross-decoding between unisensory auditory and visual conditions to identify modality-independent spatial location information within the neural signal. **D)** Bias in the decoded location (predicted stimulus location minus actual stimulus location) for audiovisual stimuli. Positive values indicate a greater bias towards the location of the offset stimulus. Right inset shows topographic bias across significant clusters. Inset in **A** shows electrode selection for all analyses.