Audiovisual Integration Follows Different Rules for Perceptual and Metacognitive Decisions

Perrine Porte (perrine.porte@univ-grenoble-alpes.fr)[1]

Matan Mazor (matan.mazor@all-souls.ox.ac.uk)[2]

Louise Goupil (louise.goupil@univ-grenoble-alpes.fr)[1]

Nathan Faivre (nathan.faivre@univ-grenoble-alpes.fr)[1]

Université Grenoble Alpes, CNRS, LPNC[1]; University of Oxford[2]

Abstract

Being confident in our inferences about the presence or absence of multisensory stimuli is crucial in many contexts. We investigated how humans form both amodal and modality-specific confidence judgments following the detection of audiovisual stimuli. To model this, we extended a Bayesian evidence accumulation framework. The model accurately reproduced amodal detection and modality-specific confidence judgments, despite being fitted only to amodal decisions and decision times. However, it failed to capture amodal confidence. Overall, this suggests that different integration rules apply to perceptual and metacognitive decisions.

Keywords: confidence judgments; multisensory integration; Bayesian modelling; detection

Introduction

While multisensory influences are well studied at the perceptual level, they remain poorly understood at the metacognitive level (Deroy et al., 2016). Evidence accumulation models have explained reduced response times in the presence of multisensory signals (Gondan et al., 2010; Plass & Brang, 2021). To account for accuracy effects, some models have simulated decision-making processes using a drift diffusion model with separate decision thresholds (Blurton et al., 2014). Crucially, in these tasks, a stimulus was always present – a target or a distractor. As a result, no direct comparison could be made between decisions about the presence vs absence of sensory evidence: judgments that rely on partly different mechanisms.

A recent study (Mazor et al., 2025) successfully reproduced behavioral patterns in a visual detection task with a model integrating both factual and counterfactual evidence: evidence is accumulated only for presence (factual evidence), and absence is inferred from counterfactual detectability (i.e., "*I would have perceived it if it were present*"). Here, we extend this model to audiovisual detection to account for both decisions and confidence in the presence or absence of multiple sensory cues.

Model

In the original model by Mazor et al. (2025), agents observe the activation of a single "presence sensor" that samples either a 1 (activation) or a 0 (inactivation) at each time point. Activation probabilities are captured by model parameters $\theta_{present}$ (i.e., if a target is present) and θ_{absent} (i.e., if a target is absent). Importantly, this model assumes agents hold beliefs about the probability of sampling a 1, if the target is present, and if it is absent. These beliefs are captured by model parameters $\bar{\theta}_{present}$ and $\bar{\theta}_{absent}$ and reflect the degree to which agents believe that they would have perceived the target if it was present.

To extend this model, we introduced separate visual and auditory sensors and applied a disjunctive integration rule: $p(x \text{ or } y)=p(x)+p(y)-p(x \text{ and } y)$, reflecting that a stimulus can be present in only one modality. The model also included prior beliefs about the probability of the presence of a stimulus in each modality. The model was fitted only to amodal detection and decision time data. Still, it made predictions about amodal confidence based on the probability of being correct at the time of the decision. We also derived model-based predictions about modality-specific effects, by computing the probability of presence per modality at decision time.

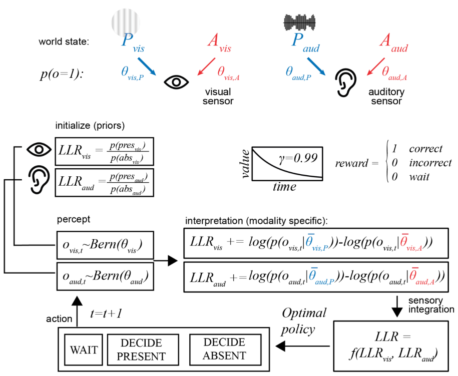

Figure 1: Model architecture. The agent has access to two sensors, probabilistically tuned to the presence of visual and auditory evidence. The agent observes their activations and updates their beliefs about the presence of a signal in each modality separately, using Bayes' rule. These beliefs are then combined into an amodal belief about the presence of a target. Based on this belief, they decide whether to commit to a decision or accumulate more evidence by following an optimal policy, derived using backward induction (Callaway et al., 2023).

Key Results

In a pre-registered experiment (https://osf.io/3nvyx), 48 participants performed an audiovisual detection task where a stimulus was present on 75% of the trials. On each trial, participants judged the presence or absence of a stimulus regardless of its modality before reporting their amodal confidence regarding their detection choice from 0 to 100. Finally, they reported their modality-specific detection and confidence judgments on a bi-dimensional (audio/visual) scale, with each axis ranging from 100% sure not perceived to 100% sure perceived and corresponding to one modality.

At the perceptual level, our model reproduced the amodal detection pattern with higher accuracy for bimodal (*true*: 73%, *simulated*: 72%) compared to unimodal trials, and for visual (*true*: 59%, *simulated*: 57%) compared to auditory (*true*: 42%, *simulated*: 41%) trials (Fig. 2A). It also reproduces the bias toward responding 'absent' (*true criterion*: 0.58, *simulated*: 0.57), and the higher accuracy in the visual modality (*true visual d'*: 1.8, *simulated*: 1.67) compared to the auditory one (*true auditory d'*: 1.16, *simulated*: 1.13).

Despite not being fitted to confidence ratings, the model captured the higher modality-specific metacognitive sensitivity (i.e. the ability to distinguish between correct and incorrect responses based on confidence judgments) for audiovisual than unimodal trials, for both auditory and visual modalities (Fig. 2C-D). However, the model predicted higher amodal confidence for presence than for absence, an effect not observed behaviorally (Fig. 1B).

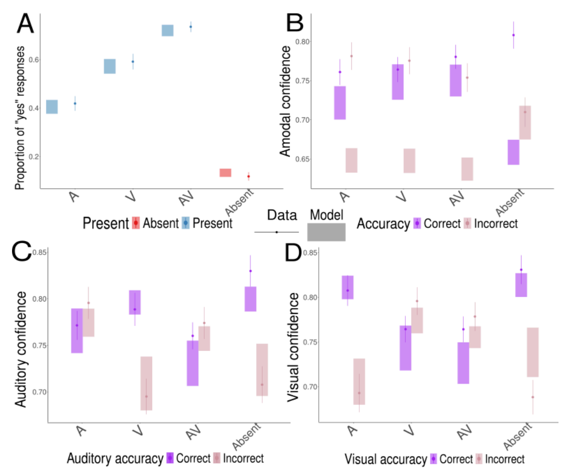

Figure 2: Results. Error bars represent the standard error from the data. Rectangles represent data simulated from the model, centered on the mean value and with height equal to the standard error. A) Percentage of "yes" responses B) Amodal confidence. C) Auditory confidence. D) Visual confidence.

Conclusion

To conclude, our results suggest that different integration rules apply to amodal decisions and confidence judgments during audiovisual detection. More work is needed to elucidate how amodal confidence is computed, be it post-decisional evidence accumulation, over-reliance on prior beliefs, or the adoption of computationally cheap heuristics to approximate Bayesian inference.

Acknowledgments

This work is supported by the ERC grant (Volta, 101125379) awarded to NF. MM is supported by a postdoctoral fellowship from All Souls College at the University of Oxford. The authors thank the IDEX for funding PP mobility grant.

References

Blurton, S. P., Greenlee, M. W., & Gondan, M. (2014). Multisensory processing of redundant information in go/no-go and choice responses. *Attention, Perception, & Psychophysics*, *76*(4), 1212-1233. https://doi.org/10.3758/s13414-014-0644-0

Callaway, F., Griffiths, T. L., Norman, K. A., & Zhang, Q. (2023). Optimal metacognitive control of memory recall. Psychological Review, No Pagination Specified–No Pagination Specified. http://doi.org/10.1037/rev0000441

Deroy, O., Spence, C., & Noppeney, U. (2016). Metacognition in Multisensory Perception. *Trends in Cognitive Sciences*, *20*(10), 736-747. https://doi.org/10.1016/j.tics.2016.08.006

Gondan, M., Götze, C., & Greenlee, M. W. (2010). Redundancy gains in simple responses and go/no-go tasks. *Attention, Perception, & Psychophysics*, *72*(6), 1692-1709. https://doi.org/10.3758/APP.72.6.1692

Mazor, M., Friston, K. J., & Fleming, S. M. (2020). Distinct neural contributions to metacognition for detecting, but not discriminating visual stimuli. *eLife*, *9*, e53900. https://doi.org/10.7554/eLife.53900

Mazor, M., Moran, R., & Press, C. (2025). Beliefs about perception shape perceptual inference : An ideal observer model of detection. *Psychological Review*. https://doi.org/10.1037/rev0000552

Plass, J., & Brang, D. (2021). Multisensory stimuli shift perceptual priors to facilitate rapid behavior. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-02566-8