

Vision can contribute to auditory perception by informing of the sound field

Kevin Y. Tsang & Damien J. Mannion
UNSW Sydney

Background

Reverberant sounds travel through the environment, eventually reaching our ears. This reverberation can inform us of the sound field of the environment. However, predictions on the properties of the sound field are not solely dependent on perceiving reverberation.

Objects and surfaces in the environment inherently carry properties that can be extracted by both the auditory and visual systems [1]. Building upon the understanding of cross-modal processes [2], we propose that the visual environment can be used as an auxiliary source of information for predicting inferences on the sound field. We investigate the role of vision in an audition-based environment matching task.

Method

Ten real-world locations were used to produce the auditory and visual environment stimuli [5].

- Auditory stimulus: utterance was convolved with the impulse response for each of the ten locations
- Visual stimulus: panoramic photograph of the location was rendered in virtual reality, so that the environment could surround the observer

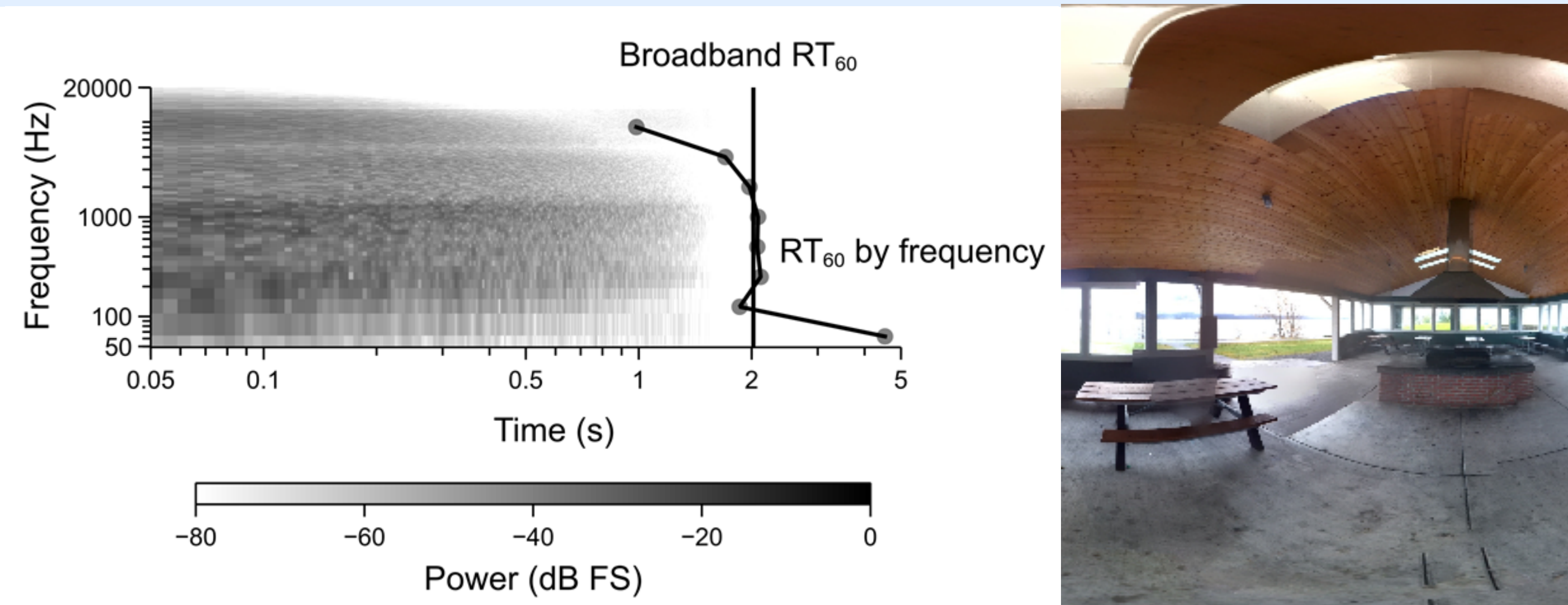
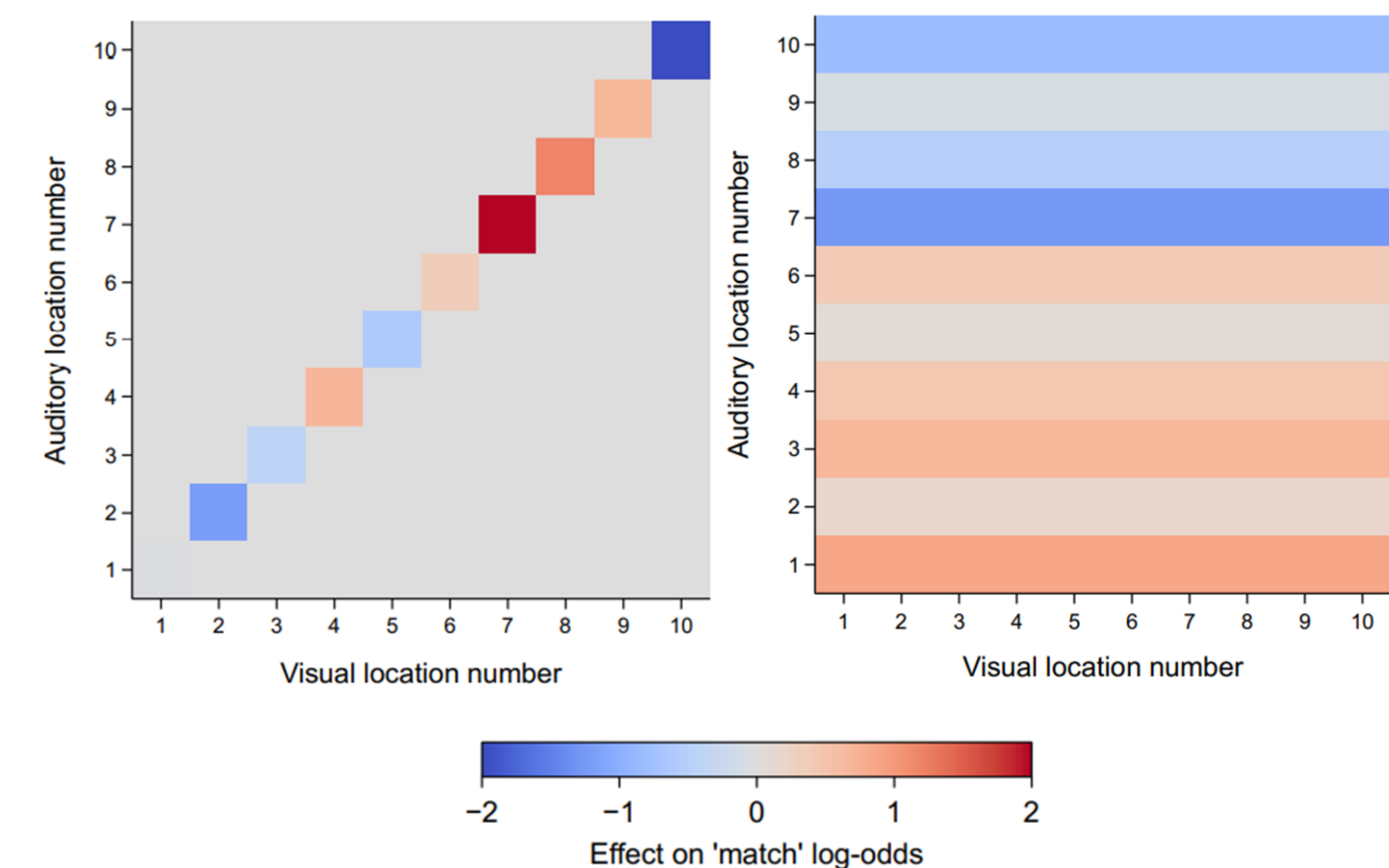
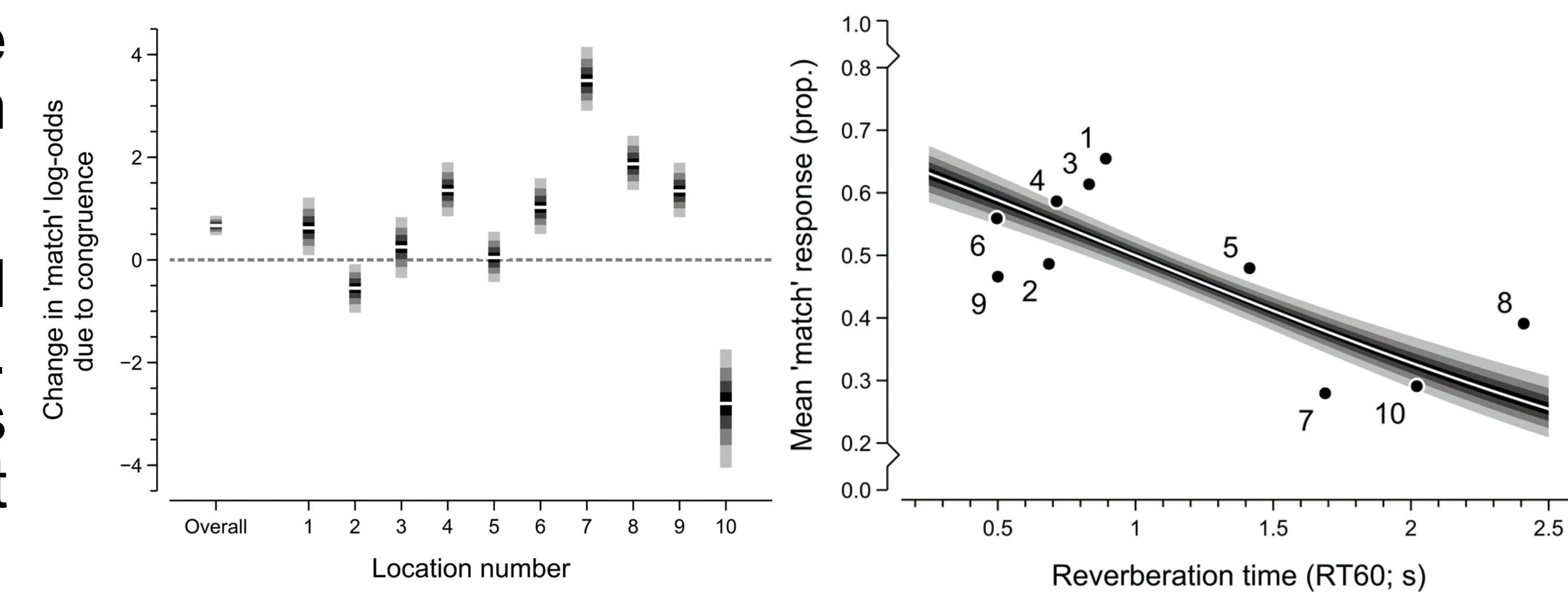
Subjects ($n = 44$) were presented with pairwise combinations of the auditory and visual environments. These environment pairs could be congruent or incongruent, depending on whether they originated from the same location. In each presentation, after searching for a visual object, subjects judged whether an utterance was produced from within same the environment they were surrounded by.

Results

We measured the change in match responses based on congruence using a Bayesian approach [3,4]. Overall, the probability of a match response was greater for congruent locations when compared to incongruent locations.

Location 10 was notable because it had comparatively lower match response probability. Despite the congruence of the pair, participants made systematically inaccurate predictions about the sound field from the image.

Averaged across all visual locations, reverberation time was negatively correlated with match probability.



Environment stimuli (Location 10)

Conclusions

- We have propensity to identify visual information that is associated with the sound field
- Matching response decreases when presented with auditory environments with long reverberation times
- Spatial structure is key to both visual and auditory perception in identifying environments

References

- [1] O'Callaghan (2008), *Philosophical Issues*, 18, 316-338.
- [2] Frassinetti et al. (2002), *Experimental brain research*, 147, 332-343.
- [3] Lee (2018), *Stevens' handbook of experimental psychology and cognitive neuroscience*, 1-48.
- [4] Wagenmakers et al. (2018), *Psychonomic Bulletin & Review*, 25, 35-57.
- [5] Echothief impulse response library (<http://www.echothief.com>)