Effect of Spatial Correlation on Causal Inference of Stochastic Audiovisual Sequences

Jiaming Xu

Sponsor and mentor: Prof. Michael Landy

Department of Psychology, New York University

September 14, 2022

Effect of Spatial Correlation on Causal Inference of Stochastic Audiovisual Sequences

In our daily lives, we are constantly bombarded with information coming from different senses. For example, imagine you are hiking in the woods. Suddenly, you hear the hissing of a snake coming from the bush in front of you and you also see some movement behind the bush. Combining the auditory and visual information will increase the probability of noticing and correctly locating the snake, which increases the chances of running away from it. However, due to internal noise in the brain and external noise in the environment, sensory signals will not be in perfect agreement, which could be manifested either as having a temporal or a spatial discrepancy. To form a coherent percept of the world, our brain integrates multisensory cues that are likely to be coming from a common source. For example, the ventriloquism effect happens when the observer integrates an auditory cue and a plausible simultaneous visual cue despite the spatial discrepancy between them, which results in a shifted location estimate of the auditory cue towards the location of the visual cue.

However, the brain does not always integrate information from different sensory modalities. One important factor determining integration is spatiotemporal proximity: if the cues from different modalities are far apart either in space or time, they are likely to be treated as if they originated from different sources and therefore segregated (Körding et al., 2007; Wallace et al., 2004; Hairston et al., 2003). In other words, when making an estimate of a stimulus property, the brain makes an inference about the causal relationship between the cross-modal signals. Many studies found that humans indeed perform causal inference in a variety of multisensory tasks (Dokka et al., 2019; Badde et al., 2020; Magnotti & Beauchamp, 2017).

The role of causal inference in multisensory perception has been studied extensively (Körding et al., 2007; Wozny et al., 2010; Cao et al., 2019), but the stimuli used in these studies were single events. However, in the natural environment, many stimuli encountered in daily life are sequences of stimuli, e.g., fireworks, footsteps, etc. Similar to integrating single sensory events from different modalities, when encountering stochastic sequences of multisensory events, the brain also performs causal inference. In this case, spatiotemporal correlation becomes the main factor for determining whether the underlying cues should be integrated. Several studies have investigated the role of temporal cross-correlation between auditory and visual signals on multisensory integration and causal inference (Locke & Landy, 2017; Parise et al., 2012, 2013; Parise & Ernst, 2016). For example, Parise and colleagues (2012) presented trains of either unimodal visual, unimodal auditory, or bimodal audiovisual stimuli, the temporal structure of which was manipulated to be correlated or uncorrelated while the spatial locations of which always coincided. Participants were asked to localize the perceived location of the stimuli. The results showed that participants' response precision in bimodal trials relative to unimodal trials was statistically optimal only when audiovisual stimuli were temporally correlated, indicating that temporal correlation was taken into account when inferring whether the audiovisual signals shared a common cause (Parise et al., 2012).

Current study

However, to date, no study has investigated the role of spatial correlation in causal inference and multisensory integration. Here, I maintain audiovisual temporally synchrony and manipulate spatial correlation between auditory and visual stimuli, from completely correlated (each pair of events within the auditory and the visual sequences are co-located in space) to uncorrelated (the location of each event within the auditory or the visual sequence is randomized within a range such that the correlation between the two sequences is 0). The spatial discrepancies between the auditory and visual sequences are introduced by varying the physical location of the centroid of each sequence. I ask participants to localize the centroid of the sequences in both modalities and ask them to make an explicit causal-inference judgment.

Hypothesis: If spatial correlation operates as a cue for causal inference, auditory and visual signals will be integrated more fully when their spatial correlation is higher due to the causal inference of a common source. I hypothesize that participants' localization responses will shift toward the location of the centroid of the other modality, compared to their localization responses in unimodal trials and they will be more likely to report that the two cues share a common cause.

Method

Apparatus and Stimuli

The experiment will be conducted in a dark, semi sound-attenuated room. Participants will be seated comfortably with their chins rested on a chin rest.

A visual event is a brief flash of a low-contrast Gaussian blob projected onto a gray background. An auditory event is a brief broadband noise burst, delivered via headphones. Each stimulus presentation is a sequence of five visual and/or five auditory events, presented every 500 ms over a 2 s temporal interval. The spatial locations of all visual and auditory events range from -18° to 18° along a horizontal line. Each sequence of visual or auditory events has a spatial window of 15°, that is, the locations of individual visual or auditory events are randomized within the range of 15° and centered on the centroid. In bimodal trials, the spatial structure of each pair of visual and auditory sequences are manipulated to have various spatial correlations between them, ranging from 0-1. The centroids of the auditory sequences are fixed at the left, right, or center locations while the centroids of the visual sequences are situated at the left or right locations, yielding six combinations with various spatial discrepancies.

To create compelling spatialized sounds, in a preliminary session the raw auditory events will be played from a loudspeaker behind an acoustically transparent screen from each of 20 spatial positions and recorded with a pair of in-ear binaural microphones placed inside the left and right ear canals of each participant. The recording for each participant will then be cut into short clips, each containing the sound of an auditory event at a specific spatial location. By performing this procedure individually for each participant, the auditory stimuli are effectively filtered by the individual's head-related transfer function, thereby providing rich and ecological cues for externalized sound localization when played back over headphones.

To find the perceptual location of each auditory event for each participant, in a preparatory experiment, participants will localize each of the auditory events recorded from 20 spatial locations 20 times played from headphones, the order of which will be randomized. Then, the perceptual location of each auditory event can be computed by calculating the mean localization response of that location. To find the perceptually corresponding locations of 20 visual events for each participant, I will fit a linear regression to the mean auditory localization responses for each participant. The perceptual locations of auditory and visual events for each participant will be used to generate their individualized auditory and visual sequences and their respective centroids in the main experiment.

Procedure

This study will consist of four preparatory experiments and two main experiments. The purposes of the preparatory experiments are to find perceptual event (auditory and visual) locations for each participant based on their individual perceptual biases, to familiarize participants with the experimental setup, the stimuli, and the task, and to obtain information about participants' memory and motor noise.

The first main experiment is a unimodal localization task. On each trial, participants will be presented with either a visual sequence or an auditory sequence with a duration of 2 s (the order of auditory and visual trials will be randomized). After each stimulus presentation, participants will localize the centroid of that sequence using the trackball on a mouse. Scrolling the trackball will move a visual cursor left and right pointing at the horizontal line where stimuli are presented. Once participants move the visual cursor to the desired location, they click a mouse button to confirm that they are satisfied with the location setting.

The second main experiment is a bimodal localization task. On each trial, participants will be presented with sequences of visual and auditory events with a duration of 2 s, simultaneously, and with various spatial discrepancies between two centroids. After each stimulus presentation, participants will localize the centroid of either the auditory sequence or the visual sequence while ignoring the other modality. Participants will be asked to localize the centroid of the auditory sequence in half of the trials and localize the centroid of the visual sequence in the other half of the trials, the order of which will be randomized. After making a localization response, participants will be asked to judge whether the two stimuli came from the same or separate sources.

Statistical analysis

To understand how the brain detects and integrates spatially related information across continuous streams of multisensory signals and how it adapts to spatial conflicts between the sensory modalities, I plan to fit the data to the multisensory correlation detector (MCD) model, which was initially adapted from the Hassenstein-Reichardt detector for visual motion perception and has successfully replicated human behavior in empirical studies (Parise & Ernst, 2016).

References

Badde, S., Navarro, K. T., & Landy, M. S. (2020). Modality-specific attention attenuates visual-tactile integration and recalibration effects by reducing prior expectations of a common source for vision and touch. *Cognition*, *197*, 104170. https://doi.org/10.1016/j.cognition.2019.104170

Cao, Y., Summerfield, C., Park, H., Giordano, B. L., & Kayser, C. (2019). Causal Inference in the Multisensory Brain. *Neuron*, *102*(5), 1076-1087.e8. https://doi.org/10.1016/j.neuron.2019.03.043

Dokka, K., Park, H., Jansen, M., DeAngelis, G. C., & Angelaki, D. E. (2019). Causal inference accounts for heading perception in the presence of object motion. *Proceedings of the National Academy of Sciences*, *116*(18), 9060–9065. https://doi.org/10.1073/pnas.1820373116

Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L., & Schirillo, J. A. (2003). Visual localization ability influences cross-modal bias. *Journal of Cognitive Neuroscience*, *15*(1), 20–29. https://doi.org/10.1162/089892903321107792

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE*, *2*(9). https://doi.org/10.1371/journal.pone.0000943

Locke, S. M., & Landy, M. S. (2017). Temporal causal inference with stochastic audiovisual sequences. *PLOS ONE*, *12*(9), e0183776. https://doi.org/10.1371/journal.pone.0183776

Magnotti, J. F., & Beauchamp, M. S. (2017). A Causal Inference Model Explains Perception of the McGurk Effect and Other Incongruent Audiovisual Speech. *PLOS Computational Biology*, *13*(2), e1005229. https://doi.org/10.1371/journal.pcbi.1005229

Parise, C. V., & Ernst, M. O. (2016). Correlation detection as a general mechanism for multisensory integration. *Nature Communications*, *7*(1), 11543. https://doi.org/10.1038/ncomms11543

Parise, C. V., Harrar, V., Ernst, M. O., & Spence, C. (2013). Cross-correlation between Auditory and Visual Signals Promotes Multisensory Integration. *Multisensory Research*, *26*(3), 307–316. https://doi.org/10.1163/22134808-00002417

Parise, C. V., Spence, C., & Ernst, M. O. (2012). When Correlation Implies Causation in Multisensory Integration. *Current Biology*, *22*(1), 46–49. https://doi.org/10.1016/j.cub.2011.11.039

Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, *158*(2). https://doi.org/10.1007/s00221-004-1899-9

Wozny, D. R., Beierholm, U. R., & Shams, L. (2010). Probability Matching as a Computational Strategy Used in Perception. *PLoS Computational Biology*, *6*(8), e1000871. https://doi.org/10.1371/journal.pcbi.1000871