Audiomotor temporal recalibration modulates feeling of control: Exploration through an online experiment and Bayesian modeling

Yoshimori Sugano[1,*]

[1] Department of Business and Marketing, Faculty of Commerce, Kyushu Sangyo University, Fukuoka, Japan

[*] Corresponding author: sugano@ip.kyusan-u.ac.jp

ABSTRACT

This online experiment aimed to replicate Sugano's (2021) findings on how exposure to delayed auditory feedback affects feeling of control (agency). Participants first adapted by repeatedly reproducing a short sequence of tones while receiving either immediate (10 ms) or delayed (110 ms) auditory feedback on their keypresses. This exposure aimed to recalibrate their timing perception. Following adaptation, they reproduced the sequence again. However, the computer would assume control of the tone pips from the participants with 50% probability. Participants judged whether their keypresses caused the tones (agency judgment: AJ) or if the tones felt synchronized with their keypresses (simultaneity judgment: SJ). Analyses revealed similar shifts in both judgments after exposure to delay, however, the AJ exhibited a greater shift compared to the SJ. In addition, the shift reflects a change in bias, not in sensitivity. The result aligns with previous in-person study, suggesting online experiments can effectively explore agency.

KEYWORDS

temporal recalibration; delayed auditory feedback; sense of agency; signal detection theory; Bayesian modeling

Introduction

Background

Artificial systems, such as computers, often provide sensory feedback signals (e.g., a beep) to indicate that a voluntary action (e.g., pressing a key) has been completed. A delay between the sensory feedback and the action can negatively impact the interaction with the system and reduce the sense of agency (SoA) (Wen, 2019). However, prolonged exposure to the delay can recalibrate the perception of simultaneity between action and feedback, as demonstrated by Stetson et al. (2006). This phenomenon is known as temporal recalibration (TR) (Vroomen et al., 2004) or lag adaptation (Fujisaki et al., 2004). As a result of this adaptive shift in subjective simultaneity, the interaction between humans and artificial systems becomes smoother, and the sense of agency (SoA) is regained (Cunningham et al., 2001).

The modulation of the sense of agency (SoA) associated with temporal recalibration is so fundamental that after removing the delay, the perception of cause and effect is reversed. Specifically, when the delay was removed, participants typically reported perceiving the sensory feedback as occurring “before” their action, even if it was presented “after” the action (Cunningham et al., 2001; Stetson et al., 2006). Recent experimental studies confirm these anecdotal reports that the sense of agency (SoA) over nondelayed sensory feedback decreases after exposure to delayed feedback (Timm et al., 2014; Haering and Kiesel, 2015, 2016; Imaizumi and Asai, 2015). This suggests that the SoA is dependent on subjective simultaneity between one's voluntary action and the sensory feedback received rather than on veridical simultaneity.

When considering the changes in the sense of agency (SoA) and the perception of simultaneity due to temporal recalibration (TR), the concept of a temporal window is frequently employed. The point of perceived simultaneity is believed to be extended over time to combine different sensory inputs into a single perception, also known as the psychological present (Michon, 1978). This is commonly referred to as the temporal window of simultaneity (TWS) (e.g., Varela, 1999). Likewise, the point of perceived agency is also extended in time, which is referred to as a temporal window of agency (TWA) (e.g., Farrer et al., 2013). The sense of agency (SoA) is strongly influenced by the perceived time between voluntary actions and sensory feedback (Franck et al., 2001; Sato and Yasuda, 2005; Rohde et al., 2014b;

Timm et al., 2014), making it highly probable that changes in the TWA will coincide with changes in the TWS. Indeed, Sugano (2021) confirmed this co-occurrence.

Primary Aim: Replicating Sugano's (2021) Findings

Sugano (2021) investigated whether exposure to delay causes a change in perceptual sensitivity and/or a change in the decision criterion of agency perception, leading to modulation of the SoA. The study revealed that the midpoint and/or width of both the TWS and the TWA changed after exposure to delayed auditory feedback (e.g., a tone pip) following a voluntary action (e.g., a key press). The analysis using signal detection theory (SDT; Green and Swets, 1966) showed that the modulation of the SoA was accompanied by a shift in the criterion measure but not a change in the sensitivity measure. This suggests that it is a product of a late stage of processing, such as a change in the decision criterion and/or a change in perceptual bias (Meyerhoff & Scholl, 2018).

However, there has been criticism of Sugano's (2021) experimental procedure. In the study, participants judged agency and simultaneity within a single trial, with the agency judgment (AJ) always preceding the simultaneity judgment (SJ). Additionally, the AJ was accompanied by a confidence rating, while the SJ was not. This procedure may have made participants more sensitive to the AJ than to the SJ, and there may be some interaction between the two judgments. More specifically, participants may judge either the AJ or the SJ without due consideration and then give the same response to the other. Sugano (2021) verified this possibility by analyzing the correlation between SJ and AJ for each participant. He found that approximately 61% of the participants tended to judge them differently—a moderate result that does not allow for a definitive conclusion. Thus, it is valuable to retest the findings presented by Sugano (2021) with a distinct group of participants who separated SJ and AJ. The main objective of this study is to achieve this goal.

Secondary Aim: Conducting an Online Experiment

The secondary objective of this study was to determine whether online experiments could replicate the modulation of the SoA and temporal recalibration, which necessitates precise control of stimulus presentation and data acquisition. Online experiments, also called web-based or internet-based experiments, are conducted entirely over the internet (e.g., Crump et al., 2013). As conducting in-person psychological experiments was made difficult by the COVID-19 pandemic, several psychologists have attempted to conduct psychological experiments via the internet (e.g., Grootswagers, 2020; Anwyl-Irvine et al., 2021; Ma & Schnupp, 2021; van der Burg et al., 2021; Schroeger et al., 2022; Tanaka, 2024). However, conducting experiments online presents a significant challenge because the experimenter has no control over the participants' environment, such as their computer setup, room lighting, ambient noise, and viewing distance from the screen.

Additionally, it is important to consider the timing precision of the personal computer (PC) used to control stimulus presentation and data acquisition in online experiments. The experimenter cannot verify the timing accuracy of the participant's PC. However, this is not an issue when the factors under investigation are within-subjects. In such cases, timing errors can be cancelled out between the experimental conditions in the within-subjects design (Kuroki, 2020).

Psychophysical experiments require millisecond-level timing precision, as many experimental paradigms rely on the precise timing of stimulus and response measurements to accurately measure perceptual and cognitive processes. This is widely recognized among researchers in the fields of psychophysics and cognitive psychology (Crump et al., 2013). Fortunately, recent developments in devices and applications for online experiments have allowed us to overcome limitations and replicate results observed in in-person experiments in the laboratory. Currently, there are many options available for performing psychophysical experiments that require stimulus control or data acquisition with millisecond temporal precision (see Crump et al., 2013; de Leeuw, 2015; Semmelmann & Weigelt, 2017). However, there are currently few online psychophysical experiments in which timing is critical. For example, Schroeger et al. (2022) explored the kappa effect, and Tanaka (2024) examined intentional binding. Therefore, it is worth conducting temporal recalibration experiments online to obtain results similar to those of laboratory experiments. This is the secondary aim of this study.

What Can the SDT Analysis Reveal?

The SDT is a method used to distinguish between perceptual sensitivity and decision-making criteria in tasks where individuals must choose between two options (Green & Swets, 1966; Wickens, 2002). However, caution should be exercised when applying SDT to perception studies.

One criticism from a technical perspective is that changes in the criterion measures of SDT, such as c or β, can sometimes reflect changes in perception itself rather than just in decision-making (Witt et al., 2015). Witt et al. (2015) used computer simulations to show that the well-known Müller–Lyer illusion, where lines appear to have different lengths due to surrounding arrows, can be explained by changes in the SDT criterion measure. They also noted that the type of task used in an experiment can affect which type of bias is being measured. For instance, in a task requiring discrimination between two similar stimuli, experimental manipulation can affect both the clarity of the signal (the target stimulus) and the clarity of the noise (the distractor stimulus). As a result, the SDT criterion measure could reflect both how the person is making decisions and how they are perceiving the stimuli.

However, in the type of task used by Green and Swets (1966), where the person just needs to identify if a signal is present, an experimental manipulation only affects the signal, not the noise. Therefore, the SDT criterion measure used here reflects only decision-making bias, not perceptual bias. This is why changes in SDT's sensitivity measure, such as d-prime, reflect changes in perception itself in this type of task (see Witt et al., 2015, for more details).

Another criticism comes from a psychological perspective (Bertelson & Radeau, 1976; Meyerhoff & Scholl, 2018). Meyerhoff and Scholl (2018) argue that if an experimental manipulation significantly alters perception, it is unlikely to solely reflect a judgment change. They suggest that genuine perceptual changes are more likely. Bertelson and Radeau (1976) specifically critique a study by Choe et al. (1975) that applied SDT to the ventriloquist effect, where sound appears to originate from a dummy's mouth. Choe et al. (1975) concluded that the effect results from decision criterion changes, not from altered perceptual sensitivity. Bertelson and Radeau counterargue that the perceived unity of voice and image in the ventriloquist effect feels too real to be solely attributed to decision bias, suggesting a genuine perceptual shift. Furthermore, the illusion's aftereffects, where sounds continue to seem misplaced even after the dummy's removal, cannot be explained by decision-making changes alone.

Therefore, changes in the criterion measure in the SDT may reflect changes that occur at both the level of perception (early stages of human information processing) and at the level of decision making (later stages) (Meyerhoff & Scholl, 2018). In this way, Sugano's (2021) conclusion that these changes only reflect later stages may be an oversimplification. The SDT criterion measure can be relevant to both the early and late stages of information processing.

Despite these limitations, applying the SDT framework to specific data types can offer valuable insights. SDT's ability to differentiate between sensitivity and bias (perceptual and/or decisional) holds promise for understanding the parameters underlying the TWS and the TWA. As suggested by Sugano (2021), the SDT criterion measure might correspond to the center of the window, while the sensitivity measure might relate to its width.

How Does Perceptual Adaptation Affect the Sensitivity and Criterion of the SDT?

The relationship between the measures of the SDT and the temporal window parameters is not as simple as it seems. Multiple possibilities exist for how the probability distributions of the noise and the signal change due to perceptual adaptation of delayed sensory feedback. Figure 1 graphically illustrates the two representative types of change. The figure displays the hit rate (HR) and false alarm rate (FAR) in the discrimination task. A participant's response is categorized as a hit if the participant responds “yes” where the signal exists and categorized as a false alarm if the participant responds “yes” where the signal does not exist. The sensitivity in the SDT is the distance between the distributions (i.e., signal vs. noise), while the decision criterion separates participants’ responses (i.e., yes vs. no).

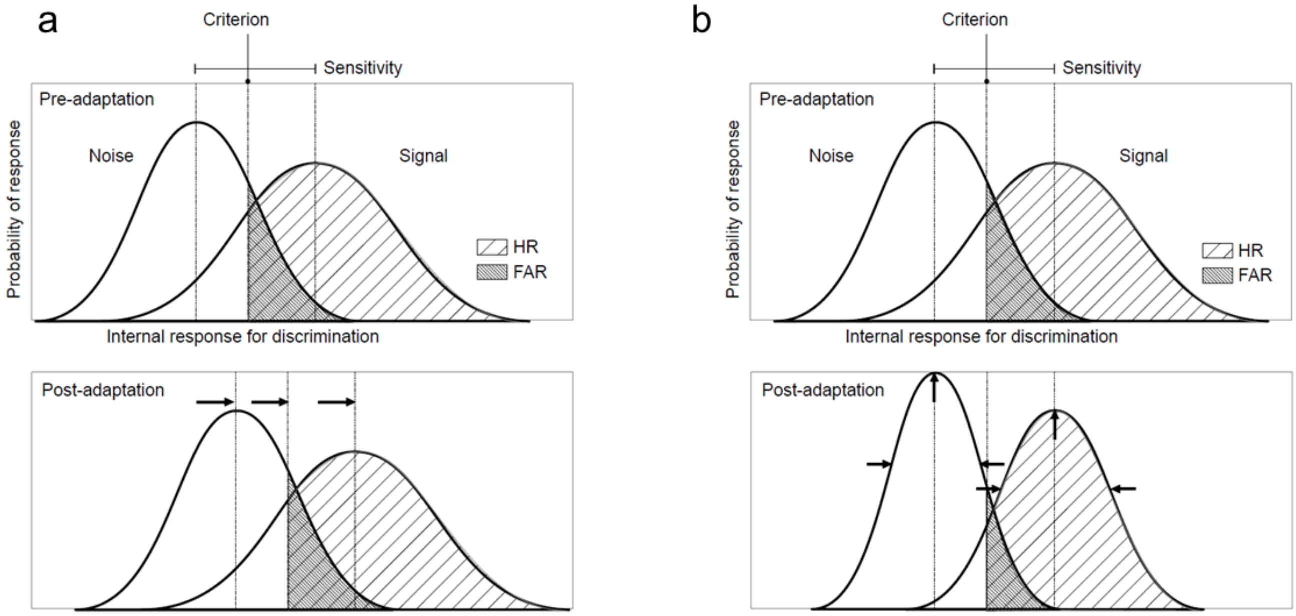


Figure 1. Perceptual adaptation effects on signal-detection theory (SDT) distributions. This figure illustrates how perceptual adaptation can alter signal and noise distributions within the SDT framework. **(a)** PSE-shifting adaptation. It depicts a shift in the point-of-subjective-equality (PSE) caused by adaptation. The centers of both the signal and noise distributions are displaced equally in the same direction, without changes in their widths. The decision criterion may adjust accordingly, but its relative position to the two distributions remains unchanged, indicating no apparent alteration. **(b)** Sensitivity-changing adaptation. It portrays sensitivity enhancement adaptation. Both distributions narrow in width, indicating improved discriminability, while their centers remain constant. The decision criterion may or may not be affected. Note that these two adaptation types are not mutually exclusive and can cooccur.

Imagine a scenario where adaptation shifts both the noise and signal distributions in the same direction equally without changing their shape. This type of adaptation, which I call a PSE-shifting adaptation (Figure 1a), affects the point of subjective equality (PSE), which corresponds to a particular point on the internal response in Figure 1a, in tasks where people judge whether two stimuli are synchronous or self-controlled. This occurs because the location of the noise distribution is not fixed when distinguishing between stimuli (Witt et al., 2015). Consequently, neither the sensitivity nor the criterion in SDT changes, even if the PSE does change (Witthoft et al., 2018). Therefore, the measures of SDT are not effective in capturing the underlying adaptation process.

Another possibility is an adaptation that alters the shape of both distributions in the same way. For example, both distributions could become narrower and taller, indicating increased sensitivity to both noise and signals. Figure 1b depicts this scenario, which I call a sensitivity-changing adaptation. In this scenario, changes occur in the sensitivity measure while maintaining the distance between the distributions. Although the criterion measure can also change, this type of adaptation mainly affects sensitivity in SDT.

Finally, the complete execution of adaptation is considered. When adaptation is complete, it is expected that the sensitivity and bias return to their initial states, indicating no lasting changes.

The adaptation process is likely a combination of these two types (PSE-shifting and sensitivity-changing) and may be incomplete due to limitations in the duration of the experiment. By analyzing how the sensitivity and bias change, we can gain insights into the type (PSE-shifting or sensitivity-changing) and extent (complete or incomplete) of the adaptation process.

Furthermore, when the adaptation is dominated by the sensitivity-changing type, such as changing the shape of the temporal window by adjusting its width or height, the sensitivity measure in the SDT should be related to parameters defining the shape of the window. Conversely, neither the sensitivity nor the criterion measure for PSE-shifting adaptations is related to the parameter defining the center of the window. Therefore, it is predicted that the criterion measure will not be correlated with the center of the window, while the shape of the window will be correlated with the sensitivity measure.

Research Questions and Predictions

Building on the discussion thus far, I addressed the following research questions:

1. How does an exposure to delayed sensory feedback affect the TWS and the TWA? These windows represent the perception of the temporal proximity required for two events (an action and its resulting feedback) to be perceived as synchronous or self-generated. Specifically, does the center of these windows shift, and/or does

their width change?

2. Does experiencing a delay affect our sensitivity to the variability of timing between our actions and feedback while still being perceived as synchronous or self-generated?
3. Does an exposure to the delay influence our sensitivity (ability to distinguish differences) or decision criteria (threshold for making a judgment) for simultaneity and agency?
4. How are the parameters of the temporal windows related to the parameters used in the SDT for analyzing simultaneity judgments (SJs) and agency judgments (AJs)?
5. Can an online experiment replicate the findings from an in-person experiment?

For each question, I made specific predictions based on existing research and my own hypotheses:

1. The center of the TWS and TWA might shift in the direction of the delay, and the width of these windows may or may not increase.
2. Exposure to the delay might make us less sensitive to variations in the timing between our action and feedback.
3. Based on previous research (Sugano, 2021), we might expect that the sensitivity to agency remains constant while the decision criterion changes. For simultaneity, the results are less clear.
4. We predict that the sensitivity is related to the width of the temporal window. However, the relationship between the decision criterion and the shift of the window center is uncertain.
5. We anticipate that an online experiment can replicate the findings of an in-person experiment.

Results

Temporal Windows of Simultaneity (TWS) and Agency (TWA)

The posterior distributions of the parameters of the hierarchical Gaussian window model for the TWS and the TWA were estimated using a Bayesian framework with Markov chain Monte Carlo (MCMC) simulations employing Hamiltonian dynamics (Hamiltonian Monte Carlo: HMC). A total of 40,000 samples were obtained from four chains, each with 11,000 samples. The initial 1,000 samples per chain were discarded as a warm-up period to ensure convergence. Convergence diagnostics confirmed satisfactory mixing of the chains, with R-hat values below 1.1 (Vehtari et al., 2021) and effective sample sizes (ESS) exceeding the minimum recommended threshold (>100 [Matsuura, 2016; Ebrahim & Cengiz, 2022]) for all parameters. This indicates that the model achieved convergence, and the remaining samples effectively approximated the posterior distributions. Subsequent statistical inferences were based on these 40,000 MCMC samples.

Figure 2 depicts the estimated hierarchical Gaussian window model. Figure 2a and 2b present scatter plots of the raw SJ data for the baseline and the delayed condition, respectively, plotted against the mean asynchrony (averaged across participants). Figure 2c and 2d present scatter plots of the raw AJ data. The probability density function (PDF) of the data is visualized as a curve on the periphery of each plot. Circles represent the mean rate of "synchronous" or "self" responses for each bin of mean asynchrony. The size of each circle reflects the number of observations within that bin, which is also indicated by the digit below the circle. The black solid line indicates the median of the MCMC samples, representing the central tendency of the estimated model parameters. Dotted lines represent the 95% equal-tailed credible interval (CI), indicating a 95% probability that the true parameter values lie within this range. Figure 2e and 2f display the expected a posteriori (EAP) estimates of the model parameters at both the individual (top panels) and group levels (middle and bottom panels) for the SJ and AJ, respectively. The 95% equal-tailed CIs for these EAP estimates are also shown as error bars.

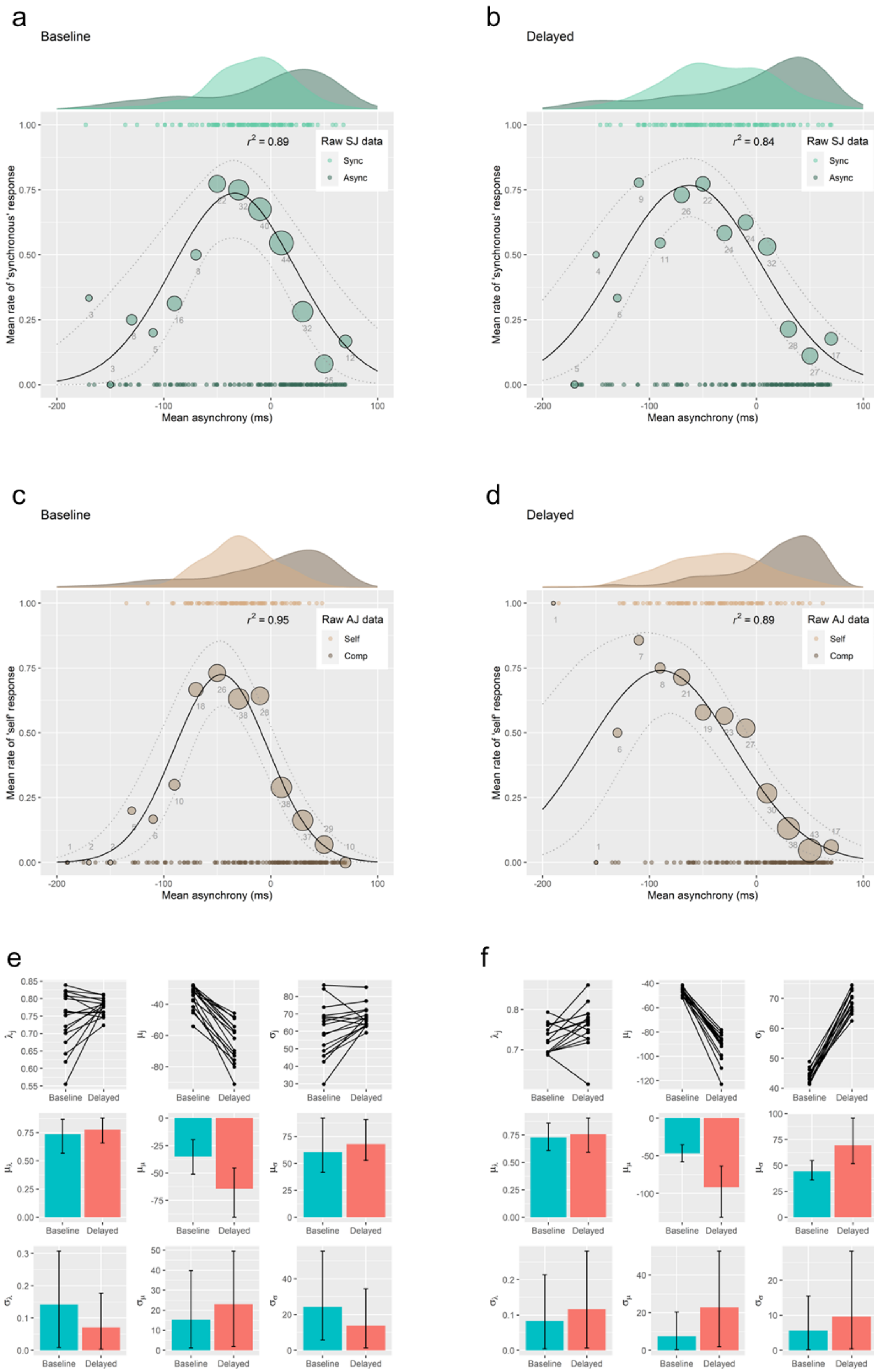

a
Baseline
b
Delayed
Raw SJ data
Sync
Async
r^2 = 0.89
r^2 = 0.84
Mean rate of 'synchronous' response
Mean asynchrony (ms)
c
Baseline
d
Delayed
Raw AJ data
Self
Comp
r^2 = 0.95
r^2 = 0.89
Mean rate of 'self' response
Mean asynchrony (ms)
e
f
Baseline
Delayed

Figure 2. Estimated temporal windows of simultaneity (TWS) and agency (TWA). **(a, b, c, d)** Mean response rate of "synchronous" or "self" responses plotted against binned mean asynchrony. The estimated Gaussian curves (solid lines) represent the median of the MCMC samples for both the baseline and delayed conditions. Dotted lines represent the corresponding 95% credible intervals. Raw SJ/AJ data with their probability density functions (PDFs) are displayed on the periphery. The goodness-of-fit for each curve was assessed using r^2 (refer to the main text for calculation details). **(e, f)** Group-level and individual-level estimates of the Gaussian model parameters for the SJ and AJ tasks plotted against mean asynchrony. The top panels depict individual-level mean estimates for each participant. The middle panels show group-level mean estimates, and the bottom panels display group-level standard deviation estimates. All parameter estimates are accompanied by their respective 95% credible intervals.

The model's goodness-of-fit was assessed using Pearson's product-moment correlation coefficient squared (r^2). This metric evaluated the correlation between the estimated rate of "synchronous" or "self" responses (i.e., the median of the MCMC samples) and the observed rate at each bin of mean asynchrony. The correlations were weighted by the number of observations within each bin (c.f., Bland and Altman, 1995). As illustrated in the figure, the r^2 values were 0.89 (baseline) and 0.84 (delayed) for the SJ task, and 0.95 (baseline) and 0.89 (delayed) for the AJ task.

For the SJ task, the EAP estimates of the mu parameter (Gaussian distribution midpoint) revealed a shift in the TWS following delayed feedback exposure (middle-center panel, Figure 2e). The mu value was -35.0 ms (95% CI: -51.0, -19.7 ms) under the baseline and shifted to -64.3 ms (95% CI: -90.3, -45.3 ms) in the delayed condition. This shift ($\Delta\mu$) was -29.3 ms (95% CI: -60.4, -2.4 ms), with a large effect size (Cohen's d = 1.99). The 95% CI not including zero and the large effect size indicate a significant TWS shift towards the delay (temporal recalibration). Notably, the $\Delta\mu$ magnitude (29.3 ms) aligns with previously reported audio-motor temporal recalibration values (30-40 ms) (Heron et al., 2009; Sugano et al., 2010, 2012; Sugano, 2021), further supporting the data. Individual participant estimates mirrored these trends (top-center panel, Figure 2e).

In contrast, the EAP estimates of the sigma parameter (Gaussian distribution width) did not support a change in TWS width after the delay (middle-right panel, Figure 2e). The sigma values were 60.4 ms (95% CI: 41.6, 92.0 ms) under the baseline and 67.9 ms (95% CI: 52.9, 90.6 ms) in the delayed condition. The $\Delta\sigma$ (7.5 ms; 95% CI: -28.7, 38.7 ms) with a small effect size (Cohen's d = 0.46) suggested no significant difference in TWS width between conditions.

Similar to the SJ task, the AJ task revealed a significant shift in the TWA following delayed feedback exposure (middle-center panel, Figure 2f). The EAP estimates of the mu parameter, representing the center of the Gaussian distribution, showed the baseline value of -46.6 ms (95% CI: -58.2, -35.3 ms) that shifted to -91.8 ms (95% CI: -131.6, -63.7 ms) in the delayed condition. This $\Delta\mu$ of -45.2 ms (95% CI: -84.8, -13.7 ms) with a large effect size (Cohen's d = 2.48) suggested a substantial shift in agency judgment towards the delay. Notably, the magnitude of $\Delta\mu$ (-45.2 ms) was greater than that previously reported (-17.0 ms; Sugano, 2021), potentially reflecting task- or population-specific differences. Individual participant estimates mirrored these trends (top-center panel, Figure 2f).

In contrast to the SJ task, the AJ task results suggest a change in the width of the TWA (middle-right panel, Figure 2f). The sigma parameter, representing the Gaussian distribution width, increased from 44.1 ms (95% CI: 36.1, 54.7 ms) at the baseline to 69.5 ms (95% CI: 51.8, 95.7 ms) in the delayed condition. The $\Delta\sigma$ of 25.3 ms (95% CI: 4.6, 52.9 ms) with a large effect size (Cohen's d = 2.05) indicates a significant widening of the TWA after the delay. This trend was also reflected in the individual-level estimates (top-right panel, Figure 2f).

Previous research suggests that the perception of agency tolerates longer delays than simultaneity (Rohde et al., 2014b). This study aimed to investigate whether the TWS and TWA differed in their representations of temporal tolerance. Specifically, I hypothesized that the EAP estimate of the mu parameter (center of the Gaussian distribution) for the TWA would be greater (i.e., reflecting a greater tolerance for delays) than that for the TWS. Similarly, I expected a larger EAP estimate of the sigma parameter (distribution width) for the TWA, indicating a broader range of tolerated delays.

Analyses revealed that the difference in the EAP of the mu parameter between the TWA and the TWS was -11.7 ms (95% CI: -30.7, 7.6 ms) for the baseline condition and -27.5 ms (95% CI: -71.3, 10.9 ms) for the delayed condition.

While the effect sizes were large (Cohen's d = 1.20 and 1.33, respectively), the 95% CIs for both conditions included zero, precluding a definitive conclusion about the mu parameter being significantly larger for TWA than TWS.

Similar results emerged for the sigma parameter. The difference in the EAP between the TWA and TWS was -16.32 ms (95% CI: -48.34, 5.66 ms), with a large effect size (Cohen's d = 1.22) for the baseline, but it was 1.5 (95% CI: -27.5, 32.1 ms), with a very small effect size (Cohen's d = 0.10) for the delayed condition. These findings do not support the hypothesis of a significantly larger sigma parameter for the TWA than for the TWS.

Impact of Delay Exposure on Sensitivity to Asynchrony Variability

The posterior distributions of the parameters of the hierarchical half-Gaussian window model regarding the sensitivity of the SJ and the AJ to the variability of asynchrony were estimated using a Bayesian framework with MCMC simulations employing HMC. A total of 40,000 samples were obtained from four chains of 11,000 samples. The initial 1,000 samples per chain were discarded as a warm-up. Convergence diagnostics confirmed by R-hat (< 1.1) and ESS (>100) for all parameters. Subsequent statistical inferences were based on these 40,000 MCMC samples.

Figure 3 depicts the estimated hierarchical half-Gaussian window model. Scatter plots (Figure 3a-3d) present the response rates of the SJ and the AJ versus the standard deviation (SD) of asynchrony for the baseline and the delayed conditions. A curve on the periphery visualizes the data distribution. The circle size reflects the number of observations per bin. The black line shows the median of the MCMC samples, and the dotted lines represent the 95% credible interval. Figure 3e and 3f display individual (top) and group-level (middle and bottom) EAP estimates of the model parameters with error bars for the SJ and AJ, respectively. As is shown in the figure, the model's r^2 values were 0.84 (baseline) and 0.64 (delayed) for the SJ task, and 0.90 (baseline) and 0.73 (delayed) for the AJ task.

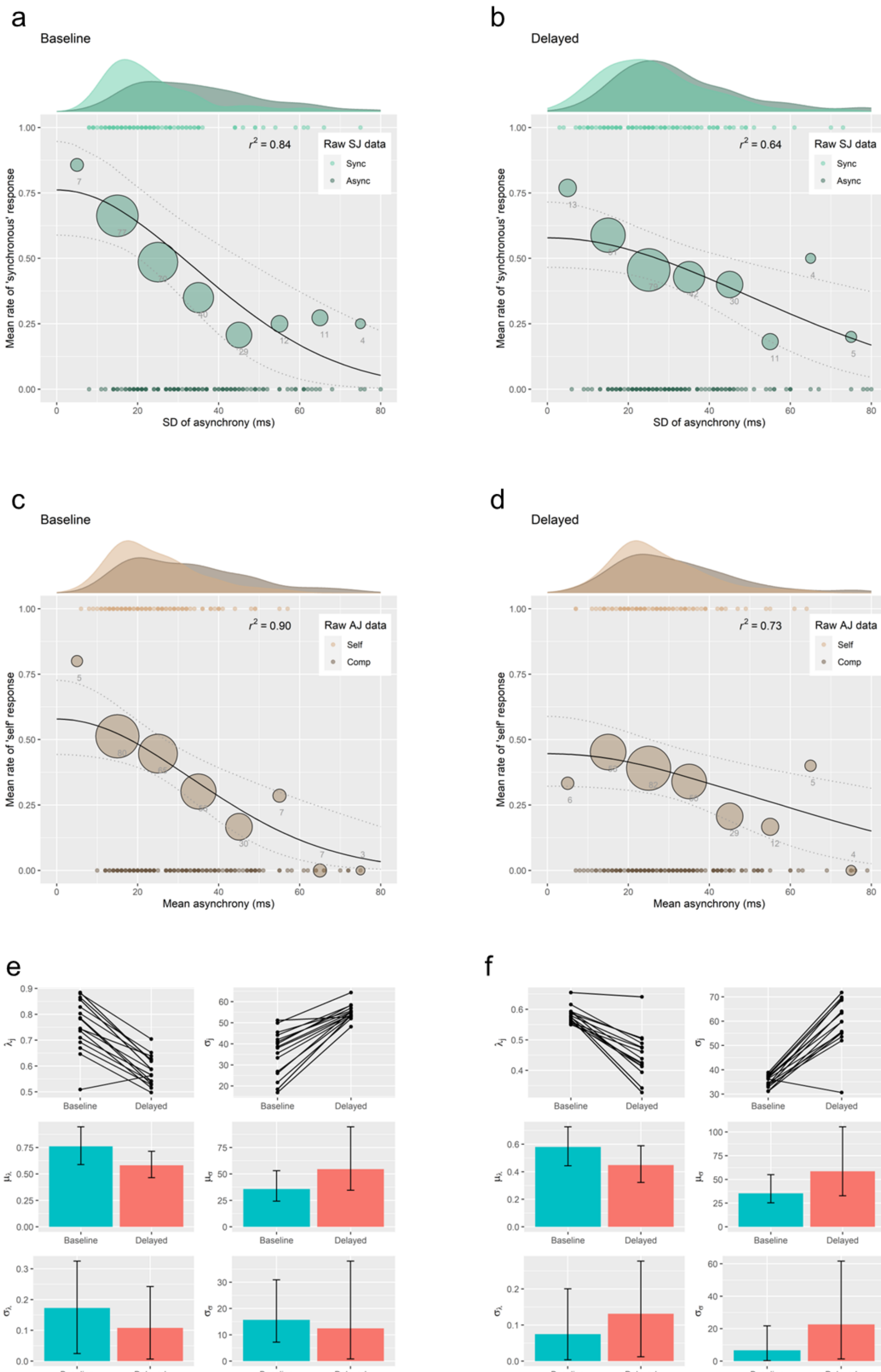

a
Baseline
r^2 = 0.84
Raw SJ data
Sync
Async
Mean rate of 'synchronous' response
SD of asynchrony (ms)
b
Delayed
r^2 = 0.64
Raw SJ data
Sync
Async
Mean rate of 'synchronous' response
SD of asynchrony (ms)
c
Baseline
r^2 = 0.90
Raw AJ data
Self
Comp
Mean rate of 'self' response
Mean asynchrony (ms)
d
Delayed
r^2 = 0.73
Raw AJ data
Self
Comp
Mean rate of 'self' response
Mean asynchrony (ms)
e
f
Baseline
Delayed

Figure 3. Influence of asynchrony variability on SJ and AJ tasks. **(a, b, c, d)** Mean response rate of "synchronous" or "self" responses plotted against the binned standard deviation (SD) of asynchrony. The estimated half-Gaussian curves (median of MCMC samples) for the baseline and delayed conditions are shown as solid lines. Dotted lines represent the corresponding 95% credible intervals. Raw SJ/AJ data with their probability density functions (PDFs) are displayed on the periphery. The goodness-of-fit for each curve was assessed using r^2 (refer to the main text for calculation details). **(e, f)** Group-level and individual-level estimates of the half-Gaussian model parameters for the SJ and AJ tasks plotted against the SD of asynchrony. The top panels depict individual-level mean estimates for each participant. The middle panels show group-level mean estimates, and the bottom panels display group-level standard deviation estimates. All parameter estimates are accompanied by their respective 95% credible intervals.

The Figure 3 reveals a clear decrease in both the mean rate of "synchronous" (SJ: Figure 3a and 3b) and "self" (AJ: Figure 3c and 3d) responses as the standard deviation (SD) of asynchrony increases. This dependence of SJ and AJ on asynchrony variability appears to be attenuated after exposure to delayed feedback, as suggested by a potential increase in the sigma parameter of the half-Gaussian window model. The estimated sigma values (representing the distribution width) for the SJ task were 35.8 ms (95% CI: 24.3, 53.1 ms) and 54.6 ms (95% CI: 34.6, 94.6 ms) in the baseline and delayed conditions, respectively. Similarly, the sigma values for the AJ task were 35.2 ms (95% CI: 25.4, 55.0 ms) and 58.5 ms (95% CI: 32.7, 106.0 ms).

Although there was a trend toward increased window width for both TWS (18.8 ms) and TWA (23.3 ms) after the delay, with large effect sizes (Cohen's d = 1.11 and 1.14, respectively), these changes were not statistically significant due to the inclusion of zero within the 95% CIs for TWS (95% CI: -8.5, 59.1 ms) and TWA (95% CI: -8.9, 71.9 ms). Therefore, we cannot definitively conclude that exposure to the delay led to a decrease in sensitivity to asynchrony variability.

Effects of Delay Exposure on Sensitivity and Response Criterion

The posterior distributions of the parameters of the hierarchical unequal variance SDT model for the SJ and the AJ were estimated using a Bayesian framework with MCMC simulations employing HMC. A total of 4,000 samples were obtained from four chains of 2,000 samples. The initial 1,000 samples per chain were discarded as a warm-up. Convergence diagnostics confirmed by R-hat (< 1.1) and ESS (>100) for all parameters. Subsequent statistical inferences were based on these 4,000 MCMC samples.

Figure 4 depicts the estimated hierarchical unequal variance SDT model. Figure 4a and 4b present scatter plots of the observed data mean with standard error (open circles with error bars) alongside the estimated receiver operating characteristic (ROC) curves for the SJ and the AJ tasks (solid lines). These ROC curves are derived from the median of the MCMC samples. The 95% CI is also displayed as the shaded area between the dotted lines. The observed data points on the ROC curves were obtained from multiple pairs of mean hit and false alarm rates, calculated using confidence ratings as the decision boundary between "synchronous" (or "self-controlled") and "asynchronous" (or "computer-controlled") responses. These pairs were constructed by varying the confidence rating threshold (e.g., 1 vs. 2-6, 1-2 vs. 3-6, etc.) (c.f., Wickens, 2002). Figure 4c and 4d display the expected a posteriori (EAP) estimates of the model parameters at both the individual (top panels) and group levels (middle and bottom panels) for the SJ and the AJ, respectively. The 95% CIs for these EAP estimates are also shown.

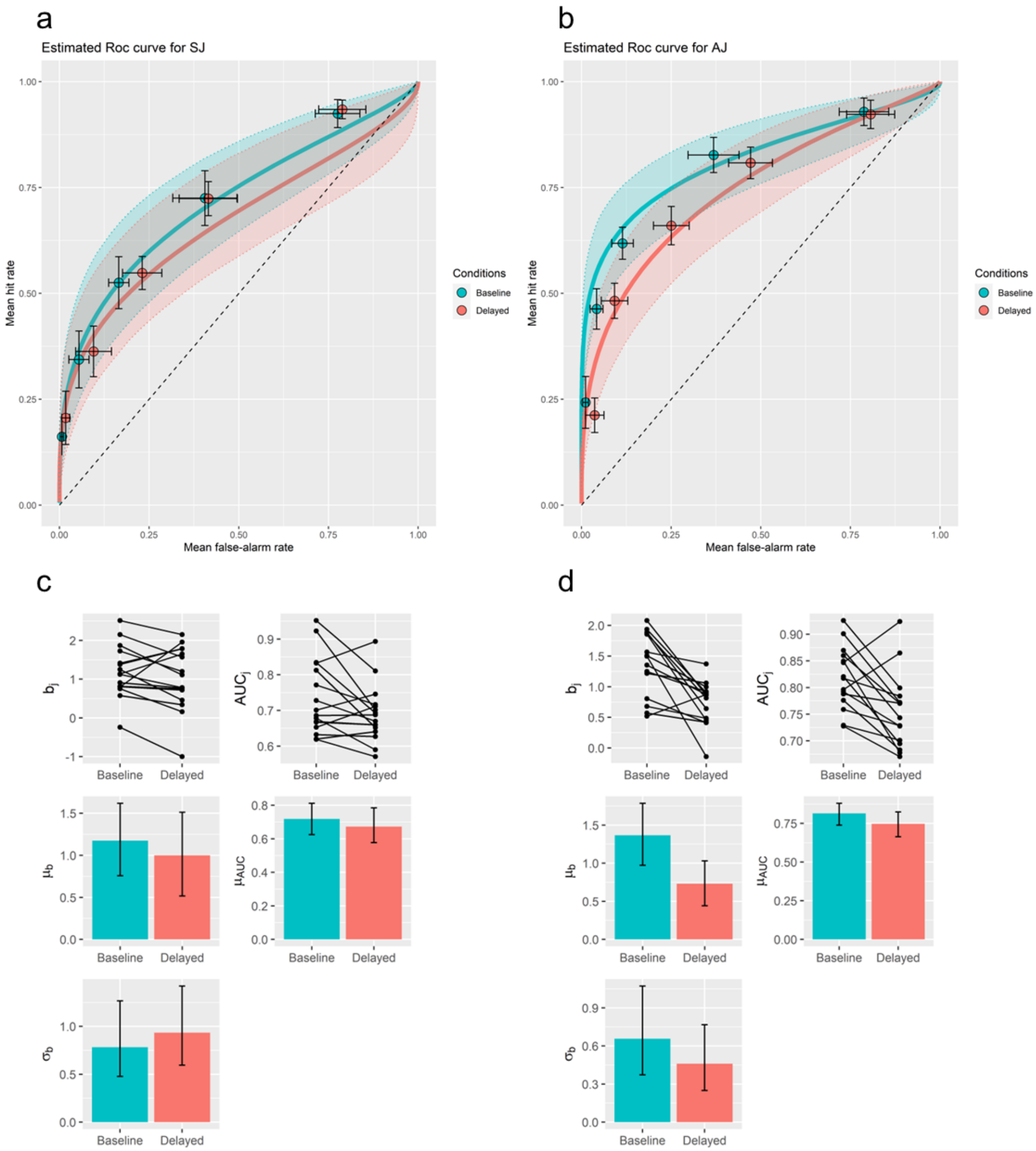


Figure 4. Results of the unequal variance Bayesian signal detection theory (SDT) analysis. **(a, b)** Scatter plots depicting the mean observed data points with standard errors (open circles with error bars) for both the SJ and AJ tasks under the baseline and delayed conditions. The estimated ROC curves (median of the MCMC samples) are shown as solid lines. Dotted lines with shaded areas represent the corresponding 95% credible intervals for the ROCs. The dashed diagonal line indicates a sensitivity of zero. **(c, d)** Group-level and individual-level parameter estimates for the unequal variance Bayesian SDT model for the SJ and AJ tasks are displayed. The top panels show individual-level mean estimates for each participant. The middle panels depict group-level mean estimates. The bottom panels represent

group-level standard deviation estimates. All parameter estimates are accompanied by their respective 95% credible intervals.

The suitability of the unequal variance SDT model was evaluated by examining the EAP estimates of the standard deviation (SD) of the signal distribution ($\sigma^{(s)}$). The model is considered appropriate if the SD of the signal distribution is demonstrably larger than the noise distribution (i.e., $\sigma^{(s)} > \sigma^{(n)} = 1.0$). Analyses revealed that the signal distribution SD for SJ data was 1.56 (95% CI: 1.24, 1.93) in the baseline condition and 1.73 (95% CI: 1.25, 2.38) in the delayed condition. Similarly, for the AJ data, the signal distribution SD was 1.96 (95% CI: 1.57, 2.45) in the baseline condition and 1.44 (95% CI: 1.15, 1.80) in the delayed condition. Importantly, none of the 95% CIs included 1.0, confirming that the signal distribution's variance consistently exceeded that of the noise distribution. This provides support for the validity of the unequal variance model across all datasets and conditions.

The estimated area under the curve (AUC) for the SJ task (middle-right panel, Figure 4c), a measure of sensitivity, was 0.72 (95% CI: 0.62, 0.81) under the baseline condition and 0.67 (95% CI: 0.58, 0.78) under the delayed condition. The $\Delta\mathrm{AUC}$ (difference in AUC between conditions) was -0.046 (95% CI: -0.181, 0.099), with a medium effect size (Cohen's d = 0.65). Despite a trend towards decreased sensitivity (negative $\Delta\mathrm{AUC}$), we cannot definitively conclude a significant decline in SJ sensitivity after exposure to the delay due to the inclusion of zero within the 95% CI of $\Delta\mathrm{AUC}$.

Similarly, the AUC for the AJ task (middle-right panel, Figure 4d) was 0.81 (95% CI: 0.74, 0.88) under the baseline condition and 0.75 (95% CI: 0.66, 0.82) under the delayed condition. The $\Delta\mathrm{AUC}$ for AJs was -0.068 (95% CI: -0.174, 0.038), with a large effect size (Cohen's d = 1.27). While there is a tendency for decreased sensitivity (negative $\Delta\mathrm{AUC}$), a significant decline in AJ sensitivity after the delay cannot be concluded due to the nonsignificant $\Delta\mathrm{AUC}$ (95% CI containing zero).

The estimated criterion measure b for the SJ task (middle-left panel, Figure 4c) revealed a trend towards a lower value in the delayed condition (1.00, 95% CI: 0.52, 1.51) than in the baseline (1.17, 95% CI: 0.76, 1.62). However, the difference (Δb) between conditions was not statistically significant (-0.17, 95% CI: -0.83, 0.49), with a medium effect size (Cohen's d = 0.52). This suggests no change in the SJ decision criterion following exposure to delayed feedback.

In contrast, the estimated b for the AJ task (middle-left panel, Figure 4d) showed a clear shift towards a lower (more conservative) criterion in the delayed condition (0.73, 95% CI: 0.44, 1.03) compared to baseline (1.37, 95% CI: 0.97, 1.79). The Δb for AJs was statistically significant (-0.64, 95% CI: -1.14, -0.14), with a large effect size (Cohen's d = 2.50). This finding provides strong evidence for a lowered decision criterion in the AJ task after exposure to delayed feedback.

Relationships Between the SDT and Temporal Window Parameters

Figures 5 and 6 present the results of Bayesian correlation analyses between the parameters derived from the SDT model and the temporal window models for the SJ and AJ tasks, respectively. The focus was on the differences in these parameters between the delayed and baseline conditions. Three specific correlations were examined, as detailed in the Data Analysis section:

1. $\Delta\mu$ (shift in the center of the Gaussian window for mean asynchrony) and Δb (change in the SDT criterion measure).
2. $\Delta\mathrm{AUW}^{(G)}$ (change in the size of the Gaussian window for mean asynchrony) and $\Delta\mathrm{AUC}$ (change in the SDT sensitivity measure).
3. $\Delta\mathrm{AUW}^{(H)}$ (change in the size of the half-Gaussian window for the SD of asynchrony) and $\Delta\mathrm{AUC}$ (change in the SDT sensitivity measure).

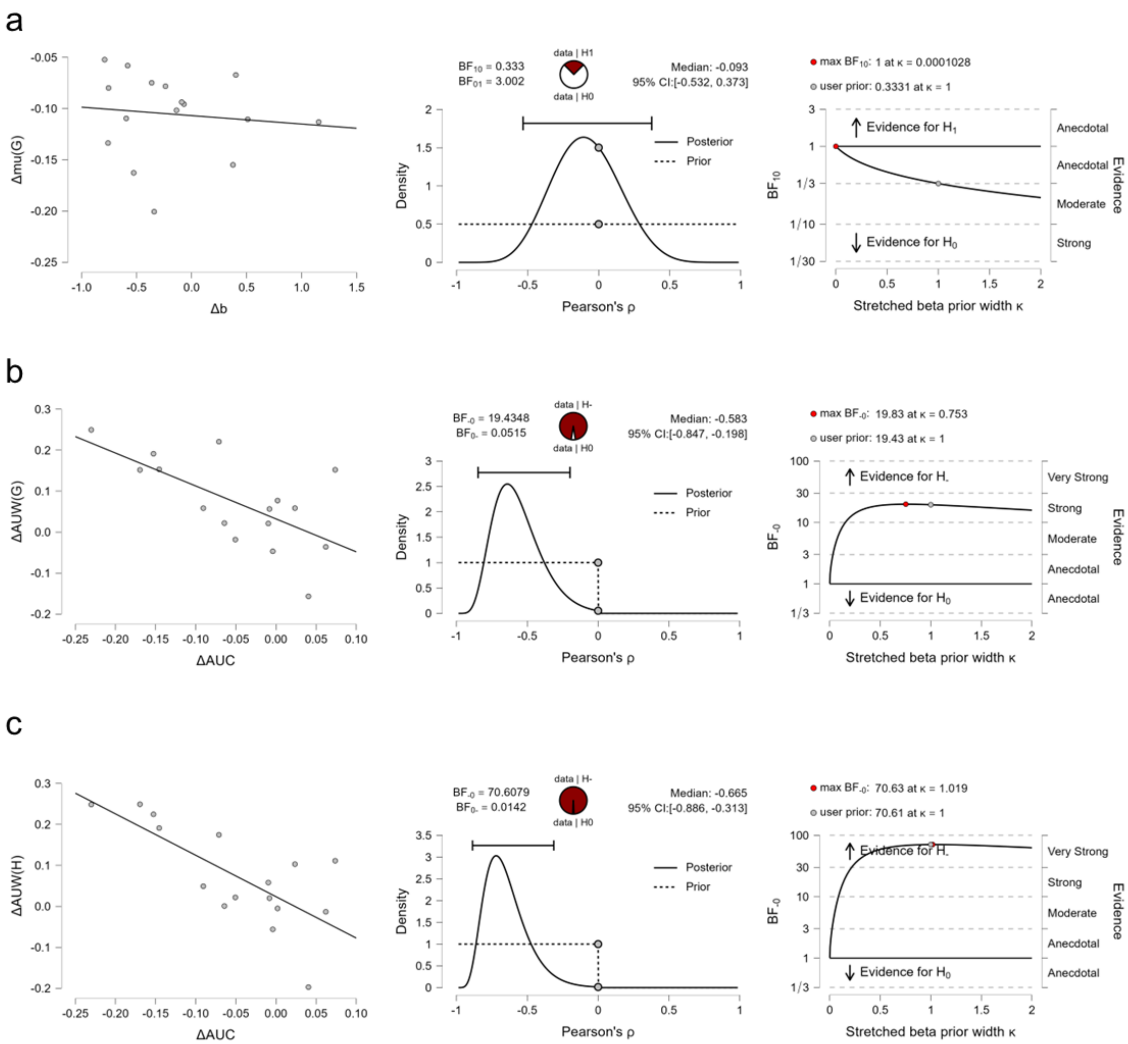


Figure 5. Bayesian correlation analyses: SJ task. This figure shows the results of Bayesian correlation analyses examining the relationships between changes in the SDT parameters (Δb and ΔAUC) and changes in the temporal window parameters ($\Delta\mu^{(G)}$, ΔAUW$^{(G)}$, and ΔAUW$^{(H)}$) for individual participants in the SJ task. **(a)** Correlation between Δb and $\Delta\mu^{(G)}$, **(b)** ΔAUC and ΔAUW$^{(G)}$, and **(c)** ΔAUC and ΔAUW$^{(H)}$. The left panels show the scatter plot showing the relationships between the compared parameters and regression lines. The center panels illustrate the prior and posterior distributions for the correlations are presented. These panels also show the posterior median, 95% credible interval (CI), and Bayes factor (BF). The left panels display the BF robustness plots illustrating how the BF changes across different values of the beta prior parameter. The gray dots represent the chosen prior parameters. A relatively stable alternative hypothesis across a wide range of prior parameter values indicates robust analysis. Figures from JASP (JASP Team, 2022).

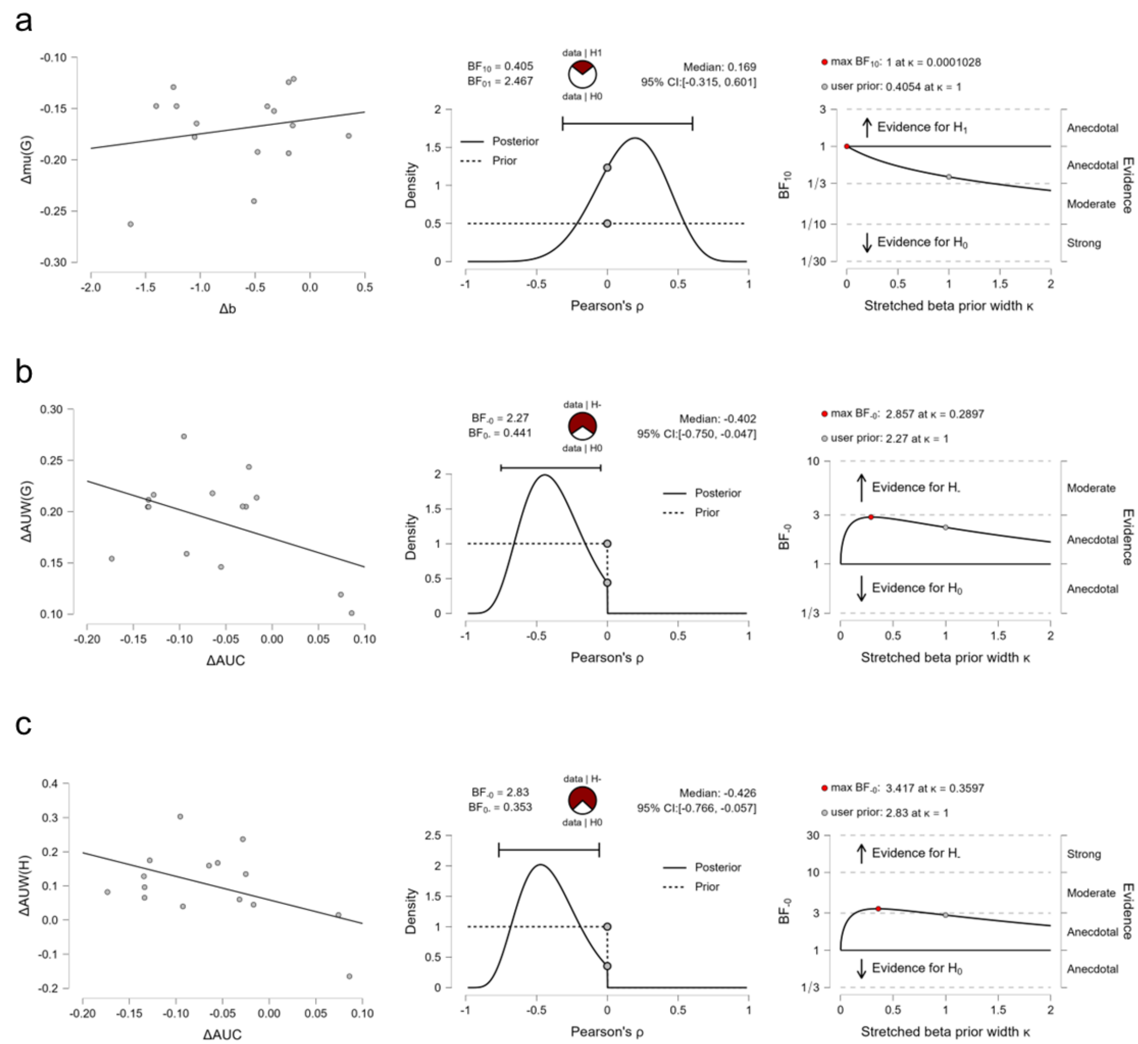


Figure 6. Bayesian correlation analyses: AJ task. This figure shows the results of Bayesian correlation analyses examining the relationships between changes in the SDT parameters (Δb and ΔAUC) and changes in the temporal window parameters ($\Delta\mu^{(G)}$, $\Delta\mathrm{AUW}^{(G)}$, and $\Delta\mathrm{AUW}^{(H)}$) for individual participants in the AJ task. **(a)** Correlation between Δb and $\Delta\mu^{(G)}$, **(b)** $\Delta\mathrm{AUC}$ and $\Delta\mathrm{AUW}^{(G)}$, and **(c)** $\Delta\mathrm{AUC}$ and $\Delta\mathrm{AUW}^{(H)}$. The left panels show the scatter plot showing the relationships between the compared parameters and regression lines. The center panels illustrate the prior and posterior distributions for the correlations are presented. These panels also show the posterior median, 95% credible interval (CI), and Bayes factor (BF). The left panels display the BF robustness plots illustrating how the BF changes across different values of the beta prior parameter. The gray dots represent the chosen prior parameters. A relatively stable alternative hypothesis across a wide range of prior parameter values indicates robust analysis. Figures from JASP (JASP Team, 2022).

Each figure utilizes a combination of panels to present the results:

- Left panels: Scatter plots depicting the relationships between the compared parameters.
- Center panels: Prior and posterior distributions of the correlations. These panels also show the posterior median, 95% credible interval (CI), and Bayes factor (BF).
- Right panels: Bayes factor robustness plots.

For the SJ task (Figure 5), a strong negative correlation emerged between the change in the SDT sensitivity (ΔAUC) and the change in TWS size (ΔAUW$^{(G)}$) (left panel, Figure 5b). The posterior median correlation coefficient (ρ) was -0.583 (95% CI: -0.847, -0.198), and the Bayes factor (BF) supported H^- (negative correlation) over H_0 (no correlation), with a value of 19.4 (center panel, Figure 5b). This evidence for H^- remained robust across a range of prior distribution parameter values (right panel, Figure 5b).

Similarly, a strong negative correlation was found between ΔAUC and the change in the half-Gaussian window size for the SD of asynchrony (ΔAUW$^{(H)}$) (left panel, Figure 5c). The posterior median ρ was -0.665 (95% CI: -0.886, -0.313), and the BF provided very strong evidence for H^- (BF_{-0} = 70.6) (center panel, Figure 5c). This result indicates that the sensitivity to simultaneity increases with a narrower window, which holds true across various prior parameter values (right panel, Figure 5c).

In contrast, no significant correlation was observed between the change in the Gaussian window center ($\Delta\mu$) and the change in the SDT criterion measure (Δb) (left panel, Figure 5a). The posterior median ρ was -0.093 (95% CI: -0.532, 0.373), with a weak BF10 of 0.3, suggesting anecdotal to moderate evidence for H_0 (no correlation) over H_1 (positive or negative correlation) (center panel, Figure 5a). Notably, the strength of evidence for H_0 in this comparison showed some variation depending on the prior distributions (right panel, Figure 5a).

The findings for the AJ task (Figure 6) mirrored those of the SJ task, but with weaker correlations. A moderate negative correlation emerged between ΔAUC and ΔAUW$^{(G)}$ (posterior median ρ = -0.402, 95% CI: -0.750, -0.047), with a

BF_{-0} of 2.27 providing anecdotal to moderate evidence for H^- (negative correlation) over H_0 (no correlation) (left and center panel, Figure 6b). This evidence remained stable across prior distributions (right panel, Figure 6b). Similarly, ΔAUC and ΔAUW$^{(H)}$ showed a moderate negative correlation (ρ = -0.426, 95% CI: -0.766, -0.057; BF_{-0} = 2.83), again suggesting H^- with anecdotal to moderate evidence (left and center panel, Figure 6c). This pattern (sensitivity increasing with a narrower window) held true across the priors (right panel, Figure 6c).

As with the SJ task, no significant correlation was observed between $\Delta\mu$ and Δb in the AJ task (ρ = 0.169, 95% CI: -0.315, 0.601; BF_{10} = 0.4). The data provide anecdotal evidence for H_0 (no correlation) over H_1 (left and center panel, Figure 6a), although the strength of evidence for H_0 varied with prior distributions (right panel, Figure 6a).

Comparison With Sugano (2021)

Tables 1 and 2 summarize the differences in temporal window parameters (Table 1) and SDT parameters (Table 2) between the current study and Sugano (2021). The tables show the changes observed in the delayed condition relative to baseline. Statistically significant changes are denoted by "changed" in boldface ($p < .05$ in Sugano [2021] or a 95% credible interval excluding zero in the present study). Conversely, "no change" indicates nonsignificant differences. Due to methodological discrepancies in the application of SDT (not applicable to SJ data in Sugano [2021]), the corresponding cells in Table 2 are left blank. In Table 2, Cohen's d effect sizes are presented due to variations in the SDT measures used between the studies.

Table 1. Changes in the temporal window parameters (delayed vs. baseline conditions)

	Simultaneity (TWS)		Agency (TWA)	
	center (μ)	width (σ)	center (μ)	width (σ)
Sugano (2021) [a]	**changed (-21.3 ms)**	no change (+4.2 ms)	no change (-7.1 ms)	no change (-7.9 ms)
Present study	**changed (-29.3 ms)**	no change (+7.5 ms)	**changed (-45.2 ms)**	**changed (+25.3 ms)**

[a] First and second session is collapsed.

Table 2. Changes in the SDT parameters (delayed vs. baseline conditions)

	SJ		AJ	
	Sensitivity	Criterion	Sensitivity	Criterion
Sugano (2021) [a]	—	—	no change (+0.05)	**changed (-0.28)**
Present study	no change (-0.65)	no change (-0.52)	no change (-1.27)	**changed (-2.50)**

Note. Values are Cohen's d.

[a] First and second session is collapsed.

A key initial observation from these tables is the larger magnitude of temporal window modulation observed in the present study following the delay-exposure compared to Sugano (2021) (Table 1). This effect appeared stronger for the TWA task. Similarly, the current study suggested a more pronounced modulation of the AJ response criterion (Table 2). The reasons for these discrepancies between the studies will be explored in the Discussion section.

Discussion

Summary of the Results

The objective of this study was to replicate the results of a previous experiment conducted by Sugano (2021) using an online approach. This study aimed to investigate whether changes in the sense of agency (SoA) and temporal recalibration (TR) were due to altered perceptual sensitivity or decision-making criteria. Participants completed a tone reproduction task by pressing the spacebar on a computer, with varying interstimulus intervals (ISIs) and tone numbers. In half of the trials, the computer controlled the tones, and participants judged whether the tones were self-controlled (agency judgment) or synchronized with their keypresses (simultaneity judgment), depending on their assigned group. Prior to the main task, participants completed another tone reproduction task, during which they experienced either delayed tone (110 ms) or a baseline condition with synchronized tone (10 ms). These conditions were blocked, and the order was counterbalanced across participants.

The study addressed the following questions:

1. How does an exposure to delayed sensory feedback affect the TWS and the TWA? These windows represent the perception of the temporal proximity required for two events (an action and its resulting feedback) to be perceived as synchronous or self-generated. Specifically, does the center of these windows shift, and/or does their width change?
2. Does experiencing a delay affect our sensitivity to the variability of timing between our actions and feedback while still being perceived as synchronous or self-generated?
3. Does an exposure to the delay influence our sensitivity (ability to distinguish differences) or decision criteria (threshold for making a judgment) for simultaneity and agency?
4. How are the parameters of the temporal windows related to the parameters used in the SDT for analyzing the SJ and the AJ?
5. Can an online experiment replicate the findings from an in-person experiment?

Predictions were made for each question based on existing research and hypotheses:

1. The center of the TWS and TWA might shift in the direction of the delay, and the width of these windows may or may not increase.
2. Exposure to the delay might make us less sensitive to variations in the timing between our action and feedback.
3. Based on previous research (Sugano, 2021), we might expect that the sensitivity to agency remains constant while the decision criterion changes. For simultaneity, the results are less clear.
4. We predict that the sensitivity is related to the width of the temporal window. However, the relationship between the decision criterion and the shift in the window center is uncertain.
5. We anticipate that the online experiment can replicate the findings of the in-person experiment.

The key findings of the present study are as follows:

1. Prediction 1 was confirmed: The center of the TWS shifted in the direction of the delay, but the window width did not change. The TWA center also shifted, and its width increased.
2. Prediction 2 was not confirmed: Sensitivity to timing variations did not change between the delayed and baseline conditions in a statistically reliable manner. However, it appeared to decrease after exposure to the delay for both the SJ and the AJ, with large effect sizes.

3. Prediction 3 was confirmed: The decision criterion for agency judgment shifted, suggesting that participants became more conservative in attributing actions to themselves. The sensitivity to agency remained unchanged. Neither sensitivity nor the decision criterion changed for simultaneity judgment.
4. Prediction 4 was partly confirmed: Sensitivity in the SDT correlated with the width of the temporal window for simultaneity judgment but not for agency judgment. The decision criterion did not correlate with the window center.
5. Prediction 5 was confirmed: The online experiment replicated the main findings of the in-person experiment, with some differences in the magnitude of the TWA shift and width changes.

The present study successfully replicated key findings from the in-person experiment (Sugano, 2021), suggesting that experiencing a slight delay in sensory feedback can affect our perception of time and our sense of agency, highlighting the dynamic nature of these experiences. However, there are several discrepancies between studies. I will discuss them in the next section.

Explaining Discrepancies with Sugano (2021)

This section explores potential explanations for the discrepancies observed between the current online experiment and the findings of Sugano (2021). As detailed in the Results section, the present study revealed a larger modulation of the temporal window and a more pronounced shift in the AJ response criterion compared to Sugano (2021). Here, I will systematically analyze potential factors contributing to these differences, followed by a discussion of less likely explanations.

Factors That Could Have Caused the Discrepancies

Two potential factors could have caused these discrepancies. The first factor was the way and the order of task execution. In the present study, participants performed either the SJ or the AJ depending on the group. However, in Sugano (2021), participants performed both the simultaneity judgment (SJ) and the agency judgment (AJ) sequentially within each trial, with the AJ consistently preceding the SJ. This sequential execution of the tasks might have led to an assimilation effect, where participants' performance on one judgment was influenced by the subsequent judgment and became similar to each other (e.g., Helson, 1964). This possibility is supported by Sugano's (2021) finding that nearly 39% of participants exhibited a high degree of response similarity between the SJ and AJ, suggesting a potential bias towards judging the AJ identical to the SJ, and vice versa. Separating the two judgments could free them from the constraint that binds them together. This could lead to a more dynamic modulation of the TWA.

The second factor was that the number of delayed tones presented during the adaptation phase varied between the studies. Sugano (2021) used 4-7 tones per trial, but the current study used 7-9 tones per trial. This resulted in an approximately 10% increase in the total number of delayed tones during adaptation in the present study (144 tones) compared to Sugano (2021) (132 tones). Consequently, the present study potentially induced a greater degree of adaptation, which could explain the observed stronger dynamic modulation of both the TWS and the TWA.

Factors That May Be Irrelevant for the Observed Differences

Four potential factors may contribute to these discrepancies, but they are unlikely to be relevant. Firstly, the age of the participants may have affected the results. In the present study, the average age of the employed participants was higher than that in Sugano (2021). Previous research has suggested that older individuals may be less susceptible to temporal recalibration (TR) (Chan et al., 2014). However, in the current study, more pronounced changes were observed between the baseline and delayed conditions. Specifically, greater changes were observed in the center and width of the temporal window of apparent motion (TWA) and the center of the temporal window of simultaneity (TWS). Therefore, while participant age variation may have played a role, the significant changes in TWS and TWA observed in this study cannot be solely attributed to the age-related findings of Chan et al. (2014) and likely had minimal impact on the overall results.

Secondly, although the potential influence of personal computer (PC) hardware and software on the findings warrants consideration, it is unlikely to have significantly impacted the results of the present study. Sugano (2021) meticulously ensured temporal accuracy and stability by isolating PCs from the internet, minimizing background processes, and employing dedicated verification software. Conversely, due to unavailable data, the specifications of the PCs used by participants in the present study remain undetermined. It is, however, reasonable to assume they were standard machines, likely incapable of the same level of temporal precision as the dedicated setup in Sugano (2021). Additionally, potential variations in these PC specifications (although unrecorded) could introduce noise into the timing data, potentially obscuring differences between experimental conditions. Interestingly, despite these potential limitations, the present study revealed a greater difference between conditions compared to Sugano (2021). This suggests that any potential influence of PC differences on the results was likely minimal, as the observed effect contradicts the anticipated direction.

Additionally, the potential influence of the response device (mouse vs. keyboard) on the findings warrants investigation. While both devices introduce inherent latencies (Shimizu, 2002; Li et al., 2010; Khitrov et al., 2013), even those under 100 ms can significantly impact human–machine interactions (Jota et al., 2013; Doherty & Sorenson, 2015; Attig et al., 2017). Consequently, it would be logical to assume a potential influence on the SJ or AJ. However, I posit that such an effect was likely negligible in the present study due to the following considerations. Input latency is the time elapsed between the keypress and device detection. Studies have shown that keyboards, particularly those with high-profile keycaps (e.g., they have additional time to contact from the moment of keypress), may have longer input latency than mice (Plant et al., 2003; Bockes et al., 2018; Wimmer et al., 2019). This potential prolongation could introduce ambiguity regarding the precise timing of key press initiation or termination, potentially attenuating the magnitude of observed temporal recalibration (TR). This is particularly relevant considering the reliance of motor-sensory TR induction on tactile feedback rather than intention (Arnold et al., 2012). Interestingly, despite this potential influence, the present study revealed a greater magnitude of TR (reflected in the difference in the TWS centers between conditions) than Sugano (2021). This suggests that any potential effect of the response device variations was likely minimal, as the observed results contradict the anticipated direction.

Finally, the data screening criteria may have affected the results. As detailed in the Data Analysis section, the current study employed a slightly more lenient approach to outlier data screening than Sugano (2021). This modification aimed to achieve comparable data exclusion rates between the two studies. To investigate the potential influence of this methodological difference on the findings, a secondary analysis employing Sugano's (2021) screening criteria was conducted on the present data. Notably, the results and conclusions from this analysis were consistent with those obtained using the criteria of the current study. The modulation of the temporal window was larger than that of the data screened with the current criteria. For example, the difference in mu was -40.3 ms for the TWS and -75.8 ms for the TWA. Additionally, the difference in sigma for the TWA was +48.7 ms.

Distinct Psychological Processes Underlying SJ and AJ?

The identical stimuli used in the SJ and AJ tasks, coupled with the observed differences in the results, suggest the potential involvement of distinct psychological processes. As previously discussed, adaptation can be categorized as either PSE-shifting or sensitivity-changing. Although both processes likely influence both SJ and AJ, their relative contributions might differ.

Further supporting this notion, Sugano (2021) proposed that the observed divergence between the TWS and TWA might be attributed to variations in how participants weight different cues during judgments. The two-stage account of the SoA (Synofzik et al., 2008) and the Bayesian cue integration framework (Moore & Fletcher, 2012) propose the use of various cues, including contextual (cause–effect), sensorimotor (perceived simultaneity), internal (motoric signals), and external (sensory inputs) cues. If the impact of these cues differs depending on the judgment type (SJ vs. AJ), the modulation of the TWS and TWA may also diverge.

Finally, it is important to consider alternative explanations beyond distinct mechanisms. The observed effects could reflect a single mechanism operating with different time courses for SJ and AJ.

Both the PSE-Shifting and the Sensitivity-Changing Might Work in SJs and AJs

The temporal window analysis for the SJ revealed a change in the center of the TWS (μ) but no change in width (AUW) across the TRs. Similarly, the SDT analysis of the SJ showed no change in sensitivity (AUC). These converging findings suggest that sensitivity to action–feedback simultaneity remained constant, potentially reflecting a PSE-shifting type of modulation in the TWS. However, the possibility of altered sensitivity cannot be entirely ruled out. Several studies have demonstrated the ability of delayed sensory feedback to influence sensitivity in simultaneity or temporal order judgments. For instance, Navarra et al. (2005, 2007) and Vatakis et al. (2008) investigated audiovisual TR, while Winter et al. (2008), Keetels & Vroomen (2012), and Sugano et al. (2014, 2016, 2021) focused on sensorimotor TR. Despite the lack of statistically significant changes in TWS width (AUW), the observed effect size (Cohen's d = 0.46) suggests a possible small-to-medium effect according to Cohen (1988, 1992). This raises the possibility that sensitivity in the TWS task may have been modulated by delayed feedback, but the effect size might be too subtle for detection with the current sample size. An alternative explanation is that sensitivity might have changed transiently following exposure to the delay but then returned to baseline as adaptation progressed during the experiment. This rapid adaptation process could have led to an underestimation of the true sensitivity change in TWS. Considering these possibilities, a more nuanced interpretation suggests that the observed modulation in TWS might be a combination of both PSE-shifting and sensitivity-changing mechanisms.

While the temporal window analysis for the AJ revealed changes in both center (μ) and width (AUW), the SDT analysis of the AJ did not show a significant change in sensitivity (AUC). This apparent discrepancy between the two measures warrants further discussion. One possibility is that a true change in AJ sensitivity might be too subtle for detection with the current sample size. Although statistically nonsignificant, the observed trend towards a decrease in ΔAUC (-0.068) with a large effect size (Cohen's d = 1.27) suggested a potential decrease in sensitivity. This finding, along with the changes in TWA parameters, points towards a possible hybrid modulation of the TWA with the TR, encompassing both PSE-shifting and sensitivity-changing mechanisms.

Differential Time Course of Adaptation?

While both tasks modulated the temporal window, the pattern differed. The change in TWA width suggested a more dynamic response to delayed feedback compared to TWS. Similarly, the SDT analysis revealed a larger effect size for the AJ sensitivity measure than for the SJ, potentially reflecting a greater impact of delay exposure on the AJ. These findings suggest a more dynamic and potentially more vulnerable response to feedback delay in the AJ than in the SJ. This contradicts Sugano's (2021) results, which might be attributed to procedural differences between the studies.

In the previous section, I discussed the potential influence of separating judgments and the increased number of trials in the present study. Although this could explain the larger modulation in AJs, it does not account for the smaller modulation observed in SJs compared to Sugano's (2021) findings.

This discrepancy may indicate a difference in the time course of adaptation between the TWS and TWA tasks. Navarra et al. (2005, 2007) proposed a two-stage process: initial sensitivity reduction followed by a PSE shift leading to restored sensitivity. This model aligns with the idea that prioritizing PSE shifts for survival is advantageous. If the SJ adaptation process is completed faster than the AJ process, this could explain why the SJ exhibits less overall modulation in the present study.

Future studies could investigate this hypothesis by examining the time course of temporal window changes. However, extending the adaptation time risks participant fatigue and inattention, particularly in online experiments. Maintaining participant focus during extended tasks remains a challenge, regardless of the online or in-person format. Exploring methods to address this issue is crucial for future research.

Limitations: Between-Subjects Design and Online Environment

The present study employed a between-subjects design, assigning participants to either the SJ or AJ task. This design raises the possibility that the observed differences between the TWS and TWA might be attributable to group-specific characteristics rather than the judgment type itself. Although the study implemented random assignment and found no significant baseline differences in demographics (age, sex, or handedness), a potential limitation remains. The online nature of the experiment limited our ability to control the environment in which participants completed the tasks, introducing potential variability between groups. Future research addressing this limitation could utilize a within-subjects design where participants perform both SJ and AJ tasks, manipulating feedback delay as the critical factor. This approach would enable a more conclusive determination of whether the observed effects are specific to judgment type or generalizable across tasks.

The online nature of the study introduced two additional limitations. First, the data exclusion rate was slightly greater than that of Sugano's (2021) in-person study. This could be attributed to factors such as minor instruction violations, excessive errors, or consistently biased responses. While measures were taken to mitigate these issues, future studies can benefit from referencing the established mitigation strategies outlined in Stewart et al. (2017), Majima (2019), and Kuroki (2020). Second, online experiments face the challenge of potential variability in timing precision across participants' personal computers (PCs). Unlike controlled laboratory settings, the experimenter cannot directly verify individual PC timing accuracy. Employing within-subjects designs, as previously mentioned, helps to cancel out potential timing errors between conditions (e.g., Kuroki, 2020).

Conclusion

In conclusion, this online study has successfully replicated key findings from the in-person experiment (Sugano, 2021), demonstrating the potential for conducting online studies requiring precise timing control and high temporal accuracy. A longstanding critique of laboratory research is its limited generalizability due to participant demographics, which are often restricted to university students (Majima, 2019; Kuroki, 2020). Online experiments offer a distinct advantage by enabling the recruitment of participants with broader age ranges, as demonstrated in this study. The successful replication of findings with a more diverse sample strengthens the generalizability of the observed phenomenon.

Materials and Methods

Participants

Number of Participants

Forty-three participants (17 females, mean age 39.3 ± 8.1 years, 2 left-handed) participated in the experiment, which was conducted from December 2021 to January 2022. Participants were assigned to either the agency judgment (AJ; n = 20) or simultaneity judgment (SJ; n = 23) task.

Sample Size Justification

The sample size was determined based on two considerations. First, it adhered to conventions for standard psychophysical experiments (e.g., ~20 participants; Klatzky et al., 2017). Second, a power analysis was conducted using R (version 4.1.3; R Core Team, 2022) to estimate the minimum number of participants needed to detect the effect of feedback delay on the SDT criterion measure (log(β)) from the previous study (Sugano, 2021). The analysis used a two-sided significance level of 0.05, a power of 0.80 (following Cohen's, 1988, 1992, guidelines), and assumed a mean difference of 0.50 and a standard deviation of 0.64 in log(β) between the baseline and delayed conditions (data from Sugano, 2021, p. 14; open data available at https://doi.org/10.6084/m9.figshare.12607697). This analysis indicated a minimum of 12 participants per condition. As the conditions were manipulated as a between-subjects factor in the design (see the Design and Procedure section), at least 24 participants were needed to achieve adequate statistical power.

Participant Demographics and Exclusion Criteria

All participants reported normal hearing and normal or corrected-to-normal vision. Forty-one were right-handed and two were left-handed. Informed consent was obtained online. The study was approved by the Kyushu Sangyo University ethics committee (2021-0016) and adhered to the Declaration of Helsinki.

Twelve participants were excluded from the final analysis due to (1) failing the instructional manipulation check (IMC; 4 participants, see the Design and Procedure section), (2) excessive retrials of “not good” trials (1 participant), (3) a streak of missing responses across three consecutive trials (suggesting potential deliberate manipulation; 1 participant), (4) extreme judgment bias (answering “self-controlled” or “computer-controlled” for most trials; 2 participants), or (5) program coding errors leading to unbalanced self-controlled and computer-controlled trials (6 participants; see the Coding Error and Data Retention section). Two participants met criteria from multiple categories.

Coding Error and Data Retention

A coding error in the program resulted in uneven presentation frequencies across the conditions for 23 participants. Some conditions were presented more than twice, while others were presented only once or not at all. This error was corrected for the remaining 20 participants. To maximize data retention for the 23 affected participants, data were included if the number of self-controlled and computer-controlled trials differed by no more than ±2 for a given participant. Based on these criteria, the data of six participants were excluded due to an excessive imbalance in trial types.

The final analysis included data from 31 participants (13 females, mean age 38.7 ± 7.9 years, 1 left-handed). The AJ task included 15 participants (mean age 37.9 ± 9.0 years, all right-handed), and the SJ task included 16 participants (mean age 39.4 ± 7.0 years, 1 left-handed). This final sample size met the power analysis criterion (>12 participants per condition) and was comparable to that of the previous study (Sugano, 2021; n = 18).

Apparatus and Stimuli

Apparatus, Stimuli, Response Collection and Masking

Participants completed the experiment on their personal computers while wearing headphones for auditory stimulus presentation. The auditory stimulus was a 2000 Hz pure tone pip with a 30 ms duration and 2 ms rise/fall ramps. Inquisit Player version 6.4.2 (Millisecond Inc.) running on participants' computers controlled the experiment.

Participants pressed the spacebar key on their keyboards to trigger tone presentation. White noise was continuously presented to mask the faint sound of keypresses. The volume of both the tone and noise was adjustable by participants to a comfortable level. On average, participants set the noise level 8.93 dB lower than the tone level (SD = 1.32 dB, range: 0 dB to -66 dB).

Timing Verification and Rationale for Delay Values

Direct monitoring and recording of the precise timing of both tone output and participant keypresses were not possible due to the online nature of the experiment. However, the experimenter's computer employed a multitrace oscilloscope (PicoScope 2203, Pico Technology Ltd.) to verify the timing of these events (details in the Design and Procedure section).

The feedback delays for the tone were set to either 10 ms (baseline) or 110 ms (delayed) (see the Design and Procedure for details). These delays were shorter than those used in the previous study (80 ms vs. 180 ms; Sugano, 2021). This difference was implemented because, compared to Sugano (2021), the present study involved an anticipated additional latency for sound output due to potential factors such as internet connections, other running programs, or background services on participants' personal computers.

To confirm the actual sound output latencies, the author used a nonoptimized notebook type personal computer (Acer Swift 1 SF114-32-H14Q/S) for testing. The observed baseline latency (delay set at 10 ms) was 74.9 ± 7.2 ms (n = 17), and the observed delayed latency (delay set at 110 ms) was 175.9 ± 7.1 ms (n = 13). These latencies were comparable to those observed in the in-person laboratory experiment with an optimized environment in Sugano (2021): 80.3 ± 0.7 ms (baseline; n = 30) and 180.4 ± 0.6 ms (delayed; n = 30) (not reported in paper). It is important to acknowledge that the timing precision of participants' personal computers remains unknown, and these values serve as a reference point only.

Design and Procedure

Design

The experiment employed a mixed factorial design with one between-subjects factor and four within-subjects factors. The between-subjects factor was judgment type (simultaneity judgment [SJ] vs. agency judgment [AJ]). The four within-subjects factors were feedback delay (baseline vs. delayed), agency over tone output (self-controlled vs.

computer-controlled), interstimulus interval (ISI) of the tone sequence (500, 600, and 700 ms), and number of tones (7, 8, and 9). This combination resulted in 36 unique conditions, each presented twice for a total of 72 trials.

Procedure

The experiment consisted of two blocks (baseline and delayed) with constant feedback delay within each block (10 ms or 110 ms). Each block comprised 36 trials formed by two sets of 18 unique conditions presented in random order. These conditions included all possible combinations of the within-subjects factors (agency, ISI, and number of tones).

The order of the baseline and delayed blocks was counterbalanced across participants. The blocks were separated for at least 5 minutes rest period to minimize carry-over effects. The entire experiment, including instructions, practice and main sessions, and postexperimental procedures (e.g., finishing code registration), lasted approximately 50 minutes.

A schematic illustration of a single trial is presented in Figure 7. The overall procedure was similar to that of Sugano (2021) with some modifications. To avoid task interaction, the AJ and SJ tasks were performed by separate participant groups. The SJ was accompanied by a confidence rating, as was the AJ. Considering the potential difficulties for online participants compared to university students, the task difficulty was reduced by increasing the number of presented tones. The response device was changed from a computer mouse to a keyboard due to the assumption that participants might not have a mouse readily available.

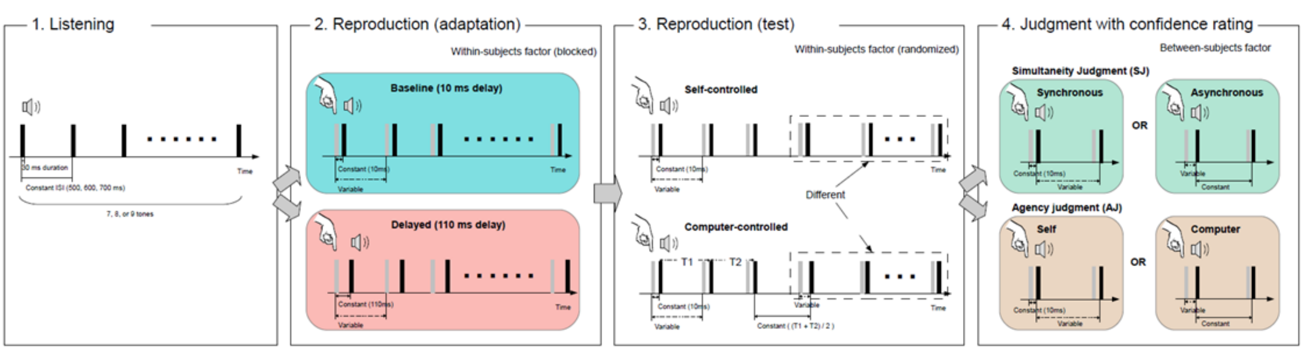


Figure 7. Schematic illustration of a single trial. The experiment comprised three sections: listening, reproduction

(adaptation), and reproduction (test). During listening, participants memorized a standard sequence of tones with a specific interstimulus interval (ISI). In the reproduction (adaptation) section, participants reproduced the standard sequence by pressing the spacebar key. The delay of the tone output after the keypress was manipulated to be either 10 ms (baseline) or 110 ms (delayed). Finally, the reproduction (test) section required participants to reproduce the sequence again. However, in this section, their keypresses, except for the first three, were ignored with a 50% chance, and the tones were presented at equal intervals, which used the average interstimulus interval (ISI) based on the first three keypresses. Following reproduction, participants in the agency judgment group judged whether the tones were presented by the computer, while those in the simultaneity judgment group judged whether the tones were presented in synchrony with their keypresses. Each judgment was combined with their confidence rating.

Before the experiment, participants provided demographic information (age, gender, handedness, visual/auditory abilities) and received general task instructions and precautions. They then completed checklists to confirm adequate headphones (soundproof) and keyboards (quiet). An instructional manipulation check (IMC) verified participants' understanding of the instructions by asking them to answer “No” to irrelevant task-related statements. Following this, participants adjusted the tone and noise volume to a comfortable level.

Participants completed a brief practice session (nine randomly selected trials from the baseline condition's 18 unique conditions) to familiarize themselves with the tasks. These trials mirrored those in the main session but included performance feedback (tapping stability and accuracy of their judgment about simultaneity or agency).

The main session consisted of three sections: listening, reproduction (adaptation), and reproduction (test). During the listening section, participants were presented with a standard sequence of equal-interval tones for later reproduction. They were required to remember the number of tones (7, 8, or 9) and the ISI (500, 600, or 700 ms).

The reproduction (adaptation) section followed, where participants reproduced the standard tone sequence from the listening section by pressing the spacebar. In this section, the tone timing was manipulated. In the baseline condition, the tone was presented 10 ms after the keypress. For the delayed condition, it was presented 110 ms after the keypress.

This section allowed participants to adapt to the feedback delay (100 ms; approximately 3 minutes with 288 delayed tones), as research suggests that adaptation can occur with exposure to delayed feedback for a few minutes and a few hundred repetitions (Heron et al., 2009; Sugano et al., 2010).

Finally, the reproduction (test) section required participants to reproduce the standard tone sequence again. Here, the tone delay was set to 10 ms after the keypress in both the baseline and delayed conditions. Critically, in half of the trials, the computer ignored the participant's keypress and presented the tones at equal intervals from the fourth to the last tone (computer-controlled trial). The tone interval was set to the average interval of the first three participant-initiated tones. In the other half of the trials, the tone output was controlled by the participant's keypress (self-controlled trial). The computer-controlled and self-controlled trials were presented in a pseudorandom order.

For both self-controlled and computer-controlled trials, participants were required to repeat the trial if the standard deviation of the intertap interval (time between keypresses) exceeded 200 ms. In computer-controlled trials, additional repetition criteria included a standard deviation of asynchrony exceeding 120 ms (time difference between keypress and tone), missing taps, or extra taps.

Following the reproduction (test) section, participants rated their combined judgment and confidence level on a unique 6-point Likert scale presented on the computer screen. This scale assessed either agency judgment (AJ) or simultaneity judgment (SJ), depending on the participant's group assignment. Agency judgment (AJ) pertained to the participant's perception of who controlled the tone output—either by themselves or the computer. Simultaneity judgment (SJ) focused on whether the participant perceived their keypress and the tone as synchronous (occurring at the same time) or asynchronous (not occurring at the same time). The scale structure reflected a combined judgment, ranging from high confidence in one judgment (right side) to high confidence in the other judgment (left side). Intermediate points on the scale captured varying levels of confidence between the judgments. Critically, the order of the labels was counterbalanced across participants to control for potential order effects associated with scale presentation.

Data Analysis

Overview

Practice data were excluded from the analysis. Only data from the reproduction (test) section were analyzed. These data included:

1. Asynchrony: Time difference between the participant's keypress onset and computer-controlled tone onset. Negative values indicated keypresses preceding the tone.
2. Simultaneity Judgment (SJ): Binary judgment (0 = asynchronous, 1 = synchronous) with confidence levels (1-3).
3. Agency Judgment (AJ): Binary judgment (0 = computer, 1 = self) with confidence levels (1-3).

The mean and standard deviation (SD) of asynchrony in computer-controlled trials were calculated for each trial and participant. The first three keypresses were omitted because they were self-controlled.

The participants' subjective rating score for simultaneity or agency (1-6) was calculated by combining their binary judgment (SJ or AJ) and confidence rating (1-3). For example, "asynchronous" with confidence "1" resulted in a rating value of "1" (least confident asynchronous). Conversely, "synchronous" with confidence "3" resulted in a rating value of "6" (most confident synchronous).

Trials were included in the analysis if they met all the following criteria:

1. There were no missing taps in computer-controlled trials.
2. The ratio of the observed intertap interval (ITI) to the interstimulus interval (ISI) of the tone sequence was between 0.8 and 1.25.
3. Mean ITI ranged within ±100 ms from the ISI of tones.
4. The standard deviation of the ITI is less than 100 ms.
5. The mean asynchrony between the participants' tap and the tone onset ranged from -200 to 70 ms in computer-controlled trials.
6. The standard deviation of asynchrony was less than 80 ms in computer-controlled trials.

The other trials were excluded from the analysis.

These criteria were nearly identical to those of Sugano (2021), who adapted the original criteria from Knoblich and Repp (2009). However, criteria #5 and #6 were more lenient than Sugano (2021) in achieving an average exclusion rate (5.0% per participant) comparable to that of the study. This suggests that online participants exhibited slightly less stability in maintaining consistent ITI between keypresses than did in-person participants. After the data were screened, an average of 4.4 trials (6.2%) per participant were excluded.

The data analysis consisted of three steps:

1. Relationship between asynchrony and judgments: A hierarchical Bayesian statistical model was used to examine the relationship between mean asynchrony and participant's judgments (SJ or AJ) to assess how the temporal window of simultaneity (TWS) and agency (TWA) changed after exposure to the sensory feedback delay. This model estimates parameters at the individual and group levels, allowing group-level distributions to constrain individual estimates (Gelman et al., 2014). Additionally, the relationship between the standard deviation (SD) of asynchrony and the judgments was analyzed, as asynchrony fluctuations are known to be another sensorimotor cue for SJs and AJs (Knoblich & Repp, 2009; Sugano, 2021).
2. Signal detection theory (SDT) analysis of judgments: A hierarchical Bayesian statistical model based on SDT (Green & Swets, 1966) was used to analyze SJ and AJ. This model accounts for potential unequal variances between signal and noise distributions (Selker et al., 2019; Greene & Rhodes, 2022). The analysis aimed to determine how sensitivity to simultaneity or agency and the response criterion of judgment changed after exposure to the sensory feedback delay. Participant judgments were categorized as hit, false alarm, miss, or correct rejection based on their response (synchronous/self vs. asynchronous/computer) and trial type (self-controlled vs. computer-controlled) (see Table 3 for details).
3. Bayesian correlation analysis: This analysis examined the relationship between parameters defining the temporal window (shape and location) and the SDT parameters (criterion and sensitivity) for each participant.

Table 3. Response categorization for the signal detection theory (SDT) analysis.

	Response

		Self/Synchronous	Computer/Asynchronous
Trial	Self-controlled	correct rejection (CR)	false alarm (FA)
	Computer-controlled	miss (M)	hit (H)

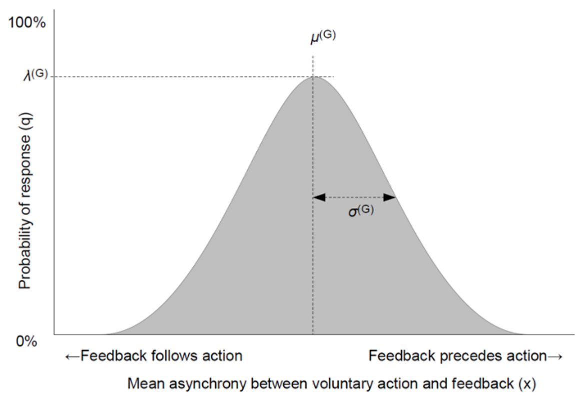
Bayesian statistical modeling was performed using the R environment version 4.1.3 (R Core Team, 2022) with RStan version 2.21.8 (Stan Development Team, 2023): the R interface to Stan (Carpenter et al., 2017). Bayesian correlation analysis was performed using JASP version 0.16.3 (JASP Team, 2022).

Estimating Temporal Windows of Simultaneity (TWS) and Agency (TWA)

To estimate how exposure to sensory feedback delay modulates the TWS and TWA, a Gaussian temporal window model was applied to the relationship between participants' judgments (SJ or AJ) and mean asynchrony in computer-controlled trials (Rohde et al., 2014a, 2014b; Sugano, 2021; Timm et al. 2014; Yarrow et al., 2011, 2013; van Eijk et al., 2008). This model, widely used in psychophysics, assumes that participants perceive the relationship between their action (keypress) and the sensory feedback (tone) as synchronous if the onsets fall within a specific temporal window with a Gaussian-like shape (TWS). Similarly, agency judgments are modeled as a Gaussian-like temporal window (TWA).

A hierarchical Bayesian statistical model was implemented. This model allows individual-level parameters (specific to each participant) to be constrained by group-level hyperparameters (representing the population distribution of individual parameters) (Greene & Rhodes, 2022). Figure 8 schematically illustrates the model (Figure 8a) using plate notation (Figure 8b) (Lee, 2008; Selker et al., 2019; Toyoda, 2017). The observed variables are shaded nodes, the unobserved variables are unshaded, the discrete variables are squares, the continuous variables are circles, the stochastic variables have single borders, and the deterministic variables have double borders (Figure 8b).

a



b

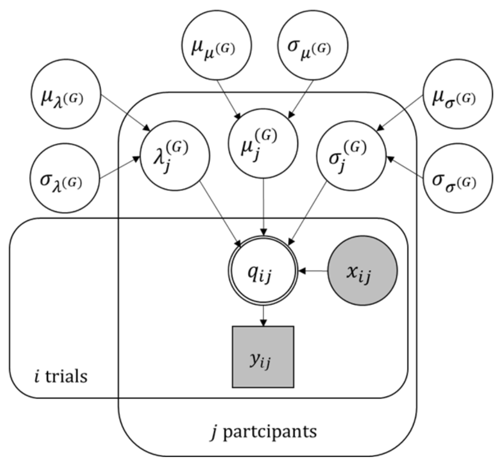


Figure 8. Hierarchical Gaussian model for simultaneity and agency judgments. **(a)** Schematic illustration: This panel depicts a Gaussian temporal window model for simultaneity judgment (SJ) or agency judgment (AJ) as a function of mean asynchrony. **(b)** Plate notation: This panel presents the hierarchical extension of the Gaussian model for SJ or AJ using plate notation. Shaded nodes represent observed variables, unshaded nodes represent unobserved variables, squares denote discrete variables, circles denote continuous variables, single-bordered nodes indicate stochastic variables, and double-bordered nodes represent deterministic variables.

The j-th participant's binary judgment (SJ: sync/async or AJ: self/computer) for the i-th trial (y_{ij}) is modeled as a Bernoulli variable with parameter q_{ij}. This parameter reflects the participant's hypothetical probability of responding "sync" (SJ) or "self" (AJ). It is determined by a Gaussian function of the mean asynchrony (x_{ij}), characterized by three parameters: midpoint ($\mu_j^{(G)}$), width ($\sigma_j^{(G)}$), and height ($\lambda_j^{(G)}$) of the curve. To improve the efficiency of Markov chain Monte Carlo (MCMC) sampling, the mean asynchrony (x_{ij}), which was originally restricted between -200 ms and 70 ms, was rescaled to a range from 0.0 to 1.0 (Reparameterization: Gelman, 2004; Matsuura, 2016; Stan Development Team, 2024).

$$y_{ij} \sim \text{Bernoulli}\left(q_{ij}\right). \tag{1}$$

$$q_{ij} = \lambda_j^{(G)} \times \exp\left[-\frac{1}{2} \times \left(\frac{x_{ij} - \mu_j^{(G)}}{\sigma_j^{(G)}}\right)^2\right]. \tag{2}$$

Individual-level parameters ($\lambda_j^{(G)}$, $\mu_j^{(G)}$, and $\sigma_j^{(G)}$) were assumed to follow a normal distribution with two group-level hyperparameters: mean ($\mu_{\lambda^{(G)}}$, $\mu_{\mu^{(G)}}$, and $\mu_{\sigma^{(G)}}$) and standard deviation ($\sigma_{\lambda^{(G)}}$, $\sigma_{\mu^{(G)}}$, and $\sigma_{\sigma^{(G)}}$).

$$\begin{aligned} \lambda_j^{(G)} &\sim \text{Normal}\left(\mu_{\lambda^{(G)}}, \sigma_{\lambda^{(G)}}\right) \\ \mu_j^{(G)} &\sim \text{Normal}\left(\mu_{\mu^{(G)}}, \sigma_{\mu^{(G)}}\right) \\ \sigma_j^{(G)} &\sim \text{Normal}\left(\mu_{\sigma^{(G)}}, \sigma_{\sigma^{(G)}}\right). \end{aligned} \tag{3}$$

Informed by previous research on the TWS and the TWA (Heron et al., 2009; Rohde et al., 2014b; Sugano et al., 2010; Sugano, 2021; Timm et al., 2014), the group-level mean parameter ($\mu_{\mu^{(G)}}$) was restricted to the range of observed mean asynchrony, assuming that the optimal probability of simultaneity or agency judgments occurs within this range for all participants. The group-level width parameter ($\mu_{\sigma^{(G)}}$) had a lower bound of zero but no upper limit. The group-level height parameter ($\mu_{\lambda^{(G)}}$) was constrained to a range of 0 to 1, as it represents a rate.

For the group-level mean parameters ($\mu_{\lambda^{(G)}}$, $\mu_{\mu^{(G)}}$, and $\mu_{\sigma^{(G)}}$), a normal distribution was used as a weakly informative prior for these parameters. The mean of the prior distribution was set to the average of the minimum (i.e., 0.0) and maximum (i.e., 1.0) observed values in the data (i.e., 0.5), and the standard deviation was set to half the difference between these values (i.e., 0.5). Note that because $\mu_{\sigma^{(G)}}$ does not have an upper limit, the mean and the standard deviation (SD) were set to the same values as $\mu_{\mu^{(G)}}$.

$$\mu_{\lambda^{(G)}} \sim \text{Normal}(0.5,\ 0.5)$$
$$\mu_{\mu^{(G)}} \sim \text{Normal}(0.5,\ 0.5)$$
$$\mu_{\sigma^{(G)}} \sim \text{Normal}(0.5,\ 0.5). \tag{4}$$

For the group-level standard deviation parameters ($\sigma_{\lambda^{(G)}}$, $\sigma_{\mu^{(G)}}$, and $\sigma_{\sigma^{(G)}}$), a half-Cauchy distribution was chosen as a weakly informative prior due to its robustness to outliers (Gelman, 2006; Greene & Rhodes, 2022; Matsuura, 2016). The half-Cauchy distribution is a variant of the Cauchy distribution constrained to positive values, which aligns with the inherent nonnegativity of variance parameters. The location parameter of the half-Cauchy priors was set to zero, reflecting an absence of prior bias towards specific variances. The scale parameter was set to half the range of the observed data (i.e., 0.5), which is a reasonable setting as a weakly informative prior distribution.

$$\sigma_{\lambda^{(G)}} \sim \text{Cauchy}^{+}(0,\ 0.5)$$
$$\sigma_{\mu^{(G)}} \sim \text{Cauchy}^{+}(0,\ 0.5)$$
$$\sigma_{\sigma^{(G)}} \sim \text{Cauchy}^{+}(0,\ 0.5). \tag{5}$$

Investigating the Impact of Delay Exposure on Sensitivity to Asynchrony Variability

The analysis focused on how exposure to a delay influences participants' sensitivity to the variability (standard deviation) of temporal asynchrony between action and feedback in simultaneity judgment (SJ) and agency judgment (AJ). A half-Gaussian window model was employed to analyze the relationship between SJ/AJ and asynchrony variability in a computer-controlled trial. This model assumes that the probability of a “synchronous” or “self” judgment is highest at zero variability and progressively decreases as asynchrony variability increases.

The model was implemented as a hierarchical Bayesian statistical model, as shown in Figure 9. The j-th participant's binary judgment (y_{ij}) for the i-th trial, indicating either simultaneity (SJ) or agency (AJ), is modeled as a Bernoulli variable with parameter q_{ij}. This parameter represents the hypothetical probability of a “synchronous” or “self” judgment for that specific trial-participant combination. It is further determined by a half-Gaussian function of the corresponding trial's standard deviation of asynchrony (x_{ij}). This function incorporates two individual-level parameters: width ($\sigma_j^{(H)}$) and height ($\lambda_j^{(H)}$) of the curve. The standard deviation of asynchrony (x_{ij}), which originally ranged from 0 ms to 80 ms, was rescaled to a 0.0 to 1.0 range for computational efficiency.

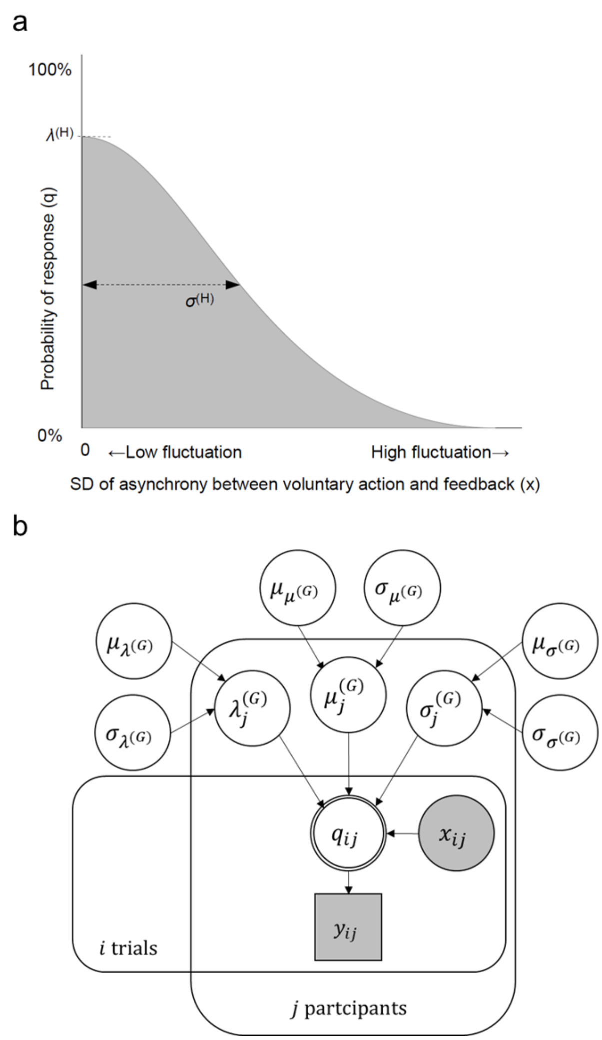


Figure 9. Hierarchical half-Gaussian model for simultaneity and agency judgments. **(a)** Schematic illustration: This panel depicts a half-Gaussian temporal window model for simultaneity judgment (SJ) or agency judgment (AJ) as a function of the standard deviation (SD) of asynchrony. **(b)** Plate notation: This panel presents the hierarchical extension of the half-Gaussian model for SJ or AJ using plate notation. Shaded nodes represent observed variables, unshaded nodes represent unobserved variables, squares denote discrete variables, circles denote continuous variables, single-bordered nodes indicate stochastic variables, and double-bordered nodes represent deterministic

variables.

$$y_{ij} \sim \text{Bernoulli}\left(q_{ij}\right). \tag{6}$$

$$q_{ij} = \lambda_j^{(H)} \times exp\left[-\frac{1}{2} \times \left(\frac{x_{ij}}{\sigma_j^{(H)}}\right)^2\right],\ x_{ij} \geq 0. \tag{7}$$

The individual-level parameters ($\lambda_j^{(H)}$ and $\sigma_j^{(H)}$) are assumed to follow a normal distribution with two group-level parameters for the mean ($\mu_{\lambda^{(H)}}$ and $\mu_{\sigma^{(H)}}$) and the standard deviation ($\sigma_{\lambda^{(H)}}$ and $\sigma_{\sigma^{(H)}}$). The group-level height parameter ($\sigma_{\lambda^{(H)}}$) is constrained to the range [0, 1] to reflect a valid probability range. The group-level width parameter ($\sigma_{\sigma^{(H)}}$) is restricted to positive values but has no upper limit.

$$\begin{gathered} \lambda_j^{(H)} \sim \text{Normal}\left(\mu_{\lambda^{(H)}}, \sigma_{\lambda^{(H)}}\right) \\ \sigma_j^{(H)} \sim \text{Normal}\left(\mu_{\sigma^{(H)}}, \sigma_{\sigma^{(H)}}\right). \end{gathered} \tag{8}$$

Weakly informative priors were assigned to all group-level parameters. Specifically, normal distributions were used for the mean parameters ($\mu_{\lambda^{(H)}}$ and $\mu_{\sigma^{(H)}}$), while half-Cauchy distributions were chosen for the standard deviation parameters ($\sigma_{\lambda^{(H)}}$ and $\sigma_{\sigma^{(H)}}$).

$$\begin{gathered} \mu_{\lambda^{(H)}} \sim \text{Normal}(0.5,\ 0.5) \\ \mu_{\sigma^{(H)}} \sim \text{Normal}(0.5,\ 0.5) \\ \sigma_{\lambda^{(H)}} \sim \text{Cauchy}^{+}(0,\ 0.5) \\ \sigma_{\sigma^{(H)}} \sim \text{Cauchy}^{+}(0,\ 0.5). \end{gathered} \tag{9}$$

Investigating the Effects of Delay Exposure on the Sensitivity and Response Criterion

The analysis examined how exposure to a sensory feedback delay influences participants' sensitivity and response criterion in simultaneity judgment (SJ) and agency judgment (AJ) tasks. To analyze these changes, a hierarchical SDT model with unequal variances for signal and noise distributions was employed (Selker et al., 2019; Greene & Rhodes, 2022). The model's schematic illustration and plate notation are presented in Figure 10.

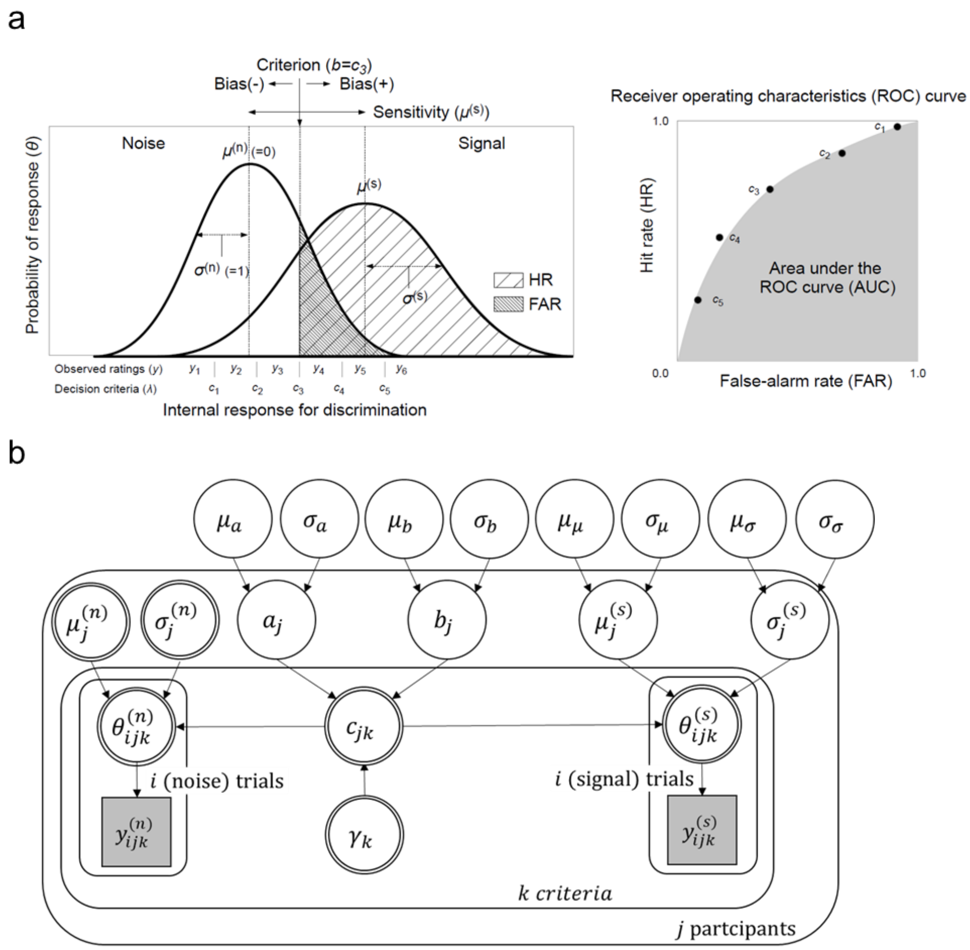


Figure 10. Hierarchical unequal-variance signal detection theory (SDT) model. **(a)** Schematic illustration: This panel depicts a schematic illustration of the model based on unequal-variance signal detection theory (SDT). **(b)** Plate notation: This panel presents the hierarchical extension of the unequal-variance SDT model using plate notation (Selker et al., 2019; Greene & Rhodes, 2022). Shaded nodes represent observed variables, unshaded nodes represent unobserved variables, squares denote discrete variables, circles denote continuous variables, single-bordered nodes indicate stochastic variables, and double-bordered nodes represent deterministic variables.

The model assumes that the j-th participant's observed probability (y_{ijk}) of selecting the k-th rating category (1-6) for the i-th trial regarding simultaneity or agency is determined by a categorical variable with six categories. This variable is parameterized by θ_{ijk}, which represents the hypothetical probability of each rating value.

$$y_{ijk} \sim \text{Categorical}(\theta_{ijk}). \tag{10}$$

The θ_{ijk} are derived from the participant's internal response about simultaneity or agency, separated by their decision criterion (c_{jk}). The decision criterion, c_{jk}, is assumed to be constant across trials for each participant.

$$\theta_{ijk} = \Phi(c_{jk} | \mu_j, \sigma_j) - \Phi(c_{j(k-1)} | \mu_j, \sigma_j). \tag{11}$$

The internal response is modeled as a normal distribution with mean ($\mu_j^{(n)}$ and $\mu_j^{(s)}$) and standard deviation ($\sigma_j^{(n)}$ and $\sigma_j^{(s)}$). Under noise trials (self-controlled), $\mu_j^{(n)}$ is fixed at zero, and $\sigma_j^{(n)}$ is fixed at one. For signal trials (computer-controlled), $\mu_j^{(s)}$ follows a normal distribution with a group-level mean (μ_μ) and standard deviation (μ_σ), and $\sigma_j^{(s)}$ follows another normal distribution with a group-level mean (σ_μ) and standard deviation (σ_σ).

$$\log \mu_j \sim \begin{cases} 0 \\ \text{Normal}(\mu_\mu, \sigma_\mu) \end{cases} \text{if} \begin{matrix} \text{noise } \left(\mu_j^{(n)}\right) \\ signal \left(\mu_j^{(s)}\right) \end{matrix}. \tag{12}$$

$$\log \sigma_j \sim \begin{cases} 1 \\ \text{Normal}(\mu_\sigma, \sigma_\sigma) \end{cases} \text{if} \begin{matrix} \text{noise } \left(\sigma_j^{(n)}\right) \\ signal \left(\sigma_j^{(s)}\right) \end{matrix}. \tag{13}$$

Notably, the logarithm function serves as a link function, imposing the constraint that the SDT parameters be positive (Selker et al., 2019).

The j-th participant's response criterion for the k-th rating category (c_{jk}) is modeled as a linear transformation of an unbiased criterion (γ_k) with an intercept (a_j) and a slope (b_j), both allowed to vary across participants. The unbiased criterion (γ_k) is constructed by dividing the range 0-1 into six equal intervals and then is transformed by a logit function to span negative to positive infinity (Selker et al., 2019). In this study, the response criterion of interest was the slope parameter (b_j), corresponding to the c_{j3} category (see Figure 10a).

$$c_{jk} = a_j \gamma_k + b_j. \tag{14}$$

$$\gamma_k = \log\left(\frac{\frac{k}{K}}{1 - \frac{k}{K}}\right), \tag{15}$$

where K denotes the number of rating categories (i.e., six).

Weakly informative priors were assigned to all group-level parameters. Specifically, normal distributions with a mean of 0 and a standard deviation of 1 (μ_μ, μ_σ, and μ_a) or 2 (μ_b) were used for the mean parameters. Half-Cauchy distributions with a mean of 0 and a standard deviation of 1 (σ_μ, σ_σ, and σ_a) or 2 (σ_b) were used for standard deviation parameters. The selection of these priors aligns with previous research (Greene & Rhodes, 2022). Note that the logarithm function is employed to constrain the a_j parameters to be positive.

$$\begin{aligned}
\log a_j &\sim \text{Normal}(\mu_a, \sigma_a) \\
b_j &\sim \text{Normal}(\mu_b, \sigma_b) \\
\mu_\mu &\sim \text{Normal}(0,\ 1) \\
\sigma_\mu &\sim \text{Cauchy}^+(0,\ 1) \\
\mu_\sigma &\sim \text{Normal}(0,\ 1) \\
\sigma_\sigma &\sim \text{Cauchy}^+(0,\ 1) \\
\mu_a &\sim \text{Normal}(0,\ 1) \\
\sigma_a &\sim \text{Cauchy}^+(0,\ 1) \\
\mu_b &\sim \text{Normal}(0,\ 2) \\
\sigma_b &\sim \text{Cauchy}^+(0,\ 2).
\end{aligned} \tag{17}$$

Sensitivity in this analysis was quantified using the area under the receiver operating curve (AUC) (Wickens, 2002;

Selker et al., 2019). The AUC is derived from the model's estimated mean ($\mu^{(s)}$) and standard deviation parameters of the signal distribution ($\sigma^{(s)}$). A higher AUC indicates a greater ability of the participant to discriminate between the signal (computer-controlled trial) and noise (self-controlled trial).

$$\mathrm{AUC}_j = \Phi\left(\frac{\mu_j^{(s)}}{\sqrt{1+\sigma_j^{(s)^2}}}\right), \tag{18}$$

where Φ denotes the cumulative Gaussian function and j represents each participant.

Investigating Relationships between SDT and Temporal Window Parameters

Building on the predictions from the introduction, Bayesian correlation analyses using Pearson's correlation coefficient (ρ) were conducted to explore relationships between the EAP estimates from the SDT and temporal window models for each participant in the SJ and AJ tasks. I hypothesized that the size of the temporal window, reflected by the area under the window (AUW), would correlate with the SDT's sensitivity measure (AUC), while the location parameter (μ) was not expected to be associated with the SDT's criterion (b).

The area under the window (AUW) was calculated based on the height (λ) and width (σ) parameters. This AUW served as a proxy for the window's sensitivity. The formula for calculating $\mathrm{AUW}^{(G)}$in the Gaussian window model for TWS and TWA is provided below:

$$\mathrm{AUW}_j^{(G)} = \lambda_j^{(G)} \times \sigma_j^{(G)} \times \sqrt{2\pi}, \tag{19}$$

where $\lambda_j^{(G)}$ is the height and $\sigma_j^{(G)}$ is the width of the Gaussian window for the j-th participant.

Similar to the Gaussian model, the area under the window (AUW) was calculated for the half-Gaussian window model of the SJ and AJ tasks in relation to the standard deviation (SD) of asynchrony. The formula for $\mathrm{AUW}^{(H)}$ is provided below:

$$\mathrm{AUW}_j^{(H)} = \frac{\lambda_j^{(H)} \times \sigma_j^{(H)} \times \sqrt{2\pi}}{2}, \tag{20}$$

where $\lambda_j^{(H)}$ represents the height and $\sigma_j^{(H)}$ represents the width of the half-Gaussian window for the j-th participant.

To isolate the effects of delay manipulation on parameter estimates, I calculated difference scores for each participant on four key parameters: location parameter ($\mu_j^{(G)}$) of the Gaussian window model, response criterion (b_j) from the SDT model, sensitivity measure (AUC_j) from the SDT model, and area under the window (AUW) for both the Gaussian ($\text{AUW}_j^{(G)}$) and half-Gaussian ($\text{AUW}_j^{(H)}$) models. These difference scores capture the change between the delayed and baseline conditions for each parameter (e.g., $\Delta\mu_j^{(G)}$ represents the change in the location parameter).

$$\begin{gathered}
\Delta\mu_j^{(G)} = \mu_{j(Delayed)}^{(G)} - \mu_{j(Baseline)}^{(G)} \\
\Delta b_j = b_{j(Delayed)} - b_{j(Baseline)} \\
\Delta\text{AUC}_j = \text{AUC}_{j(Delayed)} - \text{AUC}_{j(Baseline)} \\
\Delta\text{AUW}_j^{(G)} = \text{AUW}_{j(Delayed)}^{(G)} \text{-} \text{AUW}_{j(Baseline)}^{(G)} \\
\Delta\text{AUW}_j^{(H)} = \text{AUW}_{j(Delayed)}^{(H)} \text{-} \text{AUW}_{j(Baseline)}^{(H)}.
\end{gathered} \tag{21}$$

To evaluate the hypothesized relationships between the temporal window parameters and the SDT measures, I conducted the following Bayesian correlation analyses with two key hypotheses:

1. Location and Criterion (H_0 vs. H_1): The null hypothesis (H_0) posits no correlation between the location parameter difference score ($\Delta\mu^{(G)}$) and the response criterion difference score (Δb) (i.e., $\rho = 0$). The alternative hypothesis (H_1) predicts a significant correlation between these difference scores (i.e., $\rho \neq 0$).
2. Shape and Sensitivity (H_0 vs. H^-): The null hypothesis (H_0) posits no correlation between the area under the window difference scores ($\Delta\text{AUW}^{(G)}$ and $\Delta\text{AUW}^{(H)}$) and the sensitivity measure difference score (ΔAUC) (i.e., $\rho = 0$). The one-tailed alternative hypothesis (H^-) predicts a negative correlation between these difference scores (i.e., $\rho < 0$). This directional prediction aligns with the expectation that a narrower temporal window (smaller AUW) would be associated with greater sensitivity (larger AUC).

A Bayes factor (BF) hypothesis test was employed to quantify the evidence for or against the hypothesized effects. A

Bayes factor BF_{10} represents the evidence for the two-tailed alternative hypothesis (H_1) of an effect (positive or negative) compared to the null hypothesis (H_0) of no effect. Conversely, BF_{-0} represents the evidence for the one-tailed alternative hypothesis (H^-) of a negative effect compared to H_0 (van Doorn et al., 2021).

Author Contributions

YS designed the study, conducted the experiment, analyzed the data, and wrote the manuscript.

Funding

This work was funded by JSPS KAKENHI Grant Number 17K00218 and 22K12757.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data Availability Statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://doi.org/10.6084/m9.figshare.26047888.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

For this paper, I utilized generative AI/machine translation tools such as ChatGPT, Gemini, Google Translate, and DeepL Write to improve clarity, comprehensibility, and readability. Afterward, I carefully reviewed and revised the content, taking full responsibility for the published work.

References

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods*, *53*(4), 1407–1425. https://doi.org/10.3758/s13428-020-01501-5

Arnold, D. H., Nancarrow, K., & Yarrow, K. (2012). The critical events for motor-sensory temporal recalibration. *Frontiers in Human Neuroscience*, *6*(235), 1–6. https://doi.org/10.3389/fnhum.2012.00235

Attig, C., Rauh, N., Franke, T., & Krems, J. F. (2017). System latency guidelines then and now – Is zero latency really considered necessary? In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics 2017: Vol. 10276 LNAI* (Issue May, pp. 3–14). Springer International Publishing AG. https://doi.org/10.1007/978-3-319-58475-1_1

Bertelson, P., & Radeau, M. (1976). Ventriloquism, sensory interaction, and response bias: Remarks on the paper by Choe, Welch, Gilford, and Juola. *Perception & Psychophysics*, *19*(6), 531–535. https://doi.org/10.3758/BF03211222

Bland, J. M., & Altman, D. G. (1995). Calculating correlation coefficients with repeated observations: Part 2-correlation between subjects. *BMJ*, *310*(6980), 633.

Bockes, F., Wimmer, R., & Schmid, A. (2018). LagBox—Measuring the Latency of USB-Connected Input Devices. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3170427.3188632

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Chan, Y. M., Pianta, M. J., & McKendrick, A. M. (2014). Reduced audiovisual recalibration in the elderly. *Frontiers in Aging Neuroscience*, *6*(226), 1–9. https://doi.org/10.3389/fnagi.2014.00226

Choe, C. S., Welch, R. B., Gilford, R. M., & Juola, J. F. (1975). The "ventriloquist effect": Visual dominance or response bias? *Perception & Psychophysics*, *18*(1), 55–60. https://doi.org/10.3758/BF03199367

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates Hillsdale, NJ.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/14805-018

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE*, *8*(3). https://doi.org/10.1371/journal.pone.0057410

Cunningham, D. W., Billock, V. A., & Tsou, B. H. (2001). Sensorimotor adaptation to violations of temporal contiguity. *Psychological Science*, *12*(6), 532–535. http://www.ncbi.nlm.nih.gov/pubmed/11760144

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Doherty, R. A., & Sorenson, P. (2015). Keeping users in the flow: Mapping system responsiveness with user experience. *Procedia Manufacturing*, *3*, 4384–4391. https://doi.org/10.1016/j.promfg.2015.07.436

Ebrahim, E. A., & Cengiz, M. A. (2022). Predicting verbal learning and memory assessments of older adults using Bayesian hierarchical models. *Frontiers in Psychology*, *13*(855379). https://doi.org/10.3389/fpsyg.2022.855379

Farrer, C., Valentin, G., & Hupé, J. M. (2013). The time windows of the sense of agency. *Consciousness and Cognition*, *22*(4), 1431–1441. https://doi.org/10.1016/j.concog.2013.09.010

Franck, N., Farrer, C., Georgieff, N., Marie-Cardine, M., Daléry, J., D'Amato, T., & Jeannerod, M. (2001). Defective recognition of one's own actions in patients with schizophrenia. *American Journal of Psychiatry*, *158*(3), 454–459. https://doi.org/10.1176/appi.ajp.158.3.454

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*(7), 773–778. https://doi.org/10.1038/nn1268

Gelman, A. (2004). Parameterization and Bayesian modeling. *Journal of the American Statistical Association*, *99*(466), 537–545. https://doi.org/10.1198/016214504000000458

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*(3), 515–533.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis third edition (with errors fixed as of 15 February 2021)*. Taylor & Francis. http://www.stat.columbia.edu/~gelman/book/BDA3.pdf

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley New York.

Greene, N. R., & Rhodes, S. (2022). A tutorial on cognitive modeling for cognitive aging research. *Psychology and Aging*, *37*(1), 30–42. https://doi.org/10.1037/pag0000637

Grootswagers, T. (2020). A primer on running human behavioural experiments online. *Behavior Research Methods*, *52*(6), 2283–2286. https://doi.org/10.3758/s13428-020-01395-3

Haering, C., & Kiesel, A. (2015). Was it me when it happened too early? Experience of delayed effects shapes sense of agency. *Cognition*, *136*, 38–42. https://doi.org/10.1016/j.cognition.2014.11.012

Haering, C., & Kiesel, A. (2016). Time perception and the experience of agency. *Psychological Research*, *80*(2), 286–297. https://doi.org/10.1007/s00426-015-0654-0

Helson, H. (1964). *Adaptation-level theory.* Harper & Row.

Heron, J., Hanson, J. V. M., & Whitaker, D. (2009). Effect before cause: Supramodal recalibration of sensorimotor timing. *PloS One*, *4*(11), e7681. https://doi.org/10.1371/journal.pone.0007681

Imaizumi, S., & Asai, T. (2015). Dissociation of agency and body ownership following visuomotor temporal recalibration. *Frontiers in Integrative Neuroscience*, *9*(35). https://doi.org/10.3389/fnint.2015.00035

JASP Team. (2022). *JASP (Version 0.16.3)* [Computer software]. https://jasp-stats.org/

Jota, R., Ng, A., Dietz, P., & Wigdor, D. (2013). How fast is fast enough? A study of the effects of latency in direct-touch pointing tasks. *Conference on Human Factors in Computing Systems - Proceedings*, 2291–2300. https://doi.org/10.1145/2470654.2481317

Keetels, M., & Vroomen, J. (2012). Exposure to delayed visual feedback of the hand changes motor-sensory synchrony perception. *Experimental Brain Research*, *219*(4), 431–440. https://doi.org/10.1007/s00221-012-3081-0

Khitrov, M. Y., Laxminarayan, S., Thorsley, D., Ramakrishnan, S., Rajaraman, S., Wesensten, N. J., & Reifman, J. (2013). PC-PVT: A platform for psychomotor vigilance task testing, analysis, and prediction. *Behavior Research Methods*, *46*(1), 140–147. https://doi.org/10.3758/s13428-013-0339-9

Klatzky, R. L., Thompson, W. B., Stefanucci, J. K., Gill, D., & McGee, D. K. (2017). The perceptual basis of vast space. *Psychonomic Bulletin and Review*, *24*(6), 1870–1878. https://doi.org/10.3758/s13423-017-1265-0

Knoblich, G., & Repp, B. H. (2009). Inferring agency from sound. *Cognition*, *111*(2), 248–262. https://doi.org/10.1016/j.cognition.2009.02.007

Kuroki, D. (2020). Advantages and disadvantages of conducting Web-based experiments and the knowledge required to create the programs [In Japanese]. *The Japanese Journal of Psychonomic Science*, *38*(2), 250–257. https://doi.org/10.14947/psychono.38.37

Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin and Review*, *15*(1), 1–15. https://doi.org/10.3758/PBR.15.1.1

Li, X., Liang, Z., Kleiner, M., & Lu, Z.-L. (2010). RTbox: A device for highly accurate response time measurements. *Behavior Research Methods*, *42*(1), 212–225. https://doi.org/10.3758/BRM.42.1.212

Ma, K. S. T., & Schnupp, J. W. H. (2021). The unity hypothesis revisited: Can the gender incongruent McGurk effect be disrupted by priming ? *PsyArXiv*. https://doi.org/10.31234/osf.io/y7uvk

Majima, Y. (2019). Should I, as a cognitive scientist, use crowdsourcing in my research? [In Japanese]. *Cognitive Studies*, *26*(2), 272–281. https://doi.org/10.11225/jcss.26.272

Matsuura, K. (2016). *Bayesian statistical modeling with Stan and R [In Japanese, translated by the author of this article.]*. Kyoritsu Shuppan Co., LTD.

Meyerhoff, H. S., & Scholl, B. J. (2018). Auditory-induced bouncing is a perceptual (rather than a cognitive) phenomenon: Evidence from illusory crescents. *Cognition*, *170*, 88–94.

https://doi.org/10.1016/j.cognition.2017.08.007

Michon, J. A. (1978). The making of the present: A tutorial review. *Attention and Performance VII*, *1538*, 89–111. https://doi.org/10.1016/j.brainres.2013.09.031

Moore, J. W., & Fletcher, P. C. (2012). Sense of agency in health and disease: A review of cue integration approaches. *Consciousness and Cognition*, *21*(1), 59–68. https://doi.org/10.1016/j.concog.2011.08.010

Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Adaptation to audiotactile asynchrony. *Neuroscience Letters*, *413*(1), 72–76. https://doi.org/10.1016/j.neulet.2006.11.027

Navarra, J., Vatakis, A., Zampini, M., Soto-Faraco, S., Humphreys, W., & Spence, C. (2005). Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Cognitive Brain Research*, *25*(2), 499–507. https://doi.org/10.1016/j.cogbrainres.2005.07.009

Plant, R. R., Hammond, N., & Whitehouse, T. (2003). How choice of mouse may affect response timing in psychological studies. *Behavior Research Methods*, *35*(2), 276–284. http://www.ncbi.nlm.nih.gov/pubmed/12834085

R Core Team. (2022). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rohde, M., Greiner, L., & Ernst, M. O. (2014). Asymmetries in visuomotor recalibration of time perception: Does causal binding distort the window of integration? *Acta Psychologica*, *147*, 127–135. https://doi.org/10.1016/j.actpsy.2013.07.011

Rohde, M., Scheller, M., & Ernst, M. O. (2014). Effects can precede their cause in the sense of agency. *Neuropsychologia*, *65*, 191–196. https://doi.org/10.1016/j.neuropsychologia.2014.10.011

Sato, A., & Yasuda, A. (2005). Illusion of sense of self-agency: Discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition*, *94*(3), 241–255. https://doi.org/10.1016/j.cognition.2004.04.003

Schroeger, A., Grießbach, E., Raab, M., & Cañal-Bruland, R. (2022). Spatial distances affect temporal prediction and interception. *Scientific Reports*, *12*(15786). https://doi.org/10.1038/s41598-022-18789-2

Selker, R., van den Bergh, D., Criss, A. H., & Wagenmakers, E. J. (2019). Parsimonious estimation of signal detection models from confidence ratings. *Behavior Research Methods*, *51*(5), 1953–1967. https://doi.org/10.3758/s13428-019-01231-3

Semmelmann, K., & Weigelt, S. (2017). Online psychophysics: Reaction time effects in cognitive experiments. *Behavior Research Methods*, *49*(4), 1241–1260. https://doi.org/10.3758/s13428-016-0783-4

Shimizu, H. (2002). Measuring keyboard response delays by comparing keyboard and joystick inputs. *Behavior Research Methods, Instruments, & Computers*, *34*(2), 250–256. http://www.ncbi.nlm.nih.gov/pubmed/12109021

Stan Development Team. (2023). *RStan: The R interface to Stan. R package version 2.21.8.* [Computer software]. https://mc-stan.org/

Stan Development Team. (2024). *Stan user's guide version 2.34*. https://mc-stan.org/docs/2_34/stan-users-guide-2_34.pdf

Stetson, C., Cui, X., Montague, P. R., & Eagleman, D. M. (2006). Motor-sensory recalibration leads to an illusory reversal of action and sensation. *Neuron*, *51*(5), 651–659. https://doi.org/10.1016/j.neuron.2006.08.006

Stewart, N., Chandler, J., & Paolacci, G. (2017). Crowdsourcing samples in cognitive science. *Trends in Cognitive Sciences*, *21*(10), 736–748. https://doi.org/10.1016/j.tics.2017.06.007

Sugano, Y. (2021). Audiomotor temporal recalibration modulates decision criterion of self-agency but not perceptual sensitivity. *Frontiers in Psychology*, *12*(580441). https://doi.org/10.3389/fpsyg.2021.580441

Sugano, Y., Keetels, M., & Vroomen, J. (2010). Adaptation to motor-visual and motor-auditory temporal lags transfer across modalities. *Experimental Brain Research*, *201*(3), 393–399. https://doi.org/10.1007/s00221-009-2047-3

Sugano, Y., Keetels, M., & Vroomen, J. (2012). The build-up and transfer of sensorimotor temporal recalibration measured via a synchronization task. *Frontiers in Psychology*, *3*(246). https://doi.org/10.3389/fpsyg.2012.00246

Sugano, Y., Keetels, M., & Vroomen, J. (2014). Concurrent sensorimotor temporal recalibration to different lags for the left and right hand. *Frontiers in Psychology*, *5*(140). https://doi.org/10.3389/fpsyg.2014.00140

Sugano, Y., Keetels, M., & Vroomen, J. (2016). Auditory dominance in motor-sensory temporal recalibration. *Experimental Brain Research*, *234*(5), 1249–1262. https://doi.org/10.1007/s00221-015-4497-0

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition*, *17*(1), 219–239. https://doi.org/10.1016/j.concog.2007.03.010

Tanaka, T. (2024). Evaluating the Bayesian causal inference model of intentional binding through computational modeling. *Scientific Reports*, *14*(2979). https://doi.org/10.1038/s41598-024-53071-7

Timm, J., Schönwiesner, M., Sanmiguel, I., & Schröger, E. (2014). Sensation of agency and perception of temporal order. *Consciousness and Cognition*, *23*, 42–52. https://doi.org/10.1016/j.concog.2013.11.002

Toyoda, H. (2017). *Practical Bayes modeling: Analysis techniques and cognitive models [In Japanese, translated by the author of this article.]*. Asakura Publishing Co., LTD.

van der Burg, E., Toet, A., Brouwer, A. M., & van Erp, J. B. F. (2021). Serial dependence of emotion within and between stimulus sensory modalities. *Multisensory Research*, *35*(2), 151–172. https://doi.org/10.1163/22134808-bja10064

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E. J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin and Review*, *28*(3), 813–826. https://doi.org/10.3758/s13423-020-01798-5

van Eijk, R. L. J., Kohlrausch, A., Juola, J. F., & Van De Par, S. (2008). Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Perception and Psychophysics*, *70*(6), 955–968. https://doi.org/10.3758/PP.70.6.955

Varela, F. J. (1999). The specious present: A neurophenomenology of time consciousness. In J. Petitot, F. J. Varela, B. Pachoud, & J.-M. Roy (Eds.), *Naturalizing phenomenology: Issues in contemporary phenomenology and cognitive science* (pp. 266–314). Stanford University Press.

Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2008). Audiovisual temporal adaptation of speech: Temporal order versus simultaneity judgments. *Experimental Brain Research*, *185*(3), 521–529. https://doi.org/10.1007/s00221-007-1168-9

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Burkner, P. C. (2021). Rank-normalization, folding, and localization: An improved R-hat for assessing convergence of MCMC (with discussion). *Bayesian Analysis*, *16*(2), 667–718. https://doi.org/10.1214/20-BA1221

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research*, *22*(1), 32–35. https://doi.org/10.1016/j.cogbrainres.2004.07.003

Wen, W. (2019). Does delay in feedback diminish sense of agency? A review. *Consciousness and Cognition*, *73*(102759). https://doi.org/10.1016/j.concog.2019.05.007

Wickens, T. D. (2002). *Elementary signal detection theory*. Oxford University Press.

Wimmer, R., Schmid, A., & Bockes, F. (2019). On the latency of USB-connected input devices. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3290605.3300650

Winter, R., Harrar, V., Gozdzik, M., & Harris, L. R. (2008). The relative timing of active and passive touch. *Brain Research*, *1242*, 54–58. https://doi.org/10.1016/j.brainres.2008.06.090

Witt, J. K., Taylor, J. E. T., Sugovic, M., & Wixted, J. T. (2015). Signal detection measures cannot distinguish perceptual biases from response biases. *Perception*, *44*(3), 289–300. https://doi.org/10.1068/p7908

Witthoft, N., Sha, L., Winawer, J., & Kiani, R. (2018). Sensory and decision-making processes underlying perceptual adaptation. *Journal of Vision*, *18*(8). https://doi.org/10.1167/18.8.10

Yarrow, K., Jahn, N., Durant, S., & Arnold, D. H. (2011). Shifts of criteria or neural timing? The assumptions underlying timing perception studies. *Consciousness and Cognition*, *20*(4), 1518–1531. https://doi.org/10.1016/j.concog.2011.07.003

Yarrow, K., Sverdrup-Stueland, I., Roseboom, W., & Arnold, D. H. (2013). Sensorimotor temporal recalibration within and across limbs. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(6), 1678–1689. https://doi.org/10.1037/a0032534