

A Separate Mechanisms Investigation of Rapid Recalibration to Audiovisual Asynchrony

Candidate nr.: 10049

PSY2900 Bachelor thesis in psychology

16 May 2022 Hamar

Dawn M. Behne

Foreword

As a starting point for this project, the advisor introduced students to the project's research question and some related issues, together with initial supporting literature. Further literature was identified by the students and shared with the group, and occasionally supplemented by the project advisor. Hypotheses were formulated by the students with supervision, based on the research question and issues presented. Students had the possibility to focus on one or all of the hypotheses in their reports. The experiment was created by the advisor. The students carried out all phases of data collection for the experiment. Data handling was arranged by the advisor and students participated in the process. Statistical analyses and their interpretation were discussed as a group. Students have had the datafile and could run additional/alternative analyses if they chose.

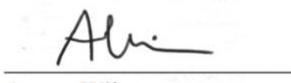
The group had regular seminars, discussions, and close supervision throughout the semester, as well as optional feedback on writing. Students worked as a group to carry out all phases of the project. Literature and materials related to the experiment were stored on a wiki, shared by everyone on the project.

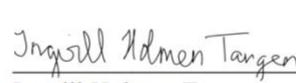
With this basis, each student submits a report (written individually) which has the form and style of a journal article. Students are allowed and encouraged to work together, but the final product must be their own. The report can be in Norwegian or English.



Ane Kristine Eggen
Date: 10.05.2022



Bente Mari Aakvik
Date: 10.05.2022

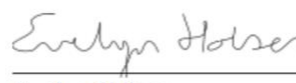

Karoline Hatlen
Date: 11.05.2022



Angus Wilson
Date: 10.05.2022



Ingvill Holmen Tangen
Date: 10.05.2022


Vegard Dahn
Date: 10.05.2022

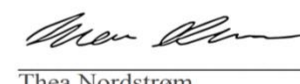

Astrid Brøvig Silde
Date: 10.05.2022


Evelyn Holsen
Date: 10.05.2022


Advisor: Dawn M. Behne
Date: 10.05.2022


Benjamin Bornø
Date: 10.05.2022


Linda Marie Leirvik
Date: 11.05.2022


Thea Nordstrøm
Date: 10.05.2022

Abstract

Integration of temporally misaligned sensory information is crucial for constructing a unified perception of our environment. Investigation into the processes facilitating temporal integration of audiovisual stimuli has uncovered that our brain adapts to compensate for cross-modal delays (i.e. temporal recalibration). Previous studies on audiovisual temporal recalibration have observed that synchrony judgement is affected by the modality order of a preceding exposure to a greater degree when vision, rather than audio, was the leading modality. This study aimed to investigate this asymmetry by considering the possibility that temporal integration of audiovisual stimuli is mediated by independent mechanisms as a function of leading modality. Here, participants responses on a simultaneity judgement task (SJ) were compared based on the modality order of a preceding trial. The stimulus used for the experiment was audiovisual alignments of the syllable /ba/. A separate mechanisms interpretation of temporal integration was supported if subjective synchrony of audio-lead and video-lead asynchronies were affected differently by the preceding modality order. The findings of this study were inconclusive. Results show that subjective synchrony judgements of audio-lead asynchronies were affected to a larger degree when the preceding modality order was audio-lead, compared to video-lead. The opposite was found for subjective synchrony of video-lead asynchronies. These findings indicate that effects of rapid temporal recalibration to audiovisual asynchronies is mediated by independent shifts of timing criteria for either modality order.

Introduction

Perception is not passive; we construct coherent representations of our environment by actively navigating the ambiguous and incomplete information received by our sensory organs. Integral to this endeavor is information organization: identifying relevant information and parsing it into discrete entities. This phenomenon is commonly referred to in the neuropsychological literature as “the binding problem” (e.g., Revonsuo & Newman, 1999; Treisman, 1996; von der Malsburg, 1995). Temporal coincidence between sensory modalities is one factor, among several, that are known to inform us about whether different sensory information belong together (Keetels & Vroomen, 2012). Events in our environment, however, will rarely produce information that is received simultaneously by our different sensory organs. The speed of light, for example, is much faster through air than sound (300 000 000 m/s and 300 m/s, respectively). One might therefore expect that audiovisual (AV) information from a single source will only be perceived as synchronous when it is at a distance from the recipient at which the information reaches our primary sensory cortices simultaneously (“the horizon of simultaneity,” ca. 10 m from observer) (Pöppel, Schill, & von Steinbüchel, 1990). Perception of synchrony is not limited to such events (Keetels & Vroomen, 2012), so our perceptual processes must in some way account for naturally occurring lags between the senses.

One such way is that our brain seems to be lenient in what it regards as synchronous, showing a degree of tolerance for temporal misalignments between sensory modalities (Vroomen & Keetels, 2010; Wallace & Stevenson, 2014). Indeed, our perception has a temporal binding window (i.e., the time interval within which separate sensory cues are likely to be integrated and perceptually unified) of a few hundred milliseconds for which AV stimuli will be perceived as synchronous (e.g., Hay-McCutcheon, Pisoni, & Hunt, 2009; McGrath & Summerfield, 1985; Stevenson et al., 2014). Interestingly, the window is not static. The window within which temporally misaligned AV signals will be integrated is contingent on several factors, one of which is stimulus complexity. Perceptual sensitivity deteriorates with increase in stimulus complexity: the temporal binding window is narrower for simple AV stimulus pairs (flash/beep stimuli) compared to more complex stimuli (speech) (Boer, Eussen, & Vroomen, 2013). Temporal binding is also proposed to be adaptive or malleable, contingent on prior experience (Hillock-Dunn & Wallace, 2012; Keetels & Vroomen, 2012; Vroomen & Keetels, 2010). Both

long-term experience with our environment and short-term contextual information will affect our sensitivity to stimulus asynchronies. For example, Alm and Behne (2013) found that middle-aged adults show less tolerance for audio-lead AV asynchronies than young adults, suggesting that this is a manifestation of audio-lead asynchronies occurring less frequently in our natural environment than visual-lead asynchronies, thus becoming a more familiar experience with age. An interpretation of the apparent malleability of our temporal binding window is that our brain makes statistical inferences about the likelihood that information streams from different sensory modalities originate from the same event, and makes adjustments accordingly (Wallace & Stevenson, 2014). In other words, variability in perceptual sensitivity to AV asynchronies can be viewed as an attempt by our brain to continuously remain stringent enough to segregate asynchronous stimuli too great to have originated from the same source, yet lenient enough to integrate those asynchronies that are to be naturally expected.

Investigation of how the brain deals with temporal binding has proven not only to be important for understanding the human perceptual process, but also for understanding the implications of atypical processing. Research has shown that the issue of temporal binding is associated with both autism spectrum disorder and schizophrenia, both demonstrating a widened AV temporal binding window (Stevenson et al., 2014; Zhou et al., 2021). Efforts should therefore be made to further elucidate the ways in which our brain reconciles temporal discrepancies between sensory information originating from the same source. The aim of this study was just that. In this study, temporal binding was elucidated by investigating one of several known perceptual mechanisms that facilitates temporal binding of AV stimuli, specifically rapid temporal recalibration. In this endeavor, this study attempted to address previously suggested alternatives to interpreting results of AV synchrony judgement experiments, inspired by Yarrow, Jahn, Durant, and Arnold (2011) and Cecere, Gross, and Thut (2016), and discuss implications of possible findings on how temporal recalibration is understood, both in relation to the way in which this process occurs in the brain and how such explanations can be transferrable and applicable to the process of temporal binding in general.

Temporal recalibration

Recalibration, when related to perception, refers to an adaptive strategy employed by our brain in which perceptual systems are adjusted to compensate for misalignment between information

received by different sensory modalities. Perceptual recalibration was first introduced by von Helmholtz when he demonstrated that participants with displaced visual fields would adapt to the displacement with prolonged exposure (Helmholtz & Southall, 1962). His experiment was set up so that observers would point at an object while wearing prismatic goggles that artificially displaced the visual field. He discovered that while initial pointing attempts were off target in the same direction as the visual shift, after a series of repeated attempts, this error diminished. Additionally, when the participants removed their goggles, a negative after-effect was observed. Pointing attempts were now erroneous in the opposite direction from the previous visual displacement. The conclusion was that the visual displacement caused by wearing the prismatic goggles induces misalignment between the observers visual and proprioceptive spatial maps, causing the observer to unconsciously realign those maps. When the goggles were removed, realignment to the previous displacement persists, causing overcompensation.

Recalibration has become a fundamental paradigm for understanding cross-modal integration, temporal binding included. Fujisaki, Shimojo, Kashino, and Nishida (2004) were among the first to demonstrate recalibration of AV simultaneity judgement. In their experiment, participants were exposed to a fixed AV time lag over several minutes before completing a classical simultaneity judgement (SJ) task. In the SJ task, participants were subjected to various stimulus onset asynchronies (SOA) of either beep before flash, or flash before beep. They were instructed to judge whether the signals were synchronous or asynchronous. After the task was completed, the percentage of synchronous responses were plotted for each SOA and data was fitted with a gaussian function, averaged over all participants. The point of subjective simultaneity (PSS) was defined as the center of the gaussian curve. Results showed that the PSS shifted based on the adaptation procedure participants were subjected to prior to conducting the SJ task. The PSS was significantly more visual-lead when the adaptation procedure was visual-lead, compared to when the adaptation procedure was audio-lead. The conclusion was that a lag adaptation shifts our subjective simultaneity in the direction of the lag. Interestingly, prolonged adaptation procedures are not necessary to elicit recalibration of subjective simultaneity. In fact, only a single audiovisual event will effect subsequent judgements (Van der Burg, Alais, & Cass, 2013; Van der Burg & Goodbourn, 2015). Van der Burg et al. (2013) used the classical SJ task with different SOAs ranging from -800 ms audio-lead to +800 ms video-lead. Their results showed that the PSS on a given trial (n) was contingent on the asynchrony of the preceding trial

(n-1), with the PSS of video-lead n-1 SOA being significantly greater than PSS of audio-lead n-1 SOA. These results suggest that temporal recalibration can occur rapidly.

Van der Burg et al. (2013) point out that their results show asymmetry in the magnitude of effect of recalibration across SOA n-1. Their results indicate that the degree to which the PSS is affected is not equal across the SOA range. Visual-lead SOA n-1 seem to affect the PSS for a subsequent trial to a greater degree than audio-lead SOA n-1. Several other studies show the same trend (Roseboom, 2019; Van der Burg et al., 2013; Van der Burg & Goodbourn, 2015). Their explanation for this phenomenon is that asymmetry in temporal recalibration between audio- and video-lead asynchronies reflect naturally occurring lags between vision and sound. Indeed, based on the horizon of simultaneity, which is at ca. 10 m, any events occurring at a farther distance will be visual-lead and any event occurring closer is audio-lead. The range at which audio can naturally lead vision is therefore much narrower than for vision leading audio. The increased magnitude of effect in temporal recalibration seen for video-lead asynchronies can therefore be interpreted as a strategy for optimizing perception, where perceptual adaptability constrains itself to natural possibilities. Following that logic, one can assume that the degree to which our brain is willing to adjust its synchrony criteria, contingent upon preceding AV alignments, should be less for instances where audio leads video than for the opposite.

Temporal audiovisual integration mediated by separate mechanisms

Several studies have revealed that sensitivity to asynchronous AV stimuli is dependent on which modality comes first (Alm & Behne, 2013; Behne & Wang, 2018; Cecere et al., 2016).

In general, our brain is more sensitive to audio-lead asynchronies than for video-lead asynchronies (Cecere et al., 2016). Additionally, the apparent malleability of our temporal binding window, as previously discussed, shows asymmetry, in that the outer boundary of the window at the audio-lead side is more static, while the outer boundary on the video-lead side is more flexible (Cecere et al., 2016). Accordingly, when observing malleability in the temporal binding window, one can assume that it is most often facilitated by a response to video-lead asynchronies. Indeed, Cecere et al. (2016) found that sensitivity to video-lead AV asynchronies is more easily trainable than sensitivity to audio-lead asynchronies. Additionally, training effects on sensitivity to either audio-lead or video-lead asynchronies, were not transferrable to the other. These findings, in addition to those from other studies that indicate that audio-lead rather than

video-lead sensitivity improves with long-term experience (Alm & Behne, 2013; Behne & Wang, 2018), has inspired some to propose that AV synchrony judgement is mediated by two distinct perceptual mechanisms, in which audio-lead and video-lead temporal binding are processed independently of each other (Cecere et al., 2016; Yarrow et al., 2011).

Assuming a dual mechanism interpretation of AV synchrony judgement, a question then arises of whether using a central tendency measure, such as PSS, extracted from a continuous function fit to averaged responses, is the best measure when investigating AV synchrony judgement. If we are to assume audio-lead and video-lead as processed separately, a two-criterion assumption to synchrony judgement is fitting. Instead of assuming that respondents operate with a single-decision criteria (a stimulus pair is simultaneous or not), one can assume that respondents operate with two (an AV stimulus pair is asynchronous either because audio leads video, or video leads audio). Accordingly, Yarrow et al. (2011) proposed that two criteria on each side of the SOA range, one for audio-lead AV alignments and another for video-lead AV alignments, constitute the outer boundaries within which AV alignments are perceived as synchronous. Thus, an observed deviation in subjective synchrony from objective synchrony can be interpreted as a shift in either or both of these criteria. For example, a PSS that is more video-lead when SOA $n-1$ was video-lead compared to when SOA $n-1$ was audio-lead, can very well be caused by a shift in only one of these criteria. Using the PSS of a Gaussian function as reference for estimating synchrony judgement effects, as is common in previous research on the topic, does not allow for identification of asymmetrical shifts in decision criteria (Yarrow et al., 2011). Fitting two psychometric functions to SJ responses, one for audio-lead SOAs, another for video-lead SOAs, enables the identification of information regarding potential asymmetry in synchrony judgement on either side of the SOA range, audio-lead or video-lead. Yarrow et al. (2011) suggested that the points of audio-lead and video-lead thresholds (ALT and VLT) should be the parameters used to investigate synchrony judgement, as they reflect the two hypothesized decision criteria used. These parameters mark the two points of maximum uncertainty (50% synchrony responses) of subjective synchrony, ALT for audio-lead alignments and VLT for video-lead alignments. Extracting these parameters from two Sigmoid curves fitted to participants responses, allows for estimation of independent shifts in decision criteria on either the audio-lead or the video-lead side of the SOA range.

Aim of the study

The primary objective of this experiment was to further elucidate rapid temporal recalibration by investigating asymmetry across SOA n-1 magnitude of effect. The literature on this topic has primarily relied on PSS as reference point. Assuming a two criterion/dual mechanism interpretation of synchrony judgement, using ALT and VLT as reference points for the audio-lead and video-lead sides of the SOA spectrum, respectively, is more logical. By investigating differences in the degree to which ALT and VLT are influenced by preceding modality order, identification of asymmetry in audio-lead and video-lead perception is allowed. Fitting two sigmoid curves to SJ responses is thus more appropriate than the more conventional Gaussian function (Yarrow et al., 2011). In order to see whether using ALT and VLT provides any additional information than previous studies using PSS, a baseline must first be established. Accordingly, this experiment replicated previous findings of rapid temporal recalibration with PSS. PSS is expected to be more video-lead when n-1 SOA is video-lead, compared to when n-1 SOA is synchronous or audio-lead. Trends in previous experiments showed that visual-lead n-1 asynchronies produce larger effects on PSS than audio-lead n-1 asynchronies (Roseboom, 2019; Van der Burg et al., 2013; Van der Burg & Goodbourn, 2015). Alm and Behne (2013) suggested that audio-lead and video-lead perception are mediated by different variables. Additionally, Cecere et al. (2016) proposed that AV synchrony judgement is determined through two distinct perceptual mechanisms, one for audio-lead alignments, the other for video-lead alignments. Based on this research, serial dependence in AV synchrony judgement is predicted to primarily effect VLT, not ALT. Results were expected to support this prediction if the VLT is more video-lead when n-1 SOA is video-lead compared to perceptually synchronous n-1 SOA, while ALT remains unaffected.

Methods

Design

A classical SJ task was used in this study to measure perception of AV synchrony. Using a repeated measures design, participants were presented with an AV stop consonant syllable /ba/ with varying degrees of discrepancy between the audible and visual signals, and were instructed to judge whether they were synchronous or asynchronous. Based on the percentage of “synchronous” responses for each SOA, curves can be fitted to the data for each participant, allowing extraction of relevant parameters PSS, ALT and VLT. 21 different SOAs were used; 10 SOAs being audio-lead up to a maximum of 400ms discrepancy, 1 physically synchronous SOA, and 10 video-lead SOAs up to a maximum of 400ms discrepancy. The aim of this study was to investigate the effects of preceding audiovisual stimuli on subsequent synchrony judgements, so each parameter extracted from the SJ task must be compared based on the modality order of the previous trial. In order to do this, all 21 SOAs must precede each individual SOA. This is because parameters PSS, VLT and ALT are extracted using the entire SOA range, averaged across participants. Accordingly, each participant judged the relative simultaneity of 21 SOAs x 21 n-1 SOAs, 441 unique trials in total.

Participants

Thirty-three students at NTNU were recruited for the experiment, of which 29 were included in the final analysis. Three of those recruited were excluded from the experiment because of failure to satisfy requirements. Additionally, data from one male participant was excluded from analysis because of seemingly random responses. The participants were all between the ages 20-28 years ($M= 22,86$). Out of the sample used for analysis ($N=29$), 21 were female (72%), 7 were male (24%) and 1 didn't respond (3%). Participants in the experiment were all right-handed, had Norwegian as their native language, and demonstrated adequate hearing and vision. Participants also provided written consent.

Age, gender and native language were determined by having participants answer a self-administered questionnaire. The questionnaire also included questions regarding alcohol consumption in the last 24 hours, musical experience, quality of sleep the night prior, time spent playing video games, and use of medication. Only age and native language, however, was relevant for inclusion in the experiment. Handedness was established by answering a revised

version of the Edinburgh Handedness Inventory (Oldfield, 1971). Only participants who were right-handed were included in the study. Adequate vision was established by use of the Snellen chart. Binocular visual acuity of 20/25 or better was required for inclusion. In addition to the Snellen test, eye dominance was also identified (Miles, 1929), but this was not a criterion. Adequate hearing was established with an audiometry evaluation. Successful identification of sounds with frequencies between 250 to 4000Hz of a dB higher than 15 for sounds was required for inclusion. Both pre-tests and the experiment were conducted in the speech-lab of the Department of Psychology at The Norwegian University of Science and Technology (NTNU), Trondheim.

An *a priori* power analysis was conducted using SPSS version 27 to determine the sample size needed for adequate statistical power. The sample size required to achieve a statistical power of 95%, predicting a medium effect size and using a significance criterion of $\alpha=.05$, was $N = 6$. The obtained sample size of $N = 29$ was therefore more than adequate for this experiment.

Materials

The material used in this study was repurposed from recordings originally produced and edited by Alm and Behne (2013). The material is an audiovisual recording of a female speaker uttering a stop consonant syllable /ba/. Audiovisual speech was used as stimuli instead of, for example, beep/flash stimuli, because speech provides visual information, such as movement of articulators, that predict auditory input. The specific syllable /ba/ was used because the labial stop [b] provides a visually noticeable cue that can be used as a temporal reference point in synchrony judgement (Alm & Behne, 2013). The recording was then edited to create various asynchronous alignments between the video and audio signals, of which 21 were used for this experiment.

The AV recording was performed at the Speech Laboratory at the Department of Psychology, NTNU. The recording was of a young female, native Norwegian speaker, filmed from the shoulders and above. Distracting objects were removed from the frame, such as earrings, necklaces and glasses. Stress and pitch of voice was neutral, flat in intonation. Visual distractors such as eye movements and blinks, and other facial gestures were kept to a minimum. The recording was conducted in a sound-insulated room. Video was captured by a PDWF800

Sony Professional XDCAM HD422 Camcorder, positioned approximately 2 m in front of the speaker. Audio was captured by two Røde NT1-A microphones positioned in front of the speaker at the height of her knees, one connected to the camera, the other connected through a RME FIREFACE 400 to an Apple Macintosh G5 computer. Using Praat version 5.1, two audio channels were recorded at a 48 kHz sampling rate from the external microphone.

The recordings resulted with 10 iterations of a 30 fps, 1920 x 1200 pixel resolution MPEG-4 video file with corresponding internal audio, segmented by use of AVID Media Composer 3.5 software into 1400 ms video clips. The one used in this experiment was rated independently by two appointed judges as best fit considering various criteria. External audio was segmented and edited by use of Praat version 5.1. Using Logic Pro 8.0.2 digital audio workstation, audio from the external microphone was synchronized with that from the internal audio of the video cameras microphone. AVID Media Composer was then used to replace the internal audio with the external, and further to manipulate the audio onset to create asynchronous AV alignments. AV alignments were made by moving the audio segment in 40 ms increments up to a total misalignment of 440 ms before and after original onset. In total, 23 video files with each their own unique AV alignment was made, of which 21 were used for this experiment.

Procedure

The experiment was conducted in the Speech Laboratory at the Department of Psychology, NTNU, Trondheim. Participants were seated facing an iMac monitor (27 in., 5120 x 2880 pixel resolution, 60 Hz refresh rate), at a distance of approximately 70 cm. They were instructed to sit comfortably but try to keep movements to a minimum and to continuously lean on the back rest. Audio signals were conveyed through AKG K271 studio headphones, at 68 dBA. Stimuli were presented and responses logged using Superlab version 6.2, responses were given using a Cedrus RB-740 response pad. The response pad has 7 buttons, but only 2, labeled “sync” and “async,” were used in the experiment. Participants were instructed to press the “sync” button if they perceived the audiovisual stimuli as synchronous or press the “async” button if not. The order at which the labeled buttons were presented, left to right, was switched between each participant to control for any potential extraneous effects.

The experiment consisted of 450 trials divided into 3 parts, with each part further divided into 3 blocks. Because of these breaks, parts 2 and 3 started with the same SOA with which the

previous part ended. The same applied for breaks within each part. This was to ensure that the hypothesized effect of preceding stimuli on subsequent judgement did not disappear in the process of starting up the next part of the experiment. Accordingly, although the experiment only consisted of 441 unique trials, the total amount of trials for each participant was 450, divided into 1 part with 144 trials and 2 parts with 153 trials. Participants were given a 30 second break between each block, 6 breaks in total. Although participants were instructed to give a response as quickly as possible after stimuli presentation, no upper time limit for a response was specified. The next trial began only after a response was given or after the 1400 ms video clip finished, whichever came last. The experiment started with an introduction explaining the task, including 4 practice trials. Practice trials were used to familiarize the participants with the response pad. The experiment took approximately 30 minutes to complete, with an additional 30 minutes for pretests and questionnaires.

Results

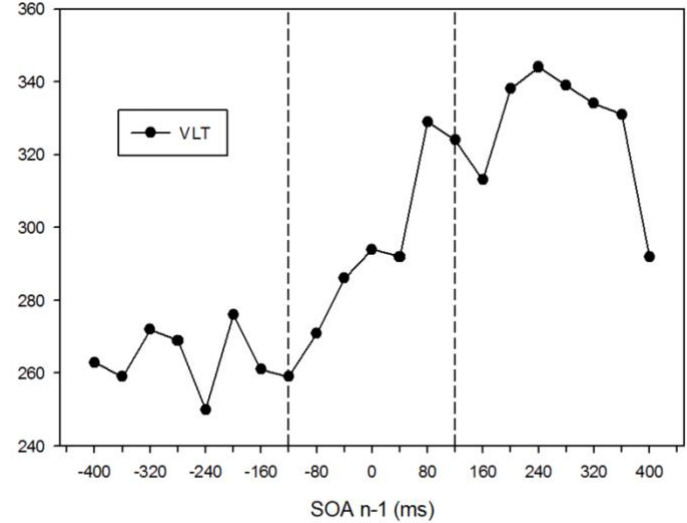
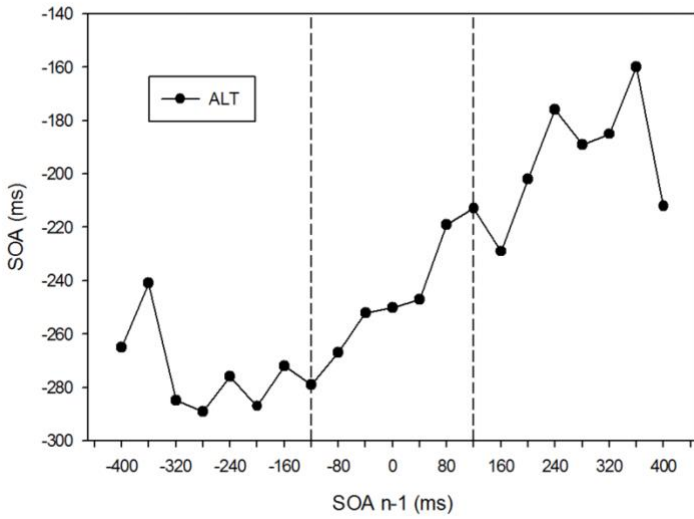
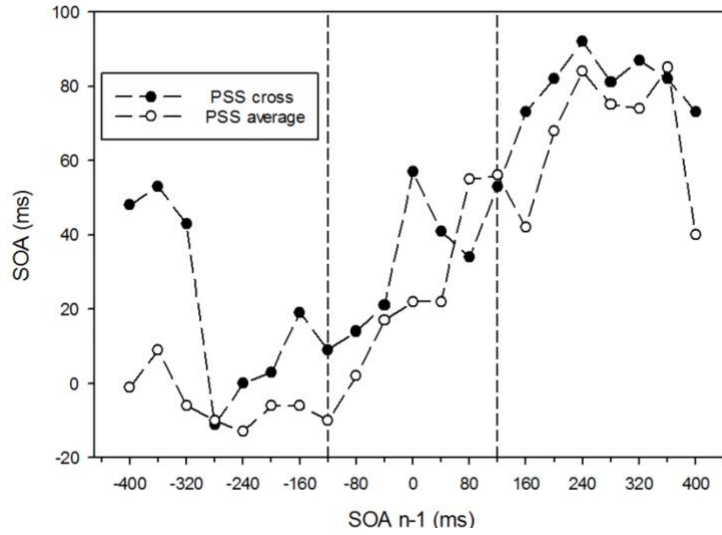
Using Matlab R2021b, data from each participant was de-randomized and reformatted. Percentages of synchronous responses for each participant was plotted for all 21 AV alignments. Two sigmoid curves were fitted to the data, one for audio-lead responses, the other for video-lead responses. The conventional approach to modelling SJ-responses is to use a single psychometric function, specifically a Gaussian one. However, as this study aimed to enshrine the possibility of dual mechanisms for recalibration, one for audio-lead, another for video-lead, and thus assumed a two-criterion approach to synchrony judgement, fitting two independent psychometric functions to respective sides of the SOA range was logical (Yarrow et al., 2011). Four parameters were then extracted from the curves: ALT, VLT, PSS_{cross} (the point at which the S-curves cross) and $PSS_{average}$ (the average between ALT and VLT) as per Yarrow et al. (2011).

In order to investigate how the modality order of a previous trial (n-1 SOA) affects perceived synchrony on a current trial, the distribution of percent synchronous responses were compared across n-1 SOA range. Figure 1 illustrates how synchrony judgement, reflected through parameters PSS_{cross} , $PSS_{average}$, ALT and VLT, changed based on the SOA of a preceding trial. In this study, the n-1 SOA range was divided in 3 windows: audio-lead asynchrony (-400

ms to -120 ms), perceived synchrony (-120 ms to 120 ms) (based on the approximate width of the temporal binding window), and video-lead asynchrony (120 ms to 400 ms). Using a repeated measures ANOVA, differences in group means between these 3 SOA n-1 windows could be analyzed with relevant parameters ALT, VLT, PSS_{cross} and PSS_{average}, the windows being the independent variable, the parameters being the dependent variables.

Figure 1

Changes in PSS_{cross} , $PSS_{average}$, ALT and VLT as a function of SOA $n-1$



Note. Each dot represents the specific AV alignment (y -axis) for which PSS_{cross} , $PSS_{average}$, ALT or VLT were defined, averaged across 29 participants, based on a preceding AV alignment (x -axis).

^a Dashed vertical lines reference the divisions that were made to the SOA $n-1$ range, that define the 3 windows used for analysis.

Using IBM SPSS Statistics version 27, 4 repeated measures ANOVAs were conducted, comparing the mean differences, for relevant parameters, between when SOA n-1 was audio-lead, perceptually synchronous and video-lead, averaged across all participants. As expected, the mean difference in PSS_{average} was significant, $F(1.49, 41.85) = 22.11, p < .001$, between SOA n-1 intervals. Bonferroni post hoc test results show that PSS_{average} was significantly more video-lead when SOA n-1 was video-lead ($M= 26\text{ms}, SD= 44$), compared with when SOA n-1 was audio-lead ($M= -10\text{ms}, SD= 64$), $\Delta M = |36|, p < .001$, or perceptually synchronous ($M= 8\text{ms}, SD= 54$), $\Delta M = |18|, p = .001$. For PSS_{cross} , however, difference between means was not significant, $F(2, 56) = 0.92, p = .403$. PSS_{cross} was non-significantly more video-lead when SOA n-1 was video lead ($M= 36, SD= 47$), compared with when SOA n-1 was audio-lead ($M= 21\text{ms}, SD= 69$), $\Delta M = |16|, p = .753$, or perceptually synchronous ($M= 33\text{ms}, SD= 53$), $\Delta M = |4|, p = 1.000$. The mean difference in VLT was significant, $F(1.44, 40.41) = 16.77, p < .001$. VLT was significantly more video-lead when SOA n-1 was video-lead ($M= 277\text{ms}, SD= 68$), compared with when SOA n-1 was audio-lead ($M= 246\text{ms}, SD= 81$), $\Delta M = |30|, p < .001$, or perceptually synchronous ($M= 260\text{ms}, SD= 74$), $\Delta M = |17|, p = .001$. The mean difference in ALT was also significant, $F(1.55, 43.30) = 15.26, p < .001$. ALT was significantly more video-lead when SOA n-1 was video-lead ($M= -224\text{ms}, SD= 74$), compared with when SOA n-1 was audio-lead ($M= -266\text{ms}, SD= 109$), $\Delta M = |43|, p < .001$, or perceptually synchronous ($M= -243\text{ms}, SD= 86$), $\Delta M = |19|, p = .008$.

The most noticeable results from the experiment, considering our hypotheses, are first and foremost that all parameters were more video-lead when preceding stimuli were video-lead, rather than audio-lead. These findings corroborate previous observations on rapid temporal recalibration in AV synchrony judgement (Roseboom, 2019; Van der Burg et al., 2013; Van der Burg & Goodbourn, 2015). Surprising, however, is that no significant differences were found for PSS_{cross} , while significant differences were found for PSS_{average} . Additionally, significant difference was found in ALT between all SOA n-1 windows.

Discussion

Several studies have revealed an asymmetry of the AV temporal binding window, with it appearing more malleable to video-leading stimuli, while static to audio-leading stimuli (Alm & Behne, 2013; Behne & Wang, 2018; Cecere et al., 2016). Indeed, several studies on rapid

temporal recalibration has found that recalibration effects appear greater for video-lead asynchronies, than for audio-lead asynchronies (Roseboom, 2019; Van der Burg et al., 2013; Van der Burg & Goodbourn, 2015). Furtherer investigation into the mechanisms facilitating this asymmetry is important, considering that dysregulation of the temporal binding window has been associated with neurodevelopmental disorders such as autism spectrum disorder and schizophrenia (Stevenson et al., 2014; Wallace & Stevenson, 2014; Zhou et al., 2021). Accordingly, the aim of this study was to investigate the possibility that synchrony judgement of AV stimuli is mediated by two independent mechanisms, one for audio-lead asynchronies, the other for video-lead asynchronies, as proposed by Cecere et al. (2016). In order to accomplish this, two independent S-curves were fit to participants responses, as opposed to the more conventional Gaussian curve, so that independent shifts in video-lead and audio-lead synchrony judgement could be identified. ALT and VLT was then extracted to reflect decision criteria on either side of the SOA range (Yarrow et al., 2011). The objectives of this study were first to replicate findings on rapid temporal recalibration of AV stimuli, using two independent S-curves instead of a Gaussian. The second, and primary objective, was to compare the degree to which ALT and VLT was affected by preceding modality order. Asymmetry in the degree to which these two parameters were affected by rapid temporal recalibration, would support a separate mechanisms interpretation of temporal AV integration.

Rapid temporal recalibration replicated

The first hypothesis for this study was that the PSS would be more video-lead when the preceding modality order was video-lead, as opposed to audio-lead or perceptually synchronous. The findings of this study support this hypothesis, with average PSS_{average} being 26 ms when SOA n-1 was video-lead, compared to 8 ms for perceptually synchronous SOA n-1 and -10 ms for audio-lead SOA n-1. A similar trend, however less pronounced and non-significant, was found for average PSS_{cross} , being 36 ms for video-lead SOA n-1, compared to 33 ms and 21 ms for perceptually synchronous and audio-lead SOA n-1, respectively. These findings indicate that a singular exposure to an AV stimulus pair affects subsequent synchrony judgement, replicating previous research on rapid temporal recalibration (Roseboom, 2019; Van der Burg et al., 2013; Van der Burg & Goodbourn, 2015). Specifically, recent exposure to AV asynchrony elicits

negative aftereffects, wherein subsequent judgements of synchrony are shifted towards the previous experience (i.e. as SOA $n-1$ becomes more video-lead, so does the PSS).

The reasons for the difference in the degree to which PSS_{average} and PSS_{cross} are affected by preceding modality order, can be many. A larger between-participant variation could explain why differences between average PSS_{cross} were not significant. However, the preferred interpretation of this study is that, because PSS_{cross} is defined as the crossing point of two fitted S-curves, the effects of serial dependence could, in part, be mediated by a difference in the steepness of the slopes for audio- and video-lead asynchronies. For example, it is entirely possible that a shift in PSS_{average} could be caused by the steepness of the curve of the video-lead side of the SOA range evening out and increasing the SOA range it encompasses, while the slope of the audio-lead side maintaining a fixed angle. For this scenario, in theory, no shifts in PSS_{cross} would be observed. Perhaps, then, future research on rapid temporal recalibration should aim to capture a measure of the steepness of the audio- and video-lead slopes, if two psychometric functions are used. For example, using the time-intervals within which audio- and video-lead asynchronies are judged as 25% and 75% likely to be synchronous. Differences in the degree to which these intervals expand or contract can then be observed between audio- and video-lead asynchronies, affording valuable information about the asymmetry of AV temporal binding.

The fact that this study has replicated findings on rapid temporal recalibration is important as it serves as baseline for which we can determine if using ALT and VLT serves greater purpose in examining temporal AV binding than using PSS. It is assuring to know that the differences between SOA $n-1$ windows that have been discovered in this study most likely reflect an already established phenomenon, and that the sample we have used that has produced the data for this study, likely does not differ from those used in other studies on the same phenomena. Accordingly, if any salient characteristics in the way ALT and VLT were affected by SOA $n-1$, that differ in comparison to PSS, were discovered, we could confidently interpret that as a manifestation of some undiscovered aspect of how rapid temporal recalibration works, rather than explanations stemming from differences in sample, experimental design or materials used.

Differences between ALT and VLT

The second hypothesis of this study was that the predicted effects of a preceding AV modality order on subsequent synchrony judgement, would primarily be mediated by shifts in VLT, not in ALT. Results from this study does not support this hypothesis. Both ALT and VLT were significantly more video-lead when SOA n-1 was video-lead, as opposed to audio-lead or perceptually synchronous. Accordingly, findings from this study do not affirm a theory of separate mechanisms mediating temporal integration of AV stimuli. However, this study does not invalidate such a theory either. Indeed, asymmetry across the SOA range was still detected. For instance, between participant variance was larger for ALT than for VLT, which could indicate that ALT is specific for each participant, reflecting variance in the degree to which participants are experienced with AV stimuli. This fits well with previous research on ALT (Alm & Behne, 2013; Behne & Wang, 2018). One could therefore speculate whether the fact that our sample consisted of relatively young participants causes larger discrepancy in between-participant ALT. Perhaps if we used older participants, who are then assumed to have more experience with AV stimuli, and thus a more fine-tuned ALT, would yield results that favor a separate-mechanisms interpretation. Future research should compare rapid temporal recalibration, using ALT and VLT as parameters, between younger and older participants.

Another observed asymmetry is that the degree to which ALT and VLT is affected by preceding modality order, was contingent on which modality came first. The difference in ALT was larger between when SOA n-1 was audio-lead and perceptually synchronous ($\Delta M = 24$ ms), than it was between when SOA n-1 was perceptually synchronous and video-lead ($\Delta M = 19$ ms). Conversely, the difference in VLT was larger between when SOA n-1 was video-lead and perceptually synchronous ($\Delta M = 17$ ms), than it was between when SOA n-1 was perceptually synchronous and audio-lead ($\Delta M = 13$ ms). These findings indicate that ALT and VLT are affected by preceding modality order primarily when the leading modality corresponds with their side of the SOA range. In other words, ALT is most affected by preceding stimuli that is audio-lead, VLT by preceding stimuli that is video-lead. This finding supports observations made by Yarrow et al. (2011). In their discussion, they concluded that when people are subjected to an asynchronous AV stimulus pair, the timing criteria that demarcates synchrony judgement for modality orders like the one to which they are subjected, is relaxed, shifting towards the perceived asynchrony, while the timing criteria for the opposite modality-order remain

unaffected. This interpretation of our findings would support the theory that video-lead and audio-lead asynchronies are processed independently of each other. Unfortunately, however, it is impossible to unequivocally distinguish between whether differences observed in synchrony judgement is caused by shifts in one or both of these criteria, as a manifestation of differences in perceptual latency, or as a combination of these two (Yarrow et al., 2011).

Conclusion

In summary, the findings of this study do not affirm a separate mechanisms interpretation of temporal binding of AV stimuli. Both ALT and VLT were affected by preceding stimuli. Still, the possibility that rapid temporal recalibration of AV stimuli is mediated by two perceptual mechanisms cannot be disregarded based on this study. Asymmetry of the SOA range was still detected, with effects of preceding stimuli being greater for ALT when SOA $n-1$ was audio-lead, and opposite for VLT. This finding indicates that rapid temporal recalibration to AV asynchrony could be caused by independent shifts in timing criteria on either side of the SOA range.

Additionally, differences in PSS_{average} and PSS_{cross} were observed, indicating differences in the total range of AV alignments covered by the two curves for audio- and video-lead asynchronies. This finding suggests, at the very least, that it is reasonable to fit two independent curves to each side of the SOA range, as these curves might change independently of each other, affording valuable information for future deliberation on whether shifts in synchrony judgement is caused by changes in neural processing times or as shifts in timing criteria.

References

- Alm, M., & Behne, D. (2013). Audio-visual speech experience with age influences perceived audio-visual asynchrony in speech. *The Journal of the Acoustical Society of America*, *134*(4), 3001-3010. doi:10.1121/1.4820798
- Behne, D., & Wang, Y. (2018). Does native language temporal experience transfer to audio-visual synchrony perception? *The Journal of the Acoustical Society of America*, *144*(3), 1717-1717. doi:10.1121/1.5067613
- Boer, L., Eussen, M., & Vroomen, J. (2013). Diminished sensitivity of audiovisual temporal order in autism spectrum disorder. *Frontiers in Integrative Neuroscience*, *7*. doi:10.3389/fnint.2013.00008
- Cecere, R., Gross, J., & Thut, G. (2016). Behavioural evidence for separate mechanisms of audiovisual temporal binding as a function of leading sensory modality. *European Journal of Neuroscience*, *43*(12), 1561-1568. doi:<https://doi.org/10.1111/ejn.13242>
- Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nat Neurosci*, *7*(7), 773-778. doi:10.1038/nn1268
- Hay-McCutcheon, M. J., Pisoni, D. B., & Hunt, K. K. (2009). Audiovisual asynchrony detection and speech perception in hearing-impaired listeners with cochlear implants: A preliminary analysis. *International Journal of Audiology*, *48*(6), 321-333. doi:10.1080/14992020802644871
- Helmholtz, H. v., & Southall, J. P. C. (1962). *Helmholtz's treatise on physiological optics : 1-2* (Vol. 1-2). New York: Dover Publications.
- Hillock-Dunn, A., & Wallace, M. T. (2012). Developmental changes in the multisensory temporal binding window persist into adolescence. *Developmental Science*, *15*(5), 688-696. doi:<https://doi.org/10.1111/j.1467-7687.2012.01171.x>
- Keetels, M., & Vroomen, J. (2012). *Frontiers in Neuroscience*
Perception of Synchrony between the Senses. In M. M. Murray & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes*. Boca Raton (FL): CRC Press/Taylor & Francis

- McGrath, M., & Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *The Journal of the Acoustical Society of America*, 77(2), 678-685. doi:10.1121/1.392336
- Miles, W. R. (1929). Ocular dominance demonstrated by unconscious sighting. *Journal of experimental psychology*, 12(2), 113-126. Retrieved from <http://dx.doi.org/10.1037/h0075694>
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97-113. doi:[https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4)
- Pöppel, E., Schill, K., & von Steinbüchel, N. (1990). Sensory integration within temporally neutral systems states: a hypothesis. *Naturwissenschaften*, 77(2), 89-91. doi:10.1007/bf01131783
- Revonsuo, A., & Newman, J. (1999). Binding and Consciousness. *Consciousness and Cognition*, 8(2), 123-127. doi:<https://doi.org/10.1006/ccog.1999.0393>
- Roseboom, W. (2019). Serial dependence in timing perception. *J Exp Psychol Hum Percept Perform*, 45(1), 100-110. doi:10.1037/xhp0000591
- Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory Temporal Integration in Autism Spectrum Disorders. *The Journal of Neuroscience*, 34(3), 691-697. doi:10.1523/jneurosci.3615-13.2014
- Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, 6(2), 171-178. doi:[https://doi.org/10.1016/S0959-4388\(96\)80070-5](https://doi.org/10.1016/S0959-4388(96)80070-5)
- Van der Burg, E., Alais, D., & Cass, J. (2013). Rapid Recalibration to Audiovisual Asynchrony. *The Journal of Neuroscience*, 33(37), 14633-14637. doi:10.1523/jneurosci.1182-13.2013
- Van der Burg, E., & Goodbourn, P. T. (2015). Rapid, generalized adaptation to asynchronous audiovisual speech. *Proceedings of the Royal Society B: Biological Sciences*, 282(1804), 20143083. doi:doi:10.1098/rspb.2014.3083
- von der Malsburg, C. (1995). Binding in models of perception and brain function. *Current Opinion in Neurobiology*, 5(4), 520-526. doi:[https://doi.org/10.1016/0959-4388\(95\)80014-X](https://doi.org/10.1016/0959-4388(95)80014-X)
- Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics*, 72(4), 871-884. doi:10.3758/APP.72.4.871

- Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia*, *64*, 105-123. doi:<https://doi.org/10.1016/j.neuropsychologia.2014.08.005>
- 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. doi:<https://doi.org/10.1016/j.concog.2011.07.003>
- Zhou, H.-y., Cui, X.-l., Yang, B.-r., Shi, L.-j., Luo, X.-r., Cheung, E. F. C., . . . Chan, R. C. K. (2021). Audiovisual Temporal Processing in Children and Adolescents With Schizophrenia and Children and Adolescents With Autism: Evidence From Simultaneity-Judgment Tasks and Eye-Tracking Data. *Clinical Psychological Science*, *0*(0), 21677026211031543. doi:10.1177/21677026211031543