

Candidate number: 10109

Investigating how synchrony perception of audiovisual speech stimulus is affected by rapid temporal recalibration

Bachelor's thesis in Psychology - PSY2900

Supervisor: Dawn M. Behne

May 2022

Candidate number: 10109

Investigating how synchrony perception of audiovisual speech stimulus is affected by rapid temporal recalibration

Bachelor's thesis in Psychology - PSY2900

Supervisor: Dawn M. Behne

May 2022

Norwegian University of Science and Technology

Faculty of Social and Educational Sciences

Department of Psychology



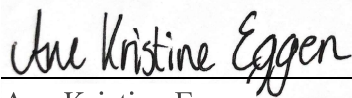
Kunnskap for en bedre verden

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



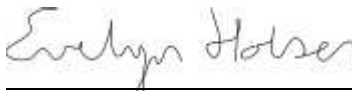
Vegard Dahn

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

Acknowledgements

I would like to thank Professor Dawn M. Behne, Norwegian University of Science and Technology (NTNU), for being a great supervisor of the bachelor project, and for answering and discussing questions and problems throughout the project. I would also like to extend a thank to Professor Peter Svensson (NTNU) and Dr. Darren Rhodes, Nottingham Trent University (NTU), for contributions and help with developing and running the scripts in MATLAB to enable the experiment, and attain parameters for statistical analyses. Also a big thanks to the project group for cooperation.

Abstract

To investigate rapid temporal recalibration, a simultaneity judgement task (SJ-task) was utilized. An audiovisual speech stimulus of the syllable /ba/, was presented with different amounts of audio and video lag, to examine how the previous trial affects the one coming after. The experimental data was split into three divisions. The study employed two cumulative Gaussian curves to examine how four dependent parameters; point of subjective simultaneity (PSS-average and PSS-cross), audio lead threshold (ALT) and video lead threshold (VLT), was affected by recalibration. Based on findings by Roseboom (2019) and Van der Burg et al. (2013), PSS-average, PSS-cross and VLT was expected to be dependent on the previous trial, while ALT was expected to be independent.

The analyses found that PSS-average, ALT and VLT had significant differences across all three divisions, with an increase towards video lead. The analysis of PSS-cross did not find significant differences across the divisions, but showed approximately the same trend as PSS-average. The resulting PSS-average and VLT were as expected, but ALT differed from expectations. An explanation for this might be that when hypothesizing the ALT it was not taken into consideration that (1) the method for fitting the curves might have a larger impact on ALT than PSS and VLT, and (2) in contrast to Roseboom (2019) and Van der Burg et al. (2013) the prior trials were grouped into three divisions, not two. Future research utilizing cumulative Gaussian curves is recommended to apply ALT and VLT together with PSS to build a stronger case for which measure of synchrony perception is preferable. Another suggestion is to examine how not only the previous trial affects the current, but also trials prior to the previous one to better understand how recalibration works and follow up on the findings by Van der Burg et al. (2013).

Introduction

In a multisensory world the brain must combine information from different modalities. Perception of synchrony of events is dependent on how our minds performs this task. Studies have shown that the perception of audiovisual synchrony is flexible (Roseboom, 2019; Van der Burg et al., 2013). Recalibration is one theory for how the brain combines multisensory information by adjusting the "temporal window" dependent on previous experiences (Keetels & Vroomen, 2012). This report set out to seek how rapid temporal recalibration affects the perception of audiovisual synchrony.

Perception of audiovisual information

Human senses continuously receive information about the world. This sensory information, whether from the eyes, ears, nose, mouth or skin, travels to the brain for interpretation, which is how humans perceive the world. When sensory information comes from different modalities such as the ears and eyes (audiovisual), it can be perceived as coming from one single multisensory event, or from several different events.

The term for when the brain connects information from different modalities to perceive it as belonging to one event, is known as the unity assumption (Welch & Warren, 1980). Connecting facial and lip movements with the sound of speech and realise it comes from a single person (i.e. from one multisensory event) is an example of how humans integrate information from different modalities. To assume unity of a sensory event, certain criteria must be fulfilled. One such criterion is timing; a visual and auditory stimuli must arrive to the recipient at about the same time. Another criterion for the assumption of unity is experience (Welch & Warren, 1980). Lightning and thunder are examples of how experience can lead to a strong unity assumption; even though the auditory and visual stimuli arrive at different times, humans still connect the two to one single event.

Light travels much faster than sound in air. Whereas the speed of light is about 300 000 m/s, the speed of sound is approximately 330 m/s. In contrast, neural processing times are usually faster for auditory information (10 ms), than visual (50 ms) (Keetels & Vroomen, 2012). This entails that only audiovisual information occurring at 10 – 15 m will arrive in the brain at the same time. This limit is known as the horizon of simultaneity. Audiovisual events occurring further away than 15 m will due to the relative speeds, have the visual signals arriving before the auditory signals, and events happening within about 15 m, auditory signals will precede visual ones (Keetels & Vroomen, 2012). The difference in arrival times of visual and auditory information makes it necessary for the brain to handle lags between the sensory information.

Beep-flash and audiovisual speech stimuli are commonly used stimuli in audiovisual experiments (Keetels & Vroomen, 2012). Beep-flash stimuli typically consists of a sharp tone and a flash of light presented at the same time or with different amounts of lag between the flash and the sound.

Audiovisual speech stimuli are commonly presented as a video of a person speaking a syllable, a word or a sentence, with the video presented as physically synchronous or asynchronous. Beep-flash stimuli are less complex than audiovisual speech stimuli (Keetels & Vroomen, 2012). In the physical world an audiovisual speech event is more common to occur than a sharp tone and a flash of light. Even though audiovisual speech is more familiar to us, studies have found that such stimuli typically has a larger "temporal window" (i.e. a greater range of stimulus onset asynchronies (SOAs) are perceived as synchronous) than beep-flash. This indicates that the more complex the stimuli, the wider a tolerance for synchrony (wider temporal window) (Keetels & Vroomen, 2012).

Experimental metrics

Stimulus onset asynchronies (SOA), measured in milliseconds (ms), is the amount of asynchrony a stimulus is presented at in an experimental task. $SOA = 0$, is when audio and video are physically synchronous. SOAs are typically considered positive for video lead, and negative for audio lead (Keetels & Vroomen, 2012; Van der Burg et al., 2013; Yarrow et al., 2016). SOA on the previous trial, not the current, is henceforth denoted by $SOA-1$. Accordingly will $SOA-2$ express the trial before the previous trial, and so on. Other researchers have denoted the previous trial ($SOA-1$) as $n - 1$ trial SOA (Roseboom, 2019) or Trial ($t - 1$) (Van der Burg et al., 2013).

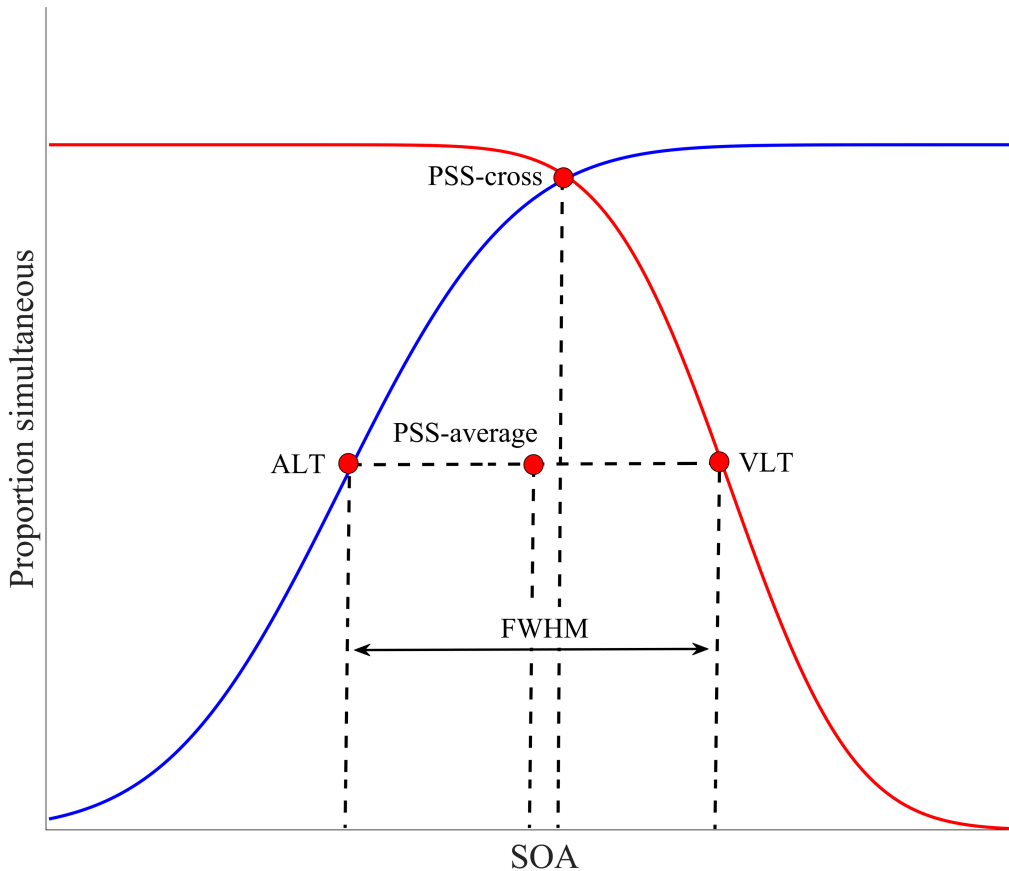
The point of subjective simultaneity (PSS) is presumed to be the SOA a person experiences most synchrony (Keetels & Vroomen, 2012). PSS can be determined from one or two curves, depending on how experimenters choose to best fit the experimental data. Commonly used is the Gaussian distribution (Alm & Behne, 2013; Keetels & Vroomen, 2012; Van der Burg et al., 2013). For this bell shaped curve the maximum value corresponds to PSS (Keetels & Vroomen, 2012). Due to the inherent symmetry of the distribution, the midpoint between each corresponding side point of the bell shape will also give PSS. Roseboom (2019) employs cumulative Gaussian curves as Yarrow et al. (2011) describes. Here PSS is found as the difference between two cumulative Gaussians (i.e. the midpoint between them). In the current study, which applies the same method as Roseboom (2019) described in Yarrow et al. (2011), PSS can be calculated by two different approaches; PSS-cross denoted by the crossing point of the two curves, or PSS-average as the midpoint between the audio lead threshold (ALT) and video lead threshold (VLT).

ALT is the point on the audio lead cumulative Gaussian curve where a person is equally likely to experiences synchrony as asynchrony. VLT is the equivalent on the video lead cumulative Gaussian curve.

Full width at half maximum (FWHM) is a mathematical term for the width of a curve at 50% of the the maximum value. In the research field on perception of audiovisual synchrony, FWHM corresponds to the "window" where synchrony is experienced, and lies between ALT and VLT (Alm & Behne, 2013). Figure 1 illustrates the positioning of PSS-average, PSS-cross, FWHM, ALT and

VLT when two cumulative Gaussian curves were applied.

Figure 1: Graphical representation of two cumulative Gaussian curves.



Note. Illustration of how PSS-average, PSS-cross, FWHM, ALT and VLT places on two cumulative Gaussian curves. The difference in procedure to determine PSS-average and PSS-cross may lead to different SOA values as illustrated for the two.

Perspectives on audiovisual synchrony recalibration

Keetels and Vroomen (2012) presents four different theories, not mutually exclusive, for the brains maintenance of temporal coherence; a window of temporal integration, compensation for external factors, temporal ventriloquism and temporal recalibration. The term recalibration describes a mechanism that adjusts the perception of sensory inputs to better fit with a persons experience. The theory suggests that the brain applies recalibration to minimize discrepancies between the senses, a technique for the brain to handle lags in arrival and processing time (Keetels & Vroomen, 2012). It is a flexible and experience based adaption to what is perceived as synchronous. Recalibration in

the spatial domain was introduced in a book by von Helmholtz (1867), where he demonstrated the flexibility in the visual-motor system by shifts in the visual field induced by wedge prisms (Keetels & Vroomen, 2012). On the other hand, temporal recalibration is a fairly new research field, first studied by Fujisaki et al. (2004) and Vroomen and de Gelder (2004). Both studies exposed participants to a period of several asynchronous sound and light flashes, and both found that PSS shifted towards the exposure lag (i.e. if exposed to visual lead stimuli, visual lead trials would be perceived as synchronous) (Keetels & Vroomen, 2012). The underlying mechanism for temporal recalibration is unknown. A theory suggests a shift in the criterion for simultaneity in the adapted modalities, while another suggests that stimuli that were once synchronous, before adaptation, can become asynchronous after adaptation (Keetels & Vroomen, 2012).

There are several perspectives or models that try to explain temporal recalibration, one such model suggests that only the previous trial has an effect on the current a "n-1" or "SOA-1" model. Roseboom (2019) and Van der Burg et al. (2013) found evidence for such model, but the experiments conducted were set up in a way that if rapid recalibration occurred, this model would fit. The model does not however exclude the option that other models might explain rapid temporal recalibration better. A bayesian model approach would assume that not only the previous trial, but trials before as well, effects how a current is perceived (Vilares & Kording, 2011). Van der Burg et al. (2013) did also analyse their data to look at SOA-2, and found that SOA-2 did affect the perception of the current SOA, which indicates that the "n-1" or "SOA-1" model does not show the whole picture regarding temporal recalibration.

Measures of audiovisual synchrony recalibration

Rapid temporal recalibration, studied by Van der Burg et al. (2013), Roseboom (2019) and Yarrow et al. (2015) among others, showed that recalibration does not need a prolonged period of adaption, but can happen after one single, brief exposure. In these studies the current trial was analysed dependent on the stimulus value on the previous trial (Roseboom, 2019; Van der Burg et al., 2013). The current study will apply the same approach. Both Van der Burg et al. (2013) and Roseboom (2019) employed simultaneity judgement tasks (SJ-tasks) with the audiovisual stimuli being a beep/flash type. Roseboom (2019) also examined rapid recalibration using a temporal order judgement (TOJ) and a magnitude judgement (MJ).

Van der Burg et al. (2013) found that temporal recalibration happens when participants are exposed to a single, short asynchronous audiovisual stimuli, and that not only did the previous trial affect the current, but also the one before that (SOA-1 and SOA-2). The study also found that PSS was best represented by an asymmetrical Gaussian model, because audio lead and video lead SOA-1 altered PSS differently (Van der Burg et al., 2013). Roseboom (2019) found results supporting Van der Burg et al. (2013) in terms of the occurrence of rapid temporal recalibration.

Only a small number of studies have employed cumulative Gaussian curves to fit the experimental data, including Roseboom (2019), Yarrow et al. (2011), Yarrow et al. (2015), and Yarrow et al. (2016), compared with the number of studies whom used Gaussian distributions, for example Alm and Behne (2013), Fujisaki et al. (2004), Van der Burg et al. (2013), and several others. It is unknown if any studies have employed PSS-cross as an alternative to PSS-average (the more commonly used PSS for cumulative Gaussian curves), which is why the current study sets out to use cumulative Gaussian curve fitting, and to employ two methods for determining PSS.

A shortcoming in the current field is that most studies use PSS to evaluate how synchrony is perceived. As Yarrow et al. (2011) suggests, PSS might not be the best measure of synchrony perception. ALT and VLT as Alm and Behne (2013) employs in their study, might be a better measure, due to synchrony being perceived within a range of SOAs, within the "window" between ALT and VLT, and not simply at one single SOA. The current study will therefore employ ALT and VLT as additional parameters to the two PSS.

Current study

The aim of this study is to investigate how recalibration affects the perception of audiovisual synchrony. The experiment uses a simultaneity judgement task (SJ-task) with an audiovisual speech stimulus. The study will employ cumulative Gaussian curves (a type of s-curve) instead of the Gaussian normal curve to best fit the experimental data. It is hypothesized that when SOA-1 is video lead, PSS will be more video lead compared to synchronous and audio lead SOA-1. The study will investigate this by determining PSS by two different mathematical methods. It will also look at VLT and ALT as alternative measurements of perception of audiovisual synchrony. It is expected that when SOA-1 is video lead, VLT will be more video lead compared to synchronous SOA-1. And that there will be no difference between ALT when SOA-1 is asynchronous or synchronous.

Method

Design

A simultaneity judgement task (SJ-task) of an audiovisual speech stimulus of the syllable /ba/ were used to compare participant responses for three different divisions of SOA-1. In the experiment the audio and video were either synchronous, audio preceded video or video preceded audio with a total of 21 different SOAs. Each division of SOA-1 consisted of seven individual SOAs. The study was registered at the Norwegian Centre for Research Data (NSD).

Participants

A power analysis was conducted using the software G*Power to estimate how large a sample size was needed. With an intermediate effect, and an alpha level of 0.5, the analysis showed that the experiment would need a minimum of six participants. Our study obtained thirty native Norwegian speaking participants (21 females, 8 males, 1 undefined) in the age group 20-28 years ($M = 23$ yrs, $SD = 2$ yrs). Participants were recruited at the Norwegian University of Science and Technology (NTNU), and signed up for the experiment through a QR-code or a link that lead to a Google-form for participation. Recruits were then contacted to schedule a suitable time to participate in the experiment.

Certain criteria and pre-tests were evaluated and performed before participation in the experiment. Age, native language, handedness, vision and hearing were criteria for participation. All recruits gave written consent before the pre-tests and experiment were started. The recruits gave information about age and native language in a questionnaire. Participation required an age between 20-30 years, and Norwegian as a native language. Handedness was self reported and evaluated by the Handedness Calculation Tool based on the Edinburgh Handedness Inventory (Oldfield, 1971). Only right-handed was qualified for the experiment. Visual acuity, with the criteria of normal or corrected-to-normal vision, was evaluated with the Snellen test (Watt, 2003), adapted to presentation on a 21.5 in. iMac monitor with ATI Radeon HD 5670 512 MG graphics and a resolution of 1920×1080 pixels. The test measured 13.9×9.8 cm on the screen. Participation in the experiment required binocular visual acuity of at least 20/25 (one flawless attempt at finishing each line of letters down to and including line 7). Hearing acuity was evaluated using a standard pure tone audiometric test in accordance with the audiometric descriptions of the British Society of Audiology (2018). Recruits needed to have a hearing threshold level at 15 dB or below across the frequencies 250, 500, 1000, 2000, and 4000 Hz to be qualified for participation. Other information such as gender, and information that could help explain discrepancies in the experiment data such as tiredness, medication use, musical experience and alcohol intake was collected in a questionnaire, but was not criteria for participation. An eye dominance test, Miles test, was also performed with results being irrelevant for participation. Three recruits were excluded before participation based on the participation criteria. One due to not having Norwegian as a native language, another by being outside the required age group, and a third due to unsatisfactory results for visual acuity.

Materials

The audiovisual stimulus used in the experiment was developed by Alm and Behne (2013) for a study investigating the effect of audiovisual experience, comparing young adults to middle aged adults. The recording, recorded in the Speech Laboratory at the Department of Psychology, NTNU,

was of a young female speaker with an urban East-Norwegian dialect. She spoke the syllable /ba/. Distractions such as jewelry and glasses were removed, and the speaker was instructed to talk with a flat intonation and keep facial movements and expressions to a minimum.

The video was recorded with a PDWF800 Sony Professional XDCAM HD422 Camcorder (Tokyo, Japan) camera positioned approximately 2 m in front of the speaker in a sound-insulated room. The sound was recorded with two Røde NT1-A microphones (Sydney, Australia) positioned in knee height in front of the speaker. One microphone was connected to the camera, the other was fed through a RME FIREFACE 400 (Haimhausen, Germany) to an Apple Macintosh G5 computer (Cupertino, CA), where two audio channels were recorded at a sampling rate of 48 kHz using the software Praat version 5.1 (Boersma and Weenik, 2009, as cited in Alm and Behne, 2013).

The audiovisual stimulus /ba/ was repeated by the speaker ten times in sequence. The video file, in MPEG-4 format, had a resolution of 1920×1200 pixels, and a frame rate of 30 frames/s. Each syllable in the sequence was rated based on 12 criteria (Alm & Behne, 2013), and one were chosen as the best. The audio was edited in Praat, and the syllable /ba/ had a length of 404 ms, measured from consonant release to the end of the vowel.

The stimulus was created by importing the video file into AVID Media Composer 3.5, and substituting the auditory signals recorded by the video camera's microphone with the auditory signals recorded by the external microphone. The video was segmented to make the consonant release during the 13th frame (between 480 and 520 ms). It was then cut to a total length of 1400 ms. The audio signal from the external microphone was synchronized with the audio signal from the video camera's microphone in Logic Pro 8.0.2, before substitution was done. Asynchronous stimuli were created by moving the audio in increments of 40 ms, resulting in 21 different SOAs, ten audio lead asynchronous with maximum of 400 ms, one simultaneous, and ten video lead asynchronous with maximum of 400 ms.

The use of the syllable /ba/ was based on it having a good temporal reference point. The visually salient burst of the syllable is better suited for judgement of audiovisual synchrony compared to for example the syllable /ga/ (Alm & Behne, 2013).

Procedure

The experiment was carried out in the Speech Laboratory at the Department of Psychology, NTNU. The room was kept as quiet as possible. One single or two participants performed the experiment at the same time in different locations in the lab. Participants were asked to remove objects from their mouth (chewing gum, snus, etc.) and to leave their phones or other distractions outside the lab. Participants were seated on a four legged chair at a table facing a 27 in. iMac monitor. The monitor was placed in a 70 cm distance from participants face when participants sat with their back against the backrest of the chair. The iMac monitor had a resolution of 5120×2880 pixels

with AMD Radeon R9 M295X 4GB graphics. The audio signals were presented to participants by studio headphones of type AKG K271 stereo closed dynamic circumaural (Vienna, Austria) at a constant noise level of 68 dBA (corresponding to a frontal incident free-field sound pressure level around 68 dBA).

Participants responded by pressing a key on response box Cedrus RB-730 or Cedrus RB-740. They were allowed to adjust the box inside a pre-drawn frame before the experiment began. The response boxes contained seven keys in a row, only two of which were active during the experiment. Participants were instructed to use the index finger of the left hand for the left key, and the right index finger for the right key. There were two layouts of the response boxes, one with "async" to the left and "sync" to the right, the other with "sync" to the left and "async" to the right. Randomization of which participant received which response box were conducted, with approximately an equal number of participants using each box. Instructions to press "sync" when the participant perceived the video and audio as synchronous, and "async" when not, were given as an instructional text on the monitor before the experiment started. Participants were also instructed to answer their immediate response, and not to take time to think about it. One or two experimenters were seated in the lab, out of sight of the participants, during the experiment to monitor the participants and occurring events. A logbook with information about the experiment and deviations during, were kept. Notes of feedback were also taken after the experiment was finished.

The software MATLAB_R2021b was used to randomize the SOA-1 to make sure all SOA-1 were presented once. Peter Svensson (NTNU) helped prepare the script for this randomization. SuperLab version 6.2 was used to present the randomized order of SOA-1 to the participant and collect the responses. The SJ-task with a video of a woman speaking the syllable /ba/, was repeated throughout the experiment with different amounts of video lead, audio lead and physical synchronous. 21 different SOAs were presented with 40 ms intervals between each SOA, starting at 0 ms going up to 400 ms audio/video lead. The experiment consisted of 21×21 different stimuli, making a total of 441 unique trials used for data analysis. The experiment was split into three parts, pt.1, pt.2 and pt.3, with two sea-life video breaks in each part, and two larger breaks between the parts (a total of eight breaks). After each break the last stimulus before the break was repeated to ensure that all SOA-1 had been presented immediately after one another. Each part had to be started manually by the experimenter. Pt.1 consisted of 144 experimental trials, pt.2 and pt.3 of 153 experimental trials (450 experimental trials in total). Instructions and information about the experiment, together with four example stimuli, were presented on the monitor before the experiment began. The experiment and pre-tests took approximately an hour, the SJ-task alone about 25 min.

Results

Data handling

After the experimental data was collected, it had to be handled before further analyses could be run and results extracted. First MATLAB_R2021b was employed to derandomize SOA based on SOA-1. Then a percentage of synchronous responses were calculated for SOA across all SOA-1, and a Gaussian curve was fit to the data. ALT and VLT were extracted from the curve. Based on the criteria for exclusion of participants; all must obtain an ALT and a VLT, one participant were excluded. This resulted in $n = 29$.

Table 1

Within-Subject Independent Variable Divisions.

SOA-1 divisions	Lower bound [ms]	Upper bound [ms]
SAL	-400	-160
SS	-120	+120
SVL	+160	+400

Note. Divisions; subjective audio lead (SAL), subjective synchrony (SS) and subjective video lead (SVL).

The experimental data were grouped into three divisions, each containing seven different SOA-1. The divisions consist of; subjective audio lead (SAL), subjective synchrony (SS) and subjective video lead (SVL). Table 1 shows the range [ms] of SOA-1 values for each division. The range of the divisions were based on results from previous studies, where each division approximately corresponds to a participants perception of asynchronous audio lead, subjective synchronous and asynchronous video lead (Keetels & Vroomen, 2012; Van der Burg et al., 2013).

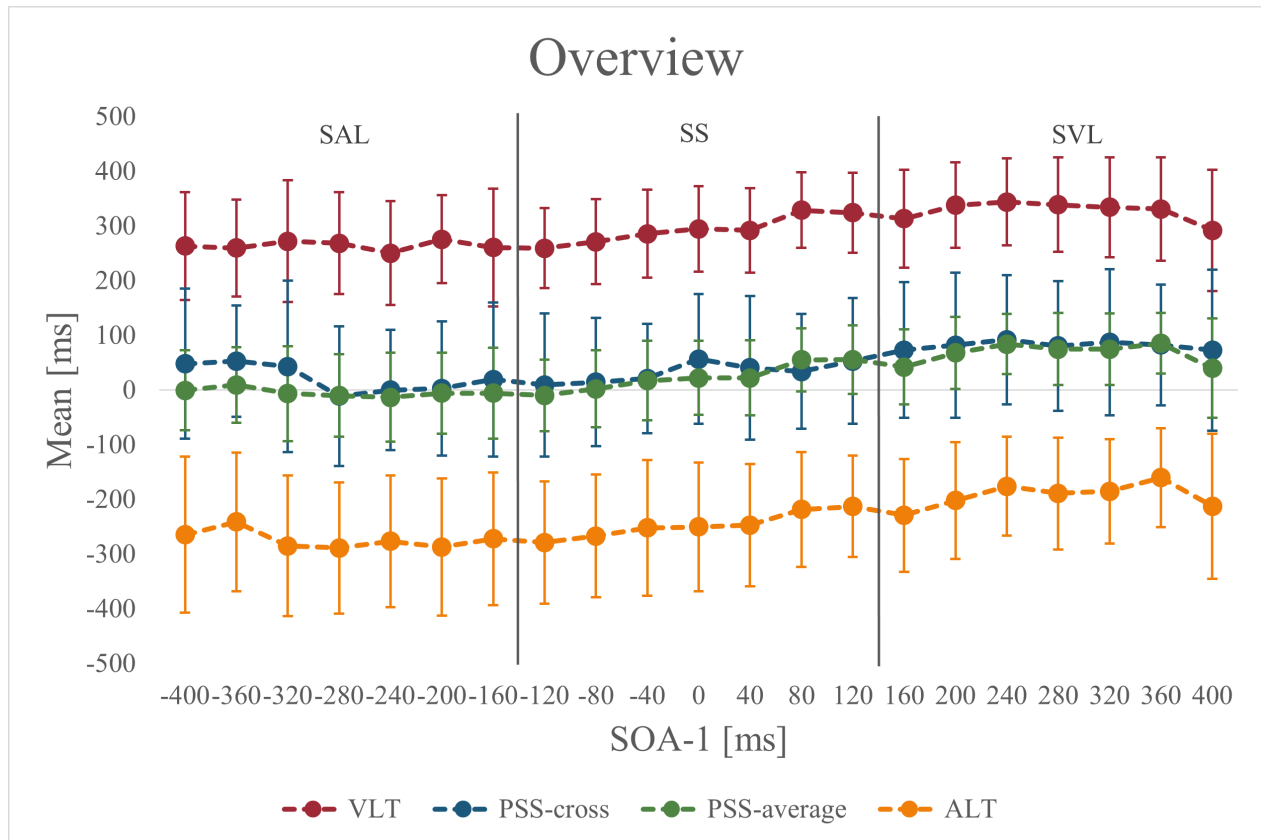
MATLAB_R2021b was used to find the set of two best fitting cumulative Gaussian curves (s-curves) for each participant. Two s-curves curves were used instead of one Gaussian to attain independent values for ALT and VLT, and to extract two different PSS for comparison. The percent synchronous responses was calculated for SOA for each SOA-1 division. For each division, two s-curves curves were fit based on the percent synchronous responses (y-axis) for the SOAs in that division (x-axis). Seed values, pseudo-random values, were used in the process to fit the s-curves. ALT and VLT were extracted for each division (SAL, SS, SVL). ALT from the audio lead cumulative Gaussian curve, VLT from the video lead cumulative Gaussian curve. ALT and VLT, when synchronous and asynchronous responses are equally likely to occur, corresponding to the x-values for which the y-value reaches 50%. Based on ALT and VLT, the other parameters PSS-average, PSS-cross and FWHM were calculated. FWHM is the difference between VLT and ALT. PSS-average can be calculated by $ALT + \frac{FWHM}{2}$. PSS-cross is the x-value for which the video lead

cumulative Gaussian curve and the audio lead cumulative Gaussian curve are equal. One parameter value for each participant in every SOA-1 division was extracted. The script for the curve fitting and extraction of the dependent parameters was prepared and run by Darren Rhodes, Nottingham Trent University (NTU).

Data analysis

Four separate repeated measures analysis of variance (ANOVA) were used to analyse how the within-subject independent variable SOA-1 (SAL, SS and SVL) affected the dependent variables PSS-average, PSS-cross, ALT and VLT. The analyses were done in IBM SPSS Statistics 27. Figure 2 shows a graphic representation of how PSS-average, PSS-cross, ALT and VLT change in respect to SOA-1, and how they compare to each other.

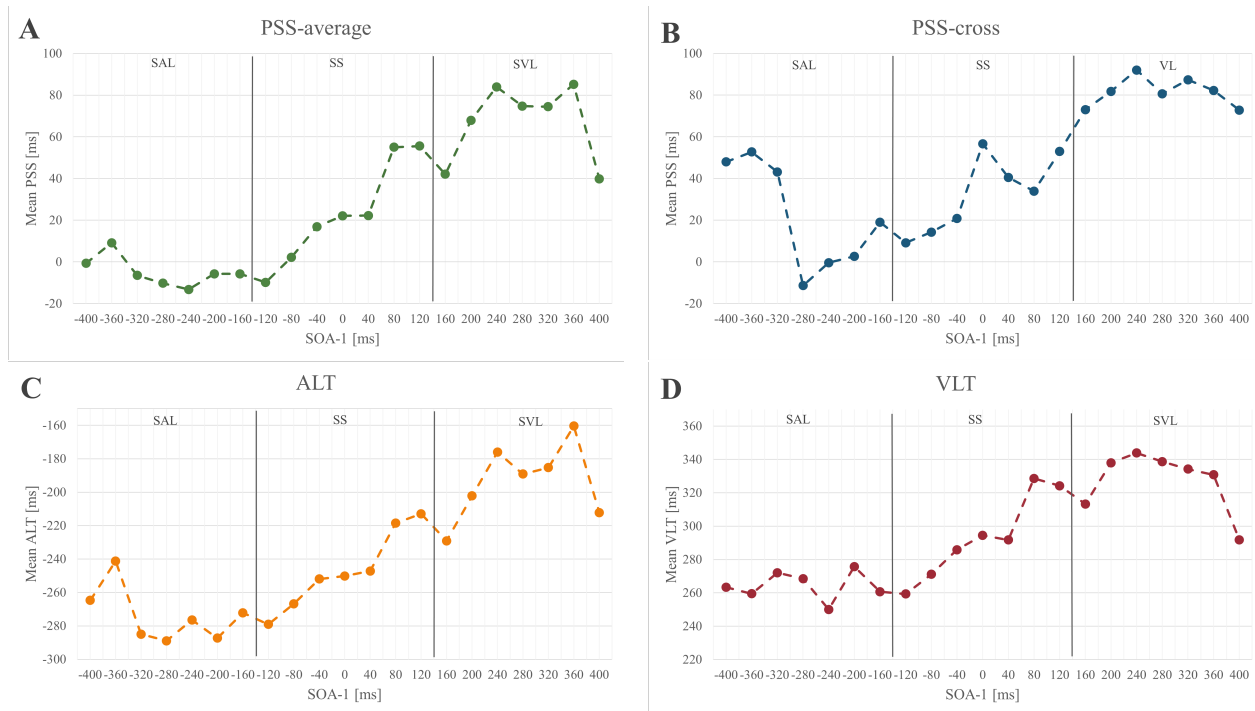
Figure 2: Overview of PSS-average, PSS-cross, ALT and VLT for different SOA-1 values.



Note. The y-axis displays mean values [ms] across participants with error bars for the four dependent variables PSS-average, PSS-cross, ALT and VLT. The x-axis presents the different SOA-1 [ms] with the three divisions; SAL, SS and SVL.

In the analysis of PSS-average Mauchly's test of sphericity indicated that the assumption of sphericity had been violated. A Greenhouse-Geisser correction was applied to correct the degrees

Figure 3: Separate presentation of **A**: PSS-average, **B**: PSS-cross, **C**: ALT and **D**: VLT.



Note. The y-axis displays mean values [ms] across participants. The x-axis presents the different SOA-1 [ms] with the three divisions; SAL, SS and SVL.

of freedom, and the analysis showed a significant difference in PSS-average across divisions with a large effect size, $F(1.49, 41.85) = 22.11, p < .001, \eta_p^2 = .441$. Figure 3 A illustrates the difference in PSS-average across SAL, SS and SVL. Bonferroni post hoc tests indicated that the largest significant difference, $\Delta M = -35.9, p < .001$, was between divisions SAL ($M = -9.6, SD = 64.1$) and SVL ($M = 26.3, SD = 43.7$). The second largest significant difference, $\Delta M = -18.1, p = .003$, was between divisions SAL and SS ($M = 8.5, SD = 53.6$). The smallest significant difference, $\Delta M = -17.8, p < .001$, was between divisions SS and SVL. The analysis showed a general increase in PSS-average across SOA-1 divisions from SAL to SVL.

For the analysis of PSS-cross, sphericity can be assumed based on Mauchly's test of sphericity. No significant differences across SOA-1 divisions was observed for PSS-cross, $F(2, 56) = 0.92, p = .403, \eta_p^2 = .032$. Figure 3 B shows the difference in PSS-cross across divisions SAL, SS and SVL.

Examination of ALT indicated that the assumption of sphericity in Mauchly's test of sphericity had been violated. A Greenhouse-Geisser correction was applied to correct the degrees of freedom, and a significant difference in ALT across divisions, with a large effect size, was found, $F(1.55, 43.30) = 15.26, p < .001, \eta_p^2 = .353$. Figure 3 C illustrates the difference in ALT across SAL, SS and SVL. Post hoc analyses using Bonferroni showed significant differences across the three

divisions. The largest difference, $\Delta M = -42.6$, $p < .001$, was between SAL ($M = -266.5$, $SD = 108.5$) and SVL ($M = -223.9$, $SD = 74.4$). The second largest difference, $\Delta M = -23.5$, $p = .013$, was between divisions SAL and SS ($M = -243.0$, $SD = 86.0$). The smallest difference, $\Delta M = -19.0$, $p = .008$, was between divisions SS and SVL. The analysis showed a general increase in ALT across SOA-1 divisions from SAL to SVL.

In the analysis for VLT Mauchly's test of sphericity indicated that the assumption of sphericity had been violated. A Greenhouse-Geisser correction was applied to adjust the degrees of freedom. The analysis found a significant difference for VLT across divisions, with a large effect size, $F(1.44, 40.41) = 16.77$, $p < .001$, $\eta_p^2 = .375$. Figure 3 D shows the difference in VLT across divisions SAL, SS and SVL. Bonferroni post hoc tests showed significant differences between all SOA-1 divisions. The largest, $\Delta M = -30.1$, $p < .001$, was between group SAL ($M = 246.5$, $SD = 81.5$) and SVL ($M = 276.6$, $SD = 68.5$). The second largest difference $\Delta M = -16.6$, $p = .001$, was between sections SS ($M = 259.9$, $SD = 74.2$), and SVL. The smallest difference, $\Delta M = -13.5$, $p = .015$, was between sections SAL and SS. The analysis showed a general increase in VLT across SOA-1 divisions from SAL to SVL.

Summary of findings

The analyses found that three of the dependent variables, PSS-average, ALT and VLT all had significant differences across all three SOA-1 divisions, with an increase in the mean values from divisions SAL to SVL. For PSS-cross a significant difference across SOA-1 divisions was not found. Figure 3 illustrates the findings.

Discussion

Recalibration was examined by conducting an SJ-task investigating how previous SOAs can affect the perception of the current one. The discussion aims to; (1) interpret the results of the statistical analyses based on the hypotheses, (2) debate the method for handling the experimental data and the extraction of the dependent parameters, and (3) present suggestions for improvements in future studies of recalibration.

Interpretation of analysis results

It was hypothesized that when SOA-1 was video lead, PSS would also be more video lead compared to synchronous and audio lead SOA-1. The ANOVAs of PSS-average and PSS-cross showed that this hypothesis gave a mixed result. While the ANOVA of PSS-average found a significant dif-

ference across SOA-1 divisions with PSS being more video lead when SOA-1 was video lead, the ANOVA of PSS-cross did not find significant differences across divisions.

Comparing the PSS results in Van der Burg et al. (2013) with our PSS-average one can see a clear resemblance in how PSS distributes based on different SOA-1; video lead recalibrates for a larger interval of SOA-1 than audio lead. Van der Burg et al. (2013) show that PSS for SOA-1 is modeled best by an asymmetrical model. This indicates that video lead and audio lead are affected differently by rapid temporal recalibration. This asymmetry was due to a decline in magnitude of recalibration, which decreased slower for video lead than for audio lead, Van der Burg et al. (2013) explains. A reason for this asymmetry in rapid temporal recalibration could be the layout of the brain and physical world. As previously introduced, sound and light travel at different speed rates. For sound to present itself before visual stimuli, audio lead, the stimuli must be within approximately 15 m of the recipient (Keetels & Vroomen, 2012). This is a relatively small area, with little flexibility for audio lead to vary largely. On the other hand, video lead has a much wider area to occur from, approximately 15 m and further away. This leads to larger variations in video lead audiovisual stimuli that could be perceived as synchronous, and can thus be an explanation of why video lead has a wider temporal recalibration field than audio lead and subjective synchronous. It is worthy of note that Van der Burg et al. (2013) used Gaussian distributions to model their experimental data and extract PSS, while this study used two cumulative Gaussian curves and calculated PSS by two different methods.

Although significant differences were not found in the analysis of PSS-cross, the same trend in the data (Figure 2 and Figure 3 B) can be seen as with PSS-average, and PSS from Van der Burg et al. (2013).

It was hypothesized that when SOA-1 was video lead VLT would be more video lead compared to synchronous SOA-1. The ANOVA of VLT showed results in agreement with this assumption. SOA-1 division SVL, subjective video lead, had a significantly larger mean value than division SS, subjective synchrony. This indicates that rapid temporal recalibration occurred, VLT became more video lead as SOA-1 was video lead compared to synchronous. Little research have been found on VLT as a dependent parameter. The hypothesis was consequently based on the asymmetrical window found in Van der Burg et al. (2013). The motivation for using Van der Burg et al. (2013) when hypothesizing the outcome for VLT was the mathematical determination of PSS. For Gaussian distributions, PSS, ALT and VLT are related to FWHM. This relation indicates that ALT and VLT can be evaluated from PSS when Gaussian distributions are applied. This relation exists for PSS-average used in this study as well, but PSS-cross does not mathematically relate to FWHM in the same way.

The analysis of VLT also showed that the largest significant difference was between SOA-1 division SAL and SVL, in accordance with Van der Burg et al. (2013). As previously introduced,

theories such as $n-1$ and the bayesian approach suggests that the previous trial affects the perception of the current one. This study did not set out to seek whether the recalibration effects discovered was due to $n-1$ or the bayesian approach. Although it did not, one may conclude that $n-1$ recalibration effects can be seen. With a different experimental setup the bayesian approach could be investigated, this would require to examine SOA-2, SOA-3 and perhaps even former SOAs.

The hypothesis regarding ALT gave unexpected results. No difference between SOA-1 divisions was assumed, but significant differences were found. The hypothesis based itself on the findings from studies by Roseboom (2019) and Van der Burg et al. (2013). Both used a SJ-task and found similar distributions of PSS dependent on SOA-1. The slopes for audio lead were quite gradual. This lead to the assumption; a change in ALT dependent on SOA-1 will be too small to detect. One can argue that since the hypothesis was based on one study that used Gaussian distributions, and this study used cumulative Gaussian curves, discrepancies may occur. When best fitting cumulative Gaussian curves are employed, ALT and VLT become independent of one another.

Our results could then indicate that when Gaussian distributions are used, and ALT and VLT are forced into a symmetrical relation, VLT has a larger effect on the distribution than ALT. With two cumulative Gaussian curves this forced symmetrical relation of ALT and VLT is avoided, and the slopes of each curve is independent of the other. This could be an explanation for why differences between the three SOA-1 divisions were found in this study, but have not been apparent in Van der Burg et al. (2013). It does however not explain why Roseboom (2019) did not find such differences. An explanation could be that Roseboom (2019) groups the SOA-1 into two divisions, while the current study applies three. With three divisions, one gets a "window" in the middle representing perceived synchrony, which is a better fit with previous findings in audiovisual perception (Keetels & Vroomen, 2012), hence the analyses could better represent the experimental data, and lead to the unexpected change in ALT.

The analyses also found that ALT had larger mean differences between SOA-1 divisions than VLT did. It is worth mentioning that ALT also had larger standard deviations (see error bars in Figure 2). These findings implies that the VLT is more consistent across participants and divisions compared to ALT. This larger difference in ALT supports the theory about VLT having a larger effect on PSS when Gaussian curves are employed.

Method of data handling and extraction of dependent parameters

Our findings indicate that the two different mathematical ways to determine PSS-average and PSS-cross, based on two cumulative Gaussian curves, give results that differ. PSS-average, the midpoint between ALT and VLT, can be found by; $ALT + \frac{FWHM}{2}$. PSS-cross, the maximum value, equals the crossing point of the cumulative Gaussian curves, the point where the participant supposedly experiences most synchrony. The analysis of PSS show that the variance in the data is larger for

PSS-cross than PSS-average (see error bars in Figure 2). This could indicate that PSS-cross gives a more individualistic representation of maximum perceived synchrony, and thus could be a better representation of PSS than PSS-average. Previous research such as Van der Burg et al. (2013) operates with PSS based on the maximum value of Gaussian distributions, which equals the midpoint between ALT and VLT for such symmetrical distributions. The PSS used in Van der Burg et al. (2013) is consequently closely related to PSS-average our study in terms of symmetry. In Roseboom (2019) the experimental data was fit with a difference of cumulative Gaussians as introduced by Yarrow et al. (2011). This approach gives different slopes for two curves cumulative Gaussian curves. PSS found using this method is the average of the synchrony criteria, similar to both PSS in Van der Burg et al. (2013) and PSS-average in the current study. This makes PSS comparable between the studies.

One could also debate whether PSS-average adds any necessary, new information about the data that ALT and VLT does not already provide. Considering how PSS-average is calculated, when analyses of ALT and VLT find significant differences, PSS-average should also show significant results. In accordance with previous research, humans have a wider range of timings where synchrony is experienced, a range of SOAs between ALT and VLT, and not a specific point (Van der Burg et al., 2013; Yarrow et al., 2011). One could therefore debate whether PSS, in any form, is a suitable measurement of synchrony perception, and propose that ALT and VLT are more suited measurements, independent of the preferred method of fitting the data.

The decision to employ two cumulative Gaussian s-shaped curves, instead of one bell-shaped Gaussian curve, to capture the data was based on the findings Yarrow et al. (2011). They suggest that a single Gaussian curve has been widely used due to convenience, not due to its capabilities of capturing the data. Yarrow et al. (2011) found that a version of the general threshold model represents the experimental data for recalibration better than for example the general independent channels model (Yarrow et al., 2011). A benefit of this version of the general threshold model is that the version makes explicit where the difference in slopes arise. A difference in slopes between audio lead and video lead was an important feature the current study wished to adapt. This, to be able to investigate how the difference between ALT and VLT as independent parameters, would be affected by recalibration, and to examine how different versions of PSS could lead to different interpretations of the findings.

During the data handling and extraction of dependent parameters in MATLAB_R2021b, seed values were used in order to find the best fitting cumulative Gaussian curves. An issue was discovered in the use of these. The seed value used for the data set imported to Excel for graphic representation of the experimental data, Figure 2 and Figure 3, was different from the seed value used to extract one value for each participant of PSS-average, PSS-cross, ALT and VLT. This led to the means in the SOA-1 divisions being different for the dependent parameters when comparing

the graphical representation and the statistical analysis done in SPSS. A solution for this issue would be to extract the data set for graphical representation, and the analytical data set from the same output data. Another possible solution would be to program the script to give the same pseudo-random seed value each time.

The audiovisual stimuli used in the experiment consisted of one female speaker only. Considering that the video was repeated an excessive number of times (450 times) within about 25 min, it is plausible to think that concentration was difficult to keep the whole time. Feedback from participants supported this idea. An argument for choosing only one speaker was to keep the variables in the experiment to a minimum. However, for future studies, it might be preferable to have more than one to keep the experiment more interesting for the participant and thus extend the level of concentration.

The same type of issue arise when considering the age limitation of the study. Participants had to be between 20-30 years old in the current study. For future studies, it would be interesting to examine different age groups as well. As Alm and Behne (2013) find; ALT and PSS were dependent on experience. Considering Alm and Behne (2013) found that ALT became more conservative with age, one may suspect that the larger difference in ALT compared to VLT found in the current study could be due to the young age, and low experience level the participants had.

Conclusion

The results of this study support a theory of rapid temporal recalibration. The findings indicate that the previous trial does affect the current, even with no prolonged exposure time. It found that PSS, ALT and VLT increase across SOA-1 divisions, with the most notable finding being the mean change in ALT across divisions. The PSS distribution for the current study is in agreement with results from Roseboom (2019) and Van der Burg et al. (2013). Suggestions for further study is to investigate not simply the previous trial, but trials before as well (SOA-2, SOA-3, etc.) in order to be able to better understand to what degree past experience affects perception of current events (a followup on the findings Van der Burg et al. (2013)). Further use of ALT and VLT in comparison with PSS would be recommended to examine which experimental parameters explains the perception of synchrony best. A final suggestion for future studies would be to investigate the neurological mechanisms for rapid temporal recalibration, which was not the intent of this study.

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. <https://doi.org/10.1121/1.4820798>
- British Society of Audiology. (2018). *Recommended procedure: Pure-tone air-conduction and bone conduction threshold audiometry with and without masking*. Retrieved May 23, 2022, from <https://www.thebsa.org.uk/wp-content/uploads/2018/11/Recommended-Procedure-Pure-Tone-Audiometry-August-2018-FINAL.pdf>
- Fujisaki, W., Nishida, S., Shimojo, S., & Kashino, M. (2004). Recalibration of audiovisual simultaneity. *Nature neuroscience*, *7*(7), 773–778. <https://doi.org/10.1038/nn1268>
- Keetels, M., & Vroomen, J. (2012). Perception of synchrony between the senses. In M. M. Murray & M. T. Wallace (Eds.), *The neural bases of multisensory processes*. CRC press/Taylor Francis.
- Oldfield, R. (1971). The assessment and analysis of handedness: The edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4)
- Roseboom, W. (2019). Serial dependence in timing perception. *Journal of experimental psychology: Human perception and performance*, *45*(1), 100–110. <https://doi.org/10.1037/xhp0000591>
- Van der Burg, E., Alais, D., & Cass, J. (2013). Rapid recalibration to audiovisual asynchrony. *The Journal of neuroscience*, *33*(37), 14633–14637. <https://doi.org/10.1523/JNEUROSCI.1182-13.2013>
- Vilares, I., & Kording, K. (2011). Bayesian models: The structure of the world, uncertainty, behavior, and the brain: Bayesian models and the world. *Annals of the New York Academy of Sciences*, *1224*(1), 22–39. <https://doi.org/10.1111/j.1749-6632.2011.05965.x>
- von Helmholtz, H. (1867). *Handbuch der physiologischen optik*. Leipzig: Leopold Voss.
- Vroomen, J., & de Gelder, B. (2004). Perceptual effects of cross-modal stimulation: The cases of ventriloquism and the freezing phenomenon [Pagination: 950]. In G. Calvert, C. Spence, & B. Stein (Eds.), *Handbook of multisensory processes* (pp. 141–150). MIT.
- Watt, W. S. (2003). *How visual acuity is measured*. Retrieved May 23, 2022, from <https://lowvision.preventblindness.org/2003/10/06/how-visual-acuity-is-measured/>
- Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological bulletin*, *88*(3), 638–667.
- 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., Martin, S. E., Di Costa, S., Solomon, J. A., & Arnold, D. H. (2016). A roving dual-presentation simultaneity-judgment task to estimate the point of subjective simultaneity. *Frontiers in psychology, 7*, 416–416. <https://doi.org/10.3389/fpsyg.2016.00416>
- Yarrow, K., Minaei, S., & Arnold, D. H. (2015). A model-based comparison of three theories of audiovisual temporal recalibration. *Cognitive psychology, 83*, 54–76. <https://doi.org/10.1016/j.cogpsych.2015.10.002>

