Bachelor's thesis

Candidate number: 10105

May 2022

NTNU
Norwegian University of Science and Technology
Faculty of Social and Educational Sciences
Department of Psychology

NTNU
Norwegian University of
Science and Technology

Bachelor’s thesis

2022

NTNU
Norwegian University of
Science and Technology

Candidate number: 10105

Bachelor's thesis
May 2022

NTNU
Norwegian University of Science and Technology
Faculty of Social and Educational Sciences
Department of Psychology

NTNU
Norwegian University of
Science and Technology

Brief exposure to audiovisual asynchrony affects perception of audiovisual synchrony in speech

Candidate number: 10105

PSY2900 Bachelor thesis in psychology - Perceived simultaneity in audiovisual speech perception

30. may 2022 Trondheim

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

I wish to thank Dawn M. Behne for the quality of guidance and supervision we have received during the bachelor thesis in perception of synchrony in audiovisual speech production.
I wish to thank Peter Svensson and Darren Rhodes for help in data formatting.
I further wish to thank the project group for exemplary teamwork during the bachelor thesis.

Abstract

Perception of audiovisual (AV) synchrony is affected by experience with temporal delays between auditory and visual signals. Recent studies using a “simultaneity judgment” (SJ) task, indicate similar changes in perception of AV synchrony following repeated and brief exposure, often referred to as temporal recalibration and rapid recalibration. The brain adapts perception of AV synchrony by altering the temporal binding window to audiovisual experience. Studies show that the perceiver is more sensitive to audio-lead asynchrony than video-lead asynchrony, indicating an asymmetrical TBW biased towards video-lead input. The current study researched the replicability of rapid recalibration and the TBW thresholds´ adaptability to brief AV experience using a SJ task. The stimulus consisted of audio-video-justifications of the audio /ba/. Previous studies indicate that simultaneity judgements are represented more precisely when fitted two cumulative Gaussians than the usually used Gaussian. Two cumulative Gaussians were therefore fitted SJ-responses in the current study. Four different one-way repeated measures ANOVAs showed rapid recalibration effects for ALT, VLT and point of subjective simultaneity (PSSaverage). The analysis showed that brief experience with AV speech affects ALT more than VLT and PSSaverage. Results indicate that two cumulative Gaussians are more sensitive to differences in the SJ-data than a Gaussian. The present study's findings can be further used to develop training programs to adjust the ALT and the VLT; narrowing broadened temporal binding windows correlated with neurological disorders.

Brief exposure to audiovisual asynchrony affects perception of audiovisual synchrony in speech

Perception consists of the brain integrating sensory information from multiple sensory modalities into a unified, coherent and representative percept (Sorati & Behne, 2019). The multisensory nature of perception is exemplified through audiovisual (AV) speech, where the observer's perception is shaped by what is seen and heard (Keetels & Vroomen, 2012). Multisensory integration is dependent on the temporal relationship between multisensory signals (Keetels & Vroomen, 2012). However, the signals do not have to not be physically simultaneous to be perceived as the same event (Conrey & Pisoni, 2006). The brain attenuates to repeatedly experienced sensory input, a process termed sensory adaptation (O´mahony, 1986). Prior experience with AV temporal relationships influences the perceivers current perception of AV synchrony, a phenomenon termed serial dependence (Cicchini et al., 2018). These processes enable the brain to compensate for temporal delays between auditory and visual signals. In a constant multisensory environment these processes are crucial to prevent an overload of irrelevant sensory information and to achieve a coherent AV perception, even with the constant asynchrony in the perceiver´s environment.

Perception of audiovisual synchrony in speech

The mechanisms behind how the brain integrates AV signals into a unified percept remain enigmatic as there is no sensory sense for time (Keetels & Vroomen, 2012). A proposed mechanism is that cortical neural activity synchronizes when multisensory signals are connected to the same event (Shadlen & Movshon, 1999). How the brain identifies which signals derive from the same sensory event is argued to be based on an assumption of unity. The assumption of unity implies that multisensory integration is dependent on the amount of shared amodal properties between modalities (Feenders & Klump, 2018; Keetels & Vroomen, 2012). A paramount amodal property is the relative timing between the auditory and visual signal (Radeau, 1994; Keetels & Vroomen, 2012). An approximately simultaneous arrival between AV signals will lead to AV signals being perceived as deriving from the same event. A sufficient delay between auditory and visual signals will result in perception of two separate AV events (Keetels & Vroomen, 2012). Perceiving AV signals deriving from the same event as synchronous is advantageous as it increases perceptual precision and speed (Van der Burg et al., 2013), as well as reliability and saliency (Keetels, & Vroomen, 2012). Consequently, these advantages may increase learning, differentiation and reaction timing to the signal (Keetels, & Vroomen, 2012).

For AV sensory events to be perceived as synchronous they have to occur within a specific time frame, termed the temporal binding window (TBW) (Zhou et al., 2020). The AV TBW is found to be a few hundred milliseconds in width (Alm & Behne, 2013; Conrey & Pisoni, 2006: van Wassenhove et al., 2006). AV signals with a delay beyond the TBW will be perceived as asynchronous. The width of the TBW is termed full-width-half-maximum (FWHM), representing the distance between the thresholds for perceived AV synchrony and asynchrony (Behne et al., 2013). The audio-lead threshold (ALT) represents the threshold for equal probability of perceiving audio-leading AV input as synchronous and asynchronous. The video-lead threshold (VLT) is the threshold for equal probability of perceiving video-leading AV input as synchronous and asynchronous (Conrey & Pisoni, 2006). The width of the TBW can vary between individuals (Stevenson et al., 2012) as a consequence of the TBW´s flexibility to adapt to experience, with, for example, perception and music (Alm & Behne, 2013). A broadened TBW leads to perceiving unrelated AV signals as occurring from the same event (Zhou et al., 2020). Enlarged TBWs are correlated with neurological disorders, such as autism and dyslexia (Zhou et al., 2020), and impaired verbal (Woynaroski et al., 2013) and non-verbal (Noel et al., 2017) communication abilities.

The point at which AV stimuli is most likely to be perceived as synchronous, referred to as the point of subjective simultaneity (PSS) (Keetels & Vroomen, 2012), is found when visual input precedes auditory input in the presented stimuli onset asynchronies (SOAs) (Alm & Behne, 2013; Conrey & Pisoni, 2006; Dixon & Spitz, 1980; Hay-McCutcheon et al., 2009; Roseboom, 2019; Van der Burg et al., 2013). SOA represents the milliseconds of temporal lag between presented auditory and visual signals. Auditory input can precede visual input (audio-lead), visual input can precede auditory input (video-lead), or the signals can be physically synchronous (0 ms. delay) (Keetels & Vroomen, 2012; Recio et al., 2019). That the PSS is more video-lead than audio-lead SOA indicates that the TBW is asymmetrical as perception of AV synchrony is biased visual signals.

Studies report an asymmetrical TBW as ALT is found closer to physical synchrony (SOA = 0 ms.) than VLT, causing the perceiver detecting audio-lead asynchrony easier than video-lead asynchrony (e.g., Alm & Behne, 2013; Conrey & Pisoni, 2006; Hay-McCutcheon et al., 2009). ALT and the VLT are further found to differ in adaptability to AV experience (e.g., Alm & Behne, 2013; Conrey & Pisoni, 2006; Hay-McCutcheon et al., 2009). Alm and Behne (2013) researched the effect of age-related AV experience on AV asynchrony perception in speech, controlling for sensory acuity and cognitive processing speed. Results showed that age-related AV experience affects ALT more than VLT, by which ALT, and not VLT, fine-tunes AV experience with speech over time. Behne and colleagues (2013) studied if musical experience could similarly affect ALT. Findings were consistent with Alm and Behne´s (2013) study, showing a fine-tuning of ALT, not VLT, in response to AV experience. However, studies finding ALT to be more affected than VLT by AV experience have researched cumulative AV experience, not changes in the thresholds following brief exposure (e.g., Alm & Behne, 2013; Behne et al., 2013). When researching rapid recalibration, Roseboom (2019) and Van der Burg and colleagues (2013) found evidence against a difference in PSS between the stimulus in the previous trial (SOA-1) was audio-lead and SOA-1 = 0 ms., whilst found evidence supporting PSS being larger for video-lead SOA-1 than SOA-1 = 0 ms. The results indicate that VLT is more affected by experience than ALT, although the studies compared shifts PSS and not ALT and VLT. As studies support different hypotheses regarding ALT and VLTs adaptability to sensory input, further research is crucial. The differences between thresholds in adaptability to AV experience and why the brain has adapted perception of AV synchrony towards a natural preference for visual signals can be explained by the perceivers' environment.

Underlying mechanisms behind a visually-biased perception

The primary, proposed underlying mechanism behind findings indicating a visually-biased perceptual system is the brain's ability to adapt to the differences in transmission timing between sound and light (Alm & Behne, 2013). In the physical (outside of the body) environment, light travels faster through air than sound (300 000 000 m/sec vs. 330 m/sec) (Baskent & Bezo, 2011). Neurally (inside the body), auditory input travels approximately 40 ms faster than visual input to the relevant primary cortices (Keetels & Vroomen, 2012). Auditory input and visual input will arrive approximately simultaneously at the primary sensory cortices when the perceiver and sensory event are distanced approximately 10 m. apart, a threshold termed the horizon of simultaneity (Pöppel et al., 1990). Within 10 m. auditory input arrives before visual and thus audio-leading signals will be perceived. Beyond 10 m. visual input arrives before auditory input, producing perception of a video-leading signal, which will increase with rising distance (Pöppel et al., 1990). Audio-lead asynchrony is delimited by neural transmission timing, whilst video-lead asynchrony will vary depending on the distance between the perceiver and the AV event (Alm & Behne, 2013).

Another potential underlying mechanism behind the visually-biased perceptual system is that visual input is used as a predictor for auditory input in AV speech (Grant et al., 2004). Visual information of articulation precedes the auditory speech production with 10-500 ms. (Smeele, 1994), hence visual signals are perceived before auditory signals and can be used to predict the auditory input. Alm and Behne (2013) propose that visual signal´s predictive nature could represent a limit of the perceptual system, whereby visual input has to come before auditory input to achieve AV binding. AV binding entails integration of AV signals to perceive a coherent percept (Odegaard et al., 2017).

The perceptual system may be visually biased, because the brain has attenuated PSS towards the reoccurring, varying video-lead asynchrony to achieve perception of AV synchrony. Alm and Behne (2013) use these underlying mechanisms to explain why ALT is more conservative than VLT. Audio-lead asynchrony is more predictable than video-lead asynchrony, and the increased predictability makes audio-lead asynchrony easier for the brain to learn in comparison to video-lead asynchronies (Alm & Behne, 2013). Video-leading signals are not delimited and vary with distance between perceiver and AV event, thus VLT cannot be as conservative as ALT (Alm & Behne, 2013). How the brain adjusts the TBW to the perceivers environment can be explained by temporal recalibration.

Temporal recalibration

The brain assumes that asynchronous AV signals that repeatedly co-occur origin from the same AV event, and will establish fixed temporal relationships to achieve perception of AV synchrony: a phenomenon termed temporal recalibration (Di Luca et al., 2009; Harrar & Harris, 2008; Vroomen et al., 2004). After repeated exposure to a specific AV temporal delay, the perceiver judges the AV stimulus in the previous trial as less similar to the AV stimulus in the given trial, an effect Roseboom (2019) refers to as a negative aftereffect.

Fujisaki and colleagues (2004) researched temporal recalibration using a simultaneity judgment (SJ) task, where participants judge whether exposed AV signals are perceived as synchronous or asynchronous. Nine participants were repeatedly exposed to a tone pip and a ring flash, with various AV time delays for 3 minutes. After being repeatedly exposed to a specific AV temporal relationship, the perceiver´s PSS shifted towards the leading sense, indicating temporal recalibration.

Fujisaki and colleagues (2004) conducted a second experiment to test if the temporal recalibration effect was due to changes in decision criteria or changes in perception. The experiment consisted of exposing participants to an auditory-induced visual illusion by which participants did not explicitly judge AV synchrony. Temporal recalibration effects were found to affect the temporal tuning of the illusion. Fujisaki and colleagues (2004) argue that this finding suggests that recalibration is due to sensory adaptation, not changes in decision criteria. In support of the notion of recalibration reflecting sensory adaptation, Roach and colleagues (2011) found that recalibration arises from changes in neurons that selectively respond to AV stimuli. Studies therefore indicate that repeated exposure to AV asynchrony alters fundamental sensory properties to achieve perception of AV synchrony (Fujisaki et al., 2004; Roach et al., 2011).

Temporal recalibration has been reported using every relative timing task, including SJ (Fujisaki et al, 2004; Vroomen et al., 2004), temporal order judgment (TOJ)(Vroomen et al., 2004) and magnitude estimation judgment (MJ) task (Roach et al., 2011). In a TOJ task, participants judge the modality order of the AV stimulus, whilst in a MJ task, participants judge the temporal distance between the AV signals. Temporal recalibration has also been demonstrated in perception of speech (Asakawa et al., 2012; Vatakis et al., 2007) and non-speech (Fujisaki et al., 2004). Recalibration is advantageous as it enables the brain to adapt the TBW to achieve perception of AV synchrony for related asynchronous AV signals (Van der Burg et al., 2013) caused by AV signals having dissimilar transmission timing. Consequently, the previously mentioned advantages of perceiving AV synchrony follow.

Rapid temporal recalibration

Recent studies have reported negative aftereffects following brief exposure to AV asynchrony, termed rapid temporal recalibration (Roseboom, 2019; Van der Burg et al., 2013). Rapid recalibration is researched under a serial dependence approach, examining whether the participant's perception of the AV stimulus on a previous trial affects the participant's simultaneity judgment of the AV stimulus on a given trial (Roseboom, 2019).

Van der Burg and colleagues (2013) researched rapid recalibration using a SJ task where 24 participants were briefly exposed to a tone and a visual white ring, across different SOAs. Results showed that PSS shifted towards the leading sense presented in SOA-1, indicating rapid recalibration. The magnitude of shift in PSS was estimated by comparing PSS after exposure versus PSS when SOA = 0 ms. After video-lead SOA-1, AV synchrony occurred when visual input was presented approximately 35 ms before auditory input in the next trial. After audio-lead SOA-1, AV synchrony occurred when visual input was presented approximately 15 ms before auditory input in the next trial (Van der Burg et al., 2013). The finding indicates that the recalibration is asymmetrical, as the magnitude of the shift in PSS was smaller when SOA-1 was audio-lead compared to when SOA-1 was video-lead. Further, audio-lead SOA-1 only significantly shifted PSS for the shortest SOA, whilst video-lead SOA-1 shifted PSS for every SOA (Van der Burg et al., 2013). The results further showed that recalibration does not occur for extremely large SOAs as great delays are most likely from two different AV events and recalibration is therefore not needed.

Roseboom (2019) argued that finding rapid recalibration in perception of AV synchrony was unexpected for two main reasons. First, when comparing Fujisaki and colleagues (2004) and Van der Burg and colleagues (2013) results, both temporal recalibration and rapid recalibration show a similar magnitude of effect, indicating that repeated exposure is not necessary to achieve the effect. Second, positive aftereffects are usually found after brief exposure to AV asynchrony (Alais et al., 2015; Bliss et al., 2017; Corbett et al., 2011; Fischer & Whitney, 2014). Positive aftereffects imply that the perceiver judges the previously experienced AV stimulus as more similar the exposed AV stimulus (Roseboom, 2019). Serial dependence producing positive aftereffects can be accounted for by Bayesian decision models, describing how a given judgment is attracted towards prior experience (Petzschner et al., 2015). Why serial dependence produces negative aftereffects using a SJ is, however, still disputed.

Roseboom (2019) conducted a study assessing the replicability of Van der Burg and colleagues (2013) results by researching whether rapid recalibration occurred using the SJ, TOJ and MJ tasks. Results from the SJ showed that PSS was larger for video-lead SOA-1 than audio-lead SOA-1. Results from TOJ and MJ did not find PSS to be larger for video-lead SOA-1 than audio-lead SOA-1, but rather the opposite effect: PSS was larger for audio-lead SOA-1 than video-lead SOA-1. Roseboom´s (2019) results showed negative aftereffects following brief exposure using a SJ task, whilst positive aftereffects occurred using TOJ and MJ. The finding introduces the question of whether the data is sufficiently represented for all tasks, as negative aftereffects are only reported using a SJ task and all tasks are assumed to measure similar sensory processing mechanisms (Keane et al., 2020).

Representing the data through curves

The majority of research uses a SJ task when researching perception of AV synchrony (e.g., Alm & Behne, 2013; Fujisaki et al., 2004; Recio et al., 2019), and the mean simultaneity judgment responses are usually fitted a Gaussian curve. A Gaussian curve is a normal distribution, symmetrical bell-shaped curve, from which relevant parameters are extracted (Keetels & Vroomen, 2012; Yarrow et al., 2011), such as PSS, ALT and VLT. The asymmetrical TBW with a visually-biased PSS is represented by the Gaussian being more shifted towards video-lead SOA than audio-lead SOA on the graph (Cecere et al., 2016). PSS is estimated as the peak of the curve and ALT and VLT as the 50% point from the y-axis. As the curve is symmetrical, ALT and VLT are represented dependently on each other, a shift in ALT is mirrored VLT and vice versa. Yarrow and colleagues (2011) criticise using a Gaussian for SJ tasks, and argue that the use is solely based on convenience which makes PSS´s representation uncertain.

Cecere and colleagues (2016) support Yarrow and colleagues (2011) criticism towards using a Gaussian. Cecere and colleagues (2016) conducted a study researching AV temporal binding and the potential of separate audio-leading and video-leading mechanisms. Three different groups were trained in audiovisual simultaneity detection with feedback. Each group was given a different leading sense, auditory-leading, visually-leading or both. Results showed that ALT and VLT differed in adaptability to experience and did not affect each other if one leading sense was trained but the other was not. The findings indicate that ALT and VLT are independent mechanisms, which the Gaussian curve does not assume when representing simultaneity judgment responses.

Yarrow and colleagues (2011) propose a different approach, using two cumulative Gaussians. Two cumulative Gaussians represent audio-lead and video-lead as separate curves, and are not delimited by symmetry. ALT and VLT are represented independently, thus shifts in ALT are not mirrored by shifts in VLT and vica versa. Two different approaches to PSS can be extracted. PSS can be estimated based on the crossing point between the two curves, an approach termed PSScrossover. PSScrossover can move independently between ALT and VLT, not delimited by being estimated as the peak of the curve. PSScrossover is affected by the curves' slopes, which represent the perceiver´s sensitivity to AV asynchrony (Keetels & Vroomen, 2012). A steep curve indicates great sensitivity to AV asynchrony, whilst a flat curve implies a lower ability of perceiving AV asynchrony (Keetels & Vroomen, 2012). PSScrossover does not assume symmetrical slopes for the audio-lead and video-lead curves, therefore PSS becomes less dependent on ALT and VLT, than the PSSaverage which is estimated as the mean between ALT and VLT (Roseboom, 2019). Assuming Cecere and colleagues´s (2016) results are correct, the two cumulative Gaussians may better capture the data as ALT and VLT are represented separately, in addition to the possibility to extract an approach to PSS which takes into account the curves' slopes.

The present study

Recent studies show that brief exposure to temporal delays between auditory and visual signal produce similar negative aftereffects as prolonged exposure, however only when using a simultaneity judgment (SJ) task (Roseboom; 2019; Van der burg et al., 2013). The majority of previous research report positive aftereffects following brief exposure to AV temporal delays. The current study researched the replicability of serial dependence producing negative aftereffects using a simultaneity judgment (SJ) task. The present study expected to replicate findings consistent with studies that show that the TBW is asymmetrical with a visually-biased PSS (e.g., Alm & Behne, 2013; Conrey & Pisoni, 2006; Hay-McCutcheon et al., 2009). The present study therefore expected PSS to be more video-lead when SOA-1 was video-lead compared to when SOA-1 is synchronized or audio-lead.

Previous studies researching rapid recalibration have usually assessed changes in perception of AV synchrony solely based on shifts in PSS towards the leading sense in the prior presented stimulus (Roseboom, 2019; Van der Burg et al., 2013), even though a range of AV temporal delays are usually perceived as synchronous (Yarrow et al., 2011; Zhou et al., 2020). The current study researched rapid recalibration by observing shifts in ALT, VLT, PSSaverage and PSScrossover in response to different SOAs. Research indicates that using two cumulative Gaussians will better represent simultaneity judgment responses than a Gaussian, the current study therefore fitted SJ-responses with two cumulative Gaussians, rather than a Gaussian. The current study expected VLT to be more video-lead when SOA-1 was video-lead than when SOA-1 was synchronous, and no difference in ALT between when SOA-1 was synchronous or asynchronous. Based on Roseboom (2019) and Van der Burg and colleagues (2013) indicating that VLT is more affected by brief exposure to AV temporal delays, than ALT.

The present study´s findings may contribute to research on rapid recalibration and support the use of a different curve approach in research on perception of AV synchrony. Findings may further contribute in developing training programs to narrow enlarged AV TBWs by researching how ALT and VLT adapts to brief experience with AV temporal delays.

Method

Design

A simultaneity judgment (SJ) task was used to assess participants sensitivity for audiovisual asynchronous speech stimuli. Participants were instructed to judge if the audio /ba/ and corresponding visual facial articulation was perceived as simultaneous, for stimuli where audio preceded video (audio-lead), audio and video where synchronous, and video preceded audio (video-lead). Rapid recalibration was researched by exposing participants to all combinations of SOA and SOA-1, and examining whether AV stimulus in the previous trial affected the participants' simultaneity judgment in the next trial.

Participants

An a priori power analysis was performed using SPSS to estimate sample size. The significance criterion was $\alpha = .05$ and power = .95. To achieve a medium effect size for the repeated measures, within factors ANOVA, comparing three groups, the sample size had to be at least $N = 6$. Hence, a sample size consisting of $N = 6$ would suffice to research the current study´s hypotheses. 33 people were recruited at The Norwegian University of Science and Technology (NTNU). The experiment's inclusion criteria were having Norwegian as a native language, right handedness and age between 20-30 years. All recruits completed a questionnaire and pretests for assessing vision, hearing and handedness.

Vision was tested using a Snellen test with a size of 13,9 x 9,8 cm, on a 21,5-inch iMac with resolution of 1920x1080 pixels, ATI Radeon HD 5670 512 MG graphics (Watt, 2010). Recruits with a binocular visual acuity of 20/25 or more were included in the experiment. An ocular dominance test was used (Miles, 1929), but was not an inclusion criteria. Hearing was tested using an audiometric test (British Society of Audiology, 2004). Recruits were included in the experiment if their average hearing threshold level was under 15 dB for frequencies ranging from 250 to 4000 Hz. Handedness was assessed using a variant of Edinburgh Handedness Inventory (Oldfield, 1971).

Three recruits were excluded from the experiment. The first recruit was excluded for having another native language than Norwegian, the second for not meeting the criteria for age, and the third for not meeting the criteria for the Snellen test. In total, 30 participants were included in the experiment, 23 (70%) were women, 9 (27%) were men and 1 (3%) did not identify gender. The average age amongst participants was 23 years ($SD = 2$), by which the variance ranged from 8 years, where the lowest was 20 and highest was 28. A fourth participant was excluded after participation in the experiment, this will be explained in further detail under the result section. The study did not require registration from the Norwegian center for research data (NSD). Participants all signed a consent form before participation in the experiment.

Material

The present study used AV recordings of a female speaker filmed from the shoulders up, pronouncing the syllable /ba/, made by Alm and Behne (2013). The speaker had an urban Eastern-Norwegian dialect and was instructed to pronounce /ba/ with a flat intonation to avoid a decline or incline at the end of the syllable. Artificial distractions, for example, glasses and jewelry, were removed before recording. Facial gestures, namely eye blinks, were kept to a minimum.

Different syllables have dissimilar points of articulation and thus varying auditory and visual saliency (Ten Oever et al., 2013). The syllable /ba/ was used in the present study as the visual salient articulation of /ba/ is found to be a better temporal reference point when judging AV synchrony than other syllables, such as /ga/ (Alm & Behne, 2013).

The AV recordings were conducted in a sound-insulated room, filmed with a PDWF800 Sony Professional XDCAM HD422 Camcorder and two Røde NT1-A microphones. One microphone was attached to the camera, and the other was linked to an Apple Macintosh G5 computer. The computer recorded audio channels at a sampling rate of 48 kHz with Praat version 5.1. Resulting MPEG-4 video recordings had a visual quality of 30 frames with a 1920 X 1200 pixels resolution (Alm & Behne, 2013). The video recordings were imported and edited in AVID Media Composer 3.5, where the auditory stimuli recorded by the camera microphone were substituted with the auditory stimuli recorded by the external microphone. Thereafter, the auditory segment was moved in 40 ms increments, creating 21 levels of AV alignments ranging from 400 ms audio-lead to 400 ms video-lead where 10 were audio-lead, 1 was synchronous and 10 were video-lead.

Procedure

The experiment was performed in the Speech Laboratory at the Department for Psychology, NTNU. Participants were instructed to remove any chewing gum and snuff ("snus").

The experiment was conducted in a quiet room with participants seated 70 cm (-/+ 10 cm) facing an iMac with a 27-inch screen, 5120x2880 pixelation, in a four legged chair to minimize movement. Participants were instructed to sit comfortably with a straight back against the back of the chair. Usually worn glasses or lenses were worn during the experiment. AKG K271 stereo closed dynamic circumaural studio headphones were used to transfer auditory signals to ears.

Superlab 6.2 was used in the SJ task to present stimuli and gather judgements of AV synchrony responses. As unintentional asynchrony between AV signals can occur in software and hardware, AV synchrony was examined using an audio/visual device developed by Electrical Geodesics, Inc. (Eugene, OR).

Participants were instructed to give their immediate response to whether the AV stimulus was perceived as synchronous or asynchronous, through a Cedrus RB-740 or RB-730 response pad. In response to perceived synchrony participants were instructed to press "SYNC", and in response to perceived asynchrony press "ASYNC". Participants used one index finger on each response button to control for variation in accuracy between fingers and discrepancies between right and left hands.

Because of the computers limited capacity, the experiment was divided into three blocks. The first block consisted of 141 trials, and the second and third block consisted of 150 trials. As the first trial in each block and after each break was a repetition of the previous trial presented, 9 trials were thus presented but not used, resulting in the presentation of 450 trials. The experiment leader had to restart the experiment after each block. Every combination of different SOA and SOA-1 was presented to research if previous AV speech stimulus affected perception of synchrony in the next trial. The 21 AV alignments were randomized through MATLAB R2021b to ensure that the order of trials did not influence the participants' perception of synchrony. Each participant was exposed to 21 randomized AV alignments across the 3 blocks, ranging from 400 ms audio-lead to 400 ms visual-lead, where 10 were audio-lead, 1 was synchronous and 10 were video-lead.

The experiment in itself lasted approximately 20 minutes, including the pretest the experiment took approximately 1 hour to carry out.

Results

Data formatting

After data collection, the three data files for each block were merged together. The data material was scripted in Matlab R2021b by Peter Svensson and formatted by Darren Rhodes. The 21 AV justifications were derandomized for each participant. The percent synchrony responses were calculated for SOA across all SOA-1. A Gaussian curve was fitted based on the percent synchronous responses (y) for SOA (x), ranging from +400 ms video-lead and -400 ms audio-lead. ALT and VLT were extracted from the Gaussian curve. Each participant had to have a threshold for audio-lead and video-lead to be included in the data analysis. A fourth participant was excluded on this basis.

The 21 AV alignments were divided into three divisions based on SOA-1. SOA-1 division A represented audio-lead asynchrony ranging from -400 ms to -160 ms. SOA-1 division S represented perceived AV synchrony ranging from -120 ms to + 120 ms. SOA-1 division V represented video-lead asynchrony ranging from +160 ms to + 400 ms. The approximation of expected perceived AV synchrony and asynchrony was based on research showing that the TBW is a few hundred milliseconds in width (Alm & Behne, 2013; Conrey & Pisoni; 2006; van Wassenhove et al., 2006). However, individuals' TBW can vary depending on multiple factors, such as experience, age and neurodevelopmental disorders (Zhou et al., 2020).

Based on the three SOA-1 divisions, the percent synchronous responses were calculated for SOA for each SOA-1 division. For each division, two cumulative Gaussians were fitted based on the percent synchronous responses (y) for the SOAs in that division (x). From the two cumulative Gaussians, one curve represented audio-lead and the other represented video-lead. As research indicates that ALT and VLT are separate mechanisms (Cecere et al., 2016), the two cumulative Gaussians may better represent ALT and VLT individually than the Gaussian.

Matlab R2021b was further used to extract four parameters for all analyses, for each of the three divisions. Audio-lead threshold (ALT) and video-lead threshold (VLT) were extracted through ALT being represented on one curve, and VLT on the other. ALT is the threshold for equal probability of perceiving audio-leading AV stimuli as synchronous and asynchronous. VLT is the threshold for equal probability of perceiving video-leading AV stimuli as synchronous and asynchronous (Conrey & Pisoni, 2006). PSS represents the point at which AV stimuli is most likely to be perceived as synchronous (Keetels & Vroomen, 2012). Based on ALT and VLT, two different approaches to PSS were calculated: PSScrossover and PSSaverage. Dawn M. Behne suggested using PSS crossover, estimated as the point where the two cumulative Gaussians cross. PSScrossover does not assume symmetrical slopes for the audio-lead and video-lead curves. Therefore PSScrossover becomes less dependent on ALT and VLT, than the classic calculation of PPS as the midpoint between ALT and

VLT (PSS average) (Roseboom, 2019). The data files were based on each participant and the four parameters for each of the three divisions.

Statistical results

The data collected from the simultaneity task (SJ) was analyzed in IBM SPSS Statistics version 27. Four different one-way repeated measures analyses of variance were performed, by which the independent variables (within-subjects) consisted of the three SOA-1 divisions (Divisions A, S and V) and the dependent variables included the parameters ALT, VLT, PSSaverage and PSScrossover. Table 1 shows the mean differences between the three SOA-1 divisions within each parameter based on Bonferroni post-hoc analyses. The purpose of the analyses was to research rapid recalibration by examining whether the three SOA-1 divisions had an effect on the parameters value.

Mauchly's test of Sphericity was significant for ALT, $p = .009$, VLT, $p < .001$, and PSSaverage, $p = .004$, violating the assumption of sphericity. The degrees of freedom were corrected by reporting Greenhouse-Geiser *F*-ratio. Mauchly's test of Sphericity was not significant for PSScrossover, $p = .504$, meeting the assumption of sphericity, assumed *F*-ratio was therefore reported.

Differences in ALT between SOA-1 divisions

A significant difference in ALT value was found between the SOA-1 divisions ($F(1.55, 43.3) = 15.26$, $p < .001$, $\eta^2_{partial} = .352$). Thus, the expectation of ALT not being affected by SOA-1 was not supported. Figure 1 shows the mean difference in ALT across SOA-1 ranging from 400 audio-lead to 400 video-lead. Shown in Table 1, The largest difference in ALT value was found between division A ($M = -267$, $SD = 109$), and division V ($M = -224$, $SD = 74$). The mean difference in ALT value between division S ($M = -243$, $SD = 86$) and division A was larger than the mean difference between division S and division V. As illustrated in Figure 4, the mean differences in ALT value between SOA-1 divisions were greater than the mean differences for other parameters between SOA-1 divisions,

Differences in VLT between SOA-1 divisions

A significant difference in VLT value was found between the SOA-1 divisions ($F(1.44, 40.41) = 16.77$, $p < .001$, $\eta^2_{partial} = .375$). Figure 2 shows the mean difference in VLT across SOA-1 ranging from 400 audio-lead to 400 video-lead. As shown in Table 1, the largest difference in VLT value was found between division A ($M = 247$, $SD = 82$) and division V ($M = 277$, $SD = 66$). The mean difference in VLT value between division S ($M = 260$, $SD = 74$) and division V was larger than the mean difference between division S and division A. The expectation that VLT of the current trial will be more video-lead when SOA-1 is video-lead, compared to when SOA-1 is synchronous, was thus supported.

Differences in PSSaverage and PSScrossover between SOA-1 divisions

A significant difference in PSSaverage values was found between the SOA-1 divisions, ($F(1.49, 41.85) = 22.11$, $p < .001$, $\eta^2_{partial} = .441$). Figure 3 shows the mean difference in PSSaverage across SOA-1 ranging from 400 audio-lead to 400 video-lead. As Table 1 shows, the largest difference in PSSaverage value was found between division A (M = -10, SD = 64) and division V (M = 26, SD = 44). The mean difference between division S (M = 9, SD = 54) and division A was larger than the mean difference between division S and V. The expectation that PSS of the current trial would be more video-lead for video-lead SOA-1 compared to audio-lead and synchrony SOA-1 was thus supported.

The one-way repeated measures ANOVA showed a non-significant difference in PSS crossover values between the three SOA-1 divisions ($F(2, 56) = .924$, $p = .403$, $\eta^2_{partial} = .032$) (see Table 1).

Table 1

*Mean difference in parameter value between SOA-1 divisions based on Bonferroni post-hoc analyses (*N *= 29)*

Dependent variables		SOA-1 divisions		*p*-value	
		Division S	Division V	Division S	Division V
ALT					
	Division A	23.51	42.55	.013*	<.001***
	Division S	-	19.04	-	.008**
VLT					
	Division A	13.48	30.08	.015*	<.001***
	Division S	-	16.61	-	.001**
PSS average					
	Division A	18.06	35.89	.003**	<.001***
	Division S	-	17.82	-	<.001***
PSS crossover					
	Division A	12.07	15.57	.881	.753
	Division S	-	3.51	-	1.000

Note. The mean difference in ALT, VLT, PSSaverage and PSScrossover value between SOA-1 divisions A, S and V. Division A represents audio-lead ranging from -400 ms to -160 ms. Division S represents synchrony ranging from -120 ms to +120 ms. Division V represents video-lead asynchrony ranging from +160 to +400 ms.
* $p < .05$, ** $p < .01$, *** $p < .001$

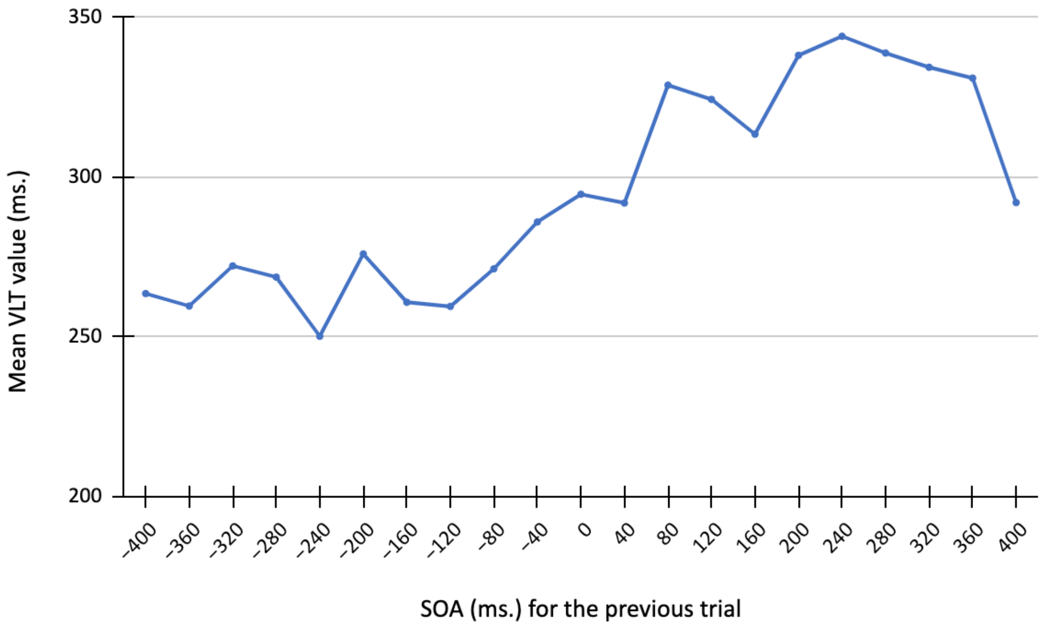
Figure 1

Mean ALT values across SOA-1 ranging from 400 audio-lead to 400 video-lead

Note. Mean ALT values in ms. (y-axis) for each SOA-1 (x-axis) amongst participants (*N* = 29).

Figure 2

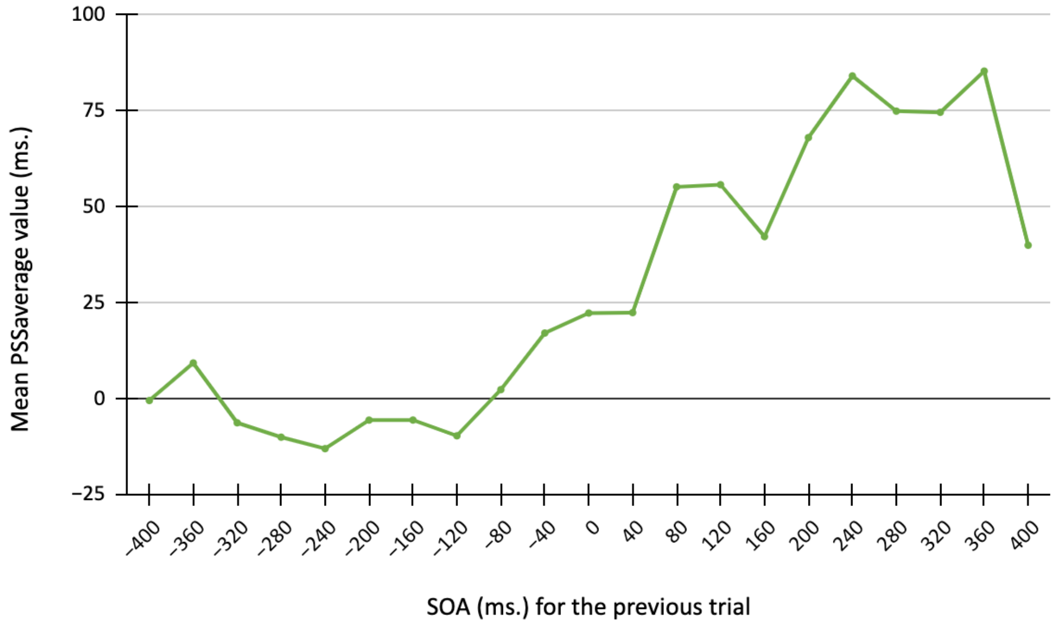
Mean VLT values across SOA-1 ranging from 400 audio-lead to 400 video-lead



Note. Mean VLT values in ms. (y-axis) for each SOA-1 (x-axis) amongst participants (*N* = 29).

Figure 3

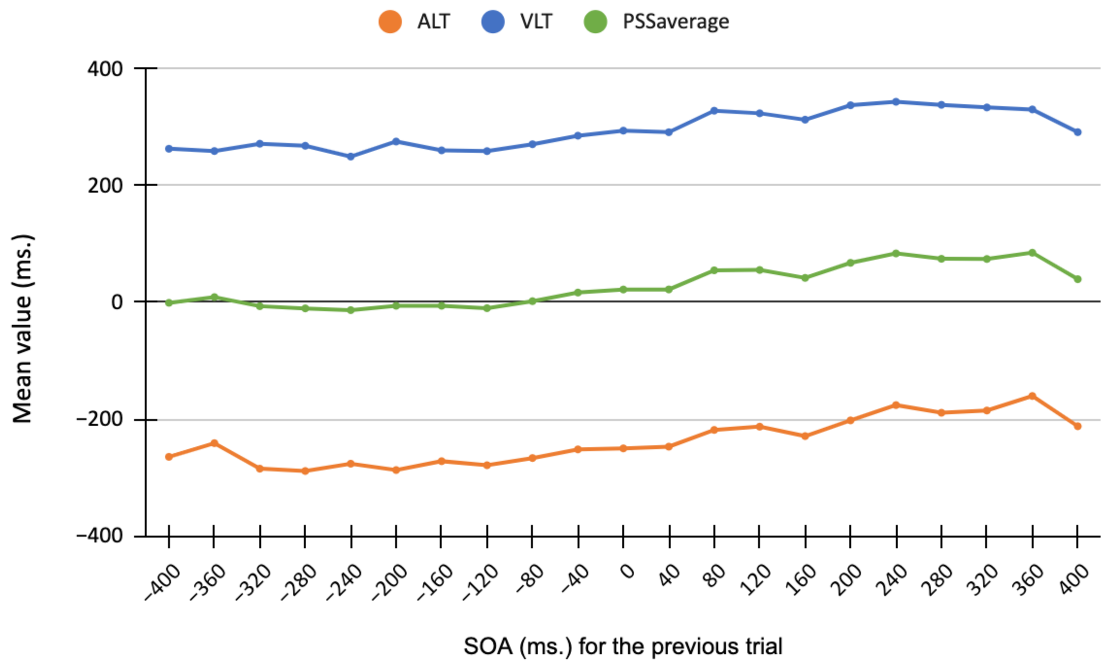
Mean PSSaverage values across SOA-1 ranging from 400 audio-lead to 400 video-lead



Note. Mean PSSaverage values in ms. (y-axis) for each SOA-1 (x-axis) amongst participants ($N = 29$).

Figure 4

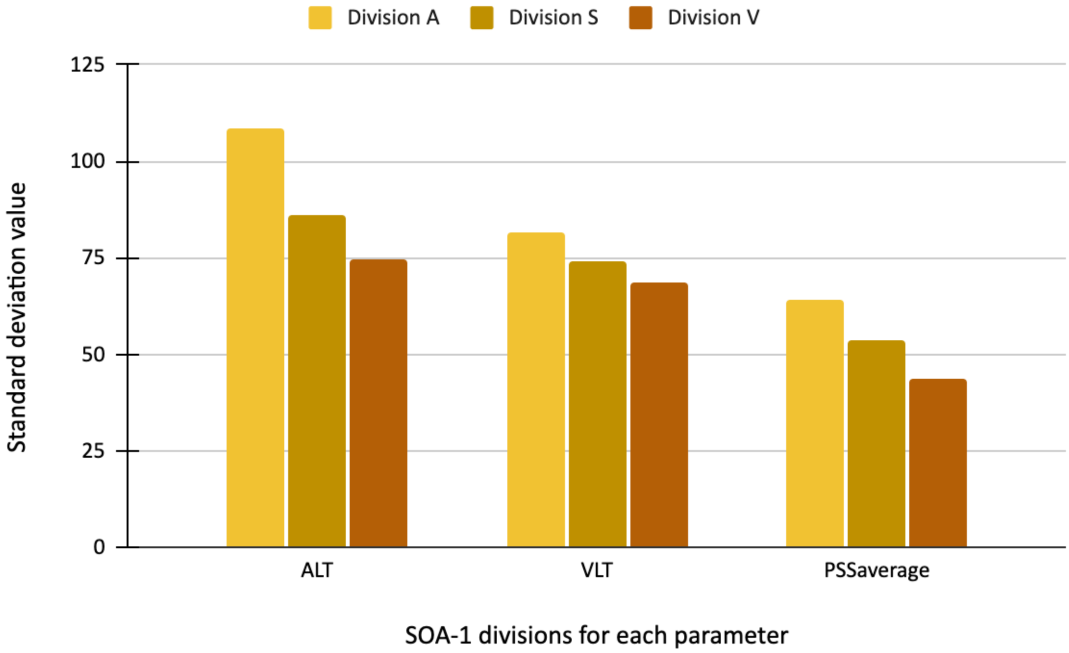
Comparison of mean differences for each parameter acros SOA-1 ranging from 400 audio-lead to 400 video-lead



Note. Mean value for ALT, VLT and PSSaverage (y-axis) for each SOA-1 (x) amongst participants ($N = 29$).

Figure 5

Comparison of standard deviation values between parameters and SOA-1 divisions



Note. Standard deviation value (y-axis) for each SOA-1 division for every parameter (x-axis) amongst participants ($N = 29$).

In summary, the results showed a significant difference in ALT, VLT and PSSaverage between the three SOA-1 divisions, as shown in Figure 4 and Table 1. The expectation of ALT not being affected by SOA-1 was not supported. ALT showed the largest mean differences between SOA-1 divisions, compared to VLT and PSSaverage. For VLT, the mean difference was greater between division S and division V than the mean difference between division S and division A. The expectation that VLT of the current trial will be more video-lead when SOA-1 is video-lead, compared to when SOA-1 is synchronous was supported. The mean difference in PSSaverage was greater between division S and division A than between division S and division V, supporting the hypothesis that PSS of the current trial would be more video-lead for video-lead SOA-1 compared to audio-lead and synchrony SOA-1. For all parameters the mean difference was greatest between division A and division V.

Discussion

Brief exposure to AV asynchrony produces negative aftereffects

Previous research has shown that perceived AV simultaneity shifts after repeated exposure to timing delays between auditory and visual signals to achieve perception of synchrony (e.g., Di Luca et al., 2009; Fujisaki et al., 2004; Roach et al., 2011). Recent studies indicate that similar recalibration effects can occur following brief exposure to AV delays, although this effect is only found for simultaneity judgements (Roseboom; 2019; Van der Burg et al., 2013). The present study´s primary purpose was to research the replication of negative aftereffects following brief exposure, as the majority of prior research has found that serial dependence usually produces positive aftereffects (e.g., Alais et al., 2015; Bliss et al., 2017; Corbett et al., 2011).

The current study found negative aftereffects following brief exposure for simultaneity judgments, consistent with Roseboom (2019) and Van der Burg and colleagues (2013) findings. Rapid recalibration effects were reported as results showed significant differences in ALT, VLT and PSSaverage between the three SOA-1 divisions, by which video-lead SOA-1 produced larger recalibration effects than audio-lead SOA-1 (see Figure 4).

Consistent with previous literature reporting an asymmetrical TBW with a visually-biased PSS (e.g., Alm & Behne, 2013; Conrey & Pisoni, 2006; Dixon & Spitz, 1980; Hay-McCutcheon et al., 2009), the present study found PSS of the given trial to be more video-lead when SOA-1 was video-lead, compared to when SOA-1 was synchronous or audio-lead (see Figure 3). Additionally, results showed that VLT of the current trial was more video-lead when SOA-1 was video-lead, compared to when SOA-1 was synchronous (see Figure 2). Perceived AV synchrony on a current trial was contingent on the modality order in the prior trial, by which the perceiver was more inclined to perceive audio-lead asynchrony than video-lead asynchrony.

The asymmetrical AV TBW is consistent with the naturally occuring delays between auditory and visual signals. Audio-lead signals are delimited by neural transmission timing, whilst video-lead signals are dependent on the distance between the perceiver and the AV event (Keetels & Vroomen, 2012). The perceptual system may have adapted the PSS towards video-lead SOA to compensate for the more prevalent and less predictable video-leading asynchrony. In AV speech, articulation of the sound occurs before sound production. Smeele and colleagues (1994) found that video-leading signals enables the perceiver to use visual signals as predictors of auditory input. A visually-biased perception may thus be a result of the brain compensating for video-lead asynchrony in non-speech and speech to achieve perception of AV synchrony.

Although the current study replicated the finding of rapid recalibration when using a SJ task, the question still remains why negative aftereffects following brief exposure do not occur for TOJ or MJ tasks. This introduces the discussion of what rapid recalibration reflects: sensory adaptation or serial dependence?

Rapid recalibration: sensory adaptation or serial dependence?

Previous studies indicate that temporal recalibration is consistent with sensory adaptation (Fujisaki et al., 2004; Roach et al., 2011), but similar processes do not necessarily account for rapid recalibration. Roseboom (2019) found that negative aftereffects following brief exposure only occurred using a SJ task, whilst positive aftereffects occurred for TOJ and MJ tasks. Arguably, if rapid recalibration reflects sensory adaptation it should occur in all tasks, because the tasks are assumed to measure the same sensory processing mechanisms (Keane et al., 2020). A task-dependent rapid recalibration indicates that there is a difference between sensory adaptation and serial dependence for AV relative timing (Roseboom, 2019). Why previous research (Roseboom, 2019; Van der Burg et al., 2013) and the present study find negative aftereffects following brief exposure in simultaneity judgements, as those following repeated exposure, is disputed.

Roseboom (2019) argues that rapid recalibration reflects changes in decisional processes that manifest differently for SJ and TOJ and MJ, not changes in the perceiver's relative timing estimation. Negative aftereffects following brief exposure for simultaneity judgements occur as a result of the placement of decision criteria being biased towards the exposed AV input. Resulting in PSS shifting in a similar manner as when the relative timing estimate changes. Perceived synchrony being between -80 ms. audio-lead SOA to 120 ms. video-lead SOA, shifts the PSS towards video-lead SOA, and the exposed AV relative timing. The positive aftereffects following brief exposure in TOJ and MJ tasks are consistent with relative timing perception being assimilated towards prior experience. Prior experience with video-leading AV signals will result in an increase in judgments of perceived video-lead when the stimulus is physically audio-lead. PSS will shift towards audio-lead SOA and not towards the timing of the exposed stimulus. Bayesian decision models describe these effects as a result of the given judgment being attracted towards prior experience (Petzschner et al., 2015). Roseboom (2019) concludes based on these findings that repeated exposure is consistent with sensory adaptation to improve perceptual discrimination and precision, and brief exposure reflects descision processes to promote perceptual stability.

Keane and colleagues (2020) argue that negative aftereffects can occur in all relative timing tasks, and that positive aftereffects following TOJ and MJ are caused by a choice-repetition bias. A choice-repetition bias involves repeating a prior judgment without regard for the relative timing of the exposed signal (Keane et al., 2020). For example, when participants were exposed to video-lead in the prior trial they judged the AV stimulus in the given trial as video-lead, without consideration of the AV stimulus temporal relationship in the given trial. Likewise, when exposed to audio-lead SOA-1, the participants were inclined to judge the next trial as an audio-lead stimulus. Perceived video-lead SOA-1 leads to PSS shifting towards audio-lead SOA, and audio-lead SOA-1 shifting PSS towards video-lead SOA. Keane and colleagues (2020) found that the data from every relative timing task in Roseboom´s (2019) study was affected by a choice-repetition bias, however how the bias manifested differed across tasks. For TOJ and MJ tasks, the bias produced shifts in the mean response distribution

(PSS). In SJ, the bias only altered the amplitude of response distribution for simultaneity judgments. The SJ task thus indicates to be more robust against bias causing incorrect rapid recalibration effects (Keane et al., 2020) than TOJ and MJ tasks. Roseboom´s (2019) hypothesis of rapid recalibration reflecting decisional processes, is based on the finding that rapid recalibration is only found in SJ tasks, and thus cannot reflect sensory adaptation. If Keane and colleagues (2020) findings are correct, rapid recalibration occurs for all relative timing tasks and may therefore also reflect sensory adaptation similarly to temporal recalibration.

Although some research may indicate that rapid recalibration reflects sensory adaptation (Keane et al., 2020), Roseboom´s (2019) reasons for why finding rapid recalibration in AV perception is unexpected remain pertinent. A similar magnitude of effect following repeated and brief exposure to AV asynchrony suggests that singular exposure is sufficient for adaptation. Assuming Keane and colleagues (2020) findings are correct regarding rapid recalibration reflecting sensory adaptation, it seems a strange assertion that a singular exposure would cause the same sensory adaptation effect as prolonged exposure. Furthermore, the effect of temporal recalibration is found to decrease when exposed to contradicting input (Alais et al., 2017; Machulla et al., 2012). If the same processes are involved in rapid recalibration, alongside a similar magnitude of effect, the purpose of repeated exposure becomes increasingly unclear. Second, the majority of previous literature has found positive aftereffects following brief exposure when researching serial dependence (e.g., Alais et al., 2015; Bliss et al., 2017; Corbett et al., 2011; Fischer & Whitney, 2014).

Van der Burg and colleagues (2013) argue that finding negative aftereffects following brief exposure to AV synchrony is not unexpected as it has an important role in perception of AV synchrony. The ability to quickly synchronize related visual and auditory signals without repeated exposure facilitates optimal multisensory integration. For instance, In AV speech, rapid recalibration is practical as quick adaptation to the initial temporal relationship in a speech stream optimizes speech comprehension for the remnant part of the stream (Sumby & Pollack, 1954; Van der Burg et al., 2013).

Further research is needed to conclude whether the current study finding rapid recalibration reflects sensory adaptation or changes in decisional processes. The likelihood of bias producing false rapid recalibration effects is relatively low in the present study as a SJ task was used, which Keane and colleagues (2020) found to be the more robust relative timing task. The present study can not conclude whether rapid recalibration reflects sensory adaptation or serial dependence, but rather promote further research within the field as studies indicate different results.

Comparing the use of two cumulative Gaussians versus a Gaussian

The study researched whether two cumulative Gaussians would better represent mean simultaneity judgment responses than a Gaussian. The majority of previous studies have fitted responses from a SJ task to a Gaussian, through which parameters are extracted to examine characteristics of perception of AV synchrony, such as the TBW (e.g., Alm & Behne, 2013; Conrey & Pisoni, 2006; Van der Burg et al., 2013). Recent research indicate that ALT and VLT are two separate mechanisms (Cecere et al., 2016), raising an issue regarding the fitting of a Gaussian as ALT and VLT are represented dependent on each other. Yarrow and colleagues (2011) propose that the two cumulative Gaussians better represent the mean simultaneity judgements as audio-lead and video-lead are represented separately. Significant findings in the current study that have not been reported in previous research with a Gaussian approach (Roseboom, 2019; Van der Burg et al., 2013), may indicate support of Yarrow and colleagues (2011)´s hypothesis.

The current study found significant differences in ALT between SOA-1 divisions, thus the expectation of ALT not being affected by SOA-1 was not supported. The expectation was based on Roseboom (2019) and Van der Burg and colleagues (2013) findings, where both studies fitted the data with a Gaussian. Roseboom (2019) and Van der Burg and colleagues (2013) found evidence against the hypothesis that PSS should be smaller when audio-lead SOA-than SOA-1 = 0 ms., whilst PSS was shown to be larger for video-lead SOA-1 than SOA-1 = 0 ms. The results indicated that there was no difference in ALT when SOA-1 was audio-lead or synchronous. The present study predicted replication of similar results. However, results in the current study were not consistent with Roseboom (2019) and Van der Burg and colleagues (2013) findings. Results in the present study showed significant differences in ALT between all SOA-1 divisions, ranging from audio-lead, synchronous and video-lead, indicating that ALT is also affected by varying SOA-1. Additionally, ALT showed the largest significant differences in value between SOA-1 divisions, compared to VLT and PSSaverage. That the current study used two cumulative Gaussians may be the reason for these unexpected results.

Roseboom (2019) and Van der Burg and colleagues (2013) data may have contained the effect, but the Gaussian may not be sensitive enough to represent the differences. In research on rapid recalibration PSS is shown to be larger for video-lead SOA-1 than when SOA-1 is audio-lead or synchronous (Roseboom, 2019; Van der Burg et al., 2013). The Gaussian will therefore shift towards video-lead SOA and force the slopes from each threshold into one curve which could pull the ALT value towards video-lead SOA, resulting in a poor representation of ALT. In contrast to the Gaussian, the two cumulative Gaussians are not delimited by forcing the slopes of the curves to be symmetrical or pulled in a certain direction. ALT and VLT are represented separately, consistent with prior research indicating that they are independent mechanisms (Cecere et al., 2016). The current study found that the slopes of the two curves differed, by which audio-lead curve was steeper than the video-lead curve (129 ms. versus 94 ms.) (see Figure 4). The curve´s different slopes are consistent with ALT showing the largest significant differences in value between SOA-1 divisions, compared to VLT. A steep audio-lead curve indicates great sensitivity to audio-leading asynchrony, whilst a flat video-lead curve indicates poorer sensitivity to video-lead asynchrony. The current study´s results indicate that two cumulative Gaussians represent the simultaneity judgment responses better than a Gaussian and offer more precise parameters when researching perception of AV synchrony.

Further, the current study supports the notion that recalibration should be researched by assessing shifts in ALT and VLT, not PSS. After a suggestion from Dawn M. Behne, the present study extracted a different approach to PSS, PSScrossover, in addition to the classical estimation of PSS as the mean between ALT and VLT. PSScrossover was expected to better capture the perceiver's point of subjective simultaneity more precisely as the parameter can move freely between the thresholds, not delimited by being estimated as the peak of the curve, and represents the slopes of the two independent Gaussians. However, PSScrossover showed no significant difference in value between SOA-1 divisions, whilst PSSaverage showed significant differences in value between SOA-1 divisions. That different approaches to PSS show dissimilar results may challenge the relevancy of the parameter, as the parameters assumption is that there is a maximal point of subjective simultaneity. Yarrow and colleagues (2011) criticize the use of PSS as its representation is misleading. Typically a range of AV temporal relationships are perceived as synchronous, not just a specific point. Therefore ALT and VLT are better to represent the shifts in perception of simultaneity in response to temporal delays between auditory and visual signals, as they capture the shifts in the TBW rather than a maximal point of subjective simultaneity.

Findings that support the use of two cumulative Gaussians rather than a Gaussian, introduce the potential issue that the majority of previous studies researching AV synchrony have fitted responses from a SJ task to a Gaussian. As the TBW is asymmetrical with a visually-biased PSS, the Gaussian will be shifted towards video-lead SOA, narrowing the audio-lead SOA and expanding the video-lead SOA (Cecere et al., 2016). Previous studies may thus have reported parameter values that are more video-lead than their actual placement. Further research on different types of curves and how they represent the parameters is needed as future research is dependent on knowledge of the limitations to studies. The current study, alongside Yarrow and colleagues (2011) and Cecere and colleagues (2016), support the use of two cumulative Gaussians rather than a Gaussian.

Audiovisual experience affects ALT more than VLT and PSSaverage

Several studies have found that ALT adapts, more than VLT, in response to SOA-1 being synchronous and asynchronous when researching how the thresholds adapt to cumulative AV experience over time (e.g., Alm & Behne, 2013; Behne et al., 2013; Conrey & Pisoni, 2006; Hay-McCutcheon et al., 2009). When researching how brief exposure to AV temporal delays affect ALT and VLT, studies find that VLT is more affected than ALT by AV experience (Roseboom, 2019; Van der Burg et al., 2013). The current study found VLT to be more video-lead for SOA-1 video-lead than SOA-1 synchrony, and significant differences in ALT between SOA-1 divisions. However, ALT showed the largest differences in value between SOA-1 divisions compared to VLT and PSSaverage, as illustrated in Figure 4. The present study indicates that ALT is more inclined to adapt to brief exposure to AV temporal delays than VLT and PSSaverage.

A possible reason for why ALT would show such large differences in value between SOA-1 divisions could be explained by Alm and Behne´s (2013) finding of experience having a systematic effect on ALT. Alm and Behne (2013) showed that age-related AV experience did not affect sensitivity to video-lead asynchrony, whilst the age-related AV experience affected the perceivers sensitivity to audio-lead asynchrony. Tolerance for audio-lead asynchrony was found to decrease with age, when comparing middle-aged adults (ALT = -165 ms) and young adults (ALT -216 ms). Large differences in ALT between SOA-1 divisions may be a result of ALT being systematically fine-tuned in response to AV experience, an effect that is not present for VLT or PSSaverage.

However, Alm and Behne (2013) researched how cumulative AV experience affected ALT and VLT, whilst the current study researched how ALT and VLT shifted after brief exposure to different AV temporal delays. Alm and Behne (2013) results can explain the current study´s finding large differences in ALT if rapid recalibration reflects both Bayesian processes and a serial dependence approach. If similar processes are involved in shifting ALT after cumulative AV experience and brief exposure to AV delays, Alm and Behne´s (2013) findings may be applied to the present study´s results.

Although ALT showed large differences across the three SOA-1 divisions, the effect SOA-1 divisions had on ALT value varied between participants. The variation in sensitivity to audio-leading signals can be demonstrated by examining the standard deviations for each parameter within every SOA-1 division. Figure 5 shows a comparison of the standard deviations between parameter and SOA-1 divisions. Standard deviation is a measure for the spread of data based on the mean (Manikandan, 2011). Results show that ALT had the greatest standard deviations for each SOA-1 division, compared to VLT and PSSaverage. The finding may reflect the great variation of AV experience that has led to different fine-tuning of ALT between participants, assuming Alm and Behne (2013) results can be applied to the current study´s findings.

As the present study finding of ALT showing the largest differences in value between SOA-1 divisions, compared to VLT and PSSaverage, is not consistent with Roseboom (2019) and Van der Burg and colleagues (2013) results, further research is needed to assess replicability of findings. Other studies using different syllables may report different findings, as the current study only used the syllable /ba/ which is not representative for speech in general. The threshold's adaptability to brief exposure to AV temporal delays should be further researched by observing shifts in ALT and VLT, rather than shifts in PSS, to achieve a more precise representation of the threshold's sensitivity to brief AV experience. The current findings may not be consistent with Roseboom (2019) and Van der Burg and colleagues (2013) because the latter studies observed shifts in PSS, whilst the current study observed shifts in PSS, ALT and VLT. The current study also fitted the data with two cumulative Gaussians, which is found to represent parameters more precisely.

Conclusion

The current study demonstrates that brief experience with AV asynchrony produces negative aftereffects using a simultaneity judgment task. Whether rapid recalibration reflects assimilation of changes in decision criteria or sensory adaptation remains elusive as previous studies support different hypotheses. The finding of TBW being asymmetrical with a visually-biased PSS was replicated, as PSS was more video-lead when SOA-1 was video-lead than when SOA-1 was synchronous and asynchronous, and VLT was more video-lead when SOA-1 was video-lead than when SOA-1 was synchronous. ALT showed the largest differences in value between SOA-1 divisions, compared to VLT and PSSaverage, indicating that ALT is more affected, than VLT, by brief experience with AV temporal delays. The current study indicates that rapid recalibration should rather be researched by observing shifts in ALT and VLT than PSS in response to brief exposure to AV temporal delays, and The current study supports that using two cumulative Gaussians better captures simultaneity judgment responses, then a Gaussian. In addition to contributing to research on perception of AV synchrony, the present findings can contribute to how training programs for individuals with broadened TBW should be designed as the study gives further insight into rapid recalibration and the adaptability of ALT and VLT.

References

Alais, D., Ho, T., Han, S. E., & Van der Burg, E. (2017). A matched comparison across three different sensory pairs of cross-modal temporal recalibration from sustained and transient adaptation. *i-Perception*, *8*(4), 1-18.

Alais, D., Orchard-Mills, E., & Van der Burg, E. (2015). Auditory frequency perception adapts rapidly to the immediate past. *Attention, Perception, & Psychophysics*, *77*(3), 896-906.

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.

Asakawa, K., Tanaka, A., & Imai, H. (2012). Audiovisual temporal recalibration for speech in synchrony perception and speech identification. *Kansei Engineering International Journal*, *11*(1), 35-40.

Baskent, D., & Bazo, D. (2011). Audiovisual asynchrony detection and speech intelligibility in noise with moderate to severe sensorineural hearing impairment. *Ear and hearing, 32*(5), 582–592.

Behne, D. M., Alm, M., Berg, A., Engell, T., Foyn, C., Johnsen, C., Srigaran, T., & Torsdottir, A. E. (2013). Effects of musical experience on perception of audiovisual synchrony for speech and music. *The Journal of the Acoustical Society of America, 133*(5), 3570.

Bliss, D. P., Sun, J. J., & D'Esposito, M. (2017). Serial dependence is absent at the time of perception but increases in visual working memory. *Scientific reports*, *7*(1), 1-13.

British Society of Audiology. (2004). Recommended procedure: Pure tone air and bone conduction threshold audiometry with and without masking and determination of uncomfortable loudness levels. http://www.thebsa.org.uk/docs/RecPro/PTA.pdf

Cecere, R., Gross, J., & Thut, G. (2016). Behavioral evidence for separate mechanisms of audiovisual temporal binding as a function of leading sensory modality. *European Journal of Neuroscience, 43*(12), 1561-1568.

Cicchini, G. M., Mikellidou, K., & Burr, D. C. (2018). The functional role of serial dependence. *Proceedings of the Royal Society B*, *285*(1890), 1-8.

Conrey, B., & Pisoni, D. B. (2006). Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *The Journal of the Acoustical Society of America*, *119*(6), 4065-4073.

Corbett, J. E., Fischer, J., & Whitney, D. (2011). Facilitating stable representations: Serial dependence in vision. *PLoS One*, *6*(1), 1-7.

Di Luca, M., Machulla, T. K., & Ernst, M. O. (2009). Recalibration of multisensory simultaneity: cross-modal transfer coincides with a change in perceptual latency. *Journal of vision*, *9*(12), 1-16.

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception, 9*(6), 719–721.

Feenders, G., & Klump, G. M. (2018). Violation of the unity assumption disrupts temporal ventriloquism effect in starlings. *Frontiers in psychology*, 1386.

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, 17 (5), 738–743.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. Y. (2004). Recalibration of audiovisual simultaneity. *Nature neuroscience*, *7*(7), 773-778.

Grant, K. W., Greenberg, S., Poeppel, D., & Van Wassenhove, V. (2004). Effects of spectro-temporal asynchrony in auditory and auditory-visual speech processing. *Seminars in Hearing*, *25*(3), 241-255.

Harrar, V. & Harris, L. R. (2008). The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. *Experimental brain research, 186* (4), 517-524.

Hay-McCutcheon, M., Pisoni, D., & Hunt, 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.

Keane, B., Bland, N. S., Matthews, N., Carroll, T. J., & Wallis, G. (2020). Rapid recalibration of temporal order judgements: Response bias accounts for contradictory results. *European Journal of Neuroscience*, *51*(7), 1697-1710.

Keetels, M., & Vroomen, J. (2012). Perception of synchrony between the senses. *The neural bases of multisensory processes*.

Machulla, T. K., Di Luca, M., Froehlich, E., & Ernst, M. O. (2012). Multisensory simultaneity recalibration: storage of the aftereffect in the absence of counterevidence. *Experimental Brain Research*, *217*(1), 89-97.

Manikandan, S. (2011). Measures of dispersion. *Journal of Pharmacology and Pharmacotherapeutics*, *2*(4), 315.

Miles, W. R. (1929). Ocular dominance demonstrated by unconscious sighting. *Journal of Experimental Psychology*, *12*(2), 113-126.

Noel, J. P., De Niear, M. A., Lazzara, N. S., & Wallace, M. T. (2017). Uncoupling between multisensory temporal function and nonverbal turn-taking in autism spectrum disorder. *IEEE Transactions on Cognitive and Developmental Systems*, *10*(4), 973-982.

Odegaard, B., Wozny, D. R., & Shams, L. (2017). A simple and efficient method to enhance audiovisual binding tendencies. *PeerJ, 5*, 1-29.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia, 9(1), 97-113

O'mahony, M. (1986). Sensory adaptation. *Journal of Sensory Studies*, *1*(3‐4), 237-258.

Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in cognitive sciences*, *19*(5), 285-293.

Pöppel, E., Schill, K., & von Steinbüchel, N. (1990). Sensory integration within temporally neutral systems states: a hypothesis. Naturwissenschaften, 77(2), 89-91.

Recio, R. S., Cravo, A. M., de Camargo, R. Y., & van Wassenhove, V. (2019). Dissociating the sequential dependency of subjective temporal order from subjective simultaneity. *Plos one*, *14*(10), 1-10.

Redaksjonen for norsk APA-stil. (2021). Norsk APA-manual: En nasjonal standard for norskspråklig APA-stil basert på APA 7th (Versjon 1.7). Unit. https://www.unit.no/tjenester/norsk-apa-referansestil

Roach, N. W., Heron, J., Whitaker, D., & Mcgraw, P. V. (2011). Asynchrony adaptation reveals neural population code for audio-visual timing. *Proceedings of the Royal Society of London B: Biological Sciences, 278*(1710), 1314–1322.

Roseboom, W. (2019). Serial dependence in timing perception. *Journal of experimental psychology: Human perception and performance, 45* (1), 100-110.

Smeele, P. M. T. (1994). Perceiving speech: Integrating auditory and visual speech (Unpublished doctoral dissertation). *Delft University of Technology*, 77–108.

Sorati, M., & Behne, D. M. (2019). Musical expertise affects audiovisual speech perception: Findings from event-related potentials and inter-trial phase coherence. *Frontiers in Psychology*, 1-19.

Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology: Human Perception and Performance, 38*(6), 1517-1529.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26,* 212–215.

Ten Oever, S., Sack, A. T., Wheat, K. L., Bien, N., & Van Atteveldt, N. (2013). Audio-visual onset differences are used to determine syllable identity for ambiguous audio-visual stimulus pairs. *Frontiers in psychology*, *4*, 331.

Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Temporal recalibration during asynchronous audiovisual speech perception. *Experimental brain research*, *181*(1), 173-181.

Van der Burg, E., Alais, D., & Cass, J. (2013). Rapid recalibration to audiovisual asynchrony. *Journal of neuroscience, 33*(37), 14633-14637.

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598-607.

Vroomen, J., Keetels, M., De Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive brain research*, *22*(1), 32-35.

Watt, S. W. (2010). How visual acuity is measured. www.mdsupport.org/library/acuity.html

Woynaroski, T. G., Kwakye, L. D., Foss-Feig, J. H., Stevenson, R. A., Stone, W. L., & Wallace, M. T.

(2013). Multisensory speech perception in children with autism spectrum disorders. *Journal of autism and developmental disorders*, *43*(12), 2891-2902.

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.

Zhou, H. Y., Cheung, E. F., & Chan, R. C. (2020). Audiovisual temporal integration: Cognitive processing, neural mechanisms, developmental trajectory and potential interventions. *Neuropsychologia*, *140*, 107396.