Published in final edited form as:

Neuropsychologia. 2016 June ; 86: 119–130. doi:10.1016/j.neuropsychologia.2016.04.015.

Electrophysiological Correlates of Individual Differences in Perception of Audiovisual Temporal Asynchrony

Natalya Kaganovich^{*,1,2} and Jennifer Schumaker¹

¹Department of Speech, Language, and Hearing Sciences, Purdue University, 715 Clinic Drive West Lafayette, IN 47907-2038

²Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907-2038

Abstract

Sensitivity to the temporal relationship between auditory and visual stimuli is key to efficient audiovisual integration. However, even adults vary greatly in their ability to detect audiovisual temporal asynchrony. What underlies this variability is currently unknown. We recorded event-related potentials (ERPs) while participants performed a simultaneity judgment task on a range of audiovisual (AV) and visual-auditory (VA) stimulus onset asynchronies (SOAs) and compared ERP responses in good and poor performers to the 200 ms SOA, which showed the largest individual variability in the number of synchronous perceptions. Analysis of ERPs to the VA200 stimulus yielded no significant results. However, those individuals who were more sensitive to the AV200 SOA had significantly more positive voltage between 210 and 270 ms following the sound onset. In a follow-up analysis, we showed that the mean voltage within this window predicted approximately 36% of variability in sensitivity to AV temporal asynchrony in a larger group of participants. The relationship between the ERP measure in the 210-270 ms window and accuracy on the simultaneity judgment task also held for two other AV SOAs with significant individual variability - 100 and 300 ms. Because the identified window was time-locked to the onset of sound in the AV stimulus, we conclude that sensitivity to AV temporal asynchrony is shaped to a large extent by the efficiency in the neural encoding of sound onsets.

Keywords

temporal binding window; audiovisual temporal processing; simultaneity judgment task; ERPs; individual differences

*Corresponding author: Department of Speech, Language, and Hearing Sciences, Purdue University, Lyles Porter Hall, 715 Clinic Drive, West Lafayette IN 47907-2038, Phone (765)494-4233, Fax (765)494-0771, kaganovi@purdue.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Temporal proximity is one of the determining factors for integrating multisensory, and more specifically audiovisual, stimuli into a coherent percept (Stein & Meredith, 1993). Importantly, a consistent finding in research on audiovisual integration is that the perception of multisensory temporal synchrony does not require that auditory and visual stimuli occur at exactly the same time. Instead, we perceive audiovisual information as synchronous as long as the onsets of the two modalities fall within a certain temporal distance from each other, termed the temporal binding window (TBW) (for reviews, see Keetels & Vroomen, 2012; Vatakis & Spence, 2010; Vroomen & Keetels, 2010).

Arguably, one of the key features of the TBW is that its size is not constant and is influenced by many factors. It is typically larger for visual-auditory (VA) sequences of stimuli compared to auditory-visual (AV) ones (Bushara, Grafman, & Hallett, 2001; Dixon & Spitz, 1980; Grant, van Wassenhove, & Poeppel, 2004; Lewkowicz, 1996; van Wassenhove, Grant, & Poeppel, 2007); it is larger for speech and other complex stimuli compared to simpler non-speech stimuli (Vatakis & Spence, 2010; Vroomen & Stekelenburg, 2011); it can be reduced through perceptual training (Powers III, Hillock, & Wallace, 2009; Stevenson, Wilson, Powers, & Wallace, 2013); it depends on one's expertise with specific audiovisual stimuli (e.g., Petrini et al., 2009); and it is affected by the task used to measure it (Stevenson & Wallace, 2013; van Eijk, Kohlrausch, Juola, & van de Par, 2008).

In most studies, the size of the TBW is evaluated at a group level. Significantly less research has been conducted on individual variability in sensitivity to audiovisual temporal asynchrony and its causes. The significance of individual variability in TBW is underlined by a number of findings. First, impairment in the ability to detect audiovisual temporal correspondences (and, as a result, a much broader than typical TBW) has been reported for multiple neurodevelopmental disorders (for a comprehensive review, see Wallace & Stevenson, 2014), such as dyslexia (Hairston, Burdette, Flowers, Wood, & Wallace, 2005), specific language impairment (SLI) (Grondin et al., 2007; Kaganovich, Schumaker, Leonard, Gustafson, & Macias, 2014), and autism (Foss-Feig et al., 2010; Kwakye, Foss-Feig, Cascio, Stone, & Wallace, 2011; Stevenson, Siemann, Schneider, Eberly, Woynaroski, et al., 2014). Importantly, at least in some of these studies, precision with which participants perceive audiovisual asynchrony predicted the degree of language and other cognitive impairments. For example, Donohue and colleagues (Donohue, Darling, & Mitroff, 2012) examined a correlation between the degree of self-reported symptoms of autism in general population and the temporal relationship between auditory and visual stimuli that leads to the most salient perception of synchrony. They found that individuals with the greater number of autism traits consistently reported as simultaneous those stimuli in which the auditory modality slightly preceded the visual one – the pattern that is opposite to the one seen in individuals with fewer autism traits. In an earlier study from our laboratory (Kaganovich et al., 2014), those children with a history of SLI who were worse at detecting asynchrony at long stimulus-onset asynchronies (SOAs) (400-500ms) also had lower core language scores as determined by the Clinical Evaluation of Language Fundamentals (CELF-4; Semel, Wiig, & Secord, 2003). Second, in healthy adults, the smaller size of the TBW was linked to greater susceptibility to the McGurk illusion (Stevenson, Zemtsov, &

Wallace, 2012) and to better comprehension of degraded audiovisual speech (Conrey & Pisoni, 2006). The relationship between the TBW and susceptibility to the McGurk illusion has also been replicated in children with autism (Stevenson, Siemann, Schneider, Eberly, Woyanowski, et al., 2014). Together, these findings suggest that sensitivity to audiovisual temporal relationships may contribute to successful development of at least some cognitive and linguistic skills.

The TBW is typically measured in a simultaneity judgment task (SJT), in which audiovisual stimuli are presented in a range of SOAs, and participants have to identify each stimulus presentation as audiovisually synchronous or asynchronous. The number of synchronous perceptions is then plotted as a function of SOA, and the results are fitted to a sigmoid function, separately for AV and VA SOAs. The TBW is determined as an estimated SOA at which participants detect asynchrony with a specific degree of certainty (typically on 50-75% of trials, depending on the study). The TBW thus provides a single measure of sensitivity to audiovisual asynchrony and is a succinct description of individuals' performance on the SJT. However, the synergistic nature of this measure may also be its weakness under some circumstances. More specifically, the shape of the sigmoid function (and, as a consequence, the size of the TBW) may be determined to a greater degree by some SOAs than by others. Yet, this information is, for the most part, lost in a single TBW measure. Careful examination of published work and research in our own laboratory show that while at short (e.g., 100 ms or less) and long¹ (e.g., 400 ms or more) SOAs adults largely agree in their simultaneity judgment, medium-sized SOAs lead to significant individual variability.

Several neuroimaging studies have examined the neural correlates of sensitivity to audiovisual temporal synchrony/asynchrony at a group level and reported activations in a network of regions that include both well-established multisensory areas, such as parts of the superior temporal cortex, as well as auditory and visual sensory cortices (Macaluso, George, Dolan, Spence, & Driver, 2004; Powers III, Hevey, & Wallace, 2012; Stevenson, Altieri, Kim, Pisoni, & James, 2010) and the right insula (Bushara et al., 2001). In an insightful addition to this literature, Powers and colleagues (Powers III et al., 2012; Powers III et al., 2009) have demonstrated that the reduction in the size of TBW following perceptual training leads to decreased activation in the key elements of the network, such as posterior superior temporal sulcus and auditory and visual cortices, as well as enhanced connectivity among them. This line of research reveals the complexity of the neural mechanisms engaged during audiovisual temporal processing and suggests that individual variability in any number of neural functions – from early sensory encoding to actual multisensory integrative mechanisms – may potentially contribute to observed individual differences in sensitivity to audiovisual temporal asynchrony.

One other aspect of audiovisual temporal processing deserves a special mention because of its relevance to the findings of the current study – namely, neural activity underlying temporal audiovisual processing can be modified not only by the physical properties of

¹The length of what may be considered “short” and “long” SOAs will of course depend to some degree on the stimuli used. The numbers given are not meant to be absolute values but serve as an example based on our own work with non-speech stimuli.

stimuli (e.g., whether the auditory and visual components of a stimulus in fact occurred at the same time) but also by the subjective perception of such properties (e.g., whether audiovisual stimuli were perceived as synchronous or asynchronous). This distinction was clearly demonstrated by the study of Stevenson and colleagues (Stevenson, VanDerKlok, Pisoni, & James, 2011), who presented their participants with ambiguous audiovisual stimuli, which were perceived as synchronous in approximately half of all trials and asynchronous in another half. They identified two distinct areas of the multisensory superior temporal cortex (mSTC) that responded differently to physical synchronicity and to perceptual fusion – the synchrony-defined mSTC was activated by true audiovisual synchrony regardless of how it was perceived, while the bimodal mSTC responded significantly only to subjective perception of synchrony, regardless of whether the stimulus that elicited the perceptual fusion was synchronous or asynchronous.

In the current study, we combined the SJT with event-related potential recordings (ERPs) in order to focus on the timing of the neural processes engaged during the detection of audiovisual temporal asynchrony. More specifically, we asked at which point in time brain responses of individuals who are better detectors of asynchrony (i.e., good performers) differ from brain responses of those individuals who are worse detectors of asynchrony (i.e., poor performers), with the expectation that the outcome of this comparison would be informative as to the perceptual and cognitive processes that underlie individual variability in sensitivity to temporal asynchrony. Earlier ERP studies of audiovisual integration reported the attenuation of the auditory N1 and/or P2 component to audiovisual as compared to the sum of auditory only and visual only stimuli (Baart, Stekelenburg, & Vroomen, 2014; Besle, Fort, Delpuech, & Giard, 2004; Kaganovich & Schumaker, 2014; Knowland, Mercure, Karmiloff-Smith, Dick, & Thomas, 2014; Stekelenburg & Vroomen, 2007; van Wassenhove, Grant, & Poeppel, 2005). However, given significant design differences between the SJT and the above studies, focusing on just N1 and P2 in our analyses was not justifiable. Instead, in an approach similar to the region of interest analyses used in fMRI research, we first defined our temporal windows of interest based on ERPs elicited by auditory only (a pure tone) and visual only (a flash of light) stimuli. These windows included all visible ERP components elicited by the onset of the stimuli. We then used these windows to analyze ERPs to an audiovisually asynchronous presentation of the same stimuli at the SOA that led to the largest variability in synchronous perceptions (200 ms). We compared ERPs elicited in good and poor performers in a series of t-tests conducted on each consecutive measurement point within the window of interest and used the false discovery rate (FDR) correction to control for type I error due to multiple comparisons. Following this initial step, we extended our analyses in two ways. First, in order to determine that the identified ERP differences between good and poor performers did in fact relate to their ability to detect asynchrony on a SJT, we conducted a linear regression analysis on a larger group of participants with the ERP measure as a predictor and the number of synchronous perceptions at a 200 ms SOA as an outcome. Second, to ascertain that our finding can be generalized to other SOAs with substantial individual variability, we conducted similar regression analyses between ERP measures and the number of synchronous perceptions for the 100 and 300 ms SOAs.

2. Materials and Methods

2.1 Participants

Thirty-six young adults participated for payment. Data from 4 participants were excluded from analysis because their accuracy on at least one of the 3 longest SOAs fell 2 standard deviations above the mean (with the higher number indicating worse performance). One more participant was excluded due to atypical ERP responses. The average age of the remaining 31 participants was 22 years (range 18-40; 5 female²). All had normal non-verbal intelligence (TONI-4, Brown, Sherbenou, & Johnsen, 2010), no history of neurological disorders, normal or corrected to normal vision, and passed a hearing screening at 20 dB HL at 500, 1000, 2000, 3000, and 4000 Hz. All gave their written consent to participate in the experiment. The study was approved by the Institutional Review Board of Purdue University, and all study procedures conformed to The Code of Ethics of the World Medical Association (Declaration of Helsinki) (1964).

2.2 Stimuli and Design

The stimuli and design of this study were identical to those described in two previous publications from our laboratory (Kaganovich, 2016; Kaganovich et al., 2014). A 2 kHz pure tone and a flash of light (shaped as a cartoon explosion, see Figure 1) served as stimuli. Both were 200 ms in duration. The tone was presented at 60 dB SPL via a sound bar located at the bottom of the computer monitor. In order to avoid visual after-effects, the explosion-shaped figure was shown at a slightly different location on the monitor on consecutive trials. On each trial, the auditory and visual stimuli appeared either synchronously (the SYNC condition) or at one of the following stimulus-onset asynchronies – 100, 200, 300, 400, or 500 ms. In half of all audiovisual trials, the sound preceded the flash of light (AV trials), while in another half, the flash of light preceded the sound (VA trials). Additionally, on some trials only the auditory stimulus (A) or only the visual stimulus (V) was presented. Each participant completed 10 blocks of trials, with each block containing 5 repetitions of 13 different trial types (sync, A, V, 5 AV SOAs, and 5 VA SOAs) presented in a random order. This yielded 50 responses for each trial type.

The task was presented as a game because we were collecting data from both children and adults. In this study, we focus on adult data only. Children's performance on the task was described in earlier studies from our laboratory (Kaganovich, 2016; Kaganovich et al., 2014). At the start of the session, participants watched a video with instructions and practiced the task until it was clear. They saw a dragon in the middle of the screen and a boy and a girl with futuristic-looking weapons at the top left and right corners of the screen. Participants were told that the dragon lives on the planet Cabula, where dragons are friendly but like to raid apple orchards. The only way to keep dragons away is to use special weapons – one of them shoots lights while the other shoots sounds. If the light and the sound “hit” the dragon at the same time, the dragon runs away. However, if either the sound or the light

²Originally, our participants were split into two groups – avid video gamers and non-gamers. Because video gamers were predominantly male, we recruited mostly male participants into the non-gamer group. However, the two groups did not differ either in behavioral or ERP measures. We therefore pooled them together in order to increase the overall number of participants and evaluate individual variability in sensitivity to audiovisual asynchrony regardless of the video gaming status.

arrives first, even by the tiniest amount, or if one of the children forgets to shoot his/her weapon (and we only see the light or only hear the sound), the dragon does not get scared away. The images of children were present at the start of each block prior to the onset of trial presentation. To avoid excessive eye movement and the shift of attention away from the dragon and the visual stimulus, there were no images of children during trial presentation. All participants sat approximately 4 feet from the monitor inside a dimly-lit sound-attenuating booth. Presentation of trials was controlled by the Presentation software (www.neurobs.com). Participants pressed one button if the dragon was scared away (i.e., the sound and the flash of light were perceived as occurring synchronously) and another button if the dragon was not scared away (i.e., either the sound and the explosion figure were perceived as asynchronous or only one of the two stimuli was detected (A and V trials)). Responses were recorded within a 2200 ms response window time-locked to the onset of the first stimulus in each pair or to the onset of a single stimulus. The response window was followed by an inter-trial interval varying randomly among 4 values: 350, 700, 1050, and 1400 ms. Hand to response button mapping was counterbalanced across participants. Participants were instructed to respond as fast as they can while also trying to maintain good accuracy. We collected both behavioral (percent of synchronous perceptions) and electroencephalographic (EEG) data for each trial type.

2.3 EEG Recordings

The EEG data were recorded from the scalp at a sampling rate of 512 Hz using 32 active Ag-AgCl electrodes secured in an elastic cap (Electro-Cap International Inc., USA). Electrodes were positioned over homologous locations across the two hemispheres according to the criteria of the International 10-10 system (American Electroencephalographic Society, 1994). The specific locations were as follows: midline sites: FZ, CZ, PZ, and OZ; mid-lateral sites: FP1/FP2, AF3/AF4, F3/F4, FC1/FC2, C3/C4, CP1/CP2, P3/P4, PO3/PO4, and O1/O2; and lateral sites: F7/F8, FC5/FC6, T7/T8, CP5/CP6, and P7/P8; and left and right mastoids. EEG recordings were made with the Active-Two System (BioSemi Instrumentation, Netherlands), in which the Common Mode Sense (CMS) active electrode and the Driven Right Leg (DRL) passive electrode replace the traditional “ground” electrode (Metting van Rijn, Peper, & Grimbergen, 1990). During recording, data were displayed in relationship to the CMS electrode and then referenced offline to the average of the left and right mastoids (Luck, 2014). The Active-Two System allows EEG recording with high impedances by amplifying the signal directly at the electrode (BioSemi, 2013; Metting van Rijn, Kuiper, Dankers, & Grimbergen, 1996). In order to monitor eye movement, additional electrodes were placed over the right and left outer canthi (horizontal eye movement) and below the left eye (vertical eye movement). Horizontal eye sensors were referenced to each other, while the sensor below the left eye was referenced to FP1 in order to create electro-oculograms. Prior to data analysis, EEG recordings were filtered between 0.1 and 30Hz. Individual EEG records were visually inspected to exclude trials containing excessive muscular and other non-ocular artifacts. Additionally, the trials on which blinks occurred during the visual stimulus presentation were excluded. Ocular artifacts were corrected by applying a spatial filter (EMSE Data Editor, Source Signal Imaging Inc., USA) (Pflieger, 2001). Similar to the independent component analysis (ICA), this method is able to separate ocular artifacts from brain activity

without constructing a head model and without subtracting ocular activity from EEG channels. However, unlike the ICA, it bypasses the need to subjectively determine which components need to be excluded by relying on a representative set of artifacts and clean data segments. ERPs were epoched starting at 200 ms pre-stimulus and ending at 1000 ms post-stimulus onset. On trials in which the auditory and visual stimuli were not completely synchronized, the stimulus onset was the onset of the first stimulus in a pair (for example, when the sound preceded the appearance of the explosion-shaped figure, ERPs were averaged to the onset of the sound). The 200 ms prior to the recording onset served as a baseline.

2.4 Data Analysis

2.4.1 Primary Analyses—Our data analysis included several distinct steps. First, in agreement with earlier reports, different SOAs led to different degrees of variability in participants' performance (see Table 1). We selected the 200 ms SOA for both the AV and VA trials as the primary focus of our analyses because it elicited the largest range in the number of synchronous perceptions and the largest standard error of the mean. We ranked all participants according to the number of synchronous perceptions, separately for the AV200 and the VA200 SOAs, with the lower number indicating greater sensitivity to audiovisual temporal asynchrony. Ten participants with the lowest number of synchronous perceptions formed the good performers group (for AV200, mean 15.4%, range 2-24%; for VA200, mean 26%, range 12-36%) while ten participants with the highest number of synchronous perceptions formed the poor performers group (for AV200, mean 76.2%, range 66-86%; for VA200, mean 70.2%, range 60-84%). Participants included in the AV200 good or poor performers group only partially overlapped with the participants included in the VA200 good or poor performers group.

Second, we selected the temporal windows over which to compare good and poor performers' ERPs to the AV200 and the VA200 stimuli. Because we did not have an a priori assumption about the timing of possible group differences and in order to avoid the problem of multiple implicit comparisons (Luck, 2014, p. 328), we chose measurement windows based on ERPs to A and V stimuli, in an approach similar to the region of interest analysis (ROI) used in fMRI research. More specifically, we created grand average waveforms elicited by A and V stimuli in the entire group of participants ($n=31$) and selected the boundaries of time window for analyses so that they included all clearly observable auditory or visual components following the stimulus onset in this group average as shown in Figures 2 and 3. Auditory stimuli elicited the typical auditory N1-P2-N2 sequence. These components were broadly distributed with the maximum amplitude over the fronto-central, central, and centro-parietal scalp (see Figure 2 Panel B), in agreement with a multitude of earlier reports (for a review, see Pratt, 2011). The entire sequence of peaks occurred between 100 and 380 ms post-stimulus onset (see Figure 2 Panel C). Visual stimuli elicited the P1-N1-P2-N2 sequence of ERP components (see Figure 3), with a clear maximum, as expected, over occipital and parieto-occipital sites (see Figure 3 Panel B). This sequence spanned the time window of 80-400 ms post-stimulus onset (see Figure 3 Panel C).

Third, the ERPs elicited by the AV200 and the VA200 stimuli in good and poor performers were down-sampled to 100Hz (which resulted in one measurement point for each 10 ms of recording). The groups were then compared through a series of t-tests conducted on each measurement point between 100 and 380 following the onset of the auditory stimulus and between 80 and 400 ms following the onset of the visual stimulus. To minimize the number of overall comparisons, only the sites with the largest amplitudes of auditory and visual components were included in analyses. These were FC1/2, C3/4, CP1/2 and CZ for auditory analyses and P3/4, PO3/4, O1/2, and OZ for visual analyses. In order to keep the number of sites equal for auditory and visual analyses, PZ was not included in the visual analyses. The false discovery rate (FDR) method was used to control for type I error due to multiple comparisons (Benjamini & Hochberg, 1995; Groppe, Urbach, & Kutas, 2011), and all reported p-values are FDR corrected values with $\alpha=0.05$. The outcome of this analysis provided us with a temporal window over which good and poor performers had significantly different brain responses to the 200 ms SOA.

Fourth, the outcome of the analyses described in step three above yielded a window of significant differences between good and poor performers only for the auditory responses elicited by the AV200 stimulus. In order to determine whether the ERP mean amplitude within this window is systematically related to individual sensitivity to audiovisual temporal asynchrony, we conducted a linear regression analysis with the mean amplitude within the identified window as a predictor and the number of synchronous perceptions at the AV200 SOA as the outcome. All 31 study participants were included in this analysis.

Fifth, to make sure that the relationship between the ERP measure and sensitivity to audiovisual temporal asynchrony described in step four above is not specific to the AV200 SOA, we also measured the ERP mean amplitude elicited by the AV100 and AV300 SOAs (both of which still had significant individual variability in the number of synchronous perceptions – see Table 1) within the same window identified in step 4 above (see Results) and conducted a linear regression between this mean amplitude and the number of synchronous perceptions, separately for the AV100 and AV300 SOAs.

Lastly, because all available trials to the AV200 stimulus were included into individual averages for good and poor performers, it is possible that the difference in ERPs to the AV200 stimulus reflected perception of synchrony on a trial by trial basis, rather than a more general individual difference in sensitivity to audiovisual temporal information. To tease these two variables apart (perception vs. individual sensitivity), we compared the mean amplitude between 210 and 270 ms in the ERPs elicited by the AV200 stimuli when such stimuli were perceived as synchronous and when they were perceived as asynchronous. To make sure that this comparison is conducted on a similar number of trials, we used ERP data from 11 individuals who were “average” performers – they were not included in either good or poor performer groups and perceived the AV200 stimulus both as synchronous and as asynchronous on a similar number of trials.

2.4.2 Follow-up Analyses—After the primary analyses of our data were completed, we conducted several follow-up analyses in order to better position our findings within the existing literature on audiovisual temporal processing.

First, in our primary analyses, we examined a correlation between voltage within the 210-270 ms window following the auditory stimulus at the AV200 SOA and the number of synchronous perceptions at the same SOA to determine whether our electrophysiological measure is related to performance on the SJT. This left open the question of whether our ERP measure would relate to a more synergistic way of evaluating sensitivity to audiovisual temporal asynchrony – namely, the TBW. Given that intermediate SOAs have the largest inter-individual variability, they should influence the size of the TBW significantly more than shorter or longer SOAs. To confirm this, we conducted two additional analyses. (1) We used the `glmfit` function in MATLAB (MATHEMATICS, Inc., Natick, MA) in order to fit a sigmoid function to the number of synchronous perceptions separately for the VA and AV SOAs, for each individual. The width of the right and left side of each function was then measured at 50% of its maximum. We performed a one-way ANOVA to see if good and poor performers (as defined by sensitivity to the AV200 SOA) differed in the size of the AV and VA TBW. (2) We conducted a linear regression analysis between mean voltage in the 210-270 ms time window at the AV200 SOA and the size of the AV TBW in the entire group of participants ($n=31$).

Second, visual examination of the ERPs elicited by the VA200 stimulus suggested that good performers may have had significantly larger visual N1 compared to poor performers (see Figure 6). Although the temporal interval included in the t-test analyses spanned the visual N1 window, we wanted to rule out with greater certainty that good and poor performers did not differ in the amplitude of visual N1. To this end, we measured its peak amplitude between 140 and 205 ms post-stimulus onset over PO3/4, O1/2, and OZ sites and conducted a repeated-measure ANOVA with group as a between-subject variable. Further, to ascertain that the amplitude of visual N1 did not play a role in behavioral performance on the SJT, we conducted a regression analysis with the visual N1 peak amplitude as a predictor and the percent of synchronous perceptions at the VA200 SOA as the outcome on the data from the entire group of participants.

Finally, the temporal window within which the ERP voltage was significantly correlated with sensitivity to the medium AV SOAs was time-locked to the onset of the auditory stimulus (see Results). However this window (210-270 ms post-auditory onset) did not include the auditory N1, which is known to be sensitive to sound onsets. To confirm this finding, we compared the peak amplitude and latency of the auditory N1 elicited by A only stimulus in good and poor performers (as defined by performance on the AV200 SOA) measured between 100 and 160 ms post-stimulus onset. Further, in order to investigate whether the peak amplitude of the auditory N1 is related to behavioral performance on the SJT, we conducted a linear regression analysis between its peak amplitude and the number of synchronous perceptions at two AV SOAs with the largest inter-individual variability — AV200 and AV300.

3. Results

3.1 Primary Analyses

Figure 4 shows the distribution of average synchronous perceptions across the range of SOAs for the entire group of participants. As expected, the number of synchronous

perceptions decreased with increasing SOAs. Error bars in Figure 4 and ranges and standard errors included in Table 1 show individual variability for each SOA and reveal that the largest range of synchronous responses was to the 200 ms SOA.

For both the AV200 and VA200 SOAs, the good and poor performers' responses had equivalent variances based on the Levene's test (AV200, $F < 1$; VA200, $F = 1.66$, $p = 0.214$), but differed significantly in the number of synchronous perceptions (AV200: $F(1,19) = 363.207$, $p < 0.001$, $r = 0.98$; VA200: $F(1,19) = 155.38$, $p < 0.001$, $r = 0.95$), with predictably fewer synchronous perceptions in the good performer group.

Figure 5 shows ERPs to the AV200 stimulus and Figure 6 ERPs to the VA200 stimulus overlaid for good and poor performers. Panel B in each figure specifies the time window over which consecutive t-tests were performed. None of the p-values resulting from the analysis of ERPs elicited by the VA200 stimulus survived the FDR correction. Only the comparison of the auditory components elicited by the AV200 stimulus in good and poor performers yielded results that remained significant after the FDR correction (see Figure 5, panel C). The temporal points and sites that had significant group difference are shown in Table 2. All consecutive temporal points at which more than half of electrode sites showed FDR corrected p-values of less than 0.1 were included in the final temporal window for a regression analysis. This approach yielded a time window of 210-270 ms following the onset of the auditory stimulus.

In order to determine if the ERPs to the AV200 stimulus in the window separating good and poor performers were in fact related to individuals' ability to detect audiovisual asynchrony, we conducted a linear regression analysis with the ERP mean amplitude (averaged across the 7 electrodes used for earlier analyses) as a predictor and the percent of synchronous perceptions at the AV200 SOA as an outcome. All 31 subjects were included in this analysis. The result is shown at the top of Figure 7. The regression produced significant results (Beta = -0.6 , $R = 0.6$, $F(1,30) = 16.272$, $p < 0.001$), with more positive voltage between 210 and 270 ms post-auditory onset being associated with better sensitivity to audiovisual temporal asynchrony. The mean voltage between 210 and 270 ms post-stimulus onset accounted for approximately 36% of variability in temporal synchrony judgment ($R^2 = 0.359$).

We also evaluated whether the above finding transfers to other SOAs with significant individual variability – namely, AV100 and AV300 (see Table 1 for ranges and standard errors). We conducted two separate linear regressions with the ERP mean amplitude between 210 and 270 ms post-stimulus onset to AV100 and AV300 stimuli as a predictor and the number of synchronous perceptions at each SOA as the outcome. The results are shown at the bottom of Figure 7. Both regressions were significant and displayed the same relationship between ERP and behavioral data as in the case of the AV200 stimulus (AV100: Beta = -0.37 , $R = 0.37$, $F(1,29) = 4.451$, $p = 0.044$; AV300: Beta = -0.564 , $R = 0.56$, $F(1,30) = 13.544$, $p = 0.001$).

For all of the above regression analyses, we examined the presence of outliers by evaluating the standardized DFBeta function in the SPSS Statistics program, which tests the influence of individual cases on a regression model. When the regression model is stable, excluding

any one case should not have a significant influence on the outcome. Predictor cases (ERP measures) with standardized $DF\beta$ values over 1 were considered to have a significant influence over the model and were excluded from analyses (Field, 2009). Based on this criterion, only 1 case was excluded from the AV100 regression analysis.

Because the temporal window differentiating good and poor performers was time-locked to the auditory stimulus, one interpretation of the regression analyses might be that it was the processing of the auditory stimulus by itself rather than in the context of the audiovisual task that influenced participants' performance. To evaluate this possibility, we measured the mean amplitude of ERPs between 210 and 270 ms post-stimulus onset elicited by A trials (i.e., trials on which only the auditory stimulus was presented) in good and poor performers and compared the two groups in a repeated-measures ANOVA with site (FC1/2, C3/4, CP1/2, CZ) as a within-subject variable. The group effect fell somewhat short of significance (group, $F(1,18)=3.804$, $p=0.067$, $\eta_p^2=0.174$), and there was no group by site interaction ($F(6,2.265)<1$, $\eta_p^2=0.033$).

Finally, "average" participants perceived the AV200 stimulus as synchronous on 44.9% of trials (range 26-60%, $SD=11.3$). Figure 8 overlays the ERP average of the AV200 trials when the stimulus was perceived (incorrectly) as synchronous (mean=21.7, range=13-29, $SD=5.7$) with the average of trials on which it was perceived (correctly) as asynchronous (mean=26.6, range=21-34, $SD=4.9$) in the same group of 11 subjects. In absolute terms, in the 210-270 ms window following the auditory onset, the ERP response to trials perceived as asynchronous was more negative than to trials perceived as synchronous. This pattern was opposite to that obtained from the comparison of good and poor performers, with good performers showing more positive ERPs during the same time window (compare Figures 5 and 8). A statistical comparison of ERPs elicited by trials perceived as synchronous and as asynchronous by "average" performers over the FC1/2, C3/4, CP1/2, and CZ sites showed a small trend toward significance ($F(1,10)=3.275$, $p=0.1$, $\eta_p^2=0.247$).

3.2 Follow-up Analyses

The VA and AV TBW for good and poor performers at 50% of the function's maximum are shown in Figure 9 panel A. Good performers had a significantly smaller AV TBW window than poor performers (134 ms ($SD=29$) vs. 270 ms ($SD=22$), respectively); however, the groups did not differ in the size of the VA TBW (192 ms ($SD=34$) vs. 206 ms ($SD=40$)) (modality order (VA vs. AV) by group (good vs. poor performers), $F(1,18)=34.204$, $p<0.001$, $\eta_p^2=0.655$; VA TBW, $F(1,19)<1$, $r=0.2$; AV TBW, $F(1,19)=135.068$, $p<0.001$, $r=0.94$).

The size of the AV TBW in the entire group of participants ($n=31$) was significantly correlated with mean voltage between 210 and 270 ms post-sound onset at the AV200 SOA (see Figure 9, panel B) ($B=-12.668$, $R=0.488$, $F(1,30)=9.061$, $p=0.005$), with more positive voltage being associated with better sensitivity to audiovisual temporal asynchrony.

Analysis of visual N1 showed that its peak amplitude tended to be larger in good compared to poor performers (as defined by accuracy at the VA200 SOA), although the group effect did not quite reach significance ($F(1,18)=3.615$, $p=0.073$, $\eta_p^2=0.167$). Importantly, there

was no relationship between the visual N1 peak amplitude and sensitivity to audiovisual asynchrony at the VA200 SOA ($B=1.868$, $R=0.298$, $F(1,30)=2.82$, $p=0.104$).

Finally, the N1 peak amplitude and peak latency elicited by the A only stimulus did not differ between good and poor performers (as defined by performance on the AV200 SOA) (N1 peak amplitude: group, $F(1,18)<1$, $\eta_p^2=0.029$, group by site, $F(6,46.342)<1$, $\eta_p^2=0.016$; N1 peak latency: group, $F(1,18)=2.446$, $p=0.135$, $\eta_p^2=0.12$). Furthermore, there was no correlation between either the peak amplitude or the peak latency of N1 to A only stimulus and performance on either the AV200 or AV300 SOA (N1 peak amplitude: AV200, $B=-0.965$, $R=0.118$, $F(1,30)<1$; AV300, $B=-0.238$, $R=0.064$, $F(1,30)<1$; N1 peak latency: AV200, $B=733.9$, $R=0.278$, $F(1,30)=2.432$, $p=0.13$; AV300, $B=117.43$, $R=0.098$, $F(1,30)<1$).

4. Discussion

We have determined that the ERP response between 210 and 270 ms following the auditory stimulus onset can predict up to 36% of individuals' ability to detect temporal AV asynchrony at the intermediate (100-300 ms) SOAs. More specifically, those individuals who had more positive mean voltage within this window detected the audiovisual asynchrony with higher accuracy. In contrast, none of the analyses of ERPs to VA SOAs survived the FDR correction.

Because the temporal window predictive of performance at the AV SOAs was time-locked to the onset of the sound, the most logical explanation for our finding is that the accuracy with which individuals detect audiovisual asynchrony is tied to their processing of the auditory stimulus. A more nuanced understanding of the nature of the sensory and/or cognitive mechanisms unfolding during the 210-270 ms time window will require further study. One possibility is that the enhanced positivity in good performers reflects a more detailed neural representation of the sound onset, which can significantly facilitate judgment about its temporal overlap (or the lack thereof) with the visual stimulus. In agreement with this proposition, a comparison of ERPs to the A stimulus in good and poor performers showed a strong trend ($p=0.067$) for greater positivity during the 210-270 window in good performers, suggesting that our finding is not specific to audiovisual stimuli. Alternatively, electrophysiological processes taking place during the identified window may reflect the encoding of the offset of the sound stimulus. However, because our stimulus was 200 ms in duration, the offset-elicited ERPs would have had the latency of just 1-70 ms, which is unlikely given earlier research on the latency of sound offset-elicited ERPs (Pratt, Bleich, & Mittelman, 2005). Nevertheless, to rule out this possibility more definitively, studies that use varying sound durations and/or examine a relationship between ERPs elicited by sound gaps and performance on the SJT are needed. It is noteworthy that in both visual and auditory modalities, the N1 component is thought to reflect sensory processing of stimulus onsets and is influenced by such basic properties as brightness and saturation in visual modality and intensity and frequency in auditory modality. Somewhat surprisingly, the peak amplitude of these components did not differentiate good and poor performers in our study and was not related to performance on the SJT, suggesting that early sensory encoding does not play a major role in the ability to detect audiovisual temporal asynchrony.

The strength of a relationship between ERP measures and sensitivity to audiovisual asynchrony was comparable for the AV200 and AV300 SOAs ($R^2=0.359$ and 0.318 , respectively). This relationship was still significant but markedly weaker ($R^2=0.137$) for the AV100 SOA. One reason for this finding likely lies in the less diffuse distribution of participants' responses to the AV100 SOA. A close examination of Figure 7 reveals that a large proportion of participants perceived this stimulus as synchronous on 80-100% of trials, with relatively few perceiving synchrony on fewer than 70% of trials. Therefore, responses to the AV100 SOA had overall smaller variability. Additionally, ERPs to the auditory stimulus within the AV100 SOA overlapped with the ERPs to the onset of the visual stimulus, which was presented just 100 ms after the onset of the 200 ms tone. Although the largest ERP components elicited by visual stimuli appeared over occipital and parietal-occipital sites, smaller visual responses could be observed across the entire scalp (see Figure 3), potentially introducing a greater degree of variability into auditory ERP measurements at 100 ms SOA.

If, as we suggested above, a more detailed neural representation of auditory onsets significantly facilitates detection of AV asynchrony, why did we not see the same facilitative effect in the detection of VA asynchrony? Analysis of auditory ERP responses to VA stimuli in good and poor performers did not yield any significant results, and this lack of differences is clearly visible in Figure 6. We believe these findings may be accounted for by assuming that the neural mechanisms underlying the detection of AV and VA asynchronies are at least partially different. More specifically, we hypothesize that the sensitivity to AV temporal asynchrony (as reflected by performance on the SJT) is determined to a greater extent by the efficiency in the encoding of the auditory onset and by the neural mechanisms engaged during temporal processing per se. In contrast, sensitivity to the VA temporal asynchrony (which is much more likely for natural audiovisual events) is shaped primarily by the strength of multisensory integrative mechanisms triggered by the VA stimuli. While, admittedly, this proposal requires direct testing, it fits well with multiple well-established findings in research on audiovisual temporal processing, which consistently show significant differences in behavioral and neurophysiological responses to AV and VA temporal sequences.

To start with, an asymmetry in sensitivity to AV and VA asynchrony at moderate SOAs is well-established and has been replicated numerous times (Bushara et al., 2001; Dixon & Spitz, 1980; Grant et al., 2004; Lewkowicz, 1996; van Wassenhove et al., 2007). Across different ranges of SOAs and types of stimuli, adults tend to be more accurate at detecting AV compared to VA asynchronies. The ability to detect multisensory temporal asynchrony has a prolonged developmental course, with even adolescents still performing worse than adults (Hillock-Dunn & Wallace, 2012). More importantly, perception of AV asynchrony matures before that of VA asynchrony, with 10-11 children significantly outperforming their 7-8 year old peers on a SJT, but only at AV SOAs (Kaganovich, 2016). Further, several studies have demonstrated that the size of the TBW is malleable and can be reduced through perceptual training (Powers III et al., 2012; Powers III et al., 2009; Stevenson et al., 2013). Of note, however, is the fact that in all such studies, the reduction of the TBW was due to increased accuracy on the VA sequences of the SJT only. In other words, the perception of AV and VA asynchronies responds differently to training. In a similar vein, in the study by

Stevenson and colleagues (Stevenson et al., 2012), only the size of the VA portion of the TBW significantly predicted individuals' susceptibility to the McGurk illusion. Accounting for all of the above differences is beyond the scope of the current report. However, our finding contributes to the cumulative message of this literature by showing that the neural encoding of auditory stimuli also differs during the processing of AV and VA sequences.

While the number of studies on the neural correlates of audiovisual integration is quite large, only a few focused specifically on sensitivity to audiovisual temporal synchrony/asynchrony (Bushara et al., 2001; Powers III et al., 2012; Stevenson et al., 2010). These studies have not reported any major differences in the neural networks for detecting asynchrony in AV and VA stimuli - the result that, seemingly, does not agree with our finding. In one of the earliest studies, Bushara and colleagues (Bushara et al., 2001) reported a nearly identical set of activated brain regions when their participants made simultaneity judgments about AV and VA stimuli, with the only difference being the bilateral, rather than the right unilateral, activation of the inferior frontal gyrus. However, in this analysis, neural responses to both short and long SOAs were averaged together, potentially obscuring modality order differences that are most noticeable at intermediate SOAs. In an additional analysis, these authors also identified the brain region - namely, the right insula - whose activation correlated with the difficulty of the SJT. Unfortunately, in this case, the neural activity during AV and VA trials was averaged together and no interaction between the order of modalities and the degree of the right insula activation was reported. The studies by Stevenson and colleagues (Stevenson et al., 2010) and by Powers and colleagues (Powers III et al., 2012) both focused primarily on identifying brain regions that respond preferentially to either synchronous or clearly asynchronous audiovisual stimuli, with little detail provided about potential modality order differences. The report by Stevenson and colleagues (Stevenson et al., 2011) on different neural correlates of physical and perceived audiovisual temporal synchrony used only auditory-visual sequences as stimuli. In sum, recent findings have identified important components of the neural network underlying audiovisual temporal function. However, whether AV and VA sequences lead to similar patterns in the activation of this network remains undetermined.

When evaluating ERPs elicited by AV200 and VA200 stimuli in good and poor performers, we included all trials - the ones that were perceived as synchronous and the ones that were perceived as asynchronous - in our averages. Since good performers had more trials on which they subjectively perceived asynchrony, could the ERP differences we found for the AV200 SOA reflect the subjective perception of asynchrony rather than better encoding of the auditory stimulus as suggested above? A comparison of ERPs to the AV200 stimulus when it was perceived (incorrectly) as synchronous and when it was perceived (correctly) as asynchronous in the group of "average" performers helped us address this issue. Although this analysis did not reach significance at the alpha level of 0.05, its effect size was quite large ($\eta_p^2=0.247$). Because the perception-driven analysis was based on a relatively small number of trials and included only 11 participants, it is likely that it lacked sufficient power. Yet, its outcome is suggestive, especially because it yielded a pattern of results that was opposite in polarity to that seen in the analyses based on overall performance. More specifically, while good performers at the AV200 SOA had more positive voltage 210-270 ms post-auditory onset compared to poor performers, the correct perception of asynchrony at

the AV200 SOA was associated with more negative voltage in the same window (compare Figures 5 and 8). This finding suggests that the ERP indices of efficient encoding of stimulus onsets within multisensory contexts may dissociate from the ERP indices of subjective perception of audiovisual asynchrony, perhaps in a manner similar to that reported by Stevenson and colleagues for fMRI measures (Stevenson et al., 2011). However, more studies that carefully manipulate perception of audiovisual asynchrony across participants are needed to make more definitive conclusions about differences or similarities in the ERP correlates of objective vs. subjective audiovisual temporal synchrony.

Lastly, as all studies, our report has its limitations. The auditory stimuli used in the current paradigm were pure tones with a short rise and fall time. We know that sound complexity and its "speechness" can significantly affect both the amplitude and latency of auditory ERP components (e.g., eponien, Alku, Westerfield, Torki, & Townsend, 2005). An important empirical question to address in future work is whether the time window identified in our study remains constant across different auditory stimulus types. Additionally, because our stimuli were 200 ms in duration, the amount of overlap between auditory and visual modalities varied for different SOAs. At longer SOAs (300-500 ms), the processing of both the onset and the offset of the first stimulus in a pair might have been completed prior to the onset of the second stimulus, potentially facilitating decision making at longer SOAs. Most of the earlier studies used significantly shorter stimuli in similar paradigms (5-10ms) (e.g., Grondin et al., 2007; Hillock-Dunn & Wallace, 2012). However, since many audiovisual events in real life last more than 10 ms, we believe our results are informative. Nevertheless, replicating our findings with shorter stimuli would help rule out modalities' overlap as a contributing factor to audiovisual temporal sensitivity.

To conclude, we examined electrophysiological correlates of individual variability in sensitivity to audiovisual temporal asynchrony in a SJT. We reported that good and poor performers had significantly different ERPs to the AV200 stimulus 210-270 ms following the sound onset, with greater positivity in good performers. Furthermore, the mean amplitude of voltage in this window accounted for a significant amount of variability in performance not only at the AV200 SOA but also at the AV100 and AV300 SOAs in a larger group of participants. We suggested that the ability to detect temporal asynchrony in AV sequences depends, to a large extent, on the efficient encoding of auditory onsets.

Acknowledgements

This research was supported in part by grants P30DC010745 and R03DC013151 from the National Institute on Deafness and Other Communicative Disorders, National Institutes of Health. The content is solely the responsibility of the author and does not necessarily represent the official view of the National Institute on Deafness and Other Communicative Disorders or the National Institutes of Health. We are thankful to James Hengenius, Kevin Barlow, Amanda Hampton Wray, Caryn Herring, Courtney Rowland, Jessica Huemmer, Kelly Sievert, and Casey Spelman for their assistance with different stages of this project.

References

- American Electroencephalographic Society. Guideline thirteen: Guidelines for standard electrode placement nomenclature. 1994. Retrieved from
- Baart M, Stekelenburg JJ, Vroomen J. Electrophysiological evidence for speech-specific audiovisual integration. *Neuropsychologia*. 2014; 53:115–121. [PubMed: 24291340]

- Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1995; 57:289–300.
- Besle J, Fort A, Delpuech C, Giard M-H. Bimodal speech: early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*. 2004; 20:2225–2234. [PubMed: 15450102]
- BioSemi. Active Electrodes. 2013. Retrieved from http://www.biosemi.com/active_electrode.htm
- Brown, L.; Sherbenou, RJ.; Johnsen, SK. *Test of Nonverbal Intelligence*. 4th. Pro-Ed: An International Publisher; Austin, Texas: 2010.
- Bushara KO, Grafman J, Hallett M. Neural correlates of auditory-visual stimulus onset asynchrony detection. *The Journal of Neuroscience*. 2001; 21(1)
- eponien R, Alku P, Westerfield M, Torki M, Townsend J. ERPs differentiate syllable and nonphonetic sound processing in children and adults. *Psychophysiology*. 2005; 42:391–406. [PubMed: 16008768]
- Conrey B, Pisoni DB. Auditory-visual speech perception and synchrony detection for speech and non-speech signals. *Journal of the Acoustical Society of America*. 2006; 119(6):4065–4073. doi: 10.1121/1.2195091. [PubMed: 16838548]
- Dixon NF, Spitz L. The detection of auditory visual desynchrony. *Perception*. 1980; 9:719–721. [PubMed: 7220244]
- Donohue SE, Darling EF, Mitroff SR. Links between multisensory processing and autism. *Experimental Brain Research*. 2012; 222:377–387. [PubMed: 22923209]
- Field, A. *Discovering statistics using SPSS*. 3. Sage; Washington, DC: 2009.
- Foss-Feig JH, Kwakye LD, Cascio CJ, Burnette CP, Kadivar H, Stone WL, Wallace MT. An extended multisensory temporal binding window in autism spectrum disorders. *Experimental Brain Research*. 2010; 203:381–389. doi:10.1007/s00221-010-2240-4. [PubMed: 20390256]
- Grant KW, van Wassenhove V, Poeppel D. Detection of auditory (cross-spectral) and auditory-visual (cross-modal) synchrony. *Speech Communication*. 2004; 44:43–53.
- Grondin S, Dionne G, Malenfant N, Plourde M, Cloutier M, Jean C. Temporal processing skills of children with and without specific language impairment. *Canadian Journal of Speech-Language Pathology and Audiology*. 2007; 31(1):38–46.
- Groppe DM, Urbach TP, Kutas M. Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*. 2011; 48:1711–1725. doi:DOI: 10.1111/j.1469-8986.2011.01273.x. [PubMed: 21895683]
- Hairston WD, Burdette JH, Flowers DL, Wood FB, Wallace MT. Altered temporal profile of visual-auditory multisensory interactions in dyslexia. *Experimental Brain Research*. 2005; 166:474–480. doi:10.1007/s00221-005-2387-6. [PubMed: 16028030]
- Hillock-Dunn A, Wallace MT. Developmental changes in the multisensory temporal binding window persist into adolescence. *Developmental Science*. 2012; 15(5):688–696. doi:10.1111/j.1467-7687.2012.01171.x. [PubMed: 22925516]
- Kaganovich N. Development of sensitivity to audiovisual temporal asynchrony during mid-childhood. *Developmental Psychology*. 2016; 52(2):232–241. [PubMed: 26569563]
- Kaganovich N, Schumaker J. Audiovisual integration for speech during mid-childhood: Electrophysiological evidence. *Brain and Language*. 2014; 139:36–48. [PubMed: 25463815]
- Kaganovich N, Schumaker J, Leonard LB, Gustafson D, Macias D. Children with a history of SLI show reduced sensitivity to audiovisual temporal asynchrony: An ERP study. *Journal of Speech, Language, and Hearing Research*. 2014; 57:1480–1502. doi:10.1044/2014_JSLHR-L-13-0192.
- Keetels, M.; Vroomen, J. Perception of synchrony between the senses. In: Murray, MM.; Wallace, MT., editors. *The Neural Bases of Multisensory Processes*. CRC Press; New York: 2012. p. 147-177.
- Knowland VCP, Mercure E, Karmiloff-Smith A, Dick F, Thomas MSC. Audio-visual speech perception: a developmental ERP investigation. *Developmental Science*. 2014; 17(1):110–124. [PubMed: 24176002]
- Kwakye LD, Foss-Feig JH, Cascio CJ, Stone WL, Wallace MT. Altered auditory and multisensory temporal processing in autism spectrum disorders. *Frontiers in Integrative Neuroscience*. 2011; 4(129) doi:10.3389/fnint.2010.00129.

- Lewkowicz DJ. Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance*. 1996; 22(5):1094–1106. [PubMed: 8865617]
- Luck, SJ. *An Introduction to the Event-Related Potential Technique*. second. The MIT Press; Cambridge, MA: 2014.
- Macaluso E, George N, Dolan R, Spence C, Driver J. Spatial and temporal factors during processing of audiovisual speech: a PET study. *NeuroImage*. 2004; 21:725–732. [PubMed: 14980575]
- Metting van Rijn, AC.; Kuiper, AP.; Dankers, TE.; Grimbergen, C,A. Low-cost active electrode improves the resolution in biopotential recordings; Paper presented at the 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Amsterdam, The Netherlands. 1996.
- Metting van Rijn AC, Peper A, Grimbergen CA. High-quality recording of bioelectric events. Part 1: Interference reduction, theory and practice. *Medical and Biological Engineering and Computing*. 1990; 28:389–397. [PubMed: 2277538]
- Petrini K, Dahl S, Rocchesso D, Waadeland CH, Avanzini D, Puce A, Pollick FE. Multisensory integration of drumming actions: musical expertise affects perceived audiovisual asynchrony. *Experimental Brain Research*. 2009; 198:339–352. doi:10.1007/s00221-009-1817-2. [PubMed: 19404620]
- Pfieger, ME. Theory of a spatial filter for removing ocular artifacts with preservation of EEG; Paper presented at the EMSE Workshop, Princeton University; 2001. http://www.sourcesignal.com/SpFilt_Ocular_Artifact.pdf
- Powers AR III, Hevey MA, Wallace MT. Neural correlates of multisensory perceptual learning. *Journal of Neuroscience*. 2012; 32(18):6263–6274. [PubMed: 22553032]
- Powers AR III, Hillock AR, Wallace MT. Perceptual training narrows the temporal window of multisensory binding. *The Journal of Neuroscience*. 2009; 29(39):12265–12274. doi:10.1523/JNEUROSCI.3501-09.2009. [PubMed: 19793985]
- Pratt, H. Sensory ERP components. In: Luck, SJ.; Kappenman, ES., editors. *The Oxford Handbook of Event-Related Potential Components*. Oxford University Press; Oxford: 2011. p. 89-114.
- Pratt H, Bleich N, Mittelman N. The composite N₁ component to gaps in noise. *Clinical Neurophysiology*. 2005; 116:2648–2663. [PubMed: 16221565]
- Semel, E.; Wiig, EH.; Secord, WA. CELF4: Clinical Evaluation of Language Fundamentals. 4th. Pearson Clinical Assessment; San Antonio, TX: 2003.
- Stein, BE.; Meredith, MA. *The Merging of the Senses*. The MIT Press; Cambridge, Massachusetts: 1993.
- Stekelenburg JJ, Vroomen J. Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience*. 2007; 19(12):1964–1973. [PubMed: 17892381]
- Stevenson RA, Altieri N, Kim S, Pisoni DB, James TW. Neural processing of asynchronous audiovisual speech perception. *NeuroImage*. 2010; 49:3308–3318. [PubMed: 20004723]
- Stevenson RA, Siemann JK, Schneider BC, Eberly HE, Woyanorski TG, Camarata SM, Wallace MT. Multisensory temporal integration in autism spectrum disorders. *Journal of Neuroscience*. 2014; 34:691–697. [PubMed: 24431427]
- Stevenson RA, Siemann JK, Schneider BC, Eberly HE, Woyanorski TG, Camarata SM, Wallace MT. Multisensory temporal integration in autism spectrum disorders. *The Journal of Neuroscience*. 2014; 34(3):691–697. doi:10.1523/JNEUROSCI.3615-13.2014. [PubMed: 24431427]
- Stevenson RA, VanDerKlok RM, Pisoni DB, James TW. Discrete neural substrates underlie complementary audiovisual speech integration processes. *NeuroImage*. 2011; 55:1339–1345. [PubMed: 21195198]
- Stevenson RA, Wallace MT. Multisensory temporal integration: task and stimulus dependencies. *Experimental Brain Research*. 2013; 227:249–261. doi:10.1007/s00221-013-3507-3. [PubMed: 23604624]
- Stevenson RA, Wilson MM, Powers AR, Wallace MT. The effects of visual training on multisensory temporal processing. *Experimental Brain Research*. 2013; 225:479–489. doi:10.1007/s00221-012-3387-y. [PubMed: 23307155]

- Stevenson RA, Zemtsov RK, Wallace MT. Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology: Human Perception and Performance*. 2012; 38(6):1517–1529. doi:10.1037/a0027339. [PubMed: 22390292]
- van Eijk R, Kohlrausch A, Juola J, van de Par S. Audiovisual synchrony and temporal order judgments: effects of experimental method and stimulus type. *Perception and Psychophysics*. 2008; 70(6):955–968. doi:10.3758/PP.70.6.955. [PubMed: 18717383]
- van Wassenhove V, Grant KW, Poeppel D. Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences*. 2005; 102(4):1181–1186.
- van Wassenhove V, Grant KW, Poeppel D. Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*. 2007; 45:598–607. [PubMed: 16530232]
- Vatakis, A.; Spence, C. Audiovisual temporal integration for complex speech, object-ation, animal call, and musical stimuli. In: Naumer, MJ.; Kaiser, J., editors. *Multisensory Object Perception in the Primate Brain*. Springer; New York: 2010.
- Vroomen J, Keetels M. Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics*. 2010; 72(4):871–884. doi:10.3758/APP.72.4.871.
- Vroomen J, Stekelenburg JJ. Perception of intersensory synchrony in audiovisual speech: Not that special. *Cognition*. 2011; 118(1):75–83. doi:10.1016/j.cognition.2010.10.002. [PubMed: 21035795]
- Wallace MT, Stevenson RA. The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia*. 2014; 64:105–123. [PubMed: 25128432]

Highlights

- Adults vary greatly in the ability to perceive AV asynchrony at medium SOAs
- Good performers at AV200 SOA had more positive voltage 210-270 ms post sound onset
- Voltage within the 210-270 ms window predicted 36% of accuracy at the AV200 SOA
- Relationship between ERPs and accuracy on the SJT also held for AV100 and AV300 SOAs
- Encoding of auditory onsets underlies variability in audiovisual temporal function

Figure 1. Stimuli

An explosion-shaped flash of light and a 2 kHz pure tone were both 200 ms in duration. The images of the boy and the girl were absent during trial presentation. From “Children with a History of SLI Show Reduced Sensitivity to Audiovisual Temporal Asynchrony: An ERP Study,” by N. Kaganovich, J. Schumaker, L. B. Leonard, D. Gustafson, and D. Macias, 2014, *Journal of Speech, Language, and Hearing Research*, Volume 57, p.1480-1502, <http://jslhr.pubs.Asha.org/>.

Figure 2. Auditory Responses

A. Grand average ERPs elicited by the 2kHz pure tone on A trials are shown for the entire group of participants ($n=31$). Electrodes FP1 and FP2 as well as the left and the right mastoids are not shown. Negative is plotted up. B. The distribution of voltage over the scalp at the peak amplitude of the auditory N1 component based on the grand average waveform (122 ms post-stimulus onset). Note that the auditory N1 was broadly distributed across the scalp with the largest values over fronto-central, central, and centro-parietal sites (shown in red in panel A). C. The auditory ERPs elicited by the onset of the pure tone and the window used for the auditory analysis (100-380 ms post-stimulus onset) are shown on the CZ site.

Figure 3. Visual Responses

Grand average ERPs elicited by the explosion-shaped figure on V trials are shown for the entire group of participants ($n=31$). Electrodes FP1 and FP2 as well as the left and the right mastoids are not shown. Negative is plotted up. B. The distribution of voltage over the scalp at the peak amplitude of the visual N1 component based on the grand average waveform (164 ms post-stimulus onset). Note that the visual N1 had the largest values over occipital and parietal sites (shown in red in Panel A). C. The visual ERPs elicited by the onset of the explosion-shaped figure and the window used for the visual analysis (80-400 ms post-stimulus onset) are shown on the OZ site.

Figure 4. Synchronous perceptions across SOAs

The number of synchronous perceptions (out of 50 trials) for each of the 11 SOAs is shown for the entire group of participants ($n=31$). Error bars are standard errors of the mean.

Figure 5. ERPs to the AV200 stimulus in good and poor performers

A. Grand average ERPs elicited by the AV200 stimulus is overlaid for good ($n=10$) and poor ($n=10$) performers time-locked to the onset of the tone (time 0). The explosion-shaped figure appeared 200 ms after time 0. Auditory N1 (Aud N1) is marked on the CZ site while visual N1 (Vis N1) is marked on the OZ site. Negative is plotted up. B. Temporal windows over which paired t-tests were used to compare the good and poor performer groups are shown in light grey on the CZ site for the auditory analysis and on the OZ site for the visual analysis. Note that the temporal window for the auditory analysis was time-locked to the onset of the auditory stimulus, and the temporal window for the visual analysis was time-locked to the onset of the visual stimulus. C. The time window over which ERPs of good and poor performers differed significantly to the AV200 stimulus (210-270 ms post-auditory onset) is shown in light grey on the CZ site.

Figure 6. ERPs to the VA200 stimulus in good and poor performers

A. Grand average ERPs elicited by the VA200 stimulus is overlaid for good ($n=10$) and poor ($n=10$) performers time-locked to the onset of the explosion-shaped figure (time 0). The tone appeared 200 ms after time 0. Auditory N1 (Aud N1) is marked on the CZ site while visual N1 (Vis N1) is marked on the OZ site. Negative is plotted up. B. Temporal windows over which paired t-tests were used to compare the poor and good performer groups are shown in light grey on the CZ site for the auditory analysis and on the OZ site for the visual analysis. Note that the temporal window for the auditory analysis was time-locked to the onset of the auditory stimulus, and the temporal window for the visual analysis was time-locked to the onset of the visual stimulus.

Figure 7. Regressions

An outcome of the regression analyses with the mean voltage between 210 and 270 ms post sound onset as a predictor and the number of synchronous perception on the SJT as an outcome ($n=31$).

Figure 8. ERPs elicited by the AV200 SOA when perceived as synchrony vs. when perceived as asynchrony

A. Grand average ERPs elicited by the AV200 stimulus when it was perceived (incorrectly) as synchronous is overlaid with grand average ERPs elicited by the same stimulus when it was perceived (correctly) as asynchronous within the same group of “average” participants ($n=11$). Time 0 is the onset of the tone. The explosion-shaped figure appeared 200 ms after time 0. Auditory N1 (Aud N1) is marked on the CZ site while visual N1 (Vis N1) is marked on the OZ site. Negative is plotted up.

B. The temporal window (210-270 ms post-stimulus onset) over which the mean amplitude was compared for each type of perception (synchronous vs. asynchronous) is shown in light grey on the CZ site. Compare to panel C in Figure 5.

Figure 9. Temporal binding window in good and poor performers and its relationship to the ERP measure

A: Group mean data for the left (VA) and right (AV) sides of the TBW window at 50% of the maximum are shown for good and poor performers (as defined by performance at the AV200 SOA). The distance between the vertical line at 0 ms SOA and dotted drop-down lines reflects the size of the TBW in each group. Note that groups do not differ in the size of the VA TBW, but good performers have a significantly smaller AV TBW.

B: The result of the regression analysis with the mean voltage between 210 and 270 ms post sound onset as a predictor and the size of the AV TBW as an outcome ($n=31$).

Table 1

Group Performance on the Simultaneity Judgment Task

Stimulus	Mean Synchronous Perceptions	Range	Standard Error of the Mean
VA500	1.03	4	.24
VA400	3.87	14	.79
VA300	19.55	50	2.58
VA200	48.45	72	3.49
VA100	88.26	38	1.71
SYNC	91.36	28	1.53
AV100	81.74	60	2.84
AV200	45.48	84	4.73
AV300	13.81	38	2.14
AV400	3.61	16	.83
AV500	1.42	8	.38

VA indicates a stimulus in which the explosion figure preceded the tone, and AV indicates a stimulus in which the tone preceded the explosion figure. The number following the VA and AV abbreviations reflects the SOA. For example, in the VA200 stimulus, the onset of the explosion figure occurred 200 ms prior to the onset of tone. Range indicates the difference between the smallest and the largest number of reported synchronous perceptions in the group of 31 participants. Stimuli with the largest range and standard error of the mean are shown in bold.

Table 2
Outcome of t-tests comparing good and poor performers' auditory ERP responses within the context of the AV200 stimulus

Site/ms	190- 200	201- 210	211- 220	221- 230	231- 240	241- 250	251- 260	261- 270	271- 280	281- 290	291- 300
FC1	*	*	**	**	**	**	*				
C3		*	*	**	**	**	**				
CP1			*	*	**	**	**	*			
CZ			*	*	**	**	**	*			
CP2			*	*	**	**	**	**	**		
C4		*	**	**	**	**	**	**	**	**	
FC2			**	**	**	**	*				

The top row indicates time following the onset of the pure tone within the AV200 stimulus. The ERP data was down sampled to 100 Hz; therefore, each column represents one measurement point spanning 10 ms. * indicates FDR corrected p values of 0.1 or less. ** indicates FDR corrected p values of 0.05 or less. Measurement points at which more than half electrode sites showed significance at either the 0.05 or 0.1 alpha levels were included in the final window for regression analysis (i.e., all values within a rectangle).