Neuropsychologia 86 (2016) 119–130

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Electrophysiological correlates of individual differences in perception of audiovisual temporal asynchrony

Natalya Kaganovich [a,b,*], Jennifer Schumaker [a]

[a] Department of Speech, Language, and Hearing Sciences, Purdue University, 715 Clinic Drive, West Lafayette, IN 47907-2038, United States
[b] Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907-2038, United States

ARTICLE INFO

Article history:
Received 13 November 2015
Received in revised form
14 April 2016
Accepted 15 April 2016
Available online 16 April 2016

Keywords:
Temporal binding window
Audiovisual temporal processing
Simultaneity judgment task
ERPs
Individual differences

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.

1. Introduction

Temporal proximity is one of the determining factors for integrating multisensory, and more specifically audiovisual, stimuli into a coherent percept (Stein and 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 and Vroomen, 2012; Vatakis and Spence, 2010; Vroomen and 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 et al., 2001; Dixon and Spitz, 1980; Grant et al., 2004; Lewkowicz, 1996; van Wassenhove et al., 2007); it is larger for speech and other complex stimuli compared to simpler non-speech stimuli (Vatakis and Spence, 2010; Vroomen and Stekelenburg, 2011); it can be reduced through perceptual training (Powers III et al., 2009; Stevenson et al., 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 and Wallace, 2013; van Eijk et al., 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 and Stevenson, 2014), such as dyslexia (Hairston et al., 2005), specific language impairment (SLI) (Grondin et al., 2007; Kaganovich et al., 2014), and autism (Foss-Feig et al., 2010; Kwakye et al., 2011; Stevenson 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 et al., 2012) examined a correlation between the degree

* Correspondence to: Department of Speech, Language, and Hearing Sciences, Purdue University, Lyles Porter Hall, 715 Clinic Drive, West Lafayette, IN 47907-2038, United States.
E-mail address: kaganovi@purdue.edu (N. Kaganovich).

http://dx.doi.org/10.1016/j.neuropsychologia.2016.04.015
0028-3932/

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) (400–500 ms) also had lower core language scores as determined by the Clinical Evaluation of Language Fundamentals (CELF-4; Semel et al., 2003). Second, in healthy adults, the smaller size of the TBW was linked to greater susceptibility to the McGurk illusion (Stevenson et al., 2012) and to better comprehension of degraded audiovisual speech (Conrey and Pisoni, 2006). The relationship between the TBW and susceptibility to the McGurk illusion has also been replicated in children with autism (Stevenson 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 long[1] (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 et al., 2004; Powers III et al., 2012; Stevenson et al., 2010) and the right insula (Bushara et al., 2001). In an insightful addition to this literature, Powers and colleagues (Powers III et al., 2012, 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 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 et al., 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 et al., 2014; Besle et al., 2004; Kaganovich and Schumaker, 2014; Knowland et al., 2014; Stekelenburg and Vroomen, 2007; van Wassenhove et al., 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.

[1] The 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.