Journal of Experimental Child Psychology 146 (2016) 17–33

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

The development of the perception of audiovisual simultaneity

Yi-Chuan Chen, David I. Shore, Terri L. Lewis, Daphne Maurer *

Department of Psychology, Neuroscience and Behavior, McMaster University, Hamilton, Ontario L8S 4K1, Canada

ARTICLE INFO

Article history:
Received 3 June 2015
Revised 9 January 2016
Available online 18 February 2016

Keywords:
Crossmodal development
Audiovisual simultaneity window
Simultaneity judgments
Childhood
Temporal sensitivity
Point of subjective simultaneity

ABSTRACT

We measured the typical developmental trajectory of the window of audiovisual simultaneity by testing four age groups of children (5, 7, 9, and 11 years) and adults. We presented a visual flash and an auditory noise burst at various stimulus onset asynchronies (SOAs) and asked participants to report whether the two stimuli were presented at the same time. Compared with adults, children aged 5 and 7 years made more simultaneous responses when the SOAs were beyond ±200 ms but made fewer simultaneous responses at the 0 ms SOA. The point of subjective simultaneity was located at the visual-leading side, as in adults, by 5 years of age, the youngest age tested. However, the window of audiovisual simultaneity became narrower and response errors decreased with age, reaching adult levels by 9 years of age. Experiment 2 ruled out the possibility that the adult-like performance of 9-year-old children was caused by the testing of a wide range of SOAs. Together, the results demonstrate that the adult-like precision of perceiving audiovisual simultaneity is developed by 9 years of age, the youngest age that has been reported to date.

Introduction

In everyday life, we must decide which sights and sounds come from the same event. To do so, we rely on two basic rules: spatial coincidence and temporal synchrony (Stein & Meredith, 1993; Welch &

* Corresponding author.
E-mail address: maurer@mcmaster.ca (D. Maurer).

http://dx.doi.org/10.1016/j.jecp.2016.01.010
0022-0965/

Warren, 1980). From birth, rudimentary forms of these rules are evident (Lewkowicz, Leo, & Simion, 2010; Morrongiello, Fenwick, & Chance, 1998) even though it takes many years for the precision demonstrated by adults to emerge. These rudimentary abilities aid the development of cognitive and social skills. For example, infants perceive causal relations between a visual collision event and a crashing sound (e.g., Scheier, Lewkowicz, & Shimojo, 2003) or acquire knowledge of an object by associating its visual features and the sound that it produces as well as the name that it is called (e.g., Chen & Westermann, 2012; Werker, Cohen, Lloyd, Casasola, & Stager, 1998; see Westermann & Mareschal, 2014, for a review). Integrating visual and auditory signals also aids infants' perception of speech (e.g., Lewkowicz & Hansen-Tift, 2012; Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009) and emotion (e.g., Walker-Andrews, 1986; see Walker-Andrews, 1997, for a review). In the current study, our goal was to measure a typical developmental trajectory of associating visual and auditory stimuli as a single event in terms of their temporal synchrony.

In their seminal neurophysiological studies, Stein and Meredith (1993) measured firing rates of cells in the superior colliculus of cats. Neural activity was stronger for multisensory inputs presented close in space and time compared with the sum of activity induced by sensory input from the individual modalities (see Stein, Stanford, Ramachandran, Perrault, & Rowland, 2009, for a review). In studies of human adult perception, there is considerable evidence for the importance of temporal synchrony, although the importance of spatial coincidence has been questioned recently (Spence, 2013). For example, when detecting the occurrence of a visual target, accuracy was higher when its onset was accompanied by a temporally synchronous sound than when no sound was presented (e.g., Andersen & Mamassian, 2008; Bolognini, Frassinetti, Serino, & Ladavas, 2005; Chen, Huang, Spence, & Yeh, 2011; Frassinetti, Bolognini, & Lavadas, 2002; Lippert, Logothetis, & Kayser, 2007). In addition, the presentation of a simultaneous and congruent sound enhances people's perceptual learning and memory retrieval for a visual stimulus compared with when the visual stimulus is presented alone (Flom & Bahrick, 2010; Murray, Foxe, & Wylie, 2005; Murray et al., 2004; Seitz, Kim, & Shams, 2006). Stronger evidence comes from studies demonstrating that the identification/discrimination of a visual target is enhanced by the presentation of a synchronous sound even though the sound provided no information about the accurate response (Chen & Spence, 2011; Chen & Yeh, 2008, 2009; Lu et al., 2009; Ngo & Spence, 2010a, 2010b; Olivers & van der Burg, 2008; van der Burg, Cass, Olivers, Theeuwes, & Alais, 2010; van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008; Vroomen & de Gelder, 2000).

Perceiving that a visual stimulus and an auditory stimulus are temporally synchronous is, nevertheless, a complex process. On the one hand, light travels faster than sound; on the other hand, peripheral processing time is longer in the visual system than in the auditory system (see Arrighi, Alais, & Burr, 2006). Hence, when visual and auditory signals originate from the same object or event, the neural activities they generate are likely to arrive at different times at the central mechanism decoding the timing of multisensory information, with the discrepancy depending on the distance and intensity of each signal (e.g., King, 2005). Accordingly, we learn a compromise between precision, which would lead us to miss many multisensory events, and flexibility, which would lead us to integrate inputs that are truly distinct.

One solution is to create a likelihood distribution of perceptual simultaneity according to the onset timing of the visual and auditory stimuli. This distribution is commonly called the audiovisual simultaneity window (see Vroomen & Keetels, 2010, for a review), although it is sometimes instead called the audiovisual "temporal binding window" (e.g., Hillock, Powers, & Wallace, 2011; Lewkowicz & Flom, 2014; Stevenson, Zemtsov, & Wallace, 2012). These two terms are often used interchangeably; however, there are subtle distinctions in that judging two events as simultaneous is not the same as perceiving the consequences of multisensory integration (e.g., Baart, Stekelenburg, & Vroomen, 2014; Eskelund, Tuomainen, & Andersen, 2011; Vroomen & Stekelenburg, 2011). Consider the McGurk effect (McGurk & MacDonald, 1976), a phenomenon that occurs only when visual and auditory speech information is integrated, for example. The occurrence of the McGurk effect is correlated *negatively* to the width of the simultaneity window (Stevenson et al., 2012), and the McGurk effect can still occur outside of the window in which its component stimuli were judged as simultaneous (Soto-Faraco & Alsius, 2009). Considering developmental data, the matching of audiovisual stimuli in terms of temporal synchrony develops at a younger age (by 7 years) than their matching in terms of stimulus