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 identity (Baart, Bortfeld, & Vroomen, 2015). It may be that the temporal windows for audiovisual simultaneity judgment and audiovisual integration are dissociated (see Chen & Spence, 2011; Fujisaki & Nishida, 2010). In the current article, we explicitly use simultaneity judgments; therefore, we refer to our measurement as the *audiovisual simultaneity window*.

The development of the audiovisual simultaneity window has been studied with infants using a habituation paradigm (Lewkowicz, 1996, 2010). Typically, infants are familiarized with a synchronous audiovisual event during the habituation phase (i.e., the stimulus onset asynchrony [SOA], is 0 ms). Then, in the subsequent test phase, infants are tested for dishabituation to an asynchronous audiovisual event. Each infant has been tested with only three SOAs after habituation because of the long habituation procedure and infants' limited attention. Thus, the habituation paradigm can provide only a crude measure of infants' audiovisual simultaneity window. Even so, the habituation procedure is sufficiently sensitive to show that the audiovisual simultaneity window is much wider in infants than in adults. For example, in a study using visual bouncing paired with a complex tone, Lewkowicz (1996) demonstrated that infants aged 2 to 8 months were sensitive to the asynchronous pairing when the tone led the visual rebound by 350 ms or when the visual rebound led the tone by 450 ms, whereas adults perceived asynchrony when the tone led by only 65 ms and when the visual rebound led by only 112 ms. When using speech stimuli, infants aged 4 to 10 months were sensitive to the asynchronous pairing when the sound led the visual stimulus by 666 ms rather than by 500 ms (Lewkowicz, 2010).

Another paradigm suitable for young children is a same–different discrimination task. For example, Lewkowicz and Flom (2014) familiarized 4-, 5-, and 6-year-olds with an audiovisual event consisting of a voice synchronous with a speaking face for 20 s on one of two adjacent monitors. After the 20-s familiarization, another audiovisual event consisting of a voice leading a speaking face by one of four possible SOAs (0, 366, 500, and 666 ms) was presented on the other monitor. The task on each trial was to judge whether the two events were the same or different. The results demonstrated greater precision with increasing age: children aged 4, 5, and 6 years were able to detect the asynchrony when the voice led the speaking face by 666, 500, and 366 ms, respectively. As with the habituation procedure, very few SOAs can be tested with the same–different discrimination task because of the long familiarization procedure.

A third task—simultaneity judgment—has been used with older children and adults (e.g., Hillock et al., 2011; Hillock-Dunn & Wallace, 2012; van Eijk, Kohlrausch, Juola, & van de Par, 2008; Zampini, Guest, Shore, & Spence, 2005; see Vroomen & Keetels, 2010, for a review). In this task, a visual stimulus and an auditory stimulus are presented with various SOAs. Participants are asked to judge on each trial whether or not the visual and auditory stimuli were presented at the same time. Typically, the probability of participants' simultaneous response increases when the SOA between the visual and auditory stimuli is close to 0 ms (i.e., physically simultaneous), giving rise to a Gaussian-shaped distribution of proportion of simultaneous responses when plotted against SOA. Based on this distribution, two critical perceptual parameters of the simultaneity window can be estimated: the point of subjective simultaneity (PSS) and the width of the simultaneity window. The PSS corresponds to the center of the distribution where the participant was most likely to make a "simultaneous" response. Note that the PSS does not necessarily correspond to physical simultaneity (i.e., 0 ms SOA). Instead, the PSS is typically located at the visual-leading side (i.e., maximal simultaneity responses are observed when the visual stimulus is physically presented before the auditory stimulus); it has also been described as "asymmetrical" with a wider window on the visual-leading side than on the auditory-leading side with respect to 0 ms SOA (e.g., Hillock et al., 2011; Hillock-Dunn & Wallace, 2012).

The width of the simultaneity window is defined by the range over which the participant's proportion of simultaneous response is higher than a particular criterion. The criterion to determine the width of the simultaneity window, however, is arbitrary. In studies where the data were fitted with a Gaussian function, the standard deviation of the distribution has been used to represent the width (e.g., Petrini, Russell, & Pollick, 2009; Zampini et al., 2005; see Vroomen & Keetels, 2010). In other studies where the data were fitted with two sigmoid functions, the SOA corresponding to a particular probability of a simultaneous response was estimated; however, note that the probability was set differently in different studies: 50% in Stevenson et al. (2012) and Stevenson and Wallace (2013), 70% in Stevenson and Wallace (2013), and 75% in Hillock et al. (2011) and Hillock-Dunn and Wallace (2012). The variability of the indexes used in different studies makes it hard to compare the width of the distributions directly.

Together, the current results suggest that infants' audiovisual simultaneity window is more than four times wider than that of adults (Lewkowicz, 1996). The audiovisual simultaneity window then appears to become narrower throughout childhood, reaching adult levels after adolescence (Hillock et al., 2011; Hillock-Dunn & Wallace, 2012; Lewkowicz & Flom, 2014). However, the evidence on the development of the shifting of the PSS toward the visual-leading side is equivocal. Lewkowicz (1996) demonstrated that the audiovisual simultaneity window is wider on the visual-leading side than on the auditory-leading side for a bouncing ball. However, this pattern may be restricted to bouncing balls because in adults the PSS shifts farther to the visual-leading side for such events than for simple flashes and beeps (van Eijk et al., 2008). In the studies using simultaneity judgments, Hillock et al. (2011) found that the visual- and auditory-leading sides were similar in width for children aged 10 and 11 years, whereas the visual-leading side was wider for adults. However, Hillock-Dunn and Wallace (2012) reported a reversed result; the audiovisual simultaneity window was wider on the visual-leading side for children aged 6 to 11 years and adolescents aged 12 to 17 years, whereas it was similar on both sides for adults. Such differences in outcome may be partly attributable to individual differences (see van Eijk et al., 2008, for a review).

In summary, two aspects of the developmental trajectory of the audiovisual simultaneity window are still unclear: the age when the width of the window narrows to adults' level and the age when the PSS shifts to the visual-leading side. The goal of the current study was to answer the above questions by measuring the typical developmental trajectory of the audiovisual simultaneity window using simultaneity judgments. Participants in five age groups were tested: children aged 5, 7, 9, and 11 years and a group of adults. We chose to begin with 5-year-olds because that is the youngest age at which most children can provide reliable behavioral responses in psychophysical experiments and because that was an age similar to the 4- to 6-year-olds tested recently by Lewkowicz and Flom (2014). Considering that younger children might become fatigued or be distracted more easily (see Hillock et al., 2011), we designed our experiment as follows. First, children aged 5 years were tested with a shorter version of the experiment than older participants. Second, participants in all age groups were monitored by an experimenter and given breaks as needed. Third, the outliers in each age group were excluded. We also used a recently developed model to estimate the parameters determining the probability distribution of each participant's simultaneity judgments (García-Pérez & Alcalá-Quintana, 2012a). The strength of this model is that it not only estimates the PSS and the width of the simultaneity window (i.e., the typical perceptual parameters) but also estimates sensory parameters such as sensory processing variability and arrival time difference of the two stimuli as well as decisional parameters such as response errors. Therefore, in this model, parameters at three processing levels (sensory, perceptual, and decisional) can be estimated separately, and developmental changes for each component can be assessed.

Experiment 1

Method

Participants

There were 20 participants remained in the final analysis in each of five age groups: children aged 5 years (10 boys; mean age = 5.0 years, range = 4.9–5.2), 7 years (12 boys; mean age = 7.0 years, range = 6.8–7.2), 9 years (13 boys; mean age = 9.0 years, range = 8.8–9.2), and 11 years (9 boys; mean age = 11.0 years, range = 10.8–11.3) and adults (7 men; mean age = 20 years, range = 18–31). An additional 5 participants were tested but not included in the final analysis because they did not complete the experiment (4 5-year-olds) or did not pass the visual screening test (1 adult). Children were recruited from a database of parents who volunteered to participate in research at the time of the children's births. Adults were undergraduate students at McMaster University participating in exchange for course credits. All of the participants had normal or corrected-to-normal visual acuity in each eye confirmed by visual screening tests (6/6 vision for 5-year-olds on the Cambridge Crowding Cards, 20/25 vision for 7-year-olds on the Lighthouse eye chart, and 20/20 vision for older participants on the Lighthouse eye chart). All had normal hearing by self-report. Written consent from the parents of children or adult participants, and written assent from the children aged 7 years or older, was obtained. All of the participants were naive regarding the purpose of the study. The study was approved by the McMaster University research ethics board.

Apparatus and stimuli

Participants were seated in a dimly lit room with their head on a chin rest located 50 cm from the monitor where the visual stimuli were presented. A gray ring with 2° inner diameters was displayed in the center of a black background on the monitor throughout the experiment. The visual stimulus was a 2° white disk presented in the middle of the gray ring for a duration of 16.7 ms (1 frame at the 60-Hz refresh rate). The auditory stimulus (white noise with a loudness of 57.5 dB SPL) was presented from speakers placed on either side of the monitor (the room had a background noise level of 40 dB SPL). The duration of the beep was 17 ms with 2 ms on and off ramping. Presentation of the stimuli was controlled by Matlab (MathWorks, Natick, MA, USA) and Psychtoolbox extensions (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997).

Design

Two factors were manipulated: age (5, 7, 9, or 11 years or adult) and SOA (−1200, −800, −400, −300, −200, −100, 0, 100, 200, 300, 400, 800, or 1200 ms), where negative values indicate that the auditory beep was presented first and positive values indicate that the visual flash was presented first. The SOA between the flash and beep was confirmed with an oscilloscope. Each SOA was tested twice in a block. Children aged 5 years completed 5 blocks, and older participants completed 10 blocks. Participants were allowed to take a short break between blocks and typically took 30 to 50 min to complete the experiment.

Procedure

During the experiment, participants were instructed to fixate the ring. Participants' task was to orally report "yes" if they considered that the flash and beep were presented at the same time or "no" if they considered that they were presented at different times. An experimenter sat beside participants and keyed the answers into the computer. The experimenter also ensured that participants fixated the ring and stopped for a break if they did not.

Two practice sessions were conducted prior to the main experiment. The first practice session consisted of eight trials: four with 0 ms SOA and four with −1200, −800, 800, or 1200 ms SOA. Participants needed to achieve 85% accuracy (i.e., no more than one error) in order to proceed. Five 5-year-old participants needed to repeat the first practice session, and none needed more than two repeats. The second practice session consisted of one trial for each of the 13 SOAs used in the main experiment and was designed to familiarize participants with the experimental procedure. There were no accuracy requirements for this second practice session. No feedback was given during either practice session except for general encouragement.

Results

Two criteria were set in order to exclude participants who did not understand or attend to the task or who performed as outliers in each age group. First, the simultaneous response at the ±1200 ms SOA was higher than the mean plus 2 standard deviations (SD) of that age group. Two adults' simultaneous response at the −1200 ms SOA (5%) exceeded this inclusion criterion for adults (3.3%); however, given that 5% represented only one trial, and an error could occur by making a wrong motor response (while the perception was accurate), these two participants were included for further analysis. Second, the simultaneous response at the 0 ms SOA was lower than 50% or the mean minus 2 SD of that age group. Following these two criteria, seven 5-year-olds, two 7-year-olds, two 9-year-olds, one 11-year-old, and three adults were replaced. Note that those participants who were excluded had a wider audiovisual simultaneity window than other participants in the same age groups; thus, including these participants' results would increase the estimated audiovisual simultaneity window of those age groups, especially for the 5-year-olds.

Proportion of simultaneous responses

The mean proportion of simultaneous responses at each SOA for each age group was submitted to a two-way analysis of variance (ANOVA) with a between-participant factor of age and a within-participant factor of SOA (see Fig. 1). Both main effects were significant: age, $F(4, 95) = 5.26$, $MSE = 0.11$, $p < .005$, and SOA, $F(12, 1140) = 646.21$, $MSE = 0.02$, $p < .001$. Critically, their interaction was also significant, $F(48, 1140) = 3.44$, $MSE = 0.02$, $p < .001$. Table 1 presents the percentage of simultaneous responses, the results of a one-way ANOVA on the factor of age at each SOA, and the post hoc tests at the SOAs where there was a significant simple main effect of age (Dunnett test, two-tailed). The age effect was significant at 10 of the 13 SOAs, with the exceptions being the −200, −100, and 200 ms SOAs. The age effect arose mainly because the 5- and 7-year-olds made more simultaneous responses than adults at the SOAs beyond ±200 ms SOA. On the other hand, for SOAs near 0, children made fewer simultaneous responses; this was true for children aged 5 and 7 years at the 0 ms SOA and for children aged 5 and 9 years at the 100 ms SOA.

Estimated parameters of simultaneity judgments

Individual data were fitted with the Matlab routine for the simultaneity judgments task (Alcalá-Quintana & García-Pérez, 2013). This routine is based on an independent-channels model for stimulus timing judgment tasks proposed by García-Pérez and Alcalá-Quintana (2012a). In this model, it is assumed that the signals in each sensory modality (i.e., vision and audition in the current study) are processed independently in each sensory pathway and then reach a central mechanism decoding their arrival time (i.e., perceived onset time of each signal). The difference of perceived onset time between the two signals, caused by the processing time and variance in each sensory pathway, would form a probability distribution rather than a constant. Participants' threshold of multisensory simultaneity is the criterion that determines whether a response of simultaneous or not is going to be made. Note that this model also assumes that participants may make motor errors when responding.

The peripheral processing time of each stimulus to reach a central mechanism is modeled by an exponential distribution determined by two parameters: the processing time (τ_i) and the processing variability (λ_i). In the current study, the processing time difference between two stimuli ($\tau = \tau A - \tau V$) and the processing variability of visual and auditory stimuli (λV and λA, respectively) were estimated. At a specific SOA, a bilateral exponential distribution of the difference of perceived onset time of the

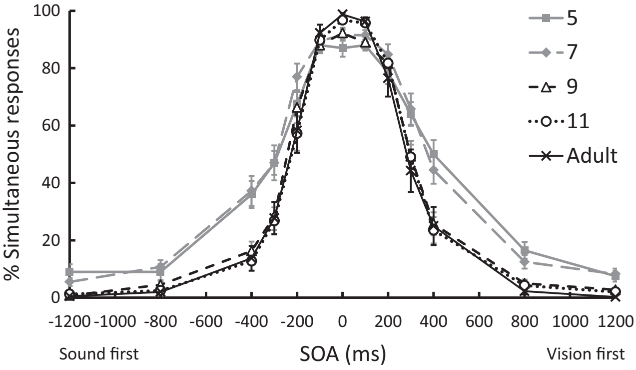

Fig. 1. Mean percentage of simultaneous responses at each SOA for the five age groups tested in Experiment 1. Error bars indicate ±1 standard error of the mean.

Table 1
Percentage of simultaneous responses in each age group and results of one-way ANOVA and post hoc tests for percentage of simultaneous responses at each SOA.

SOA (ms)	5	7	9	11	Adult	$F(4, 95)$	p	Post hoc tests ($*p < .05$; $**p < .01$)
−1200	9.0	5.5	0.8	1.3	0.5	6.30	<.001	5 > adult **; 7 > adult *
−800	9.0	10.8	4.5	2.5	2.0	4.60	<.005	5 > adult *; 7 > adult **
−400	36.0	37.3	16.3	12.8	13.8	8.38	<.001	5 > adult **; 7 > adult **
−300	47.0	47.0	28.5	26.8	27.8	4.29	<.005	5 > adult *; 7 > adult *
−200	67.5	77.0	66.3	57.3	58.3	1.93	.11	
−100	88.0	89.5	88.0	89.8	92.3	0.47	.76	
0	87.0	91.5	92.3	96.8	98.8	6.16	<.001	5 < adult **; 7 < adult *
100	88.0	91.5	89.0	95.8	96.3	4.69	<.005	5 < adult **; 9 < adult *
200	78.5	84.8	81.5	81.8	76.5	0.56	.69	
300	64.0	65.8	49.3	49.0	44.3	3.15	<.05	5 > adult *; 7 > adult *
400	50.0	44.5	25.8	23.5	25.0	6.08	<.001	5 > adult **; 7 > adult *
800	16.5	12.5	5.0	4.3	2.3	9.78	<.001	5 > adult **; 7 > adult **
1200	7.5	8.3	2.8	2.0	0.3	7.50	<.001	5 > adult **; 7 > adult **

two stimuli at the central mechanism was determined by λV, λA, and τ. A resolution parameter (δ), representing the threshold of simultaneity perception, was estimated such that a "simultaneous" response would be made when the difference of perceived onset time was smaller than the resolution parameter. Hence, the bigger the value of this parameter, the wider the audiovisual simultaneity window. Finally, the response errors that a participant misreported "simultaneous" in the auditory-leading trials (εAF) and in the visual-leading trials (εVF), as well as the "not simultaneous" in the 0 ms trials (εS), were estimated. Note that these response errors include both participants' *lapse* due to blink or inattention and their mistakes of motor responding given that these two types of errors are hard to distinguish (see García-Pérez & Alcalá-Quintana, 2012b). The starting values used to fit the data in the current study were as follows: LamBounds = [1/500 1/1]; TauBounds = [−Inf Inf]; DeltaBounds = [0 Inf]; LamTStart = [1/70 1/10]; LamRStart = [1/70 1/10]; TauStart = [−70 70]; DeltaStart = [20 150]; ErrStart = [0.05]; BiaStart = [0.5]; Model = 1; SampleSize = 1500; ConfCoef = 95; FixedSeed = true.

Each of the estimated parameters was submitted to a one-way ANOVA on the factor of age (see Table 2). The parameter δ corresponds to half of the width of the simultaneity window when the cri-

Table 2
Means (and standard errors) of estimated parameters of the simultaneity judgments task and results of one-way ANOVA as well as post hoc tests in Experiment 1.

Parameter	5	7	9	11	Adult	$F(4, 95)$	p	Post hoc tests ($*p < .05$; $**p < .01$)
δ	366.3 (24.1)	346.5 (22.0)	285.0 (11.9)	275.4 (13.2)	277.6 (21.3)	5.03	<.005	5 > adult **; 7 > adult *
PSS	58.4 (16.0)	37.2 (8.3)	24.4 (8.4)	39.1 (9.3)	36.8 (13.6)	1.13	.35	
λA	0.18 (0.07)	0.08 (0.04)	0.14 (0.06)	0.08 (0.05)	0.18 (0.08)	0.64	.64	
λV	0.25 (0.06)	0.16 (0.06)	0.11 (0.06)	0.10 (0.05)	0.11 (0.03)	1.42	.23	
τ	−99.8 (41.8)	−72.6 (28.1)	1.9 (20.2)	−49.3 (17.0)	−45.3 (22.8)	1.88	.12	
εAF	0.08 (0.02)	0.07 (0.01)	0.03 (0.01)	0.02 (0.01)	0.01 (0.01)	6.41	<.001	5 > adult **; 7 > adult **
εS	0.07 (0.02)	0.04 (0.01)	0.03 (0.01)	0.01 (0.004)	0.01 (0.002)	3.81	<.01	5 > adult **
εVF	0.10 (0.02)	0.10 (0.02)	0.04 (0.01)	0.03 (0.01)	0.01 (0.01)	8.66	<.001	5 > adult **; 7 > adult **

Note. δ, resolution (threshold of simultaneity perception); PSS, point of subjective simultaneity; λA, processing variability of auditory stimulus; λV, processing variability of visual stimulus; τ, processing time difference between visual and auditory stimuli; εAF, response errors in the auditory-leading trials; εS, response errors in the simultaneous trials; εVF, response errors in the visual-leading trials.

terion was set at 50% simultaneous response on either the auditory-leading side or the visual-leading side. There was a significant age effect for δ. Post hoc tests (Dunnett, one-tailed, because it is assumed that children have a wider window than adults; see Hillock et al., 2011) revealed that δ was larger for 5-year-olds ($p < .005$) and 7-year-olds ($p < .05$) than for adults, but not for 9- or 11-year-olds (both $ps \geqslant .69$). This pattern suggests that the width of the audiovisual simultaneity window is adult-like by 9 years of age (see Fig. 2).

The PSS is the midpoint of the audiovisual simultaneity window determined by δ. The PSS of all five age groups was positive—on the visual-leading side, 5 years: $t(19) = 3.66$, $p < .005$; 7 years: $t(19) = 4.46$, $p < .001$; 9 years: $t(19) = 2.91$, $p < .01$; 11 years: $t(19) = 4.23$, $p < .001$; adult: $t(19) = 2.70$, $p < .05$—and there was no age effect. A positive PSS (visual-leading) in all five age groups suggests that an asymmetrical audiovisual simultaneity window is established before 5 years of age.

There was no age effect for the three parameters measuring early sensory processing: auditory processing variability (λA), visual processing variability (λV), and processing time difference (τ). These results suggest that by 5 years of age, sensory processing of simple visual and auditory stimuli in order to judge their simultaneity is already adult-like. The τ was negative (suggesting that the auditory stimulus arrived first at the central processing mechanism when visual and auditory stimuli were onset simultaneously) in all groups except 9-year-olds, with a significant difference from 0 in 5-year-olds, $t(19) = -2.39$, $p < .05$, 7-year-olds, $t(19) = -2.58$, $p < .05$, and 11-year-olds, $t(19) = -2.91$, $p < .01$, and a similar trend in adults, $t(19) = -1.98$, $p = .06$.

In contrast, there was a significant age effect for the three parameters related to response errors. When the auditory stimulus was leading (εAF), post hoc tests (Dunnett, one-tailed) revealed that the response error was larger for 5- and 7-year-olds than for adults ($ps < .005$). In the simultaneous condition (εS), post hoc tests revealed that the response error was larger for 5-year-olds than for adults ($p < .005$). When the visual stimulus was leading (εVF), post hoc tests revealed that the response error was larger for 5- and 7-year-olds than for adults ($ps < .001$). These results suggest that children aged 5 and 7 years were more likely to make response errors than adults.

According to Alcalá-Quintana and García-Pérez (2013, p. 994), 15 to 20 trials at each of 10 to 12 SOAs are sufficient to estimate each parameter accurately. Participants aged 7 years and older completed 20 trials at each of 13 SOAs and, hence, followed the suggested regimen. However, this design was too long for 5-year-olds, and so they were tested with only half that number of trials (10 trails at each of 13 SOAs). Accordingly, two contrasting hypotheses can be proposed regarding the possible influence of this difference. Young children may need more practice in order to perform more accurately, resulting in a *wider* window of audiovisual simultaneity in the first half of the blocks. Alternatively, young children may become inattentive and/or fatigue easily, resulting in a *narrower* window of audiovisual simultaneity in the first half of the blocks.

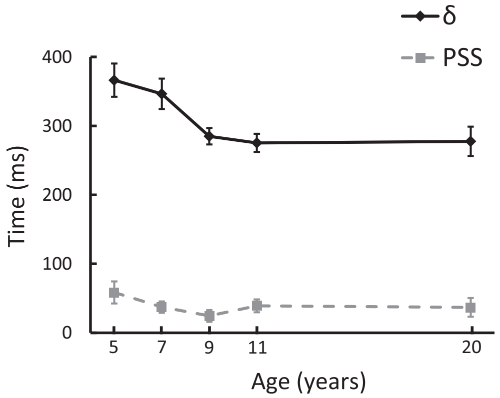

Fig. 2. Mean threshold of simultaneity perception (δ) and point of subjective simultaneity (PSS) for the five age groups in Experiment 1. Error bars indicate ±1 standard error of the mean.

When the analyses are restricted to the participants' performance using only the first half of the blocks across the five age groups (see Appendix A for details), the results still demonstrate the same pattern as reported above based on all of the data. Critically, the analysis of the mean proportion of simultaneous responses (see Fig. S1 in online supplementary material) demonstrated a significant interaction between age and SOA, $F(48, 1140) = 3.06$, $MSE = 0.02$, $p < .001$. The simple main effect of age was significant at 10 of the 13 SOAs ($Fs \geqslant 3.20$, $ps < .05$), that is, at all except the −200, −100, and 200 ms SOAs ($Fs \leqslant 2.14$, $ps \geqslant .08$). The estimated parameters demonstrated that children aged 5 and 7 years, compared with adults, had a wider window of audiovisual simultaneity (indexed by δ, $ps < .01$) and higher response errors ($ps < .05$).

In Appendix B, we compare 7-year-olds' performance in the first half versus the second half of the blocks in order to evaluate the potential consequences of testing older participants with twice as many trials as those given to 5-year-olds. Those analyses show that 7-year-olds were more likely to make "simultaneous" responses in the second half than in the first half of the blocks ($p < .005$), a pattern unlikely to arise from reduced sensitivity ($p = .82$) but instead caused mainly by their higher response errors, especially on the visual-leading side ($p < .05$). Hence, using more trials with 5-year-olds is unlikely to yield a smaller window of audiovisual simultaneity; rather it would likely widen it, as it did in 7-year-olds.

Discussion

The audiovisual simultaneity window was measured using a simultaneity judgments task in five age groups. The window was wider in younger children; children aged 5 and 7 years were more likely than adults to make simultaneous responses at SOAs larger than 200 ms in both the visual- and auditory-leading conditions. At the same time, children aged 5, 7, and 9 years were less likely to make simultaneous responses at the SOAs of 0 and/or 100 ms (visual-leading condition). The estimated parameters showed no age difference in the PSS and three sensory parameters (auditory processing variability, visual processing variability, and processing time difference). However, children aged 5 and 7 years demonstrated a wider audiovisual simultaneity window (indexed by δ) and higher response errors at both the visual- and auditory-leading sides than adults. Children aged 5 years made more response errors than adults in the simultaneous condition.

The PSS results demonstrate that it has shifted to the visual-leading side by 5 years of age (cf. Hillock et al., 2011). However, it is not until 9 years of age that the width of the audiovisual simultaneity window has narrowed to the adult size. Children in our experiment reached adult-like performance at an earlier age than reported in previous studies (Hillock et al., 2011; Hillock-Dunn & Wallace, 2012). However, one critical difference is that the SOA range used in the current study (i.e., ±1200 ms) was wider than the SOA ranges used in previous studies (±450 ms in Hillock et al., 2011; ±500 ms in Hillock-Dunn & Wallace, 2012). If adult participants tended to balance their probability of "yes" and "no" responses, then the estimated temporal window of audiovisual simultaneity would likely be larger when the SOA range was wider (see a similar concern in Zampini et al., 2005) and, hence, appears more similar to that of children. To verify that the estimate of adults' window of audiovisual simultaneity is based on their perception rather than decisional biases, in Experiment 2 we recruited a new group of adults who were tested using a narrower SOA range: ±500 ms. If adult participants responded based on their perception of audiovisual simultaneity, then there should be no significant difference between adults presented with SOA ranges of ±1200 (Experiment 1) and those presented with SOA ranges of ±500 ms (Experiment 2).

Experiment 2

Method

A new group of 20 undergraduate students (6 men; mean age = 19 years, range = 18–22) at McMaster University were tested using 15 SOAs (−500, −400, −300, −200, −150, −100, −50, 0, 50, 100, 150, 200, 300, 400, and 500 ms). An additional 3 participants were tested but excluded because their proportion of simultaneous responses was higher than the mean plus 2 SD of the group at the −500 or 500 ms SOA ($n = 2$) or lower than the mean minus 2 SD of the group at the 0 ms SOA ($n = 1$). Participants were asked to press the number key "1" when they perceived that the flash and beep were presented at the same time and to press "2" when they perceived that the flash and beep were presented at different times. Other details were the same as in Experiment 1.

Results and discussion

The data of the adult group from Experiment 1 (wide range) and the data collected in Experiment 2 (narrow range) (see Fig. 3) were compared using the nine SOAs that were tested in both groups (−400, −300, −200, −100, 0, 100, 200, 300, and 400 ms). The proportion of simultaneous responses was submitted to a two-way ANOVA with the between-participant factor of range (wide or narrow) and the within-participant factor of SOA. There was a main effect of SOA, $F(8, 304) = 150.44$, $MSE = 0.03$, $p < .001$. There was no significant effect of range, $F(1, 38) = 1.24$, $p = .27$, and the interaction between range and SOA was also nonsignificant, $F(8, 304) = 0.28$, $p = .97$. There were also no significant differences in the estimated parameters between the groups tested with the wide and narrow ranges (see Table 3). Hence, adults' performance in the simultaneity judgments task was similar when the range of SOAs was either ±1200 or ±500 ms, suggesting that adults performed this task mainly based on their perception of audiovisual simultaneity (see Zampini et al., 2005, for a similar conclusion).

General discussion

In the current study, we measured the typical developmental trajectory of the window of audiovisual simultaneity using simultaneity judgments task with children aged 5, 7, 9, and 11 years and a group of adults. Children aged 5 and 7 years judged the beep and flash to be simultaneous over a wider range of SOAs than adults both when the flash came first and when the beep came first, whereas children aged 9 and 11 years performed similarly to adults. Model fitting revealed that young children's wider distribution can be attributed to both a perceptually wider window of simultaneity and a higher rate of response error. Nevertheless, the PSS was shifted to the visual-leading side in all groups, a result suggesting that the audiovisual simultaneity window has an adult-like asymmetrical shape by 5 years of age. Combined, the results indicate that children develop an adult-like width of the audiovisual simultaneity window between 7 and 9 years of age. Nevertheless, the PSS is already on the visual-leading side by 5 years of age.

The previous developmental studies contrast with the current study by suggesting that the audiovisual simultaneity window is not adult-like at 10 or 11 years of age (Hillock et al., 2011) or even at 12 to 17 years of age (Hillock-Dunn & Wallace, 2012). However, those studies failed to determine the age at which the audiovisual simultaneity window matures in part because they either included a single

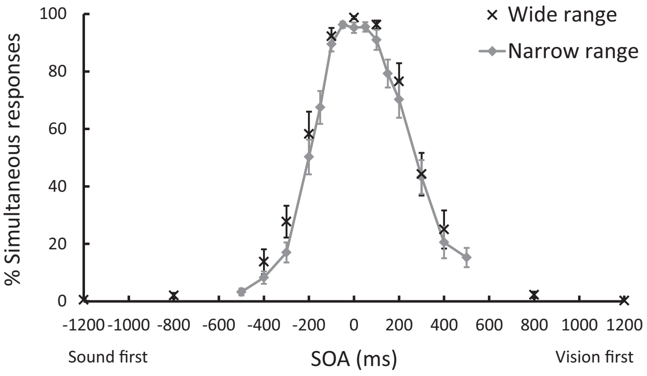

Fig. 3. Mean percentage of simultaneous responses at each SOA for two groups of adults. Black Xs represent the results for adults tested in Experiment 1 with a wide range of SOAs, and gray diamonds represent the results for adults tested in Experiment 2 with a narrow range of SOAs. Error bars indicate ±1 standard error of the mean.

Table 3
Means (and standard errors) and results of *t*-tests of estimated parameters of the simultaneity judgments task

Parameter	Wide	Narrow	*t*(38)	p
δ	237.3 (10.7)	219.2 (9.5)	1.26	.21
PSS	18.7 (8.3)	22.5 (9.1)	0.31	.76
λA	0.25 (0.08)	0.28 (0.08)	0.23	.82
λV	0.19 (0.06)	0.20 (0.06)	0.07	.94
τ	−21.2 (13.8)	−9.9 (18.2)	0.50	.62
εAF	0.12 (0.04)	0.09 (0.02)	0.70	.49
εS	0.004 (0.002)	0.02 (0.01)	1.48	.15
εVF	0.24 (0.06)	0.20 (0.06)	0.47	.64

age group (Hillock et al., 2011), averaged results over large age ranges (Hillock-Dunn & Wallace, 2012), or included a small number of participants at each age (Hillock et al., 2011; Hillock-Dunn & Wallace, 2012). Unlike previous developmental studies, we tested four age groups of children at 2-year increments starting from 5 years and reduced the variance for each age group by testing 20 participants per group. In addition, we used a modified experimental procedure suitable for children, excluded outliers in each age group, and adopted a recent model specifically designed to estimate sensory, perceptual, and decisional components that modulate the audiovisual simultaneity window (Alcalá-Quintana & García-Pérez, 2013; García-Pérez & Alcalá-Quintana, 2012a, 2012b). Perhaps for these reasons, we found that children's audiovisual simultaneity window narrowed at both the visual- and auditory-leading sides, reaching adult levels earlier than estimated previously by Hillock and colleagues (Hillock et al., 2011; Hillock-Dunn & Wallace, 2012).

PSS shifts to the visual-leading side: Developed by 5 years of age

We demonstrated that the PSS is already on the visual-leading side by 5 years of age; as in adults, the audiovisual simultaneity window is asymmetrical in that the visual-leading side is wider than the auditory-leading side. Two explanations have been proposed for why the PSS is located at the visual-leading side. The first one attributes it to the longer peripheral processing time for vision than for audition and the consequence that the presentation of the visual stimulus should lead the auditory stimulus (i.e., the positive SOAs in the current study) in order for the visual and auditory signals to arrive at the central mechanism at a similar time. The second explanation is that it is the result of adaptation to our natural environment; given that light travels faster than sound, visual and auditory signals are more likely to originate from the same source in the situation when the visual signal leads the auditory signal rather than the reversed situation (i.e., auditory signal leads visual signal). These two accounts are difficult to disassociate empirically, and they are not mutually exclusive. Nevertheless, the three sensory parameters estimated from the model for simultaneity judgments provide some insights regarding the early development of the PSS.

Specifically, we found that the parameters related to the temporal processing of visual and auditory stimuli—processing variability (λV and λA) and the processing time difference (τ)—did not differ significantly across the age range of 5 years to adulthood. This finding is in agreement with evidence that visual temporal processing, as measured by the critical frequency fusion threshold (above which separate flashes fuse into a single percept), and auditory temporal processing, as measured by the auditory gap threshold (above which two noise bursts merge perceptually), are both adult-like by 4 years of age (Ellemberg, Lewis, Liu, & Maurer, 1999; Wightman, Allen, Dolan, Kistler, & Jamieson, 1989). Together, these findings suggest that the temporal resolutions of unimodal visual and auditory processing have reached adult levels by 5 years of age, the youngest age tested in the current study. In summary, by 5 years of age children have developed peripheral processing for visual and auditory signals that is adult-like. This early maturity may partly explain why the window of audiovisual simultaneity is already asymmetrical by 5 years of age.

Width of audiovisual simultaneity window: Developed by 9 years of age

We found that the width of the audiovisual simultaneity window (i.e., the parameter δ) narrows during development and reaches adult levels by 9 years of age. Yet, the areas of the brain relevant to crossmodal temporal processing are still structurally and functionally immature at this age. For example, the neural development of white matter (e.g., neural myelination) and cortical areas associated with audiovisual integration (e.g., posterior superior temporal sulcus) continues until adolescence (Giedd et al., 1999; Gogtay et al., 2004; Paus et al., 1999; for a review, see Giedd & Rapoport, 2010). Similarly, the physical development of the size of the eyeballs (for vision) and head (for audition) are not adult-like until adolescence (Fledelius, Christensen, & Fledelius, 2014; Nellhaus, 1968), and both are critical for coordinating visual and auditory information during development (Ernst, 2008). Thus, structural and functional maturation, although partly responsible, cannot fully explain our finding that the width of the window of audiovisual simultaneity is mature by 9 years of age.

Another possible factor that modulates the width of the audiovisual simultaneity window is participants' attention. Hillock et al. (2011) demonstrated that children aged 10 or 11 years had poorer attention or fatigued more quickly than adults, leading to a wider audiovisual simultaneity window later in the testing sequence. That was not the case in the current study, where 9-year-olds performed as well as adults, perhaps because the experimenter monitored performance closely and gave breaks as necessary. However, reduced attention likely contributed to our finding that the window of audiovisual integration was wider in 5- and 7-year-olds than in adults. In fact, our measurement of errors caused by inattention, as estimated by the response error parameters in García-Pérez and Alcalá-Quintana (2012a, 2012b) model, confirmed this hypothesis; response errors decreased with age. This result is consistent with an improvement in attentional control during development. Such attentional lapses should affect performance not only in experimental settings like ours but also when children make judgments about audiovisual events in everyday life.

We consider the maturity of the audiovisual simultaneity window by 9 years of age to be consistent with the results of other developmental studies of unimodal and multimodal perception. First, maturity by 9 years of age is consistent with Gori, Sandini, and Burr's (2012) study, which used a temporal bisection task with five age groups (5–7, 8–9, 10–11, and 13–14 years and adults) to examine the development of audiovisual integration in the temporal domain. The results demonstrated that adults integrate visual and auditory information in an auditory-dominant fashion in the temporal task (rather than following the optimal integration based on the Bayesian model), and children perform similarly to adults by 8 or 9 years of age.

Second, the age of maturity is later for the audiovisual simultaneity window (by 9 years) than the visual critical frequency fusion threshold or the auditory gap threshold (by 4 years) (Ellemberg et al., 1999; Wightman et al., 1989). This finding is consistent with the notion that multimodal integration matures later and/or has higher thresholds than unimodal integration. For example, in adults, the threshold is higher for temporally separating two signals across sensory modalities (e.g., a flash and a beep) than for comparing two signals in the same modality (e.g., two flashes or two beeps) (see Fujisaki & Nishida, 2010). Developmentally, the same pattern emerges; temporal order judgments for the combination of visual and tactile stimuli become adult-like later than temporal order judgments for two tactile stimuli (Röder, Pagel, & Heed, 2013).

Linking development of audiovisual simultaneity window and other cognitive functions

The audiovisual simultaneity window, as suggested by the temporal rule, is related strongly to multisensory integration. In our daily life, for example, we find it hard to watch a movie when its video and audio are out of synch. Stevenson et al. (2012) demonstrated that the width of the audiovisual simultaneity window is positively correlated with the magnitude of the sound-induced flash illusion (Shams, Kamitani, & Shimojo, 2000) but negatively correlated with the magnitude of the McGurk effect (McGurk & MacDonald, 1976). Similarly, children have a wider audiovisual simultaneity window, a larger magnitude of sound-induced flash illusion (Innes-Brown et al., 2011; Nava & Pavani, 2013; but see Tremblay et al., 2007), and a smaller McGurk effect (Tremblay et al., 2007). Currently, it is unclear whether the correlations among these three phenomena share a common core mechanism or whether audiovisual simultaneity is causally related to the other two phenomena. This issue will provide critical insight into crossmodal cognitive functions not only in typically developing children but also in atypically developing groups such as those with autism spectrum disorders because their deficit in speech perception may be attributed to a generally lower precision than age-matched controls in audiovisual simultaneity perception (de Boer-Schellekens, Eussen, & Vroomen, 2013; de Boer-

Schellekens, Keetels, Eussen, & Vroomen, 2013; but see Stevenson et al., 2014; see Wallace & Stevenson, 2014, for a review).

Some have suggested that a wide audiovisual simultaneity window is essential for infants and young children to learn possible associations between visual and auditory stimuli (e.g., Röder et al., 2013). However, a wider audiovisual simultaneity window may lead to incorrect associations of crossmodal information that should be separate. Moreover, that account cannot explain the narrowing of the window with age. We suggest that the wide window in early life should be attributed to an immaturity in brain structure and function. Certainly, audiovisual synchrony is a critical factor for infants to extract the features and relations embedded in the co-occurring visual and auditory stimuli (Gogate, Walker-Andrews, & Bahrick, 2001; Lewkowicz et al., 2010). Furthermore, synchronous crossmodal events can draw infants' attention, which in turn enhances their crossmodal association learning (Bahrick & Lickliter, 2000). Therefore, perceiving audiovisual simultaneity plays an important role in crossmodal association learning (e.g., Mareschal, Westermann, & Althaus, 2012). After primary crossmodal associations are established, such knowledge (called the *unity assumption;* see Spence, 2007; Welch & Warren, 1980) would help to solve the problem regarding which visual and auditory stimuli should be integrated and which should not be and gradually tune the audiovisual simultaneity window into an optimal width and shape (i.e., asymmetrical). This notion implies that the processing of audiovisual simultaneity and the content of the visual and auditory stimuli are processed in two independent but highly interactive pathways (e.g., Chen & Spence, 2011; Fujisaki & Nishida, 2010; Soto-Faraco & Alsius, 2009).

An intriguing direction for future study is how visual and auditory systems are coordinated in order to develop an optimal window for audiovisual simultaneity. Recently, Burr and Gori (2012) proposed that the sensory modality which provides more precise (but not necessarily more accurate) information calibrates another sensory modality in order to coordinate two senses during development. For example, in orientation perception, vision calibrates touch during development, which is supported by evidence from early-blind children who demonstrated impaired performance in a haptic orientation discrimination task compared with sighted age-matched controls (Gori, Del Viva, Sandini, & Burr, 2008; Gori, Sandini, Martinoli, & Burr, 2010). For audiovisual interactions in the temporal domain, presumably it should be that audition calibrates vision because audition has better temporal precision (Burr, Banks, & Morrone, 2009; Gori et al., 2012; Morein-Zamir, Soto-Faraco, & Kingstone, 2003). This hypothesis can be verified by measuring the audiovisual simultaneity window of people who had early deafness and later recovered with cochlear implants. Alternatively, if the temporal calibration is bidirectional, or from vision to audition (e.g., Navarra, Hartcher-O'Brien, Piazza, & Spence, 2009), people who recovered vision after treatment for dense congenital cataracts should have an abnormal audiovisual simultaneity window.

Conclusion

We have demonstrated that the window for audiovisual simultaneity becomes asymmetrical by 5 years of age and narrows on both visual- and auditory-leading sides to the adult width by 9 years of age. In addition, children's peripheral processing of visual and auditory signals is adult-like by 5 years of age, whereas their accuracy of responding keeps improving until 9 years of age. The maturity of the audiovisual simultaneity window by 9 years of age is the youngest age reported in the literature to date and is earlier than the physical and functional maturity of areas of the brain involved in multisensory integration. By implication, by 9 years of age the system is able to recalibrate quickly when the visual or auditory system changes. Future research should address how visual and auditory systems coordinate and lead to this early maturity during childhood and how the development of the audiovisual simultaneity window conjoins with the development of other audiovisual cognitive functions.

Acknowledgments

This study was funded by the James S. McDonnell Foundation and the Natural Sciences and Engineering Research Council of Canada (NSERC) to D.M. and by a Discovery grant from NSERC to D.I.S. Y-C.

C. is currently supported by an Arts and Humanities Research Council (AHRC) Rethinking the Senses grant (AH/L007053/1). We thank Sally Stafford, Natalia Fong, Natasha Sandhu, and Tharshni Velauthapil for their great help in data collection.

Appendix A: Analysis of 5 blocks for all age groups

We analyzed participants' performance in the first half of the blocks (five blocks) using data from all five ages: the data from 5-year-olds and the data from the comparable first half of the blocks for the age groups of 7-, 9-, and 11-year-olds and adults (see Fig. S1 in supplementary material). The proportion of simultaneous responses was submitted to a two-way ANOVA with the factors of age and SOA. Both main effects were significant: age, $F(4, 95) = 5.53$, $MSE = 0.11$, $p < .001$, and SOA, $F(12, 1140) = 553.50$, $MSE = 0.02$, $p < .001$, and so was their interaction, $F(48, 1140) = 3.06$, $MSE = 0.02$, $p < .001$. The results of one-way ANOVAs with age as a factor were significant at 10 of the 13 SOAs (see Table S1 in supplementary material). The post hoc tests (Dunnett test, two-tailed) at these SOAs showed that only children aged 5 and 7 years were significantly different from adults. The estimated parameters (Table S2) demonstrate that children aged 5 and 7 years had a wider window (δ) as well as higher response errors. The results show the same pattern as reported in the text when all of the data were used.

Appendix B: Compare first and second half of blocks in the 7-year-olds

Five-year-olds were given five blocks of trials, whereas older groups were given twice as many blocks. To understand whether younger children would perform more accurately due to practice, or worse due to inattention and/or fatigue in the later five blocks, we compared performance in the first- and second-half blocks (five blocks each) of 7-year-old participants (see Fig. S2 in supplementary material).

The mean proportion of simultaneous responses at each SOA was submitted to a two-way ANOVA on the factors of order (first half or second half) and SOA. The main effect of order was significant; overall, participants were more likely to make simultaneous responses in the second half than in the first half of blocks (53.3% vs. 49.1%), $F(1, 19) = 12.71$, $MSE = 0.02$, $p < .005$. There was also a significant main effect of SOA, $F(12, 228) = 142.16$, $MSE = 0.03$, $p < .001$. However, their interaction was not significant, $F(12, 228) = 1.25$, $p = .25$.

Individual data of the first- and second-half blocks were fitted with the Matlab routine (see Table S3 in supplementary material). The only parameter that reached statistical significance was the number of response errors in the visual-leading side (εVF) being higher in the second half.

Appendix C. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jecp.2016.01.010.

References

Alcalá-Quintana, R., & García-Pérez, M. A. (2013). Fitting model-based psychometric functions to simultaneity and temporal-order judgment data: MATLAB and R routines. *Behavior Research Methods, 45*, 972–998.

Andersen, T. S., & Mamassian, P. (2008). Audiovisual integration of stimulus transients. *Vision Research, 48*, 2537–2544.

Arrighi, R., Alais, D., & Burr, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *Journal of Vision, 6*, 260–268.

Baart, M., Bortfeld, H., & Vroomen, J. (2015). Phonetic matching of auditory and visual speech develops during childhood: Evidence from sine-wave speech. *Journal of Experimental Child Psychology, 129*, 157–164.

Baart, M., Stekelenburg, J. J., & Vroomen, J. (2014). Electrophysiological evidence for speech-specific audiovisual integration. *Neuropsychologia, 53*, 115–121.

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology, 36*, 190–201.

Bolognini, N., Frassinetti, F., Serino, A., & Ladavas, E. (2005). "Acoustical vision" of below threshold stimuli: Interaction among spatially converging audiovisual inputs. *Experimental Brain Research, 160*, 273–282.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436.

Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research, 198*, 49–57.

Burr, D., & Gori, M. (2012). Multisensory integration develops late in humans. In M. M. Murray & M. T. Wallace (Eds.), *The neural bases of multisensory processes* (pp. 345–362). Boca Raton, FL: CRC Press.

Chen, Y.-C., Huang, P.-C., Spence, C., & Yeh, S.-L. (2011). The presentation of a synchronous sound enhances visual sensitivity without reducing target uncertainty. *Seeing and Perceiving, 24*, 623–638.

Chen, Y.-C., & Spence, C. (2011). The crossmodal facilitation of visual object representations by sound: Evidence from the backward masking paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 1784–1802.

Chen, Y.-C., & Westermann, G. (2012). Twelve-month-old infants learn crossmodal associations between visual objects and natural sounds in ecologically valid situations. *Seeing and Perceiving, 25*(Suppl.), 117.

Chen, Y.-C., & Yeh, S.-L. (2008). Visual events modulated by sound in repetition blindness. *Psychonomic Bulletin & Review, 15*, 404–408.

Chen, Y.-C., & Yeh, S.-L. (2009). Catch the moment: Multisensory enhancement of rapid visual events by sound. *Experimental Brain Research, 198*, 209–219.

de Boer-Schellekens, L., Eussen, M., & Vroomen, J. (2013). Diminished sensitivity of audiovisual temporal order in autism spectrum disorder. *Frontiers in Integrative Neuroscience, 7*. http://dx.doi.org/10.3389/fnint.2013.00008.

de Boer-Schellekens, L., Keetels, M., Eussen, M., & Vroomen, J. (2013). No evidence for impaired multisensory integration of low-level audiovisual stimuli in adolescents and young adults with autism spectrum disorders. *Neuropsychologia, 51*, 3004–3013.

Ellemberg, D., Lewis, T. L., Liu, C. H., & Maurer, D. (1999). Development of spatial and temporal vision during childhood. *Vision Research, 39*, 2325–2333.

Ernst, M. O. (2008). Multisensory integration: A late bloomer. *Current Biology, 18*, R519–R521.

Eskelund, K., Tuomainen, J., & Andersen, T. S. (2011). Multistage audiovisual integration of speech: Dissociating identification and detection. *Experimental Brain Research, 208*, 447–457.

Fledelius, H. C., Christensen, A. S., & Fledelius, C. (2014). Juvenile eye growth, when completed? An evaluation based on IOL–Master axial length data, cross-sectional and longitudinal. *Acta Ophthalmologica, 92*, 259–264.

Flom, R., & Bahrick, L. E. (2010). The effects of intersensory redundancy on attention and memory: Infants' long-term memory for orientation in audiovisual events. *Developmental Psychology, 46*, 428–436.

Frassinetti, F., Bolognini, N., & Lavadas, E. (2002). Enhancement of visual perception by crossmodal visuo–auditory interaction. *Experimental Brain Research, 147*, 332–343.

Fujisaki, W., & Nishida, S. (2010). A common perceptual temporal limit of binding synchronous inputs across different sensory attributes and modalities. *Proceedings of the Royal Society B, 277*, 2281–2290.

García-Pérez, M. A., & Alcalá-Quintana, R. (2012a). On the discrepant results in synchrony judgment and temporal-order judgment tasks: A quantitative model. *Psychonomic Bulletin & Review, 19*, 820–846.

García-Pérez, M. A., & Alcalá-Quintana, R. (2012b). Response errors explain the failure of independent-channels models of perception of temporal order. *Frontiers in Psychology, 3*. http://dx.doi.org/10.3389/fpsyg.2012.00094.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., ... Rapoport, J. L. (1999). Brain development during childhood and adolescence: A longitudinal MRI study. *Nature Neuroscience, 2*, 861–863.

Giedd, J. N., & Rapoport, J. L. (2010). Structural MRI of pediatric brain development: What have we learned and where are we going? *Neuron, 67*, 728–734.

Gogate, L. J., Walker-Andrews, A. S., & Bahrick, L. E. (2001). The intersensory origins of word comprehension: An ecological–dynamic system view. *Developmental Science, 4*, 1–37.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., ... Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 8174–8179.

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Current Biology, 18*, 694–698.

Gori, M., Sandini, G., & Burr, D. (2012). Development of visuo–auditory integration in space and time. *Frontiers in Integrative Neuroscience, 6*. http://dx.doi.org/10.3389/fnint.2012.00077.

Gori, M., Sandini, G., Martinoli, C., & Burr, D. (2010). Poor haptic orientation discrimination in nonsighted children may reflect disruption of cross-sensory calibration. *Current Biology, 20*, 223–225.

Hillock, A. R., Powers, A. R., & Wallace, M. T. (2011). Binding of sights and sounds: Age-related changes in multisensory temporal processing. *Neuropsychologia, 49*, 461–467.

Hillock-Dunn, A., & Wallace, M. T. (2012). Developmental changes in the multisensory temporal binding window persist into adolescence. *Developmental Science, 15*, 688–696.

Innes-Brown, H., Barutchu, A., Shivdasani, M. N., Crewther, D. P., Grayden, D. B., & Paolini, A. G. (2011). Susceptibility to the flash–beep illusion is increased in children compared to adults. *Developmental Science, 14*, 1089–1099.

King, A. J. (2005). Multisensory integration: Strategies for synchronization. *Current Biology, 15*, R339–R341.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception, 36*. ECVP Abstracts, No. 14.

Lewkowicz, D. J. (1996). Perception of auditory–visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1094–1106.

Lewkowicz, D. J. (2010). Infant perception of audio–visual speech synchrony. *Developmental Psychology, 46*, 66–77.

Lewkowicz, D. J., & Flom, R. (2014). The audiovisual temporal binding window narrows in early childhood. *Child Development, 85*, 685–694.

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences of the United States of America, 109*, 1431–1436.

Lewkowicz, D. J., Leo, I., & Simion, F. (2010). Intersensory perception at birth: Newborns match nonhuman primate faces and voices. *Infancy, 15*, 46–60.

Lippert, M., Logothetis, N. K., & Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. *Brain Research, 1173*, 102–109.

Lu, Z. L., Tse, H. C. H., Dosher, B. A., Lesmes, L. A., Posner, C., & Chu, W. (2009). Intra- and cross-modal cuing of spatial attention: Time courses and mechanisms. *Vision Research, 49*, 1081–1096.

Mareschal, D., Westermann, G., & Althaus, N. (2012). In search of the mechanisms of multisensory development. In A. J. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 342–359). Oxford, UK: Oxford University Press.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746–748.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research, 17*, 154–163.

Morrongiello, B. A., Fenwick, K. D., & Chance, G. (1998). Crossmodal learning in newborn infants: Inferences about properties of auditory–visual events. *Infant Behavior and Development, 21*, 543–553.

Murray, M. M., Foxe, J. J., & Wylie, G. R. (2005). The brain uses single-trial multisensory memories to discriminate without awareness. *NeuroImage, 27*, 473–478.

Murray, M. M., Michel, C. M., de Peralta, R. G., Ortigue, S., Brunet, D., Andino, S. G., et al (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *NeuroImage, 21*, 125–135.

Nava, E., & Pavani, F. (2013). Changes in sensory dominance during childhood: Converging evidence from the Colavita effect and the sound-induced flash illusion. *Child Development, 84*, 604–616.

Navarra, J., Hartcher-O'Brien, J., Piazza, E., & Spence, C. (2009). Adaptation to audiovisual asynchrony modulates the speeded detection of sound. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 9169–9173.

Nellhaus, G. (1968). Head circumference from birth to eighteen years: Practical composite international and interracial graphs. *Pediatrics, 41*, 106–114.

Ngo, M. K., & Spence, C. (2010a). Crossmodal facilitation of masked visual target identification. *Attention, Perception, & Psychophysics, 72*, 1938–1947.

Ngo, M. K., & Spence, C. (2010b). Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli. *Attention, Perception, & Psychophysics, 72*, 1654–1665.

Olivers, C. N. L., & van der Burg, E. (2008). Bleeping you out of the blink: Sound saves vision from oblivion. *Brain Research, 1242*, 191–199.

Paus, T., Zijdenbos, A., Worsley, K., Collins, D. L., Blumenthal, J., Giedd, J. N., ... Evans, A. C. (1999). Structural maturation of neural pathways in children and adolescents: In vivo study. *Science, 283*, 1908–1911.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Petrini, K., Russell, M., & Pollick, F. (2009). When knowing can replace seeing in audiovisual integration of actions. *Cognition, 110*, 432–439.

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 10598–10602.

Röder, B., Pagel, B., & Heed, T. (2013). The implicit use of spatial information develops later for crossmodal than for intramodal temporal processing. *Cognition, 126*, 301–306.

Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2003). Sound induces perceptual reorganization of an ambiguous motion display in human infants. *Developmental Science, 6*, 233–241.

Seitz, A. R., Kim, R., & Shams, L. (2006). Sound facilitates visual learning. *Current Biology, 16*, 1422–1427.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature, 408*, 788.

Soto-Faraco, S., & Alsius, A. (2009). Deconstructing the McGurk–MacDonald illusion. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 580–587.

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology, 28*, 61–70.

Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. *Annals of the New York Academy of Sciences, 1296*, 31–49.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.

Stein, B. E., Stanford, T. R., Ramachandran, R., Perrault, T. J., & Rowland, B. A. (2009). Challenges in quantifying multisensory integration: Alternative criteria, models, and inverse effectiveness. *Experimental Brain Research, 198*, 113–126.

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., et al (2014). Multisensory temporal integration in autism spectrum disorders. *Journal of Neuroscience, 34*, 691–697.

Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: Task and stimulus dependencies. *Experimental Brain Research, 227*, 249–261.

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*, 1517–1529.

Tremblay, C., Champoux, F., Voss, P., Bacon, B. A., Lepore, F., & Théoret, H. (2007). Speech and non-speech audio–visual illusions: A developmental study. *PLoS ONE, 2*(8), e742.

van der Burg, E., Cass, J., Olivers, C. N. L., Theeuwes, J., & Alais, D. (2010). Efficient visual search from synchronized auditory signals requires transient audiovisual events. *PLoS ONE, 5*(5), e10664.

van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Non-spatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1053–1065.

van Eijk, R. L., Kohlrausch, A., Juola, J. F., & van de Par, S. (2008). Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Perception & Psychophysics, 70*, 955–968.

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: Crossmodal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1583–1590.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics, 72*, 871–884.

Vroomen, J., & Stekelenburg, J. J. (2011). Perception of intersensory synchrony in audiovisual speech: Not that special. *Cognition, 118*, 75–83.

Walker-Andrews, A. S. (1986). Intermodal perception of expressive behaviors: Relation of eye and voice? *Developmental Psychology, 22*, 373–377.

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin, 121*, 437–456.

Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia, 64*, 105–123.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 3*, 638–667.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental Psychology, 34*, 1289–1309.

Westermann, G., & Mareschal, D. (2014). From perceptual to language-mediated categorization. *Philosophical Transactions of the Royal Society B, 369*. http://dx.doi.org/10.1098/rstb.2012.0391.

Wightman, F., Allen, P., Dolan, T., Kistler, D., & Jamieson, D. (1989). Temporal resolution in children. *Child Development, 60*, 611–624.

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio–visual simultaneity judgments. *Perception & Psychophysics, 67*, 531–544.