Combining the senses: the role of experience- and task-dependent mechanisms in the development of audiovisual simultaneity perception

Karin Petrini[1], Georgina Denis[2], Scott A. Love[3], Marko Nardini[1]

[1] Department of Psychology, University of Bath, Bath, UK
[2] UCL Institute of Neurology, London, UK
[3] PRC, INRA, CNRS, IFCE, Université de Tours, Nouzilly, France
[4] Department of Psychology, Durham University, Durham, UK

Corresponding Author:
Dr Karin Petrini
Senior Lecturer/Associate Professor
Department of Psychology
University of Bath
Claverton Down
Bath
BA2 7AY
United Kingdom
Email: K.Petrini@bath.ac.uk

The authors declare no conflict of interest.

Word count: 7990

May 2020 - in press, Journal of Experimental Psychology: Human Perception and Performance

Abstract

The brain's ability to integrate information from the different senses is essential for decreasing sensory uncertainty and ultimately limiting errors. Temporal correspondence is one of the key processes that determines whether information from different senses will be integrated and is influenced by both experience- and task-dependent mechanisms in adults. Here we investigated the development of both task- and experience-dependent temporal mechanisms by testing 7-8-year-old children, 10-11-year-old children and adults in two tasks (simultaneity judgment, temporal order judgment) using audiovisual stimuli with differing degrees of association based on prior experience (low for beep-flash vs. high for face-voice). By fitting an independent channels model to the data, we found that whilst the experience-dependent mechanism of audiovisual simultaneity perception is already adult-like in 10-11-year-old children, the task-dependent mechanism is still not. These results indicate that differing maturation rates of experience-dependent and task-dependent mechanisms underlie the development of multisensory integration. Understanding this development has important implications for clinical and educational interventions.

Keywords: experience-dependent, task-dependent, audiovisual temporal mechanism, multisensory perception, decisional processes, model-based analysis

Public Significance Statements

Combining our different senses to perceive the world underpins our abilities to learn, reason, and act. This study strongly suggests that adult-like abilities to combine different senses are achieved through a lifelong process of learning and development, in which the underlying processes develop at different rates. A better understanding of this development has clinical and educational implications for future approaches to targeting improvements in multisensory perception in children of different ages.

Introduction

The ability of the brain to integrate information from the various senses is essential for decreasing sensory uncertainty and noise (Ernst & Banks, 2002) and ultimately limiting errors in everyday tasks (e.g. understanding someone, grabbing a cup of coffee, crossing a busy road).

Temporal correspondence is one of the key factors that determines whether information from different senses will be perceived as belonging to the same event thus leading to multisensory integration (Spence & Squire, 2003; Stein, Meredith, & Wallace, 1993; Parise and Ernst, 2016). The extent to which we can tolerate a temporal misalignment between the cues and still bind them gives an estimate of how likely they are to belong together.

In adults, the ability to detect deviations in temporal correspondence or synchrony between auditory and visual information has been shown to vary greatly depending on task, stimulus type and level of prior experience (Lee & Noppeney, 2011; Love, Petrini, Cheng, & Pollick, 2013; Petrini, Holt, & Pollick, 2010; Petrini et al., 2011; Petrini, Russell, & Pollick, 2009; van Eijk, Kohlrausch, Juola, & van de Par, 2008; Vatakis, Ghazanfar, & Spence, 2008; Vatakis & Spence, 2007, 2008; Vroomen & Keetels, 2010). For example, Love et al. (2013) showed that the point of subjective simultaneity (PSS; representing the level of sensory onset asynchrony that participants perceived as most synchronous) obtained through either a synchrony judgments task or a temporal order judgements task differed and that the measures returned by the two tasks did not correlate with each other. This suggests that synchrony judgment (in which participants decide if two sensory information are in synch or not) and temporal order judgment (in which participants decide which sensory information came first or second) are supported by different mechanisms in adult participants. Neuroimaging studies have supported this suggestion by showing that synchrony judgment and temporal order judgment tasks are indeed underpinned by divergent brain mechanisms (Binder, 2015; Miyazaki et al., 2016; Love et al., 2018).

Additionally the measure of audiovisual synchrony window (ASW; representing the range of sensory onset asynchronies within which participants cannot reliably perceive asynchrony or sensory order), obtained under different levels of prior experience has been found to vary greatly in adults. Humans form assumptions through experience on whether two cues should go together (e.g. cat meowing) or not (e.g. dog meowing), a process called the ‘Unity Assumption’ or coupling prior according to Bayesian models (Chen, Shore, Lewis, & Maurer, 2016; Ernst, 2007; Petrini, Dahl, et al., 2009; Sato, Toyoizumi, & Aihara, 2007; Shams & Beierholm, 2010; van Wassenhove, Grant, & Poeppel, 2007; Vatakis & Spence, 2007, 2008). For example, Vatakis and Spence (2007) showed that participants found it more difficult to keep the auditory and visual information separate (were less sensitive to audiovisual asynchrony) when face and voice gender matched (strong unity assumption, e.g., female face with a female voice) than when they did not (weak unity assumption, e.g., female face with a male voice). In other words, the ASW in adults is usually larger for stimuli that have higher unity assumption because they are strongly coupled. This assumption of unity between auditory and visual signals can emerge very rapidly in adult participants as shown by a recent study (Habets, Bruns and Roder, 2017). Habets and colleagues (2017) found participants gave more synchrony responses (i.e. were less sensitive and had larger ASW) for rapidly learned audiovisual combinations than new combinations of the same auditory and visual stimuli. Hence, in adults, the judgement of temporal correspondence between sound and vision is a complex process affected by a number of stimuli-, task- and experience-dependent mechanisms.

We know from many studies focusing on a single multisensory mechanism that young children do not have adult-like multisensory abilities: for example, they do not combine senses optimally to reduce uncertainty as adults do (e.g., Adams, 2016; Gori, Del Viva, Sandini, & Burr, 2008; Gori, Sandini, & Burr, 2012; Nardini, Begus, & Mareschal, 2012; Nardini, Jones, Bedford, & Braddick, 2008; Petrini, Remark, Smith, & Nardini, 2014). Young children are also less sensitive to spatial and temporal correspondences between different senses (Chen et al., 2016; Hillock-Dunn & Wallace, 2012; Hillock, Powers, & Wallace, 2011; Roder, Pagel, & Heed, 2013; Stanley et al., 2019), and are less affected by prior experience or use different priors compared to adults (Chambers, Sokhey, Gaebler-Spira, & Kording, 2017; Thomas, Nardini, & Mareschal, 2010). For example, although the ability to detect lack of simultaneity between sight and sound is present in infants as young as 4 months

(Lewkowicz, 2010), children and adolescents are less sensitive to sensory asynchrony than adults (Hillock et al., 2011; Hillock-Dunn & Wallace, 2012). In fact the development of audiovisual simultaneity judgment and rapid audiovisual recalibration for simple (flash-beep) and more complex (face-voice) stimuli does not reach maturity until adolescence (Noel et al., 2016), and some multisensory processes continue to develop throughout adolescence (Brandwein et al., 2011; Downing, Barutchu, Crewther, 2014). Furthermore, evidence from different labs (using different stimuli and tasks) suggests that the age at which children show adult-like multisensory abilities is task- and sense-dependent (e.g. Gori et al., 2008; Gori et al., 2012; Petrini et al., 2014). Hence, the age for development of adult-like task- and experience-dependent audiovisual temporal mechanisms may vary (e.g. Barutchu, Crewther, & Crewther, 2009; Barutchu et al., 2010; Gori et al., 2008; Gori et al., 2012; Petrini et al., 2014), and reach their adult-like state either at similar or different ages. Knowing whether and when different audiovisual temporal mechanisms develop adult-like abilities is essential in order to provide support to the perceptual narrowing theory of multisensory development (Lewkowicz and Ghazanfar, 2009). The developmental perceptual narrowing theory of multisensory perception (Lewkowicz and Ghazanfar, 2009) states that younger infants have a broader ability to respond to different multisensory events (e.g. have the same sensitivity to asynchrony for faces and voices from native and non-native languages) while older infants can respond in the same manner to only familiar or native events (e.g. can only detect asynchrony for faces and voices from their native language). If this process of perceptual narrowing continues in childhood (and perhaps even adulthood) we would expect younger children to have less differentiated mechanisms of audiovisual simultaneity perception (e.g. their ability to detect asynchrony between auditory and visual cues should not change significantly for different stimuli or tasks). On the other hand, older children and adults should have more differentiated mechanisms and thus greater sensitivity in detecting audiovisual simultaneity depending on the task and stimulus. Furthermore, a better understanding of when different audiovisual temporal mechanisms reach near adult-like maturity is important for developing the most targeted and effective clinical and educational interventions aimed at children with deficits in these abilities (e.g. autistic and dyslexic children and children with languages impairments; Francisco, Jesse, Groen, & McQueen, 2017; Kaganovich, 2017; Stevenson et al., 2016; Stevenson, Siemann, Schneider, et al., 2014; Stevenson, Siemann, Woynaroski, et al., 2014; Wallace & Stevenson, 2014; Ye, Russeler, Gerth, & Munte, 2017).

Within a single experiment, and for the first time, we examined whether and how different mechanisms of audiovisual temporal perception develop through childhood. We also compare for the first time in children audiovisual simultaneity judgements obtained from different tasks (i.e. using both simultaneity and temporal order judgement). Differences in PSS for temporal order judgment and synchrony judgment tasks and changes in ASW for face-voice (high prior experience) and flash-beep (low prior experience) displays were examined in three different participant age groups (a group of 7-8 year-old children, a group of 10-11 year-old children and a group of adults). Importantly we applied an independent channels model (Alcala-Quintana & Garcia-Perez, 2013; Garcia-Perez & Alcala-Quintana, 2012) to the data to uncover the underlying causes of these developmental changes. In fact, measures of PSS and ASW are composite estimates of sensory, decisional and bias processes and cannot discriminate between them, thus a model-based analysis was used to obtain model parameters corresponding to sensory (e.g. rate of processing of the visual and auditory cues) and decisional processes (e.g. criterion or internal decision boundary). We examined PSS and ASW estimates in addition to model parameters (rather than focusing solely on the model parameters) as this would allow us to compare our findings with those of the few previous studies examining the development of audiovisual simultaneity perception (Hillock et al., 2011; Hillock-Dunn & Wallace, 2012; Chen et al., 2016), and showing late development of adult-like performance. The ICM has been used previously in a developmental study (Chen et al., 2016) to examine the development of audiovisual simultaneity perception using only the synchrony judgement task. Based on these few studies we predicted that both task- and experience-dependent audiovisual temporal mechanisms would mature late in childhood. Also based on evidence coming from different studies focusing on a single mechanism of audiovisual simultaneity (e.g. Stanley et al., 2019) we predicted that these two mechanisms would reach adult-like states at different ages during development.

Materials and Methods

Participants

Fifteen 7-8-year-old children, thirteen 10-11-year-old children, and fourteen adults took part in the present study. The data for one 7-8-year-old child and three 10-11-year-old children had to be excluded because either their PSS fell outside the range of asynchrony or their ASW was larger than the range of asynchrony used, indicating they could not perform the task. The data of an additional 7-8-year-old child had to be excluded because he/she did not complete the experiment. Hence we analysed the data for thirteen 7-8-year-old children (*Mean* = 7.85, *SD* = .38, 8 female), ten 10-11-year-old children (*Mean* = 10.27, *SD* = .47, 6 female), and fourteen adults (*Mean* = 24.07, *SD* = 3.12, 7 female). The children were all recruited from the same school in London. The goodness of fit of the model to the data was quantified through chi-square tests implemented in the model (Alcala-Quintana and Garcia-Perez, 2013) which returned $p>0.01$ (indicating good fit to data) for all the participants' data included in the analysis (see supplemental material for chi-square results). All participants were native English speakers, had normal or corrected to normal vision and reported no hearing difficulties. The University College London ethics committee approved the experiment and it was conducted in accordance with the ethical standards laid down in the 2013 Declaration of Helsinki.

Stimuli

Two stimulus types were used (Love et al., 2013): 1) flash-beep (low unity assumption), and 2) face-voice (high unity assumption). In flash-beep stimuli the beep was a pure tone at 2000 Hz, while the flash was a white dot (luminance: 85 cd/m^2) presented on a black background (luminance: 12 cd/m^2). The area of the white dot approximated the area subtended by the speaker's mouth region in the face-voice displays. To produce the audiovisual movies (60 Hz), the pure tone and white dot were imported in Adobe Premiere 1.5 and their duration was resized to 33 ms to create the synchronous (0 ms SOA level) condition. We used 7 SOA levels: 3 audio-leading (-333, -200, -67 ms), 3 video-leading (+333, +200, +67 ms) and 1 synchronous. The duration of asynchronous conditions increased with the increase in asynchrony level, i.e. 366, 233, 100 ms respectively for the ±333, ±200, ±67 ms. A black screen with no sound was used to fill the lag between the beep and flash in the six asynchronous SOA conditions.

Face-voice stimuli were dynamic audiovisual movies (25 Hz) of a native English speaker saying “tomorrow”. The visual speech cue contained the full face. To produce asynchronous versions the audio and visual streams were shifted along the movie timeline relative to each other using a method similar to previous research (see Love et al., 2013). This shifting produced gaps at the beginning and end of the movie timeline, which were appropriately filled with the first and last frame of either the auditory or visual stream to produce a non-speaking still face image. For speech stimuli, 7 SOA levels were used with the audio stream shifted either to begin before the video stream (-400, -240, -80 ms) or after (+400, +240, +80 ms) and 1 synchronous (duration = 1.6 s; Love et al., 2013). For face-voice stimuli, previous work (e.g., Conrey and Pisoni, 2006; Van Wassenhove, Grant, Poeppel, 2007; Stevenson et al., 2010) used a wider range of asynchrony levels than that flash-beep, which is why we used a wider range for our face-voice stimuli. Similar to flash-beep stimuli, stimulus duration can be calculated by adding the asynchrony level to the duration of the synchronous condition (1.6 s); hence, duration ranged between 1.6 seconds for the 0 asynchrony and 2 seconds for the ±400 ms asynchrony.

Apparatus and Procedure

Stimuli were presented via a MacBook Pro laptop computer running OS X 10.7.5. The visual cues were displayed on the 15-inch monitor of the laptop running at 1024x768 screen resolution and 60Hz refresh rate. Auditory cues were presented through high quality isolation headphones and the sound intensity was kept at 60 dB. Presentation was achieved using MATLAB 2010a (MATHWORKS Inc., Natick, MA) and the Psychophysics Toolbox (PTB3) extensions (Brainard, 1997; Pelli, 1997).

The experiment was split into 2 sub-experiments, one for each stimulus type. The order of these was counterbalanced across participants, with an attempt to have a similar number starting on each stimulus type. The 2 experiments were split across 2 sessions, each approximately 20 minutes, which were completed on the same day. Each experiment presented only one stimulus type and consisted of 20 blocks: half of the blocks were synchrony judgment blocks and the other half were temporal order judgment, presented in a randomised order. At the start of each experiment, participants completed 6 practice trials (3 synchrony judgment and 3 temporal order judgment) and asked any questions of clarification if needed. Participants then pressed any key to begin the experiment and the instructions as to whether the first block was an synchrony judgment or a temporal order judgment block appeared on screen for 4 seconds. The relevant task instructions were presented for 4 seconds at the start of every block. Within a block there were 7 trials: one presentation of each SOA level of the current stimulus type in a randomised order. After each trial the current task question and possible answers were displayed on screen until the participant responded, which triggered the start of the next trial. During synchrony judgment blocks participants were instructed to press ‘1’ or ‘3’ on the number pad dependent on whether they thought the audio and visual cues were synchronous or asynchronous, respectively. During temporal order judgment blocks they pressed ‘1’ if they thought the video came first and ‘3’ if they perceived the audio to come first. No feedback was given. In total participants underwent 280 trials (7 (SOA levels) x 2 (Task: synchrony judgment, temporal order judgment) x 2 (Stimuli: flash-beep, face-voice) x 10 (repetitions)).

Analysis

We used an independent channels model (ICM) to fit the temporal order judgment and synchrony judgment data jointly (with common sensory parameters for the two tasks) for each participant’s data and obtain measures of model parameters. Additionally estimates of the audiovisual synchrony window (ASW) width and point of subjective simultaneity (PSS) were obtained. The ICM model used here has been previously described and validated by Garcia-Perez and Alcala-Quintana (2012) and Alcala-Quintana and Garcia-Perez (2013) for use with synchrony judgment and temporal order judgment data. The model assumes that the arrival latencies T_V and T_A for the reference (visual cue here) and test stimulus (auditory cue here) respectively are random variables with shifted exponential distributions (Fig. 1). The model also assumes that on each trial the participant collects sensory information to judge whether the visual cue or the auditory cue arrived first, or the two cues were simultaneous (when the order of cue arrival cannot be identified).

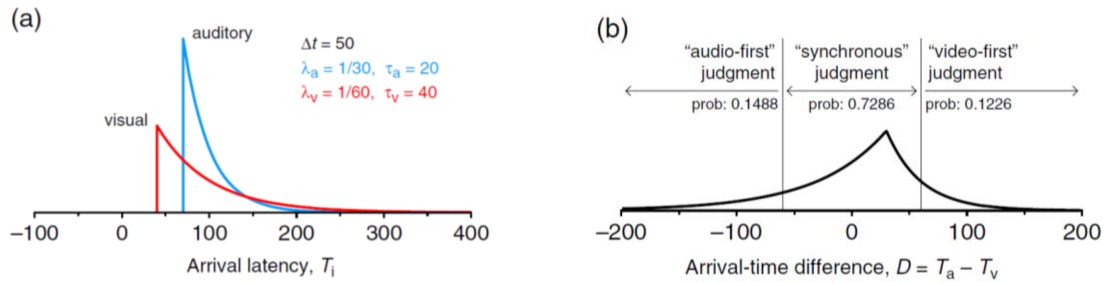

Fig. 1. (a) Example of exponential distributions for the arrival latency of a visual stimulus (red curve) presented at time 0 and an auditory stimulus (blue curve) presented at time Δt 0 = 50 ms, i.e., lagging the visual stimulus of 50ms. (b) Bilateral exponential distribution of arrival-time difference and cutpoints on the decision space (vertical lines, at $D = \pm\delta$ with $\delta = 60$), determining the probability of each judgment (taken from Garcia-Perez & Alcala-Quintana, 2012). Adapted by permission from Springer Nature: [Springer Nature] [Psychonomic Bulletin & Review] [García-Pérez, M.A., & Alcalá-Quintana, R. (2012). On the discrepant results in synchrony judgment and temporal-order judgment tasks: A quantitative model. Psychonomic Bulletin & Review, 19(5): 820e846], [Copyright © 2012, Psychonomic Society, Inc.] (2012).

Exponential distributions are commonly used to describe arrival latencies or peripheral processing times (see Alcala-Quintana and Garcia-Perez, 2013) because they do not allow the time at which the sensory signals reach a central mechanism to be before the onset of the stimulus triggering the signals. This model has been tested and validated on different sets of published data from audiovisual simultaneity perception studies (Garcia-Perez and Alcala-Quintana, 2012; Alcala-Quintana and Garcia-Perez, 2013) similar to this study, and has been used recently to test children simultaneity perception when using synchrony judgment task (Chen et al., 2016).

In contrast to psychometric functions commonly used to fit this type of data (e.g. Gaussian and Logistic) this model is generative in that it models the underlying sensory and decisional processes that lead to the pattern of responses consistently across tasks. The model includes a central mechanism that determines the judgment of temporal order or synchrony by a ternary decision rule (Fig. 1b) applied to the arrival-time difference between the two signals. This model also allows for asymmetric distribution of data which are common in these tasks (e.g. participants usually are less able to detect asynchrony when vision leads audition), and takes into consideration response errors (i.e. pressing the wrong key and participants' lapses) and response bias (see below). From the fit of this generative model it is also possible to obtain estimates of properties commonly reported in studies of multisensory processing such as the width of the ASW and the PSS for both temporal order and simultaneity judgment tasks. The notion underlying the ICM is that the generating process holds across synchrony and temporal order judgment tasks and, then, the derived psychometric functions are consistent with one another.

The model has parameters that correspond distinctly to sensory and decisional processes. The sensory parameters include those that describe the rate of processing and processing variability of the visual and auditory cues (λ_a and λ_v) and the latency difference or processing time difference at which the two stimuli arrive at the central mechanism (τ). These sensory parameters were common for the two tasks. The decisional parameters include the finest temporal resolution that can be used to detect a latency difference (δ), and the internal decision boundary or criterion for asynchrony judgments. That is, δ is a model parameter meant to capture realistic aspects of the decision process and consequently is influenced by both the resolution limit for a particular individual but also by the individual's decision to loosen up or try to narrow (through training and dedication) the decision boundary or criterion. A second decision parameter refers to the response bias parameter that is unique to Temporal Order Judgments (ξ). The smaller δ the more the participant is able and/or willing to resolve small differences in arrival latency between the cues, and thus this parameter usually correlates positively with the ASW width (larger δ = larger ASW). The ξ gives a measure of bias towards guessing auditory first ($\xi < .5$) or visual first ($\xi > .5$) when no order of arrival is perceived (i.e. the cues are perceived as simultaneous). Hence, participant responses are considered biased toward saying vision first when unsure if $\xi > .5$, while biased towards saying audio first when unsure if $\xi < .5$. The joint model fitted to the individual data had 11 parameters (λ_a, λ_v, τ, δSJ, δTOJ, εSJ2-TF, εSJ2-S, εSJ2-RF, εTOJ-TF and εTOJ-RF, ξ), where TF stands for test-first (in our case auditory-first), RF for reference-first (in our case vision-first), S for synchrony, SJ and TOJ for synchrony judgment and temporal order judgement tasks, and ε for error (all the other symbol and parameters have been explained above). Three of the parameters, as mentioned, were common to both tasks (λ_a, λ_v, τ,), while the others were not. The synchrony judgement task had three error parameters (εSJ2-TF, εSJ2-S, and εSJ2-RF), while the temporal order judgment had two (εTOJ-TF and εTOJ-RF). In addition, the temporal order judgement task had, as discussed, an additional bias parameter (ξ). Please see supplemental material for the starting values used to fit the data.

For the synchrony judgment task, the proportion of synchronous and asynchronous responses at each SOA level were fit by the ICM described above, while for the temporal order judgment task the proportion of video and audio first responses were fit with the same model. The model fitting procedure was conducted separately for each participant and stimulus combination (to see examples of the fitting procedure to individual data see Fig. 2 and Fig. 1S in the supplemental material). The PSS represents the level of SOA that participants perceive as most synchronous, and was derived from the peak (i.e., the SOA at which "simultaneous" responses are most prevalent) and middle point (the center of range of SOAs over which "simultaneous" responses prevail) for synchrony judgment and from the 50% point of ICM fit for temporal order judgment. The ASW represents the range of SOA within which participants cannot reliably perceive asynchrony or cue order. PSS and ASW were calculated from the ICM fitted parameters (see supplemental material for further details).

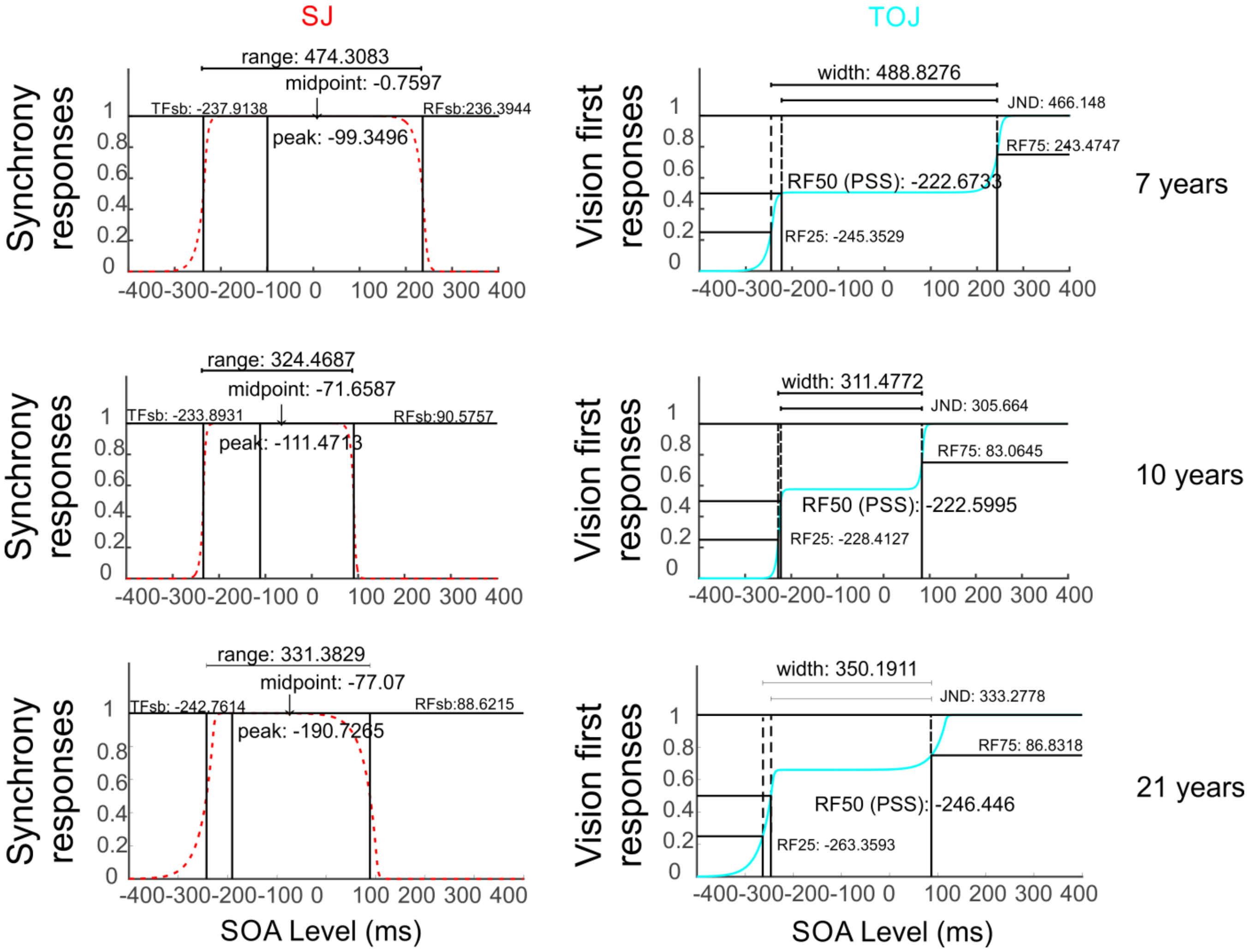

Fig. 2. The individual ICM (independent channels model) fitting results for a 7-year-old child (top panels), a 10-year-old child (middle panels) and an adult (bottom panels) in the face-voice condition. The left panels describe the results for the synchrony judgment task (red and dashed line), while the right panels for temporal order judgment (TOJ) task (cyan and solid line). Range for synchrony judgment (SJ) and width for temporal order judgment (TOJ) = audiovisual synchrony window (ASW). Midpoint and peak for synchrony judgment (SJ) and RF50 for temporal order judgment (TOJ) = point of subjective simultaneity (PSS). TFsb = Auditory-first simultaneity boundary (the 50% point on the left side of the psychometric function for simultaneity judgments); RFsb = Vision-first simultaneity boundary (the 50% point on the right side of the psychometric function for simultaneity judgments).; RF25 = The 25% point on the psychometric function for visual-first responses; RF75 = The 75% point on the psychometric function for visual-first responses; JND = The size of the just noticeable difference (JND; the distance between the 50% and the 75%

points). The y axis presented the proportion of synchrony (for synchrony judgment) or visual first (for temporal order judgment) responses. Please see Fig. 4S in the supplemental material for the same examples fitted by normal and cumulative Gaussian functions. Also see Fig. 3S for a representation of synchrony judgment and temporal order judgment average responses as a function of stimulus onset asynchronies (SOAs) for the three age-groups, tasks (synchrony judgment and temporal order judgment) and stimuli (flash-beep and face-voice).

Results

PSS and ASW

We first examined the effect of age, task and stimulus on the PSS individual estimates as assessed by the ICM model and as exemplified for three participants in Fig. 2. We carried out a mixed factorial ANOVA with age (7-8 years, 10-11 years, and adults) as between-subjects factor, and task (synchrony judgment and temporal order judgment) and stimuli (flash-beep and face-voice) as within-subjects factors. This analysis revealed a significant main effect of stimulus ($F(1, 34) = 5.244$, $p = .028$, $\eta^2 = .134$), with the PSS for face-voice stimuli (*Mean* = -1.50, *SD* = 117.82) being closer to the point of physical synchrony than that for flash-beep (*Mean* = 57, *SD* = 101.08). η^2 = partial eta squared. We also found a significant interaction between age and task ($F(2, 34) = 3.658$, $p = .036$, $\eta^2 = .177$).

No other main factor or interaction reached significance ($F \leq 1.323$, $p \geq .280$). Fig. 3a and b show the average PSSs for the interaction between age and task, and shows that while both child groups had similar PSSs for the synchrony judgment and temporal order judgment tasks, adults, as expected, had different estimates of PSS for the temporal order judgment than synchrony judgment (Fujisaki and Nishida, 2009; Love et al., 2013; Maier et al., 2011; Petrini et al., 2010; Van Eijk et al., 2008; Vatakis et al., 2008; Vroomen and Stekelenburg, 2011). Paired-samples t-tests, Bonferroni corrected, supported these observations in that 7-8 year-old children ($t(12) = -.296$, $p = .772$, 95% CI [-96.97, 73.79]), and 10-11 year-old children ($t(9) = -1.024$, $p = .333$, 95% CI [-93.35, 35.17]) had similar PSSs for the two tasks, while adults $t(13) = 2.906$, $p = .036$, 95% CI [22.91, 155.67], Cohen's *d* =0.78) did not. Independent-samples t-tests, Bonferroni corrected, showed that there were no significant differences in PSS for either temporal order judgment or synchrony judgment among age groups ($t \leq$ -2.231, $p \geq .108$). The PSS results for the middle point rather than peak returned very similar results (see supplemental material). We also carried out a correlation, separate for children (given that children do not differ in PSS) and adults, to assess whether the PSS estimates of the two tasks were positively correlated or not. Whereas we found no correlation for the adult group between the PSS estimates ($r_s = .261, p = .180$) we did find a significant correlation for the children ($r_s = .433, p = .003$).

We next examined the effect of age, task and stimulus on the ASW individual estimates as assessed by the ICM model and as exemplified for three participants in Fig. 2. We carried out a mixed factorial ANOVA with age (7-8 years, 10-11 years, and adults) as between-subjects factor, and task (synchrony judgment and temporal order judgment) and stimuli (flash-beep and face-voice) as within-subjects factors. This analysis revealed a significant main effect of stimulus ($F(1, 34) = 8.664$, $p = .006$, $\eta^2 = .203$), with the ASW for face-voice (*Mean* = 356.58, *SD* = 117.10) being larger than that of flash-beep (*Mean* = 297.32, *SD* = 96.79) stimuli, of task ($F(1, 34) = 12.596$, $p = .001$, $\eta^2 = .270$), with synchrony judgment (*Mean* = 364.70, *SD* = 98.04) having a larger ASW than temporal order judgment (*Mean* = 289.20, *SD* = 110.01), and of age X stimulus ($F(2, 34) = 3.931$, $p = .029$, $\eta^2 = .188$). No other main factor or interaction reached significance ($F \leq 1.437$, $p \geq .252$).

Fig. 3c and d display the ASWs for age x stimulus and shows that while the younger children had a similar ASW width for flash-beep (low level of experience) and face-voice (high level of experience), the older children and adults showed an enlargement of the ASW for face-voice as expected by the ‘Unity Assumption’ and shown several times for adult participants (see Chen and Spence, 2017 for a review). Paired-samples t-tests, Bonferroni corrected, support these observations in that 7-8 year-old children had similar ASWs for the two stimuli ($t(12) = .519$, $p = .613$, 95% CI [-64.22, 104.43]), while 10-11 year-old children ($t(9) = -3.053$, $p = .042$, 95% CI [-203.69, -30.29], Cohe’s $d = 0.97$) and adults ($t(13) = -2.793$, $p = .045$, 95% CI [-162.64, -20.78], Cohe’s $d = 0.75$) had not. Fig. 3c and d also show that for flash-beep stimuli adults had a smaller ASW than either older or younger children in line with previous findings (Hillock et al., 2011), however, independent-samples t-tests showed that these differences did not reach significance (7-8-year-old vs adults: *t*(25)= 1.912, *p* = .067, 95% CI [-5.59, 150.62]; 10-11-year-old vs adults: *t*(22)= 1.292, *p* = .210, 95% CI [-26.43, 113.79]). Also no significant difference was found for the face-voice stimulus (7-8-year-old vs adults: *t*(25)= -.870, *p* = .393, 95% CI [-132.38, 53.76]; 10-11-year-old vs adults: *t*(22)= 1.634, *p* = .116, 95% CI [-18.54, 156.47]).

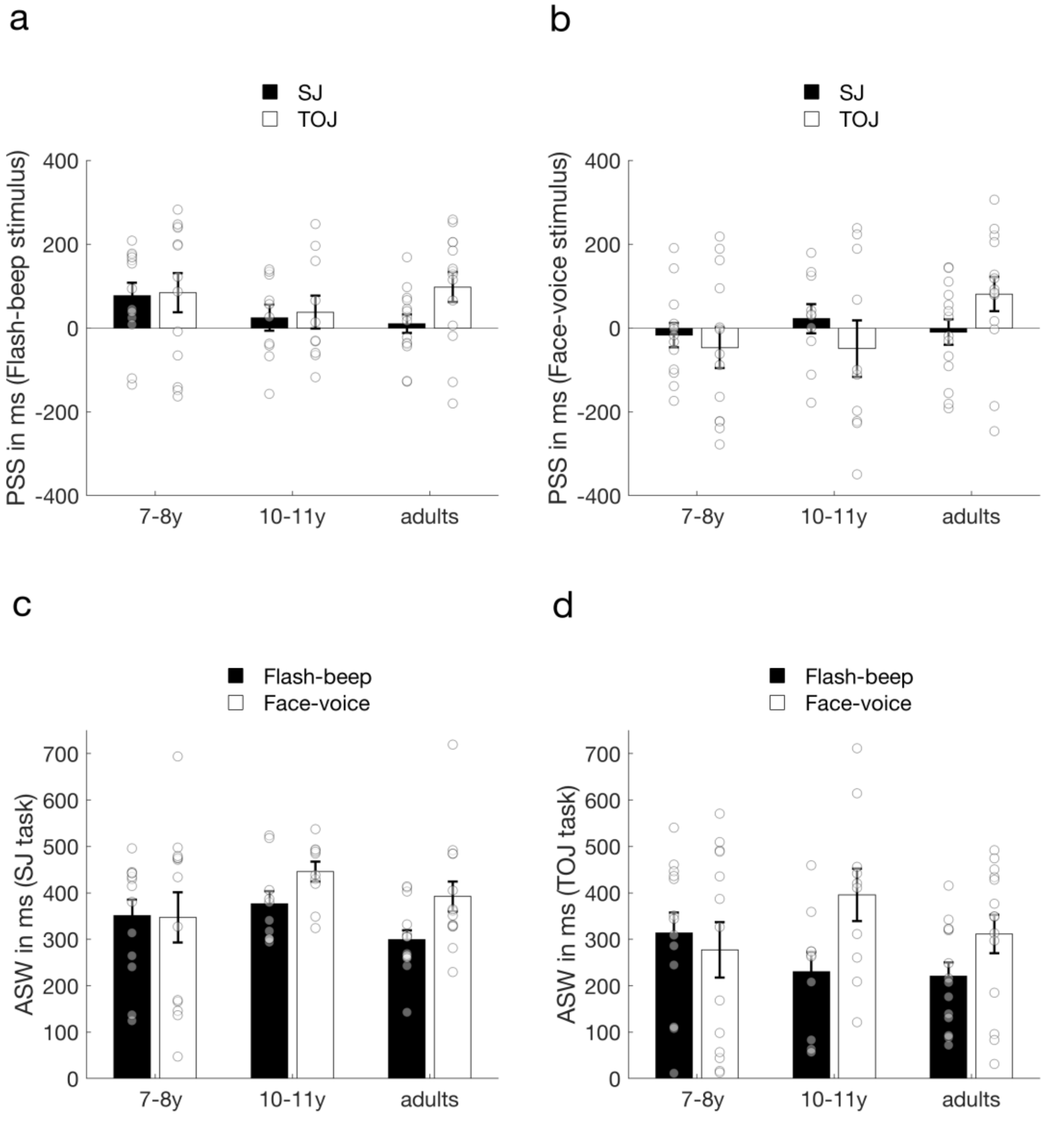

Fig. 3. Effect of age on the estimates returned by the ICM (independent channels model). (a) and (b) Interaction between age and task for the synchrony judgment (SJ) and temporal order judgment (TOJ) PSS estimates (from peak) for flash-beep stimuli on the left panel and for face-voice stimuli on the right panel. (c) and (d) Interaction

between age and stimuli for the flash-beep and face-voice ASW (audiovisual synchrony window) for synchrony judgment task on the left panel and temporal order judgment task on the right panel. The bars represent the group mean while the error bars the standard error of the mean. The circles represent the individual data. Please see Fig. 5S in the supplemental material for the same figure but with added connecting lines for the individual data, and Fig. 6S for a representation of PSS separate for tasks and of ASW separate for stimuli.

ICM Parameters

Since measures of PSS and ASW are composite estimates of sensory and decisional processes and discrimination between these processes is not possible, we also used the ICM to obtain model parameters corresponding to sensory (e.g. rate of processing of the visual and auditory cues) and decisional processes (e.g. criterion or internal decision boundary). Distinguishing between decisional and sensory processes can further explain why the experience-dependent multisensory mechanism achieves an adult-like state earlier than the task-dependent mechanism.

Fig. 4a and b display the δ for age x stimulus and shows that while the younger children had a similar δ for flash-beep (weak unity assumption) and face-voice (strong unity assumption), the older children and adults showed a greater δ for face-voice, supporting the findings for the ASW width. To test the effect of age, task and stimulus on the decision parameter (δ) of the ICM we carried out a mixed factorial ANOVA with age (7-8 years, 10-11 years, and adults) as between-subjects factor, and task (synchrony judgment and temporal order judgment) and stimuli (flash-beep and face-voice) as within-subjects factors. The smaller δ is the more the participant is able and/or willing to resolve small differences in arrival latency between the cues. This analysis revealed a significant main effect of stimulus ($F(1, 34) = 14.139$, $p = .001$, $\eta^2 = .294$), with the δ for face-voice (*Mean* = 189.91, *SD* = 51.95) being greater than that of flash-beep (*Mean* = 156.98, *SD* = 46.27) stimuli, of task ($F(1, 34) = 4.795$, $p = .035$, $\eta^2 = .124$), with synchrony judgment (*Mean* = 183.36, *SD* = 48.20) having a greater δ than temporal order judgment (*Mean* = 163.53, *SD* = 48.87), and an interaction between age and stimulus ($F(2, 34) = 5.267$, $p = .010$, $\eta^2 = .237$). No other main factor or interaction reached significance ($F \leq 1.097$, $p \geq .345$).

Paired-samples t-tests, Bonferroni corrected, support these observations in that 7-8 year-old children had similar δ for the two stimuli ($t(12)= .406$, $p = .692$, 95% CI [-29.77, 43.42]), while 10-11 year-old children ($t(9)= -3.402$, $p = .024$, 95% CI [-96.24, -19.36], Cohen's $d = 1.08$) and adults ($t(13)= -3.876$, $p = .006$, 95% CI [-81.12, -23.05], Cohen's $d = 1.04$) had not. Fig. 4a and b also shows that for flash-beep adults had a smaller δ than either older or younger children. Independent-samples t-tests, Bonferroni corrected, showed that there were no significant differences in δ for either flash-beep or face-voice among age groups ($t \leq 2.338$, $p \geq .084$).

We next examined the effect of age and stimuli on the sensory parameters that were common to both tasks (λ_a, λ_v and τ). These sensory parameters include those that describe the rate of processing or processing variability of the visual and auditory cues (λ_a and λ_v) and the latency difference or processing time difference at which the two stimuli arrive at the central mechanism (τ). We carried out a mixed factorial ANOVA for the three parameters with age (7-8 years, 10-11 years, and adults) as between-subjects factor and stimuli (flash-beep and face-voice) as within-subjects factors. This analysis did reveal a significant main effect of stimuli for λ_a ($F(1, 34) = 4.419$, $p = .043$, $\eta^2= .115$) and τ ($F(1, 34) = 28.244$, $p < .001$, $\eta^2= .454$), with these sensory parameters differing for face-voice (λ_a: *Mean* = .19, *SD* = .12; τ: *Mean* = 21.92, *SD* = 76.04) and flash-beep (λ_a:*Mean* = .14, *SD* = .12; τ: *Mean* = -49.58, *SD* = 49.83) stimuli. No other main factor or interaction was significant ($F \leq 2.921$, $p \geq .068$).

Finally, we tested the effect of age and stimuli on the bias parameter ξ for the temporal order judgment task as a change in bias could explain the found age-related changes in PSS under the temporal order judgment task. We found a significant effect of age ($F(2, 34) = 4.725$, $p = .015$, $\eta^2= .217$), with ξ changing with age (Fig. 4c and d) and resulting in a significant difference in bias between the 7-8 year-old children and the adults group (Bonferroni post hoc tests, P=.021). While the younger children group was slightly biased toward saying vision first when unsure ($\xi > .5$), the adult group was biased towards saying audio first when unsure ($\xi < .5$). No other main factor or interaction reached significance ($F \leq 2.332$, $p \geq .136$). For the analysis of the response errors please see the supplemental material. Finally, we examined whether there was a different relation between PSS for the temporal order judgment task and the measure of bias for the children and adult groups. Correlation analyses returned the same significant negative correlation between bias and PSS for the temporal order judgement task for all age groups ($r_s \geq -.664, p < .001$).

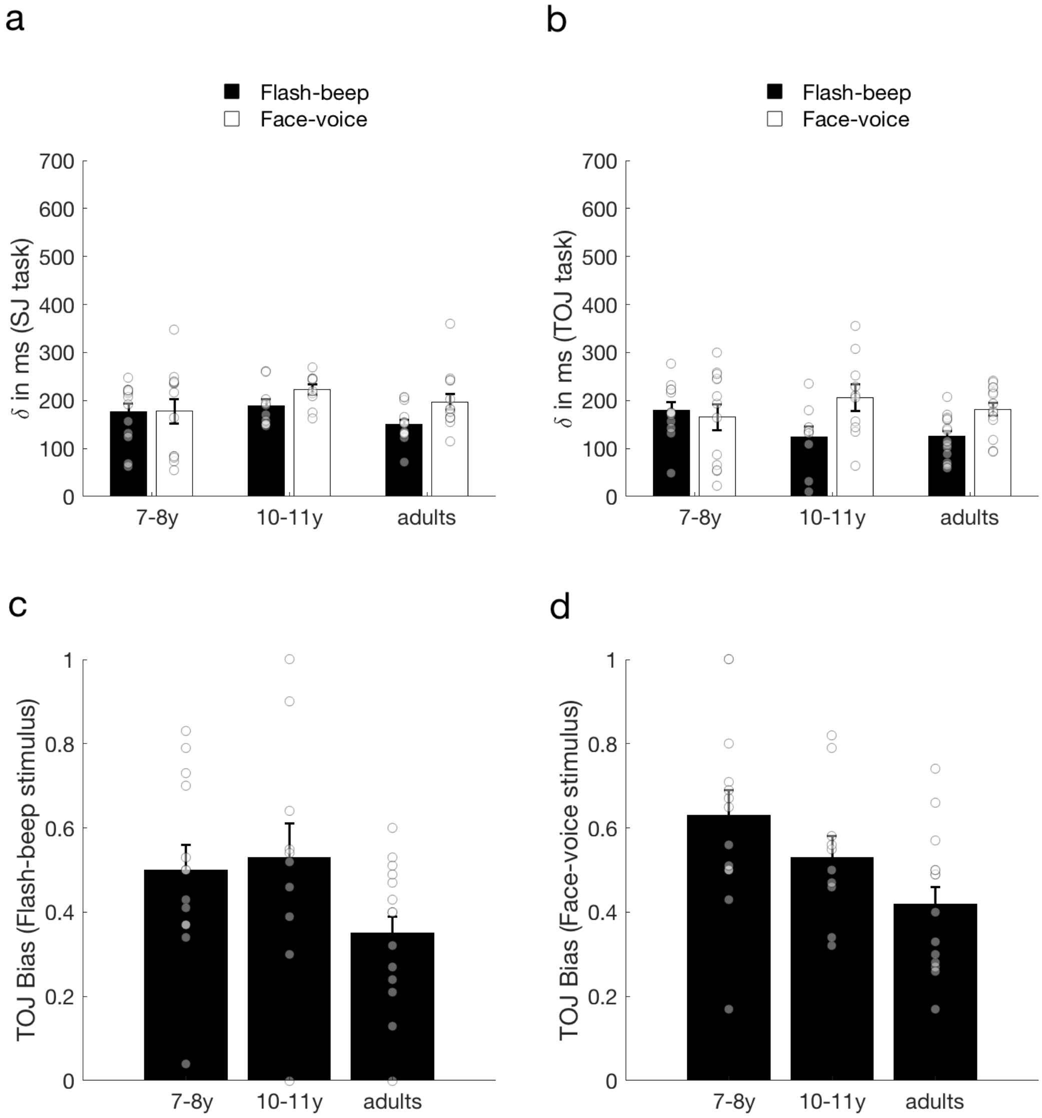

Fig. 4. Effect of age on the parameters returned by the ICM (independent channels model). (a) and (b) Interaction between age and task for flash-beep and face-voice δ (decisional parameter, i.e. the finest temporal resolution that can be used to detect a latency difference) for synchrony judgment (SJ) task on the left panel and temporal order judgment (TOJ) task on the right panel. (c) and (d) Effect of age on temporal

order judgment (TOJ) bias parameter for flash-beep stimulus on the left panel and face-voice stimulus on the right panel. Participant responses are considered biased toward saying vision first when unsure if ξ (the TOJ bias parameter) > .5, while biased towards saying audio first when unsure if ξ < .5. The bars represent the group mean while the error bars the standard error of the mean. The circles represent the individual data.

Discussion

In the present study, within a single experiment, we investigated the development of both task- and experience-dependent audiovisual temporal mechanisms, both of which have a strong influence on adults' synchrony perception (e.g., Love et al., 2013; Love et al., 2018).

Our findings show, as predicted, that both mechanisms develop late in childhood, in that 7-8-year-old children did not show adult-like characteristics in either experience- or task-dependent audiovisual mechanisms. The PSS estimates for the children did not differ for synchrony judgment and temporal order judgment tasks, while as expected they did differ for the adult group (e.g., Love et al., 2013; Love et al., 2018). In addition the ASW estimates of the 7-8-year-old children did not differ for the two stimuli (flash-beep and face-voice) while as expected they did differ in adults (Vatakis & Spence, 2007, 2008). In contrast, the ASW estimates of the 10-11-years-old children were wider for face-voice stimuli compared to flash-beep stimuli indicating that like adults they are affected by the "Unity assumption". This key marker of the experience-dependent mechanism therefore shows a sign of maturity at this age. Taken together, these points highlight that the two audiovisual temporal mechanisms investigated mature at different rates or ages. The experience-dependent mechanism shows markers of adult-like maturity at 10-11-years-old, in contrast with the task-dependent mechanism which is still immature at this age.

Analyses of the ICM parameters show that the maturity of the experience-dependent mechanism, indexed by the widening of the face-voice ASW in the older group of children, results from changes in decisional processes and not sensory ones. The results for all the sensory parameters did not show any age-related difference driven by stimuli, suggesting that the sensory mechanisms underpinning experience-dependent audiovisual temporal mechanisms are already mature in early childhood.

Finally, our results show that the development of task-dependence – i.e., the segregation of temporal order judgment and synchrony judgment processes - requires longer to fully achieve an adult-like state. That is, both groups of children, in contrast to the adult group, showed a lack of difference between PSS estimates for synchrony judgment and temporal order judgment tasks. In fact, only children's PSSs for the two tasks correlated significantly indicating a level of similarity between the two tasks, while adults' PSSs for the two tasks did not (in line with previous findings, e.g. van Eijk et al., 2008; Love et al., 2013). This delivers evidence of differentiated task-dependent mechanisms in adults for audiovisual simultaneity perception. Whereas the bias for the temporal order judgment responses does show a shift with age from reporting visual first to reporting auditory first when uncertain about the cues order, this change in bias cannot fully explain the age-related PSS results for the temporal order judgment task. That is, while 10-11-year-old children did not differ significantly in bias from the adult group they did differ significantly from the adult group in the PSS for the temporal order judgment task. In support of this argument both children and adults showed a negative relation between PSS and bias estimates for the temporal order judgment task, indicating that the bias affected the PSS estimates from this task similarly for children and adults. Hence, while changes in PSS could be the result of a change in bias when uncertain, this might not be the whole explanation for the age-related differences we found here. For the same reason, the results for the response errors (see supplemental material) made by participants cannot fully account for the age-related differences in PSS.

Previous studies (Jaskowski, 1991) suggested that the temporal order judgment task requires more cognitive resources than synchrony judgment, since temporal order judgment not only includes the perceptual processes required for synchrony judgment (detecting successive/simultaneity) but also additional perceptual processes (determination of the temporal order) and this has also been supported by neuroimaging evidence (Binder, 2015; Love et al., 2018; Miyazaki et al., 2016). Our results suggest that these task-dependent perceptual processes might remain undifferentiated and may be carried out by a general multisensory temporal mechanism in children up to at least 10-11 years of age. The pattern of cognitive and neural specialization observed in adults may therefore develop markedly late in childhood, after 10-11 years. Alternatively, it may be plausible that children deal differently with the additional demand of temporal order judgement task (i.e., guessing an order when uncertain), and consequently generate PSS estimates in the temporal order judgment task that better match those in the synchrony judgment task. To identify when adult-like behaviour for the two tasks arises, future behavioural and neuroimaging / neurophysiological studies could include older children and adolescent groups.

Only a small number of previous studies have investigated the development of audiovisual simultaneity perception using a synchrony judgment task and flash and beep stimuli, and one with flash and beep as well as face and voice (Noel et al., 2016); none to our knowledge have used the temporal order judgment task. Two studies examined the development of the ASW for audiovisual simultaneity perception (Hillock et al., 2011; Hillock-Dunn and Wallace, 2012) using a synchrony judgment task and simple ring flash and tone pip stimuli. These studies showed that children as well as adolescents were less sensitive to timing discrepancy than adults (i.e. had wider ASW than adults). A third study also applied the ICM model, similarly to the present study, to test the development of audiovisual simultaneity using a synchrony judgment task and flash and beep type of stimulus (Chen et al., 2016) and showed that children performed similarly to adults (had a similar measure of δ) at 9-11 years of age, but that children and adults did not differ in PSS. Our synchrony judgment findings with the flash and beep stimuli are in line with these previous studies. That is, our results show that adult-like performance (as measured by ASW or δ) is achieved late in childhood (Hillock et al., 2011; Hillock-Dunn and Wallace, 2012; Chen et al., 2016) and that adult-like performance for δ is reached at 10-11 years of age (Chen et al., 2016). Additionally, we show that the PSS for synchrony judgment and flash-beep stimuli did not differ across ages (Chen et al., 2016). Finally, our findings for the ASW and δ do overall show that although this mechanism of audiovisual simultaneity perception is near-adult-like in 10-11-year-old children, ASW and δ for 10-11 year-olds are not as narrow as in adults (Hillock-Dunn and Wallace, 2012). Finally, in line with our findings, in the study by Noel et al. (2016) showing a late maturation of both audiovisual simultaneity judgement and rapid recalibration, the ASW for flash-beep and face-voice stimuli start differentiating (with the ASW for face-voice stimuli being larger than that for flash-beep) in late childhood/adolescence.

Our findings additionally show that for the natural and more commonly-experienced stimuli of face and voice, the development of audiovisual simultaneity perception follows a very different trend. Whereas for flash and beep stimuli we show a narrowing of the ASW or δ as in previous studies (Hillock et al., 2011; Hillock-Dunn and Wallace, 2012; Chen et al., 2016) for face-voice stimuli we show an enlargement of these measures. Furthermore, while we show no difference between children and adults in PSS for synchrony judgment task in line with a previous study (Chen et al., 2016), we do show a difference in PSS as measured by a temporal order judgment task. Our study thus demonstrates that the developmental trend of audiovisual simultaneity perception is task- and experience-dependent.

Limitations

It should be noted that the two stimuli used in the present experiment did not only differ in level of experience but also in complexity. The face-voice stimulus is clearly more complex than the flash-beep, in addition to having a higher level of unity assumption/experience. Therefore, the differences we found between children and adults could potentially be due to the complexity of the stimuli and/or differences in experience. Our decision to use these stimuli was driven by the need to maximise the difference in experience between the stimuli and use a set of standardised stimuli for which synchrony judgment and temporal order judgment tasks have been previously judged as similarly difficult by adults (i.e. temporal order judgment was rated as more difficult than synchrony judgment similarly for the two stimuli used here; Love et al., 2013). Furthermore, we wanted to make sure that participants would be able to perform the temporal order judgment task for both stimuli. This was because it has previously been shown that modifying the flash-beep clips to match the dynamic profile of a more natural and complex stimulus greatly impaired participants ability to perform the temporal order judgment task (Love et al., 2013). Thus we used two stimuli naturally differing in experience (as it is uncommon to experience a face and voice for few milliseconds or a flash and beep for more than few milliseconds) as well as complexity. Our model-based approach helped distinguish between the influence of these factors. If stimulus complexity was influencing participants' synchrony judgements, an age-related differences in sensory processes for the two stimulus types would have been found. That is, if levels of complexity rather than experience-dependent mechanisms were driving the age-related effect we found here for the two stimuli, then we would expect to find a difference between children and adults in sensory processes for the two types of stimuli chosen, but we do not. Furthermore it would be difficult to explain why no difference in ASW and decision parameter (δ) measures between flash-beep and face-voice stimuli were found in the younger children if the complexity was driving the differences. Indeed, we should have found this effect of complexity either across all age-groups (with ASW and δ being larger for face-voice than flash-beep for children and adults) or possibly decreasing with age (with adults showing a smaller difference in ASW and δ for the two stimuli compared to young children). However, we found the opposite result. Finally, a recent study by Barutchu et al. (2019) also shows near adult-like audiovisual processes with familiar verbal stimuli with no semantics (e.g. "jat" and "chel") even when the complexity of the auditory signal was controlled for. Hence, this brings further evidence that stimulus complexity is unlikely to account for our findings. For all these reasons, we conclude that the age-related changes we found are driven largely by maturation of experience-dependent mechanisms rather than differences in complexity between the stimuli used. Nevertheless, future studies could avoid differences in stimulus complexity or other characteristics besides the one of interest by having children and adults learn an association between arbitrary pairs of audiovisual features (e.g. sound frequency/color) to manipulate the level of experience with a given stimulus before testing them with different tasks.

Another point to discuss refers to the different range of audiovisual asynchrony for the two stimuli used in the present study. As mentioned in the methods section we chose the range for these two stimuli based on previous studies (i.e., Love et al., 2013). However, that means that for face-voice stimuli we had larger range of audiovisual asynchrony than for flash-beep stimuli. Although this difference in range is important to consider, it cannot fully explain the larger ASW we found for face-voice than flash-beep stimuli in older children and adults. That is, as this difference was the same across age groups it is unclear why young children did not have larger ASW for face-voice than flash-beep as we would have expected the younger children to be influenced by different ranges of asynchrony equally if not more than the older groups. Furthermore, having a larger range of asynchrony should have helped older children and especially adults to achieve higher precision (as the more the stimuli are desynchronised the more should be easy to detect asynchrony) and thus have smaller rather than larger ASW as we found in the present study.

Another limitation of this study, which is common to the field, is the small sample size of participants. Conducting experiments with hundreds of trials and repetitive psychophysics methods with children is difficult, especially in terms of maintaining children's level of attention, avoiding drop outs and obtaining meaningful data. Here we provide the results of a power analysis to help the reader understand the potential lack of power in our study design. A priori type of power analysis for an ANOVA repeated measures within-between interaction was run using G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) to estimate the required sample size. For the estimation we used a Cohen's F of 0.25 (for a medium effect size), a level of power of 0.80, 3 groups, 4 measurements, an alpha level of 0.05, and the adjustment to "Effect size specification as in SPSS". The sample size returned was 78 with at least 26 participants per group (but also see MorePower 6.0; Campbell & Thompson, 2012). Nevertheless, we replicate results from previous developmental studies as well as studies assessing only adults' performance (Hillock et al., 2011; Hillock-Dunn and Wallace, 2012; Chen et al., 2016; Love et al., 2013); and this despite using a model based analysis rather than psychometric fitting routines. Furthermore, the results for the 10-11-year-old children match closely the results of the 7-8-year-old children for the task-dependent factor, while they match closely the data for adults for the experience-dependent factor indicating that there is a good level of internal validity despite the different samples of participants. Linked to this limitation is also our use of a high number of model parameters due to our decision to include all possible error parameters to the ICM. Clearly, this can lead to an over-parameterised model given for example the low number of SOAs or trials per SOA level. Again, to minimise the testing time for children given the inclusion of two stimuli and two tasks within one study, we had to reduce the number of SOAs and repetition per SOA. However, effects of errors and biases have too often been unaccounted for in developmental research and thus we opted to include all the error parameters (similarly to a previous developmental study using simultaneity judgement task and ICM: Chen et al., 2016).

This was to better understand their link and impact on our age-related findings. We believe that showing that measures of bias and error cannot fully account for the developmental trends found in our study is an important contribution, despite the potential over-parameterisation of the model. In addition, our study has a high number of dependent variables as we wanted to report both commonly used estimates as well as model parameters (including error measures) similarly to previous developmental studies using ICM (Chen et al., 2016). However our comparisons were planned and we minimised the effect of multiple comparisons by using a Bonferroni correction and by reporting the Cohen's d showing that the effect sizes for the significant differences were large.

Conclusion

Overall our results support the theoretical viewpoint that multisensory development undergoes perceptual narrowing even during childhood (Lewkowicz and Ghazanfar, 2009). In fact, while children show similar sensitivity to asynchrony irrespective of stimulus and task, older children show a differentiation in their level of sensitivity to asynchrony for different stimuli (varying in strength of association via experience). However, older children show a broad and non-differentiated sensitivity to asynchrony, similarly to young children, for different tasks. Only adults showed a differentiation due to task. Hence, multisensory perceptual narrowing and tuning seems to be a process extending late into childhood and perhaps adulthood. Knowing when different multisensory temporal mechanisms develop and specialize is essential in order to provide the most targeted and effective clinical and educational interventions aimed at children with deficits in these abilities (e.g. autistic and dyslexic children and those with language impairments; Francisco et al., 2017; Kaganovich, 2017; Stevenson et al., 2016; Stevenson, Siemann, Schneider, et al., 2014; Stevenson, Siemann, Woynaroski, et al., 2014; Wallace & Stevenson, 2014; Ye et al., 2017). For example, understanding how younger and older children's multisensory processing is impacted by the level of experience with different stimuli could inform clinical and educational interventions on what stimuli would be most effective for children of different ages. Having baseline measurements of key components in the multisensory integration process via the ICM model also provides a basis for determining more precisely in which ways atypical populations differ, and so inform the development of new interventions.

Acknowledgements

We are very grateful for the support and help that Dulwich Hamlet Junior School London has given us during the testing. Thanks to Miguel Angel Garcia Perez for the insightful and useful discussions about the ICM.

Funding

Supported by grant 220020240 from the James S. McDonnell Foundation 21st Century Science Scholar in Understanding Human Cognition Program.

References

Adams, W. J. (2016). The Development of Audio-Visual Integration for Temporal Judgements. *PLoS Comput Biol, 12*(4), e1004865.

Alcala-Quintana, R., & Garcia-Perez, M. A. (2013). Fitting model-based psychometric functions to simultaneity and temporal-order judgment data: MATLAB and R routines. *Behav Res Methods, 45*(4), 972-998.

Barutchu, A., Crewther, D. P., & Crewther, S. G. (2009). The race that precedes coactivation: development of multisensory facilitation in children. *Dev Sci, 12*(3), 464-473.

Barutchu, A., Danaher, J., Crewther, S. G., Innes-Brown, H., Shivdasani, M. N., & Paolini, A. G. (2010). Audiovisual integration in noise by children and adults. *J Exp Child Psychol, 105*(1-2), 38-50.

Barutchu, A., Fifer, J. M., Shivdasani, M. N., Crewther, S. G. & Paolini, A. G. (2019). The Interplay Between Multisensory Associative Learning and IQ in Children. *Child Dev,* 1-18. doi:10.1111/cdev.13210

Binder, M. (2015). Neural correlates of audiovisual temporal processing--comparison of temporal order and simultaneity judgments. *Neuroscience, 300*, 432-447.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis, 10*(4), 433-436.

Brandwein, A. B., Foxe, J. J., Russo, N. N., Altschuler, T. S., Gomes, H., & Molholm, S. (2011). The development of audiovisual multisensory integration across childhood and early adolescence: A high-density electrical mapping study. *Cerebral Cortex, 21*(5), 1042–1055. https://doi.org/10.1093/cercor/bhq170

Campbell, J. I. D., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. Behavior Research Methods, 44, 1255–1265. doi:10.3758/s13428-012-0186-0

Chambers, C., Sokhey, T., Gaebler-Spira, D., & Kording, K. P. (2017). The integration of probabilistic information during sensorimotor estimation is unimpaired in children with Cerebral Palsy. *PLoS One, 12*(11), e0188741.

Chen, Y. C., Shore, D. I., Lewis, T. L., & Maurer, D. (2016). The development of the perception of audiovisual simultaneity. *J Exp Child Psychol, 146*, 17-33.

Chen, Y. C., & Spence, C. (2017). Assessing the role of the “unity assumption” on multisensory integration: A review. *Frontiers in Psychology*, 8, 445. https://doi.org/10.3389/fpsyg.2017.00445

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

Downing, H. C., Barutchu, A., & Crewther, S. G. (2014). Developmental trends in the facilitation of multisensory objects with distractors. *Frontiers in psychology* **5**, 1559, doi: 10.3389/fpsyg.2014.01559.

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *J Vis, 7*(5), 7 1-14.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429-433.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods, 39*(2), 175-191.

Francisco, A. A., Jesse, A., Groen, M. A., & McQueen, J. M. (2017). A General Audiovisual Temporal Processing Deficit in Adult Readers With Dyslexia. *J Speech Lang Hear Res, 60*(1), 144-158.

Garcia-Perez, M. A., & Alcala-Quintana, R. (2012). On the discrepant results in synchrony judgment and temporal-order judgment tasks: a quantitative model. *Psychon Bull Rev, 19*(5), 820-846.

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Curr Biol, 18*(9), 694-698.

Gori, M., Sandini, G., & Burr, D. (2012). Development of visuo-auditory integration in space and time. *Front Integr Neurosci, 6*, 77.

Hillock-Dunn, A., & Wallace, M. T. (2012). Developmental changes in the multisensory temporal binding window persist into adolescence. *Dev Sci, 15*(5), 688-696.

Hillock, A. R., Powers, A. R., & Wallace, M. T. (2011). Binding of sights and sounds: age-related changes in multisensory temporal processing. *Neuropsychologia, 49*(3), 461-467.

Kaganovich, N. (2017). Sensitivity to Audiovisual Temporal Asynchrony in Children With a History of Specific Language Impairment and Their Peers With Typical Development: A Replication and Follow-Up Study. *J Speech Lang Hear Res, 60*(8), 2259-2270.

Lee, H., & Noppeney, U. (2011). Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proc Natl Acad Sci U S A, 108*(51), E1441-1450.

Lewkowicz, D. J. (2010). Infant perception of audio-visual speech synchrony. *Dev Psychol, 46*(1), 66-77.

Lewkowicz, D. J., & Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences*, 13, 470–478. doi:10.1016/j.tics.2009.08.004

Love, S., Petrini, K., Cheng, A., & Pollick, F. E. (2013). A psychophysical investigation of differences between synchrony and temporal order judgments. *PLoS One, 8*(1), e54798. doi:54710.51371/journal.pone.0054798

Love, S. A., Petrini, K., Pernet, C. R., Latinus, M., & Pollick, F. E. (2018). Overlapping but divergent neural correlates underpinning audiovisual synchrony and temporal order judgments. Frontiers in Human Neuroscience, 12(274).

Miyazaki, M., Kadota, H., Matsuzaki, K. S., Takeuchi, S., Sekiguchi, H., Aoyama, T., et al. (2016). Dissociating the neural correlates of tactile temporal order and simultaneity judgements. *Sci Rep, 6*, 23323.

Nardini, M., Begus, K., & Mareschal, D. (2012). Multisensory Uncertainty Reduction for Hand Localization in Children and Adults. *J Exp Psychol Hum Percept Perform*.

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Curr Biol, 18*(9), 689-693.

Noel, J.-P., Niear, M. D., Burg, E. V., Wallace, M. T. (2016). Audiovisual simultaneity judgment and rapid recalibration throughout the lifespan. *PLoS ONE*, 11(8). doi:10.1371/journal.pone.0161698

Parise, C.V., & Ernst., M.O. (2016). Correlation detection as a general mechanism for multisensory integration. *Nat. Commun.* 7: 11543.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis, 10*(4), 437-442.

Petrini, K., Dahl, S., Rocchesso, D., Waadeland, C. H., Avanzini, F., Puce, A., et al. (2009). Multisensory integration of drumming actions: musical expertise affects perceived audiovisual asynchrony. *Exp Brain Res, 198*(2-3), 339-352.

Petrini, K., Holt, S. P., & Pollick, F. (2010). Expertise with multisensory events eliminates the effect of biological motion rotation on audiovisual synchrony perception. *J Vis, 10*(5), 2.

Petrini, K., Pollick, F. E., Dahl, S., McAleer, P., McKay, L. S., Rocchesso, D., et al. (2011). Action expertise reduces brain activity for audiovisual matching actions: an fMRI study with expert drummers. *Neuroimage, 56*(3), 1480-1492.

Petrini, K., Remark, A., Smith, L., & Nardini, M. (2014). When vision is not an option: children's integration of auditory and haptic information is suboptimal. *Dev Sci, 17*(3), 376-387.

Petrini, K., Russell, M., & Pollick, F. (2009). When knowing can replace seeing in audiovisual integration of actions. *Cognition, 110*(3), 432-439.

Roder, B., Pagel, B., & Heed, T. (2013). The implicit use of spatial information develops later for crossmodal than for intramodal temporal processing. *Cognition, 126*(2), 301-306.

Sato, Y., Toyoizumi, T., & Aihara, K. (2007). Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural Comput, 19*(12), 3335-3355.

Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends Cogn Sci, 14*(9), 425-432.

Spence, C., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. *Curr Biol, 13*(13), R519-521.

Stanley, B. M., Chen, Y.-C., Lewis, T. L., Maurer, D., & Shore, D. I. (2019). Developmental changes in the perception of audiotactile simultaneity. *Journal of Experimental Child Psychology, 183*, 208–221. https://doi.org/10.1016/j.jecp.2019.02.006

Stein, B. E., Meredith, M. A., & Wallace, M. T. (1993). The visually responsive neuron and beyond: multisensory integration in cat and monkey. *Prog Brain Res, 95*, 79-90.

Stevenson, R. A., Segers, M., Ferber, S., Barense, M. D., Camarata, S., & Wallace, M. T. (2016). Keeping time in the brain: Autism spectrum disorder and audiovisual temporal processing. *Autism Res, 9*(7), 720-738.

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. *J Neurosci, 34*(3), 691-697.

Stevenson, R. A., Siemann, J. K., Woynaroski, T. G., Schneider, B. C., Eberly, H. E., Camarata, S. M., et al. (2014). Evidence for diminished multisensory integration in autism spectrum disorders. *J Autism Dev Disord, 44*(12), 3161-3167.

Stevenson, R. A., Altieri, N. A., Kim, S., Pisoni, D. B., James, T. W. (2010). Neural processing of asynchronous audiovisual speech perception. *NeuroImage*, 49: 3308–18.

Thomas, R., Nardini, M., & Mareschal, D. (2010). Interactions between "light-from-above" and convexity priors in visual development. *J Vis, 10*(8), 6.

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. *Percept Psychophys, 70*(6), 955-968.

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

Vatakis, A., Ghazanfar, A. A., & Spence, C. (2008). Facilitation of multisensory integration by the "unity effect" reveals that speech is special. *J Vis, 8*(9), 14 11-11.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: evaluating the "unity assumption" using audiovisual speech stimuli. *Percept Psychophys, 69*(5), 744-756.

Vatakis, A., & Spence, C. (2008). Evaluating the influence of the 'unity assumption' on the temporal perception of realistic audiovisual stimuli. *Acta Psychol (Amst), 127*(1), 12-23.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Atten Percept Psychophys, 72*(4), 871-884.

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.

Ye, Z., Russeler, J., Gerth, I., & Munte, T. F. (2017). Audiovisual speech integration in the superior temporal region is dysfunctional in dyslexia. *Neuroscience, 356,* 1-10.

Supplemental material and figures

Independent channels model (ICM) fitting

The starting values used to fit the data in the current study were the same as in Alcala-Quintana and Garcia-Perez (2013), LamBounds = [1/200 1/3] (i.e. the search is restricted to distributions of arrival latencies whose standard deviation ranges from 3 to 200 ms, as values outside this range are unlikely); TauBounds = [-Inf Inf]; DeltaBounds = [0 Inf]; LamVStart = [1/70 1/10]; LamAStart = [1/70 1/10]; TauStart = [-70 70]; DeltaStart = [20 150]; ErrStart = [.05] (always bounded between 0 and 1); BiaStart = [.5] (always bounded between 0 and 1); Model = 1; SampleSize = 1500 (number n of bootstrap samples to be generated).

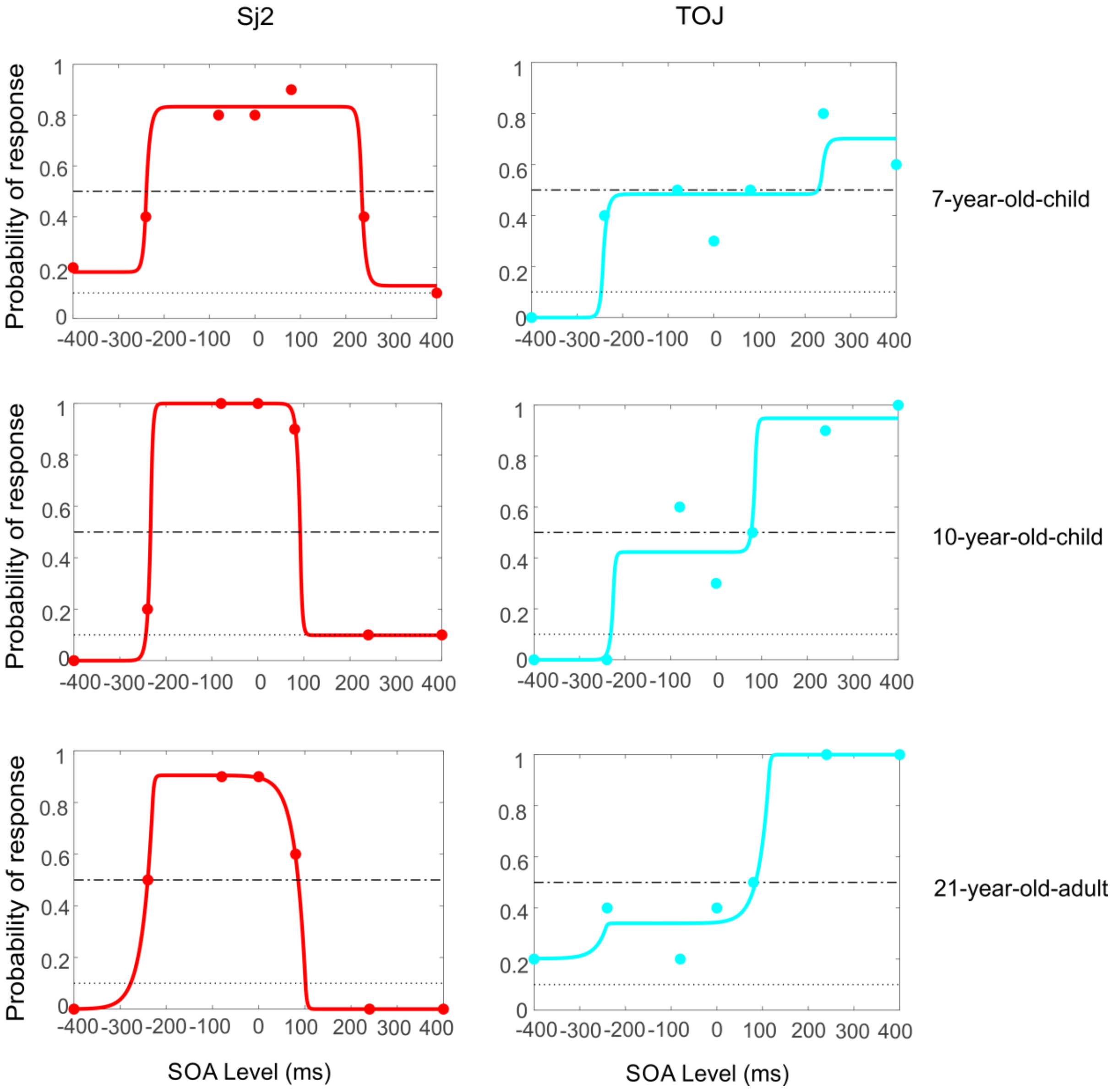

Fig. 1S. The fit of the ICM to the data for the three examples in Fig. 2 (see main manuscript). Note: the model fit shown in Fig. 1S and the resulting performance function shown in Fig. 2 in the main manuscript are not the same, e.g., the top range of the ICM fit for the SJ2 (in red) in the Fig.1S for the 7-year-old child does not reach 1 while the resulting fit for the performance measures in Fig. 2 does. This is because performance measures (point of subjective simultaneity -PSS- and range) are extracted by setting all error parameters to 0, which explains potential differences in psychometric functions across the two types of plot. The reason for excluding error parameters is that they are not part of a description of the sensory/decisional determinants of performance, as they only inform about the probability of committing response errors when reporting timing judgments. This is described in the Alcala-Quintana and Garcia-Perez (2013). This is also the reason why data points are not plotted in Fig. 2 for performance measures because this figure only aims at indicating those measures given the estimated parameters. This is intentional in the ICM, as PerformanceMeasures only takes model parameters (and task) as arguments.

We used Model 1, as it is the most general model with the largest number of free parameters for response errors in all tasks (Alcala-Quintana and Garcia-Perez, 2013). Through this model fitting to the synchrony judgement and temporal order judgement data we obtained parameter measures for sensory (λ_a, λ_v and τ), decisional (δSJ, δTOJ and ξ) processes, which were then inputted into the Matlab routines to generate PerformanceMeasures of PSS and ASW and plot the resulting figures (see Alcala-Quintana and Garcia-Perez, 2013). See Fig. 1S for an example of individual model fitting for each age group.

Comparing ICM PSS estimates for peak and middle-point

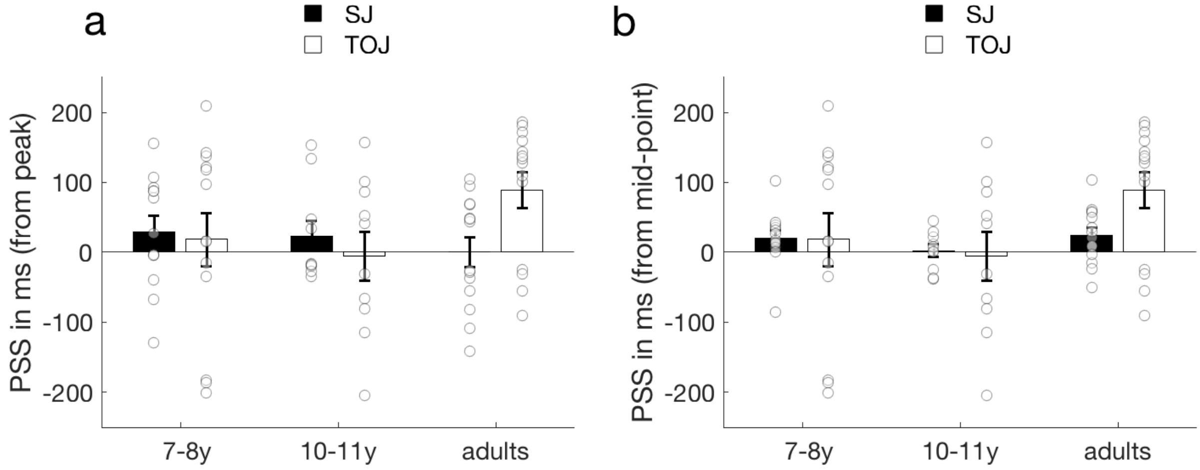

Fig. 2S. a) Interaction between age and task for the point of subjective simultaneity (PSS) estimates (from peak estimation as returned by ICM, see Fig. 2 in the main manuscript) collapsed across stimulus type (please refer to Fig. 3 and 4 in the main manuscript for the presentation of all data). b) Interaction between age and task for the PSS estimates (from middle point estimation returned by ICM, see Fig. 2 in the main manuscript) collapsed across stimulus type. Paired-samples t-tests, Bonferroni corrected, showed that the PSS results for age x task were very similar when the PSS for the synchrony judgement task was estimated from the peak or the middle-point of the distribution. That is, even for the middle-point estimation, 7-8 year-old children (*t*(12)= -.054, *p* = .958), and 10-11 year-old children (*t*(9)= -.222, *p* = .829) had similar PSSs for the two tasks, while adults *t*(13)= -2.820, *p* = .042, Cohen's *d* = 0.75) did not.

Response errors

We examined whether the response errors varied with age by analyzing whether participants misreported ‘‘simultaneous” (in synchrony judgement tasks) or “visual-first” (in temporal order judgement tasks) in the auditory-leading trials and ‘‘simultaneous” (in synchrony judgement tasks) or “auditory-first” (in temporal order judgement tasks) in the visual-leading trials, as well as the ‘‘not simultaneous” in the 0 ms trials. For the error parameters common to both tasks we carried out a mixed factorial ANOVA with age (7-8 years, 10-11 years, and adults) as between-subjects factor, and task (synchrony judgement and temporal order judgement) and stimuli (beep-flash and face-voice) as within-subjects factors.

For the auditory-leading trials this analysis revealed a significant main effect of stimulus (*F*(1, 34) = 16.809, *p* < .001, η^2= .331), with less errors for flash-beep (*Mean* = .05, *SD* = .08) than face-voice (*Mean* = .16, *SD* = .18) stimuli, of task (*F*(1, 34) = 5.939, *p* = .020, η^2= .149), with synchrony judgement (*Mean* = .09, *SD* = .12) having less errors than temporal order judgement (*Mean* = .13, *SD* = .13), of age (*F*(2, 34) = 4.417, *p* = .020, η^2= .206). We also found an interaction between task and stimulus (*F*(2, 34) = 11.863, *p* = .002, η^2= .259). No other main factor or interaction reached significance (*F* ≤ 2.412, *p* ≥ .105). The main effect of age was the results of a decrease in errors for older children (*Mean* = .07, *SD* = .09) and adults (*Mean* = .07, *SD* = .08) compared to younger children (*Mean* = .17, *SD* = .13). Bonferroni corrected independent-samples t-tests showed that only the difference in error between younger children and adults was significant (7-8-year-old vs adults: *t*(25)= 2.667, *p*

= .039, Cohen's *d* = 0.92; 7-8-year-old vs 10-11-year-old: *t*(21)= 2.075, *p* = .126; 10-11-year-old vs adults: *t*(22)= 0.270, *p* = .790). The significant interaction between task and stimulus was driven by a difference in errors for the stimuli for the temporal order judgement task (flash-beep vs face-voice: *t*(36)= 4.812, *p* < .001, Cohen's *d* = 0.79) but not for the synchrony judgement (flash-beep vs face-voice: *t*(36)= -1.793, *p* = .081). Participants made more errors for the face-voice (*Mean* = .21, *SD* = .22) than flash-beep (*Mean* = .04, *SD* = .08) stimulus in the temporal order judgement task.

For the visual-leading trials this analysis revealed a significant main effect of age (*F*(1, 34) = 4.049, *p* = .026, η^2= .192), with younger children making a higher number of errors (*Mean* = .16, *SD* = .09) than older children (*Mean* = .06, *SD* = .05) and adults (*Mean* = .09, *SD* = .08). No other main factor or interaction reached significance ($F \leq 4.025$, $p \geq .053$). Bonferroni corrected independent-samples t-tests showed that only the difference in error between younger and older children was significant (7-8-year-old vs 10-11-year-old: *t*(21)= 2.802, *p* = .033, Cohen's *d* = 1.37; 7-8-year-old vs adults: *t*(25)= 1.895, *p* = .210; 10-11-year-old vs adults: *t*(22)= -0.976, *p* = .340).

For the synchrony trials (0 ms of delay between auditory and visual information) we carried out a mixed factorial ANOVA with age (7-8 years, 10-11 years, and adults) as between-subjects factor and stimuli (beep-flash and face-voice) as within-subjects factor. This because this type of error could have occurred only for synchrony judgement task. No significant effect was found for this response error ($F \leq 1.812$, $p \geq .187$).

Since there was no significant interaction between age and task or age and stimuli, these response error results cannot fully explain the age-related PSS results. For example, younger children had a higher number of errors than older children, however the two groups of children showed a similar lack of difference between PSS measures for synchrony judgement and temporal order judgement tasks.

Average data distribution for the three age-groups and task (synchrony judgement and temporal order judgement) and stimulus (flash-beep and face-voice) condition

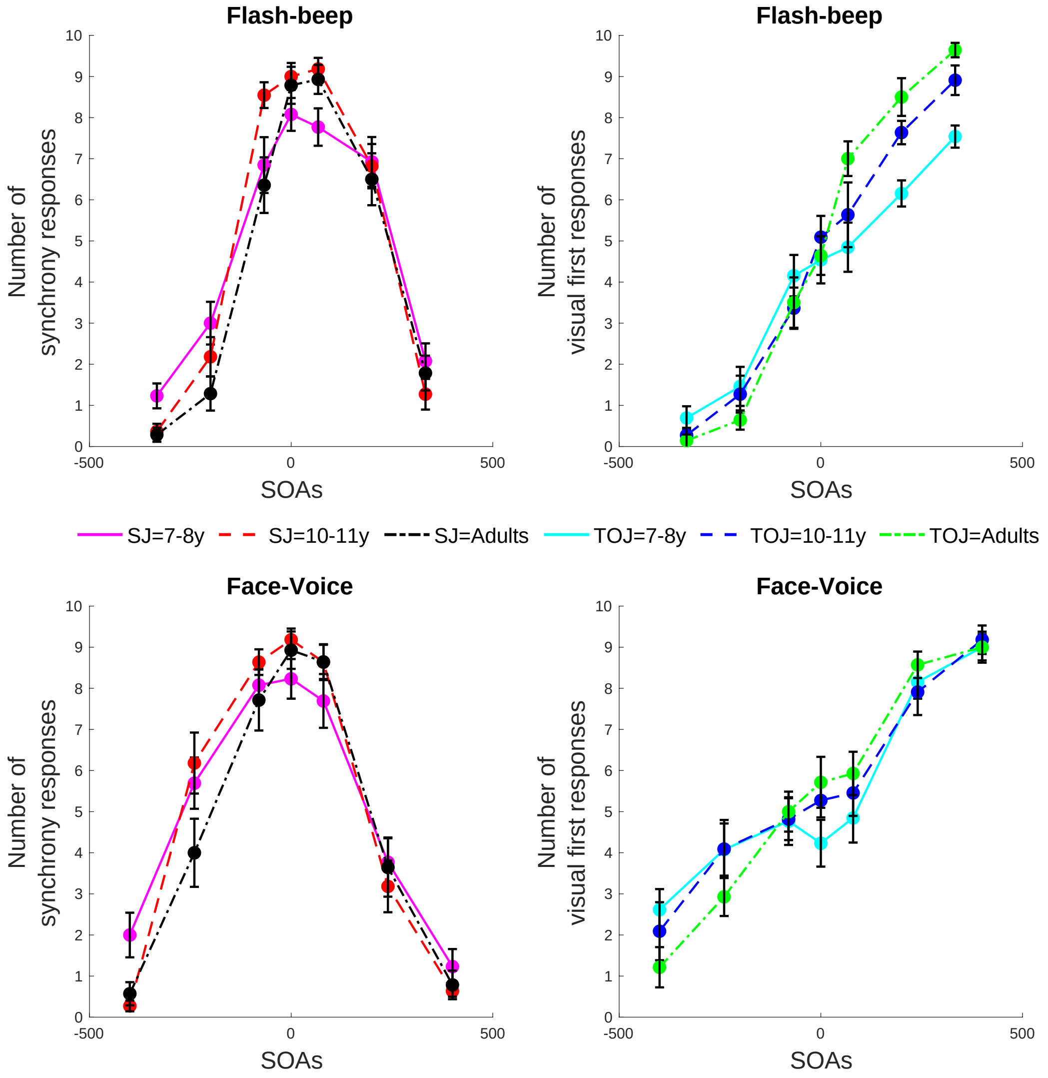

Fig. 3S. Average synchrony responses as a function of stimulus onset asynchrony (SOA) for the 7-8 years old children (magenta and cyan), the 11-10 years old children (red and blue) and the adults (black and green). The top panels represent the average responses for the flash-beep stimuli while the bottom panels for the face-voice stimuli. The left panels represent the average responses for the synchrony judgement (SJ) task while the right panels for the temporal order judgement (TOJ) task. NOTE: the independent channel model (ICM) was fitted to the individual data to obtain the point of subjective simultaneity (PSS), and the audiovisual synchrony window (ASW) and parameter estimates analysed and discussed in the main manuscript, it was not fitted to the average data represented here.

Comparison between ICM and Gaussian psychometric ASW estimates

Fig. 4S shows the psychometric Gaussian fitting for the three participants used as examples for the estimation of audiovisual synchrony window (ASW) via ICM. To this end Fig. 4S can be compared to Fig. 2 in the main manuscript and Fig.1S here.

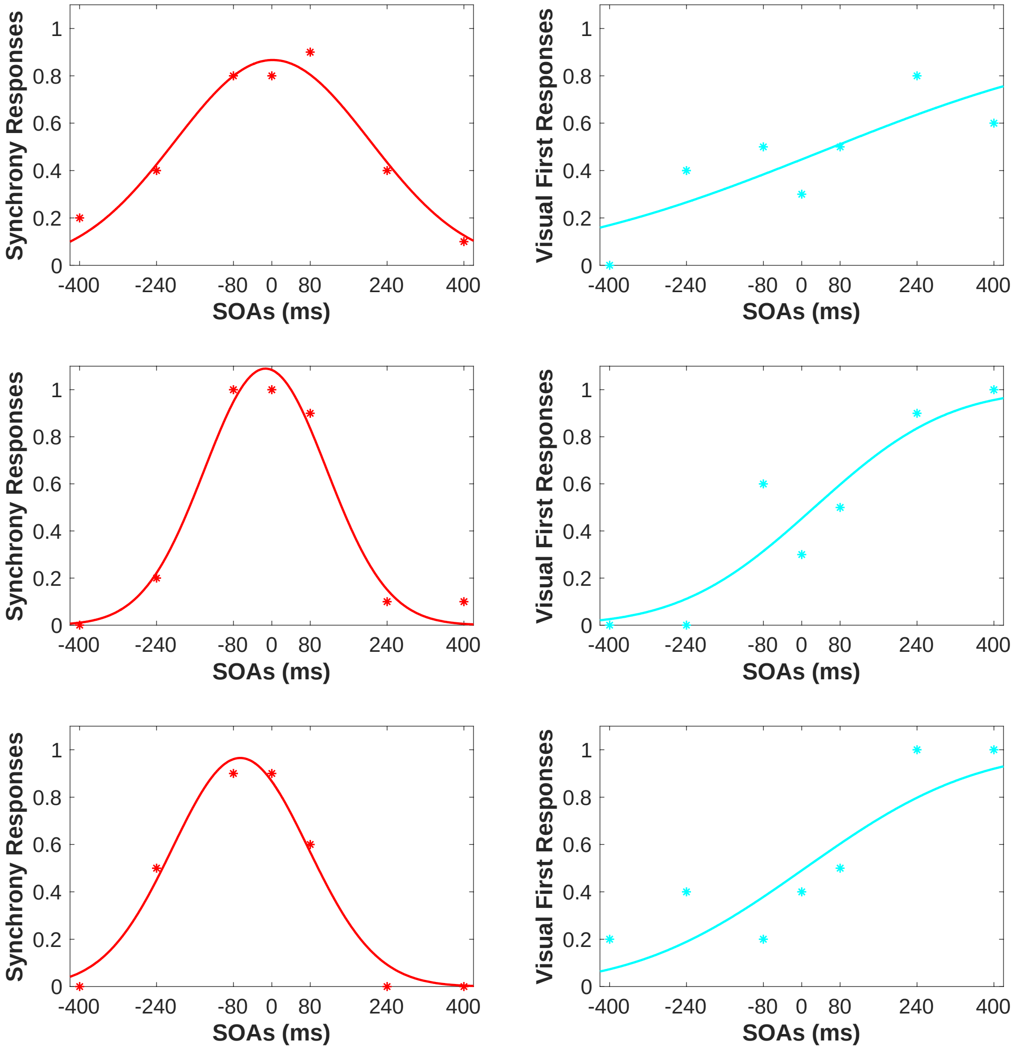

Fig. 4S. The individual psychometric fitting (using normal and cumulative functions) results for a 7-year-old child (top panels), a 10-year-old child (middle panels) and an adult (bottom panels) in the face-voice condition. The left panels describe the results for the synchrony judgement (SJ) task (red line), while the right panels for temporal order judgement (TOJ) task (cyan line). The y axis presented the proportion of synchrony (for synchrony judgement) or visual first (for temporal order judgement) responses. For comparison to the ICM fitting for these same participants and conditions please see Fig. 2 in the main manuscript and Fig. 1S here.

Furthermore, to compare the ASW estimates returned by the ICM with those returned by more commonly used psychometric functions (normal and cumulative Gaussian) we carried out a correlation analysis between the ASW estimates that met inclusion for both methods, e.g., chi-square for ICM and R-square for Gaussian fit. Four separate correlations were carried out for the four conditions (2 tasks x 2 stimulus conditions). Hence, a bivariate correlation was carried out 1) for 36 data points for flash-beep stimulus for temporal order judgment task and returned a significant positive correlation ($r_t = .517, p < .001$); 2) for 37 data points for flash-beep stimulus for synchrony judgment task and returned a significant positive correlation ($r_t = .369, p = .001$); 3) for 34 data points for face-voice stimulus for temporal order judgment task and returned a positive correlation trend ($r_t = .176, p = .071$); 4) for 37 data points for face-voice stimulus for synchrony judgment task and returned a significant positive correlation ($r_t = .255, p = .013$).

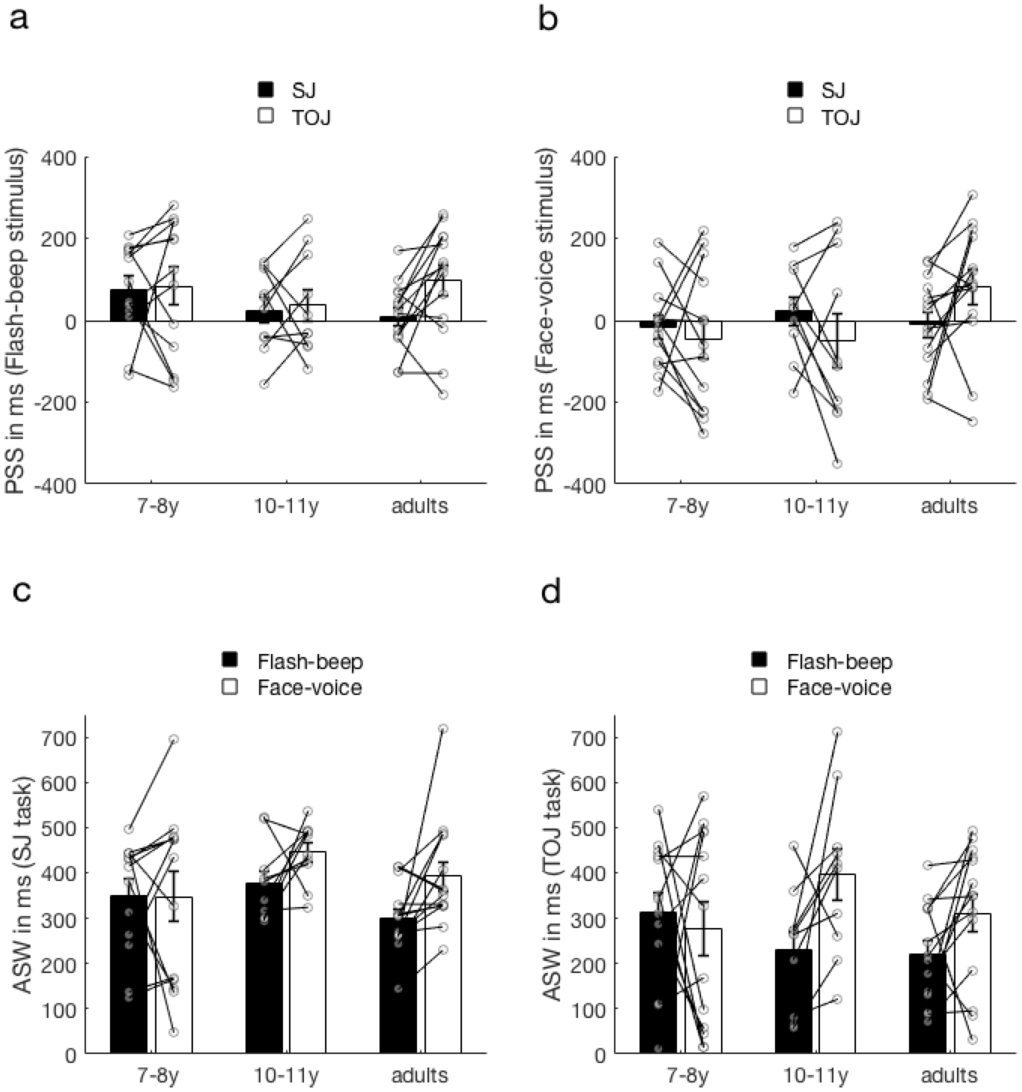

Fig. 5S. Effect of age on the estimates returned by the ICM (independent channels model). (a) and (b) Interaction between age and task for the synchrony judgment (SJ) and temporal order judgment (TOJ) PSS estimates (from peak) for flash-beep stimuli on the left panel and for face-voice stimuli on the right panel. (c) and (d) Interaction between age and stimuli for the flash-beep and face-voice ASW (audiovisual synchrony window) for synchrony judgment task on the left panel and temporal order judgment task on the right panel. The bars represent the group mean while the error bars the standard error of the mean. The circles represent the individual data. The lines connect the estimates for each individual for the two tasks in the top panels and the two stimuli in the bottom panels.

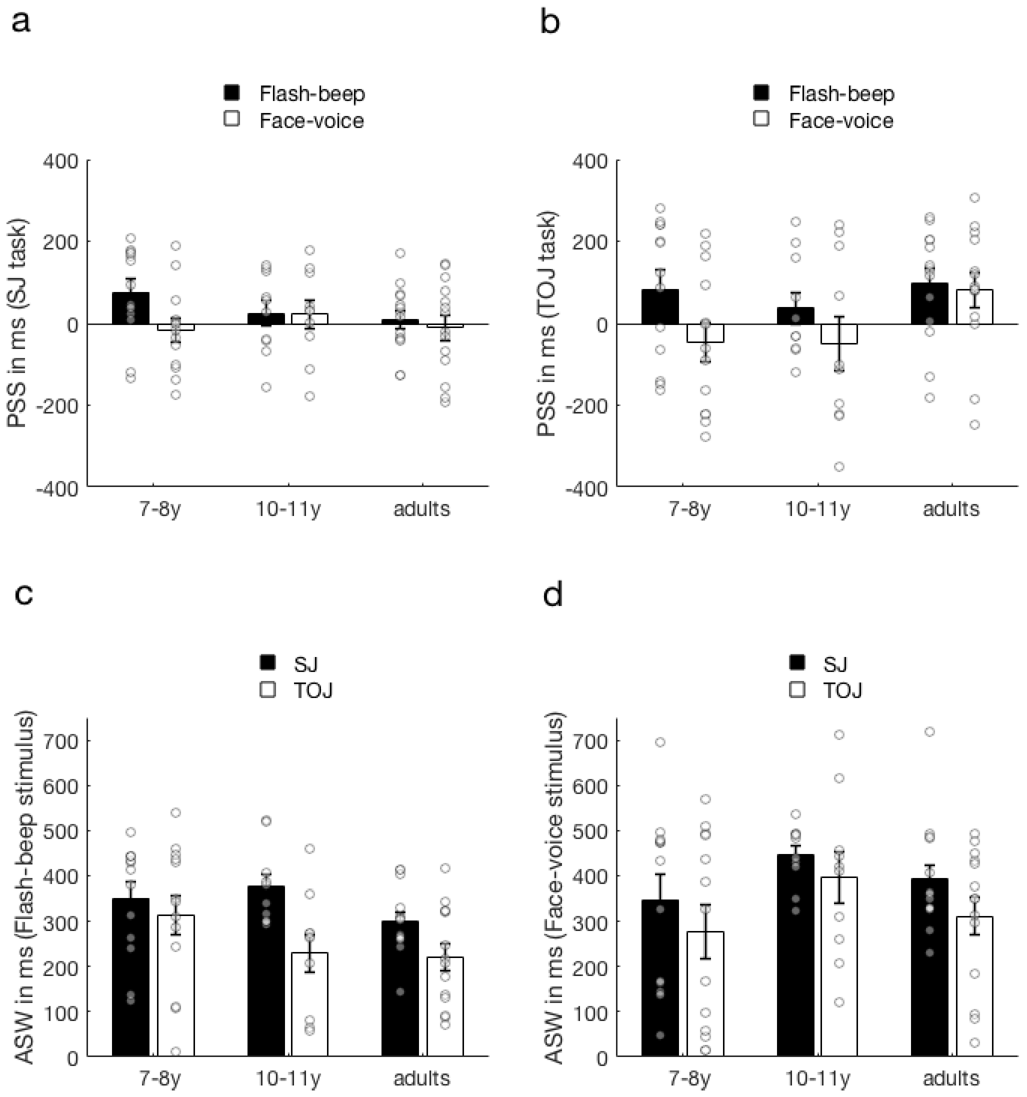

Fig. 6S. Effect of age on the estimates returned by the ICM (independent channels model). (a) and (b) Interaction between age and stimuli for the Flash-beep and Face-voice point of subjective simultaneity (PSS) estimates (from peak) for synchrony judgment (SJ) task on the left panel and for temporal order judgment (TOJ) task on the right panel. (c) and (d) Interaction between age and tasks for the SJ and TOJ ASW (audiovisual synchrony window) for Flash-beep stimulus on the left panel and Face-voice stimulus on the right panel. The bars represent the group mean while the error bars the standard error of the mean. The circles represent the individual data.

Data (ICM estimates and parameters)

PSS =point of subjective simultaneity; SJ = synchrony judgement task; TOJ =temporal order judgement task; BF = beep and flash stimulus; FV = face and voice stimulus.

PSS_TOJ_BF	PSS_SJ_BF	PSS_TOJ_FV	PSS_SJ_FV	Age group
199.67	174.20	218.28	10.23	7-8y
241.47	-134.87	1.56	55.65	7-8y
282.51	161.96	-88.45	-107.17	7-8y
239.22	154.08	-3.02	3.30	7-8y
-141.40	94.16	-222.67	-99.35	7-8y
196.81	177.48	-164.05	-2.43	7-8y
247.29	208.17	-277.61	-33.37	7-8y
-8.27	168.93	-61.68	142.66	7-8y
-148.16	44.82	-223.30	-52.44	7-8y
-65.34	23.84	94.94	190.59	7-8y
86.60	8.31	189.09	-17.34	7-8y
122.43	38.50	161.85	-174.24	7-8y
-163.38	-120.37	-239.20	-138.69	7-8y
195.75	26.03	-111.19	41.91	10-11y
-64.97	127.25	238.54	178.90	10-11y
-31.95	-157.14	-100.40	124.28	10-11y
-30.55	-42.12	-198.12	1.42	10-11y
13.27	134.43	188.65	133.16	10-11y
247.97	139.92	67.07	-178.21	10-11y
-117.97	65.54	223.21	29.70	10-11y
-59.43	-37.55	-349.52	-31.26	10-11y
159.49	56.81	-222.60	-111.47	10-11y
65.33	-67.04	-227.31	32.46	10-11y
135.55	26.93	-246.14	-191.39	adult
65.24	-22.26	306.58	110.79	adult
127.86	71.37	127.07	26.76	adult
-128.94	-128.30	78.85	54.89	adult
205.02	45.11	82.48	145.12	adult
204.67	-38.06	-2.76	-19.08	adult
141.83	-126.82	220.93	-155.31	adult
252.21	97.45	91.64	40.07	adult
184.69	169.07	84.04	-30.33	adult
258.62	39.69	16.59	-91.14	adult
-18.56	66.81	236.82	142.75	adult
5.88	16.75	-186.57	78.81	adult
114.60	-35.95	204.87	-181.28	adult
-180.09	-43.64	118.56	-67.46	adult

ASW = audiovisual synchrony window; SJ = synchrony judgement task; TOJ =temporal order judgement task; BF = beep and flash stimulus; FV = face and voice stimulus.

ASW_TOJ_BF	ASW_SJ_BF	ASW_TOJ_FV	ASW_SJ_FV	Age group
429.57	429.60	491.48	497.37	7-8y
540.43	495.88	98.34	694.11	7-8y
461.25	264.79	16.27	471.63	7-8y
309.11	314.09	43.83	145.95	7-8y
11.29	444.60	488.83	474.31	7-8y
446.73	434.51	326.20	475.38	7-8y
436.39	442.76	435.71	47.06	7-8y
351.31	414.80	13.39	479.35	7-8y
343.94	442.35	570.52	326.99	7-8y
244.34	124.43	57.11	165.50	7-8y
286.06	137.51	387.03	169.04	7-8y
108.07	240.69	168.61	434.18	7-8y
111.33	380.58	509.15	136.25	7-8y
266.02	523.57	420.27	348.83	10-11y
273.70	518.39	441.74	484.37	10-11y
82.79	406.22	121.26	492.32	10-11y
207.77	294.25	411.11	537.19	10-11y
359.31	380.10	456.37	429.84	10-11y
459.07	340.44	260.16	493.16	10-11y
63.41	386.64	614.50	419.17	10-11y
273.95	300.36	711.27	489.11	10-11y
262.88	317.54	311.48	324.47	10-11y
56.96	301.10	208.89	438.84	10-11y
342.09	331.81	348.91	331.20	adult
139.29	142.96	376.93	229.48	adult
130.58	268.90	491.95	281.52	adult
215.93	263.53	349.73	328.30	adult
88.49	243.31	95.54	405.46	adult
206.88	411.76	83.15	361.25	adult
319.40	402.55	474.62	491.61	adult
322.05	309.89	31.21	327.65	adult
248.53	306.00	313.06	329.36	adult
72.03	263.03	185.07	484.08	adult
176.36	266.10	297.18	351.35	adult
322.15	259.95	433.69	719.15	adult
92.59	304.93	450.56	484.72	adult
415.72	414.36	428.06	364.38	adult

Delta = decision parameter (δ); SJ = synchrony judgement task; TOJ =temporal order judgement task; BF = beep and flash stimulus; FV = face and voice stimulus.

Delta_TOJ_BF	Delta_SJ_BF	Delta_TOJ_FV	Delta_SJ_FV	Age group
216.73	215.71	245.74	248.69	7-8y
277.48	248.11	53.28	347.06	7-8y
232.53	132.49	87.24	235.82	7-8y
156.18	157.04	21.91	72.98	7-8y
170.02	222.30	244.42	237.15	7-8y
222.96	217.73	163.71	237.69	7-8y
222.40	221.38	258.97	55.48	7-8y
175.69	207.47	65.12	239.68	7-8y
174.91	221.17	299.22	164.56	7-8y
131.32	62.28	54.10	83.80	7-8y
143.03	68.75	193.61	84.52	7-8y
48.68	123.79	208.69	217.33	7-8y
172.76	191.46	255.43	80.73	7-8y
135.14	261.79	210.36	175.56	10-11y
136.91	259.19	252.52	243.24	10-11y
10.35	203.28	64.69	246.19	10-11y
108.71	147.15	205.55	268.59	10-11y
179.14	190.84	228.50	214.93	10-11y
234.66	170.94	134.26	246.68	10-11y
143.68	193.32	307.29	209.58	10-11y
137.03	150.18	355.76	244.55	10-11y
131.60	158.77	155.91	162.23	10-11y
31.24	150.63	144.18	220.08	10-11y
171.14	165.91	178.65	165.60	adult
69.89	71.48	193.27	114.70	adult
65.55	134.45	241.70	154.79	adult
107.98	131.80	174.78	164.34	adult
102.14	121.97	94.89	203.07	adult
114.66	205.90	118.32	180.70	adult
160.16	201.28	237.84	245.80	adult
164.07	154.94	167.88	163.82	adult
137.94	153.05	160.28	164.68	adult
140.90	132.10	92.54	242.04	adult
88.35	133.05	227.53	176.40	adult
160.78	130.72	216.92	359.57	adult
60.00	152.47	227.60	242.36	adult
207.86	207.18	214.03	182.19	adult

Lambda_A or sensory parameter for auditory cue λ_a; BF = beep and flash stimulus; FV = face and voice stimulus. This parameter was common to the two tasks.

BF	FV	Age group
0.33	0.29	7-8y
0.01	0.12	7-8y
0.33	0.09	7-8y
0.17	0.21	7-8y
0.33	0.07	7-8y
0.26	0.33	7-8y
0.32	0.01	7-8y
0.33	0.33	7-8y
0.29	0.01	7-8y
0.04	0.33	7-8y
0.30	0.17	7-8y
0.01	0.01	7-8y
0.01	0.33	7-8y
0.13	0.33	10-11y
0.27	0.24	10-11y
0.01	0.32	10-11y
0.02	0.29	10-11y
0.20	0.16	10-11y
0.10	0.01	10-11y
0.04	0.20	10-11y
0.05	0.13	10-11y
0.04	0.17	10-11y
0.02	0.33	10-11y
0.03	0.04	adult
0.15	0.16	adult
0.21	0.33	adult
0.02	0.33	adult
0.02	0.33	adult
0.02	0.01	adult
0.04	0.07	adult
0.31	0.28	adult
0.32	0.16	adult
0.02	0.08	adult
0.31	0.31	adult
0.01	0.29	adult
0.03	0.04	adult
0.11	0.28	adult

Lambda_V or sensory parameter for visual cue λ_v; BF = beep and flash stimulus; FV = face and voice stimulus. This parameter was common to the two tasks.

BF	FV	Age group
0.01	0.25	7-8y
0.19	0.12	7-8y
0.02	0.23	7-8y
0.07	0.17	7-8y
0.17	0.18	7-8y
0.01	0.33	7-8y
0.02	0.33	7-8y
0.01	0.08	7-8y
0.24	0.20	7-8y
0.30	0.02	7-8y
0.22	0.30	7-8y
0.07	0.33	7-8y
0.17	0.01	7-8y
0.18	0.01	10-11y
0.16	0.01	10-11y
0.27	0.01	10-11y
0.33	0.22	10-11y
0.01	0.02	10-11y
0.01	0.23	10-11y
0.02	0.07	10-11y
0.33	0.17	10-11y
0.05	0.28	10-11y
0.33	0.01	10-11y
0.04	0.33	adult
0.33	0.18	adult
0.20	0.01	adult
0.21	0.02	adult
0.17	0.01	adult
0.33	0.22	adult
0.33	0.33	adult
0.28	0.04	adult
0.02	0.09	adult
0.33	0.17	adult
0.31	0.01	adult
0.04	0.21	adult
0.26	0.33	adult
0.18	0.31	adult

Tau or sensory parameter τ; BF = beep and flash stimulus; FV = face and voice stimulus. This parameter was common to the two tasks.

BF	FV	Age group
29.63	8.68	7-8y
-84.90	-50.41	7-8y
-45.06	-0.58	7-8y
-84.68	4.31	7-8y
-25.78	-4.39	7-8y
22.96	2.42	7-8y
-18.42	-19.47	7-8y
20.82	2.33	7-8y
-24.50	-94.82	7-8y
-70.90	-115.90	7-8y
3.17	-5.10	7-8y
-128.96	-28.11	7-8y
-50.41	214.93	7-8y
-65.94	122.81	10-11y
-59.58	47.90	10-11y
-27.00	99.99	10-11y
-85.73	36.06	10-11y
37.13	37.31	10-11y
-6.56	-43.79	10-11y
-9.91	75.95	10-11y
-74.64	2.05	10-11y
-76.02	70.32	10-11y
-67.11	174.97	10-11y
-52.32	63.85	adult
-4.47	-116.40	adult
-67.30	122.44	adult
24.26	94.60	adult
-143.80	44.00	adult
-146.57	-138.50	adult
-32.33	-5.79	adult
-89.43	87.29	adult
-34.21	74.62	adult
-160.23	-0.78	adult
-66.21	21.01	adult
-72.18	-22.81	adult
-85.47	-6.78	adult
-11.84	57.06	adult

Bias measure for temporal order judgement task or ξ; BF = beep and flash stimulus; FV = face and voice stimulus. This parameter was common to the two tasks.

BF	FV	Age group
0.43	0.50	7-8y
0.34	0.71	7-8y
0.41	0.80	7-8y
0.37	0.50	7-8y
0.83	0.51	7-8y
0.50	0.67	7-8y
0.37	0.69	7-8y
0.53	1.00	7-8y
0.73	0.65	7-8y
0.70	1.00	7-8y
0.50	0.43	7-8y
0.04	0.17	7-8y
0.79	0.56	7-8y
0.30	0.55	10-11y
0.55	0.32	10-11y
1.00	0.79	10-11y
0.64	0.50	10-11y
0.54	0.46	10-11y
0.39	0.34	10-11y
0.90	0.47	10-11y
0.52	0.56	10-11y
0.46	0.58	10-11y
0.00	0.82	10-11y
0.47	0.66	adult
0.40	0.27	adult
0.40	0.49	adult
0.53	0.49	adult
0.13	0.28	adult
0.24	0.74	adult
0.43	0.40	adult
0.27	0.17	adult
0.32	0.30	adult
0.00	0.50	adult
0.60	0.26	adult
0.49	0.57	adult
0.21	0.33	adult
0.51	0.50	adult

Error measures for auditory (A) leading; SJ = synchrony judgement task; TOJ =temporal order judgement task; BF = beep and flash stimulus; FV = face and voice stimulus.

BF_SJ_Aerror	BF_TOJ_Aerror	FV_SJ_Aerror	FV_TOJ_Aerror	Age group
0.11	0.00	0.00	0.20	7-8y
0.00	0.21	0.50	0.60	7-8y
0.15	0.00	0.00	0.15	7-8y
0.20	0.20	0.60	0.47	7-8y
0.00	0.00	0.18	0.00	7-8y
0.00	0.00	0.00	0.10	7-8y
0.10	0.00	0.46	0.56	7-8y
0.12	0.00	0.00	0.33	7-8y
0.00	0.05	0.08	0.00	7-8y
0.16	0.00	0.60	0.45	7-8y
0.50	0.20	0.30	0.30	7-8y
0.07	0.14	0.02	0.42	7-8y
0.05	0.00	0.30	0.19	7-8y
0.00	0.00	0.12	0.84	10-11y
0.00	0.05	0.00	0.00	10-11y
0.00	0.29	0.10	0.55	10-11y
0.01	0.00	0.00	0.00	10-11y
0.00	0.00	0.00	0.00	10-11y
0.20	0.10	0.03	0.16	10-11y
0.00	0.00	0.00	0.10	10-11y
0.00	0.00	0.00	0.10	10-11y
0.00	0.05	0.00	0.00	10-11y
0.00	0.00	0.00	0.29	10-11y
0.00	0.00	0.00	0.20	adult
0.00	0.00	0.00	0.10	adult
0.00	0.00	0.00	0.11	adult
0.00	0.00	0.00	0.00	adult
0.00	0.00	0.00	0.00	adult
0.00	0.11	0.00	0.30	adult
0.00	0.02	0.00	0.00	adult
0.10	0.20	0.11	0.09	adult
0.00	0.00	0.00	0.00	adult
0.35	0.00	0.20	0.45	adult
0.05	0.00	0.00	0.00	adult
0.00	0.00	0.30	0.55	adult
0.00	0.00	0.00	0.00	adult
0.10	0.00	0.20	0.10	adult

Error measures for visual (V) leading; SJ = synchrony judgement task; TOJ =temporal order judgement task; BF = beep and flash stimulus; FV = face and voice stimulus.

BF_SJ_Verror	BF_TOJ_Verror	FV_SJ_Verror	FV_TOJ_Verror	Age group
0.00	0.27	0.00	0.10	7-8y
0.17	0.45	0.00	0.20	7-8y
0.27	0.25	0.00	0.00	7-8y
0.40	0.20	0.35	0.27	7-8y
0.30	0.20	0.11	0.32	7-8y
0.00	0.11	0.00	0.15	7-8y
0.13	0.40	0.31	0.00	7-8y
0.00	0.13	0.00	0.00	7-8y
0.00	0.10	0.00	0.00	7-8y
0.30	0.20	0.19	0.09	7-8y
0.65	0.25	0.15	0.10	7-8y
0.00	0.25	0.00	0.00	7-8y
0.00	0.18	0.46	0.10	7-8y
0.06	0.10	0.05	0.00	10-11y
0.04	0.00	0.00	0.15	10-11y
0.30	0.23	0.08	0.07	10-11y
0.00	0.05	0.00	0.00	10-11y
0.07	0.04	0.00	0.00	10-11y
0.23	0.00	0.00	0.39	10-11y
0.00	0.00	0.10	0.00	10-11y
0.00	0.00	0.10	0.10	10-11y
0.00	0.00	0.10	0.05	10-11y
0.00	0.20	0.00	0.01	10-11y
0.19	0.10	0.00	0.00	adult
0.05	0.00	0.00	0.10	adult
0.00	0.00	0.04	0.00	adult
0.25	0.00	0.00	0.00	adult
0.00	0.00	0.08	0.01	adult
0.00	0.00	0.00	0.00	adult
0.10	0.00	0.00	0.00	adult
0.20	0.10	0.15	0.34	adult
0.17	0.00	0.00	0.16	adult
0.29	0.08	0.10	0.10	adult
0.50	0.15	0.00	0.18	adult
0.30	0.00	0.39	0.20	adult
0.00	0.05	0.00	0.00	adult
0.30	0.00	0.45	0.05	adult

Error measures for synchrony; BF = beep and flash stimulus; FV = face and voice stimulus. This error was specific to synchrony judgement task.

BF_Serror	FV_Serror	group
0.15	0.07	7-8y
0.20	0.20	7-8y
0.00	0.17	7-8y
0.20	0.00	7-8y
0.10	0.17	7-8y
0.00	0.03	7-8y
0.17	0.00	7-8y
0.06	0.07	7-8y
0.42	0.00	7-8y
0.28	0.08	7-8y
0.00	0.00	7-8y
0.20	0.00	7-8y
0.00	0.00	7-8y
0.10	0.00	10-11y
0.10	0.00	10-11y
0.14	0.00	10-11y
0.00	0.10	10-11y
0.00	0.21	10-11y
0.24	0.20	10-11y
0.00	0.07	10-11y
0.00	0.07	10-11y
0.05	0.00	10-11y
0.05	0.09	10-11y
0.00	0.09	adult
0.00	0.00	adult
0.40	0.06	adult
0.00	0.00	adult
0.00	0.00	adult
0.03	0.00	adult
0.16	0.07	adult
0.23	0.00	adult
0.00	0.11	adult
0.00	0.00	adult
0.00	0.00	adult
0.04	0.24	adult
0.00	0.03	adult
0.14	0.10	adult

Chi-square results (p values) for goodness of fit for the ICM. BF=Beep and flash stimulus; FV=Face and voice stimulus; the chi square was common to synchrony and temporal order judgment tasks as it was calculated with the joint model.

BF	FV	
0.66	0.59	7-8y
0.21	0.2	7-8y
0.35	0.15	7-8y
0.21	0.02	7-8y
0.1	0.18	7-8y
0.69	0.21	7-8y
0.27	0.27	7-8y
0.22	0.02	7-8y
0.11	0.05	7-8y
0.71	0.05	7-8y
0.1	0.02	7-8y
0.76	0.66	7-8y
0.21	0.73	7-8y
0.81	0.06	10-11y
0.32	0.27	10-11y
0.26	0.06	10-11y
0.37	0.85	10-11y
0.36	0.05	10-11y
0.63	0.08	10-11y
0.68	0.4	10-11y
0.77	0.19	10-11y
0.26	0.33	10-11y
0.08	0.17	10-11y
0.49	0.74	adult
0.58	0.31	adult
0.84	0.17	adult
0.33	0.6	adult
0.61	0.13	adult
0.14	0.14	adult
0.06	0.48	adult

0.16	0.37		adult
0.33	0.13		adult
0.27	0.98		adult
0.7	0.28		adult
0.87	0.77		adult
0.57	0.33		adult
0.39	0.04		adult