Published in final edited form as:
Child Dev. 2014 March ; 85(2): 685–694. doi:10.1111/cdev.12142.

The Audio-Visual Temporal Binding Window Narrows In Early Childhood

David J. Lewkowicz and
Florida Atlantic University

Ross Flom
Brigham Young University

Abstract

Binding is key in multisensory perception. This study investigated the audio-visual temporal binding window in 4-, 5-, and 6-year-old children (total N=120). Children watched a person uttering a syllable whose auditory and visual components were either temporally synchronized or desynchronized by 366, 500, or 666 ms. They were asked whether the voice and face went together (Experiment 1) or whether the desynchronized videos differed from the synchronized one (Experiment 2). Four-year-olds detected the 666 ms asynchrony, 5-year-olds detected the 666 and 500 ms asynchrony, and 6-year-olds detected all asynchronies. These results show that the audio-visual temporal binding window narrows slowly during early childhood and that it is still wider at six years of age than in older children and adults.

Typically, the objects and events in our everyday world are specified by correlated and highly redundant combinations of multisensory attributes. Because of their redundancy, multisensory attributes increase the salience of our perceptual experiences. Reflecting this fact, nervous systems have evolved mechanisms that permit integration of multisensory inputs at the neural and behavioral levels (Ghazanfar & Schroeder, 2006; Rowe, 1999; Stein & Meredith, 1993). For example, at the neural level, it has been found that naturally occurring oscillations of neural ensembles in primary auditory cortex are amplified by the concurrent visual input normally associated with a talker's face (Hasson, Ghazanfar, Galantucci, Garrod, & Keysers, 2012; Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008). At the behavioral level, studies have shown that the adults of many species (Partan & Marler, 1999; Rowe, 1999) as well as human infants (Bahrick, Lickliter, & Flom, 2004; Lewkowicz & Kraebel, 2004) exhibit enhanced responsiveness to redundant multisensory signals as opposed to unisensory ones.

Responsiveness to the redundant nature of multisensory perceptual signals is especially important in speech and language because most human communication consists of vocalizations that are highly correlated with lip and head movements (Yehia, Rubin, & Vatikiotis-Bateson, 1998). In other words, it is quite adaptive to take advantage of the usual correlation between audible and visible speech signals through lip-reading. Consistent with this, a number of studies have found that adults and infants do, in fact, engage in a fair amount of lip-reading (Lewkowicz & Hansen-Tift, 2012; Rosenblum, 2008; Vatikiotis-Bateson, Eigsti, Yano, & Munhall, 1998; Weikum et al., 2007). Needless to say, to take full advantage of the increased perceptual salience of redundant audiovisual signals, one must be able to integrate them. Indeed, studies have found that adults and infants automatically

Address correspondence to: David J. Lewkowicz, Department of Psychology, Florida Atlantic University, 777 Glades Rd., Boca Raton, FL 33431, (561) 297-1276, lewkowic@fau.edu.

integrate their interlocutors' vocalizations and lip movements (McGurk & MacDonald, 1976; Rosenblum, Schmuckler, & Johnson, 1997). In addition, studies have found that audiovisual redundancy enhances speech comprehension in adults (Sumby & Pollack, 1954; Summerfield, 1979) and that it affects discriminative responsiveness and learning in infants (Bahrick & Lickliter, 2000; Lewkowicz, 1985, 1986, 1992).

Audio-visual (A-V) temporal synchrony is basic to the perception of multisensory coherence in general as well as to the perception of audiovisual speech coherence (Lewkowicz, 2000a; Lewkowicz & Ghazanfar, 2009; Thelen & Smith, 1994). In addition, it guides and constrains perception, learning, and memory of speech and non-speech stimuli in early development and it can scaffold cognitive, social, and language development (Bahrick et al., 2004). To illustrate, infants are sensitive to A-V speech synchrony (Lewkowicz, 2000b, 2010) and A-V synchrony has been found to promote infant detection of affect and label-object associations (Flom & Bahrick, 2007; Gogate & Bahrick, 1998), and the acquisition of speech and language (Gogate, Walker-Andrews, & Bahrick, 2001). In addition, infants' response to spatial (Neil, Chee-Ruiter, Scheier, Lewkowicz, & Shimojo, 2006), spatiotemporal (Scheier, Lewkowicz, & Shimojo, 2003), and temporal (Dodd, 1979; Lewkowicz, 1986, 1996, 2010) audiovisual relations depends on their perception of A-V synchrony. Finally, perception of synchronized audiovisual speech plays an important role in the development of speech perception and, most likely in production as well. This illustrated by findings that infants begin lip-reading when they start learning how to talk and that they rely more on lip reading when exposed to a foreign language once they have mastered their native language (Lewkowicz & Hansen-Tift, 2012). In other words, infants rely on the natural statistics of audiovisual speech signals (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009) to learn their native language and to disambiguate speech from an unfamiliar language.

Additional evidence that the ability to perceive audiovisual coherence is critical in early development comes from studies of children with developmental disabilities. These studies have shown that deficits in audiovisual integration are associated with developmental disabilities and that early experience with audiovisually integrated speech is essential for normal responsiveness to audiovisual speech later in life. For example, autistic children, who are known to have communication difficulties, exhibit deficits in audiovisual integration (Foss-Feig et al., 2010; Irwin, Tornatore, Brancazio, & Whalen, 2011; Kwakye, Foss-Feig, Cascio, Stone, & Wallace, 2010; Smith & Bennetto, 2007). Children with specific language impairment are poorer in detecting A-V synchrony relations than children without it (Pons, Andreu, Sanz-Torrent, Buil-Legaz, & Lewkowicz, 2012). Adults who were born deaf and, thus, deprived of hearing early in life, can subsequently integrate audiovisual speech if their hearing is restored with cochlear implants prior to 2.5 years of life but not if their hearing is restored after that age (Schorr, Fox, van Wassenhove, & Knudsen, 2005). Likewise, adults who were born with dense cataracts and who were, therefore, deprived of visual input for several months until the cataracts were removed exhibit audiovisual speech integration deficits (Putzar, Goerendt, Lange, Rösler, & Röder, 2007).

When do infants first exhibit sensitivity to A-V temporal synchrony relations? The answer is that starting at birth infants can detect the temporal synchrony of auditory and visual stimuli (Lewkowicz, 2000a, 2000b, 2010; Lewkowicz, Leo, & Simion, 2010). Crucially, however, sensitivity to A-V synchrony relations is considerably poorer in infants than in adults. One way to quantify such sensitivity is in terms of an intersensory temporal contiguity window (ITCW). This is a concept introduced by Lewkowicz (2000a) based on an earlier concept of the intersensory temporal synchrony (Lewkowicz, 1996). The ITCW defines sensitivity as the minimum amount of time that auditory and visual sensory inputs must be physically separated in order for participants to perceive them as asynchronous. Importantly, the

concept of ITCW is based on aggregate data and, thus, it is a group measure of a detection threshold rather than the individual type of threshold that is usually obtained in adult studies where it is possible to give hundreds of trials to obtain an individual estimate.

Findings have shown that the ITCW is much wider in infants than in adults. Specifically, in infants, the ITCW has been estimated to be around 350 ms for events where the auditory attribute of a non-speech event leads its visible action and 450 ms for events when the audible attribute follows the visible action (Lewkowicz, 1996). In adults, the corresponding values are around 80 ms and 112–187 ms, respectively (Dixon & Spitz, 1980; Lewkowicz, 1996). For audiovisual speech, the infant ITCW has been estimated to be around 666 ms for audible-leads-visible speech (Lewkowicz, 2010) and its size is currently not known for visible-leads-audible speech. The corresponding values in adults are approximately 60–200 ms for audible-leads-visible speech and 180–240 ms for visible-leads-audible speech (Grant, van Wassenhove, & Poeppel, 2004; Navarra et al., 2005; van Wassenhove, Grant, & Poeppel, 2007).

From a developmental perspective, a comparison of the infant and adult findings suggests that the ITCW window narrows during development. Of course, this conclusion should be qualified by the fact that different methods were used in infant as opposed to adult studies to obtain estimates of the size of the ITCW. Thus, differences can only be considered to be approximations of the specific size of the ITCW. Nonetheless, regardless of the specific size of the ITCW, there is little doubt that the window narrows during development.

The narrowing of the ITCW is probably due to two interacting factors. One is perceptual experience which leads to a sharpening of perceptual sensitivity (Gibson, 1969; Lewkowicz, 2010). The other is a gradual speeding up of neural transmission (Eyre, Miller, & Ramesh, 1991) which reflects, in part, increases in axon diameter, amount of myelin surrounding axons, and the number of neurons and synapses during childhood (Giedd et al., 1999; Paus et al., 1999; Yakovlev & Lecours, 1967). The developmental improvement in neural transmission time is, however, relatively slow because its underlying neural determinants change slowly during development. Thus, the ITCW is likely to narrow gradually during childhood.

Only a handful of studies have examined young children's response to A-V temporal synchrony and none have systematically investigated developmental changes in the threshold for the detection of A-V asynchrony. One study (Bebko, Weiss, Demark, & Gomez, 2006) investigated 4–6 year-olds' preferences for temporally synchronous versus asynchronous audiovisual events. In one test, synchronous and asynchronous speech was contrasted while in another test synchronous and asynchronous non-speech events were contrasted. The asynchrony in both cases was quite large (3 s). Results revealed that typically and atypically developing children preferred synchronous over asynchronous speech and non-speech events and that children with autism did not prefer synchronous speech but that they did prefer a synchronous non-speech event. Another study (Pons et al., 2012) investigated responsiveness to synchronous versus asynchronous fluent speech in children with a specific language impairment (SLI) and in children without it. The auditory speech either led or followed visual speech by 366 or 666 ms. Results indicated that neither group detected a 366 ms asynchrony regardless of direction of temporal offset, that typically developing children detected the 666 ms asynchrony regardless of direction of offset, and that children with SLI only detected the auditory-leading 666 ms asynchrony (the easier of the two). Finally, some studies have investigated responsiveness to synchronous and asynchronous flashes and tones in children, adolescents, and young adults (Hillock, Powers, & Wallace, 2011; Hillock-Dunn & Wallace, 2012). One of the most interesting findings

from these studies is that, relative to adults, even adolescents (12–17 years of age) have a higher threshold for the detection of A-V asynchrony (by approximately 100 ms).

Together, the findings from infants and children point to developmental changes in the ITCW but currently it is not known whether these changes might occur sometime between infancy and adolescence. As we have indicated earlier, neural transmission speed increases with development and experience and undoubtedly contributes in critical ways to the development of multisensory skills during early childhood. In addition, multisensory perceptual skills continue to develop and improve after infancy (Bremner, Lewkowicz, & Spence, 2012; Gori, Del Viva, Sandini, & Burr, 2008; Innes-Brown et al., 2011). All of this suggests that the ITCW is likely to change after infancy and, as a result, it is important to examine possible developmental changes in the perception of A-V temporal synchrony relations during early childhood. If the ITCW does change then this would provide important new insights into the multisensory perceptual mechanisms that must surely play a key role in the development of language, reading, social, and cognitive skills during the preschool, kindergarten, and school years.

To test the possibility that the ITCW changes in early childhood, we conducted two experiments in which we tested 4-, 5-, and 6-year-old children with different videos of a person repeatedly uttering a single syllable. The person in the videos could be heard and seen uttering the syllable synchronously or could be heard vocalizing it 366, 500, or 666 ms before she could be seen vocalizing it. In Experiment 1, we asked children to report whether the voice and face went together or not in each video (i.e., whether they were synchronous). In Experiment 2, we familiarized children to the synchronous audiovisual syllable and then presented one of the four test videos and asked children to report whether the test video differed from the familiarization video.

Experiment 1

Method

Participants—We tested three separate age groups consisting of 20 children each. They consisted of 4-year-olds (9 females; mean age: 4-years, 60 days; range: 45–52 months), 5-year-olds (10 females; mean age: 5 years, 120 days; range: 60–67 months), and 6-year-olds (8 females; mean age: 6 years, 204 days; range: 72–80 months). All children were recruited from the Brigham Young University preschool and kindergarten and tended to come from middle to upper-class families in the Provo-Orem area. No children were excluded from the analyses.

Apparatus & Stimuli—Stimuli consisted of six 28 s movies of a woman's face (named Jill) who could be seen and heard repeatedly uttering a single syllable. The movies were constructed with the aid of Premiere 6.0 (Adobe Corporation) and had the following settings: 720 × 480-pixel image, Cinepak Codec video compression, 1024 kbps audio sample rate, and 29.97 frames/s. The four movies used for testing detection of A-V asynchrony consisted of concatenated copies of a 4 s video clip of a /ba/ syllable. Each clip began with the actor holding her lips in the closed position until 1.66 s elapsed. Then, she constricted her lips, opened her mouth, phonated, and finally closed her mouth. Vocalization began at 2.033 s and ended at 2.231 s and mouth motion ended at 2.693 s. Once she stopped moving her lips, she resumed a still face with lips closed. The actor held her head still throughout the clip, had a neutral expression on her face, and the transition between each utterance was nearly imperceptible. Each clip consisted of seven repetitions of the syllable. In four of the movies (each was presented one at a time during the test trials), the actor was seen and heard uttering /ba/. These movies only differed by the onset lead-time of the audible syllable with respect to the visible syllable (i.e., 0, 366, 500, or 666 ms). Thus, the

test trials were labeled ASYNC 0, ASYNC 366, ASYNC 500, and ASYNC 666. The two remaining movies – presented during an initial practice phase – depicted a different actor (Mary) articulating either an audiovisually synchronous /da/ syllable or one where the audible /da/ preceded the visible /da/ by 1500 ms.

Testing took place in a quiet, dimly lit room. Movies were presented using a Sony DVD player and were displayed on a 32 inch flat panel monitor located 50 cm in front of the participant and the audio was played at 70 dB through a speaker located below the monitor. Two experimenters, unaware of the experimental hypotheses, were present. One sat next to the child and made sure the child attended to each movie. The second experimenter sat behind the child and the other experimenter and controlled movie presentation.

Procedure—The experiment consisted of an initial practice phase during which two practice trials were given. One practice trial consisted of the presentation of the audible and visible syllables in perfect synchrony and the other trial consisted of the presentation of an audible syllable leading the visible syllable by 1500 ms. During the two practice trials, children were told that they would watch a movie of a Mary saying something (i.e., /da/), and that we wanted them to pay close attention to what Mary was saying and that we wanted to know if Mary's voice came at the right time. The order of these trials was randomized across participants. Immediately prior to starting each practice trial children were asked to verbally report as soon as they knew whether the voice came at the right time. If the child did not respond within 15s, the experimenter asked: "Does Mary's voice 'sound right'? and/ or if needed "Do her lips match her voice?" and/or whether "Mary's voice sounded right?". At all ages, and for both the synchronous and asynchronous practice trials, children correctly reported whether the face and voice were synchronous or asynchronous.

The remainder of the experiment consisted of a test phase composed of two blocks of four trials each. During the test phase, each of the two blocks of test trials (separated by a 5-minute break) consisted of the presentation of each of the four unique test trials. Order of test trial presentation was randomized across participants except: (1) that each unique test trial occurred once as the first trial while the order of the remaining three trials was random, and (2) that neither an ascending nor descending series of asynchronies was presented. Children were given the same question(s) except that now they were told that they would watch another woman named Jill and again we wanted to know whether her voice was synchronous or asynchronous..

Results and Discussion

Because each unique test trial was given twice, i.e., once per block, it was possible for a child to have a score of 0, 1, or 2 correct responses. We chose two correct responses as the most conservative measure of performance at each degree of asynchrony and considered that as a "pass". Thus, at each age, we counted the number of children, out of the 20 tested, who passed both test trials (please note that a pass for the ASYNC 0 test trial required the children to say "sounds right" or "matches" whereas a pass for the other three test trials required the children to say "does not sound right" or "does not match"). Then, we used a non-parametric Chi-Square test to determine whether the number of children who passed at a particular level of asynchrony was greater than the number of children who did not pass (i.e., who had 1 or 0 correct responses at that asynchrony).

Figure 1 shows the number of children at each age who passed at each asynchrony level. As can be seen, for the ASYNC 0 test trial, the number of 4-year-olds who passed approached, but did not exceed, statistical significance (X^2 (1, N = 20) = 3.2, p = .06) and the number of 5- and 6-year-olds who passed significantly exceeded chance (X^2 (1, N = 20) = 9.8, p = .002; X^2 (1, N = 20) = 16.2, p < .001, respectively). For the ASYNC 366 and the ASYNC 500 test

trials, only the number of 6-year-olds who passed exceeded chance (X^2 (1, N = 20) = 9.8, p = .002; X^2 (1, N = 20) = 9.8, p = .002, respectively). Finally, for the ASYNC 666 test trial, the number of children who passed exceeded chance at all three ages (X^2 (1, N = 20) = 5.0, p = .025; X^2 (1, N = 20) = 5.0, p = .02; and X^2 (1, N = 20) = 12.8, p < .001, at each age, respectively).

As depicted in Figure 1, there were developmental differences in the number of children who passed in the critical test trials, namely the ASYNC 366 and ASYNC 500 test trials. As indicated earlier, we expected the ITCW to narrow to less than 666 ms sometime during early childhood. Therefore, we conducted a second analysis where we asked whether the number of children who passed in a given test trial increased with age. Not surprisingly, the number of children who passed did not vary with age in the ASYNC 0 test trial, X^2 (4, N = 60) = 5.7, p > .1, nor in the ASYNC 666 test trial, X^2 (4, N = 60) = 3.6, p > .1. Crucially, the developmental analysis indicated that the number of children who passed increased marginally with age in the ASYNC 500 test trial, X^2 (4, N = 60) = 8.8, p =.06, and that it increased significantly in the ASYNC 366 test trial, X^2 (4, N = 60) = 12.8, p < .01. These age-related findings are consistent with the prediction that the ITCW window narrows with development. That is, younger children should begin to find it easier to detect an asynchrony of 500 ms compared to infants who do not detect it and only the older children should be able to detect an A-V asynchrony as small as 366 ms.

In sum, these results show that, similar to infants, 4-year-olds detect an A-V asynchrony of 666 ms but that they do not detect lower asynchronies. The results also show that it is not until sometime between the fifth and sixth year of life that the ITCW starts to narrow to 366 ms.

Experiment 2

It is possible that asking children – especially the younger ones - whether the voice came at the right time to determine whether they detected the asynchrony was too challenging in Experiment 1. Consequently, in Experiment 2, we first familiarized the children with a synchronous speech event (i.e., the synchronized syllable) during the practice as well as the test trials and then examined their sensitivity to A-V synchrony by asking them to state whether the test event was the same or different from the familiarization event.

Method

Participants—Participants consisted of twenty 4-year-olds (13 females; mean age: 4 years, 86 days; range: 47–57 months), 5-year-olds (9 females; mean age: 5 years, 56 days; range: 60–68 months), and 6-year-olds (10 females; mean age: 6 years, 110 days; range: 72–82 months). One additional 4-year-old and one additional 5-year-old were tested but were excluded because they did not complete the experiment.

Apparatus, Stimuli & Procedure—The apparatus and stimuli were the same as in Experiment 1 except that here we used two side-by-side identical 32 inch monitors and used a familiarization-test procedure. Specifically, during each of the two practice trials (presented in random order) children watched an audiovisually synchronous movie of Mary on one monitor for 20 s of cumulative time Then, the children were told they would watch another movie of Mary on the other monitor (experimenter points to the other monitor). At this point, they were asked to report verbally as soon as possible whether Mary sounded the same in the second movie (the experimenter points to the other monitor) as in the first movie (i.e., the familiarization movie) or whether she sounded different in the two movies. After children reported whether the event was the same or different from familiarization – they again watched the familiarization event and were given the second practice trial. We varied

our questions from “are these two movies exactly the same – or is there something different” to “does Mary look and sound the same in both movies … or does she look and sound different”. Like in Experiment 1, children responded correctly in both practice trials and their performance did not differ as a function of type of question posed.

Following the practice phase, children were given the two blocks of 4 test trials each, with a break between them. Again, prior to the presentation of each familiarization movie, children were told that they would watch a movie of Jill saying something and they should pay close attention to what she was saying. Familiarization consisted of looking at the synchronous audiovisual event for 20 s of cumulative looking time. Following familiarization, children viewed one of the four test movies on the other monitor. Specifically, the experimenter pointed toward the adjacent monitor and told the children that they would watch another movie of Jill on this monitor and that we wanted to know whether there was any difference in how Jill sounded. They were instructed to “pay close attention to how Jill sounds” and “to pay close attention to Jill’s face and voice – and see if her face and voice are the same or different”. Like the practice trials, children were asked to verbally report as soon as they knew whether the second movie was the same or different from the first.

Results and Discussion

At each age, we compared the number of children who passed versus those who did not pass. Specifically, we examined how many children correctly stated on both test trials whether the test events at each level of asynchrony, respectively, were the same or different relative to the familiarization event. Figure 2 shows the results for Experiment 2. As can be seen, for the ASYNC 0 test trial, the number of children who passed was greater than chance at each age (X^2 (1, N = 20) = 9.8, p = .002; X^2 (1, N = 20) = 7.2, p = .007; X^2 (1, N = 20) = 16.2, p < .001, respectively). For the ASYNC 366 test trial, only the number of 6-year-olds who passed was greater than chance (X^2 (1, N = 20) = 9.8, p = .002). For the ASYNC 500 test trial, the number of 5- and 6-year-olds who passed was greater than chance (X^2 (1, N = 20) = 7.2, p = .007; X^2 (1, N = 20) = 9.8, p = .002, respectively) whereas the number of 4-year-olds who passed was only marginally greater than chance (X^2 (1, N = 20) = 3.2, p = .06). Finally, for the ASYNC 666 test trial, the number of children who passed was greater than chance at each age (X^2 (1, N = 20) = 9.8, p = .002; X^2 (1, N = 20) = 5.0, p = .025; X^2 (1, N = 20) = 9.8, p = .002; X^2 (1, N = 20) = 9.8, p = .002, respectively).

Overall, the developmental pattern found in the current experiment was similar to that found in Experiment 1 except that there were two cases of improved performance in this experiment. Specifically, as before, the 6-year-olds responded correctly in all test trials but, in contrast to Experiment 1, the 4-year-olds now responded correctly in the ASYNC 0 test trial and the 5-year-olds responded correctly in the ASYNC 500 test trial. When compared with the findings from Experiment 1, the results from this experiment suggest that familiarization facilitated performance.

To determine more directly whether developmental changes occurred, we conducted separate analyses at each asynchrony level of the number of children who passed. We found that the number of children who passed increased significantly with age in the ASYNC 500 test trial (X^2 (4, N = 60) = 22.4, p < .01) and in the ASYNC 366 test trial (X^2 (4, N = 60) = 17.1, p < .01). In addition, because children of all ages passed in the ASYNC 0 and the ASYNC 660 test trials, we found that the number of children who passed was the same across age in these two test trials (X^2 (4, N = 60) = 7.3, p > 0.10; ASYNC 660 test trial, X^2 (4, N = 60) = 1.6, p > .10).

General Discussion

As indicated earlier, there are several theoretical and empirical reasons to suspect that the perception of A-V temporal synchrony changes during early childhood. As a result, we hypothesized that the perception of A-V speech synchrony should improve during early childhood and, consistent with our prediction, we found that the ITCW narrows from an average value of 666 ms at four years of age to an average value of 366 ms at six years of age. These results are interesting for several reasons. First, they replicate findings that 4–7 year-old children can detect a 666 ms A-V asynchrony but not a 366 ms asynchrony (Pons et al., 2012). Second, the current findings suggest that the ITCW for audiovisual speech may not change during the first four years of life (i.e., it appears that its average size is approximately 666 ms). Finally, as expected, we found that the familiarization procedure enhanced detection in that following familiarization the 4-year-olds now successfully detected perfect synchrony and the 5-year-olds now detected the 500 ms asynchrony.

Importantly, the enhanced detection following familiarization was not due to children's better understanding of the task nor to the questions posed because performance during the two practice trials was at ceiling in both experiments. Consequently, the most likely explanation for the enhancing effects of familiarization was that it helped the children overcome a relatively difficult task in Experiment 1 where they were required to perceive a specific A-V temporal relation and then explicitly report whether that relation was synchronous or not. Interestingly, however, despite the fact that the task in Experiment 1 placed greater processing demands on the children, the results from this experiment are still consistent with findings from studies using implicit recognition procedures to measure children's detection of A-V temporal relations (Pons et al., 2012).

The familiarization procedure is interesting for one other reason. As indicated earlier, this procedure is closest to the procedure typically used with infants to test for asynchrony detection (Lewkowicz, 1996, 2010). Moreover, the stimulus events used here were the same as those used in a previous infant study (Lewkowicz, 2010). Therefore, it is interesting to compare the results from infants and the results from the children in the current study. The comparison reveals that the ITCW is around 666 ms in infancy suggesting that it remains at that level until four years of age, that it narrows to 500 ms by five years of age, and that it further narrows to 366 ms by six years of age. Overall, this suggests that the ITCW for audiovisual speech narrows rather dramatically during early childhood. This is certainly interesting and the possibility that the ITCW may not narrow for such a long time has important implications for the holistic processing of multisensory events. Of course, the conclusion that the ITCW narrows in this specific fashion should be treated with some caution because direct comparisons between infants and children are fraught with difficulty (Keen, 2003). Therefore, the most reasonable conclusion with regard to possible developmental changes between infancy and four years of age and/or lack thereof is that additional studies of responsiveness during this period are needed.

Bearing the same caveats regarding cross-age comparisons in mind, it is also interesting to compare our findings at six years of age indicating that the ITCW is 330 ms for an audiovisual speech event and findings from adults tested for detection of audiovisual speech asynchrony. This comparison suggests that the ITCW continues to narrow in that the adult ITCW for audiovisual speech is around 60–200 ms (Grant et al., 2004; Navarra et al., 2005; van Wassenhove et al., 2007). This conclusion is consistent with findings from studies of older children's perception of asynchrony showing that their asynchrony detection thresholds are higher than those found in adults (Hillock et al., 2011; Hillock-Dunn & Wallace, 2012). It should be noted, however, that a direct comparison of our findings with those from the older children are not possible because that latter studies tested detection of

asynchrony with non-speech audiovisual events. In addition, the studies with older children as well as those with adults generally use psychophysical testing methods whose explicit purpose is to determine an individual threshold for each subject. Thus, typically these studies use a point-of-subjective simultaneity task where subjects are given trials consisting of an auditory and visual stimulus and the subject's task is to indicate whether the stimuli are simultaneous or asynchronous (Fujisaki, Shimojo, Kashino, & Nishida, 2004; Hillock et al., 2011; Hillock-Dunn & Wallace, 2012). Obviously, the point-of-subjective simultaneity task cannot be used with young children because it is far too boring and taxing and, as a result, it is not possible to obtain individual thresholds from younger children. Because of this limitation, we must rely on group performance measures and be cognizant of the limitations of such measures when attempting to infer anything about individual thresholds. Nonetheless, it is clear that the ITCW narrows from four to six years of age. What is less clear at this point is how rapidly this narrowing occurs and when in development it occurs mainly because extant studies of younger and older subjects have used different methods and different stimulus materials.

Finally, our finding that the ITCW narrows most dramatically at the age when children enter school is intriguing. It is then that children must be able to integrate the teacher's visible and audible speech in a relatively rapid fashion so as to be able to extract useful and meaningful information. It is also then that most children begin acquiring reading skills in earnest. Learning to read requires that one be able to rapidly and accurately map the auditory attributes of one's own vocalizations onto the visible attributes of written words (Tallal, Miller, & Fitch, 1993). A narrower ITCW facilitates this process because it makes those mappings much more precise and accurate than a broader ITCW. In future studies it would interesting to see whether individual differences in reading skill correlate with the size of an individual child's ITCW.

Acknowledgments

Supported, in part, by NSF grant BCS-0751888 and grant R01HD057116 from the Eunice Kennedy Shriver National Institute Of Child Health & Human Development to DJL. The content is solely the responsibility of the author and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute Of Child Health & Human Development or the National Institutes of Health. We thank Scott Stevens, Justin Martin, Denise Free, and Jacob Jones for their assistance. Finally we thank the staff and children within BYU's School of Family Life Preschool for their assistance and participation.

References

Bahrick LE, Lickliter R. Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. Developmental Psychology. 2000; 36:190–201. [PubMed: 10749076]

Bahrick LE, Lickliter R, Flom R. Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. Current Directions in Psychological Science. 2004; 13:99–102.

Bebko JM, Weiss JA, Demark JL, Gomez P. Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism. Journal of Child Psychology and Psychiatry. 2006; 47:88–98. [PubMed: 16405645]

Bremner, AJ.; Lewkowicz, DJ.; Spence, C. Multisensory Development. Oxford: Oxford University Press; 2012.

Chandrasekaran C, Trubanova A, Stillittano S, Caplier A, Ghazanfar AA. The natural statistics of audiovisual speech. PLoS Computational Biology. 2009; 5:e1000436. [PubMed: 19609344]

Dixon NF, Spitz LT. The detection of auditory visual desynchrony. Perception. 1980; 9:719–721. [PubMed: 7220244]

Dodd B. Lip reading in infants: Attention to speech presented in- and out-of-synchrony. Cognitive Psychology. 1979; 11:478–484. [PubMed: 487747]

Eyre JA, Miller S, Ramesh V. Constancy of central conduction delays during development in man: investigation of motor and somatosensory pathways. The Journal of physiology. 1991; 434:441–452. [PubMed: 2023125]

Flom R, Bahrick LE. The Development of Infant Discrimination of Affect in Multimodal and Unimodal Stimulation: The Role of Intersensory Redundancy. Developmental Psychology. 2007; 43:238–252. [PubMed: 17201522]

Foss-Feig JH, Kwakye LD, Cascio CJ, Burnette CP, Kadivar H, Stone WL, Wallace MT. An extended multisensory temporal binding window in autism spectrum disorders. Experimental Brain Research. 2010; 203:381–389. [PubMed: 20390256]

Fujisaki W, Shimojo S, Kashino M, Nishida Sy. Recalibration of audiovisual simultaneity. Nature Neuroscience. 2004; 7:773–778.

Ghazanfar AA, Schroeder CE. Is neocortex essentially multisensory? Trends in cognitive sciences. 2006; 10:278–285. [PubMed: 16713325]

Gibson, EJ. Principles of perceptual learning and development. New York: Appleton; 1969.

Giedd JN, Blumenthal J, Jeffries NO, Castellanos FX, Liu H, Zijdenbos A, Rapoport JL. Brain development during childhood and adolescence: a longitudinal MRI study. Nature Neuroscience. 1999; 2:861–862.

Gogate LJ, Bahrick LE. Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. Journal of Experimental Child Psychology. 1998; 69:133–149. [PubMed: 9637756]

Gogate LJ, Walker-Andrews AS, Bahrick LE. The intersensory origins of word comprehension: An ecological-dynamic systems view. Developmental Science. 2001; 4:1–18.

Gori M, Del Viva M, Sandini G, Burr DC. Young children do not integrate visual and haptic form information. Current Biology. 2008; 18:694–698. [PubMed: 18450446]

Grant KW, van Wassenhove V, Poeppel D. Detection of auditory (cross-spectral) and auditory-visual (cross-modal) synchrony. Speech Communication Special Issue: Audio Visual Speech Processing. 2004; 44:43–53.

Hasson U, Ghazanfar AA, Galantucci B, Garrod S, Keysers C. Brain-to-brain coupling: a mechanism for creating and sharing a social world. Trends in cognitive sciences. 2012; 16:114–121. [PubMed: 22221820]

Hillock AR, Powers AR, Wallace MT. Binding of sights and sounds: Age-related changes in multisensory temporal processing. Neuropsychologia. 2011; 49:461–467. [PubMed: 21134385]

Hillock-Dunn A, Wallace MT. Developmental changes in the multisensory temporal binding window persist into adolescence. Developmental Science. 201210.1111/j.1467-7687.2012.01171.x

Innes-Brown H, Barutchu A, Shivdasani MN, Crewther DP, Grayden DB, Paolini A. Susceptibility to the flash-beep illusion is increased in children compared to adults. Developmental Science. 2011; 14:1089–1099. [PubMed: 21884324]

Irwin JR, Tornatore LA, Brancazio L, Whalen DH. Can children with autism spectrum disorders “hear” a speaking face? Child development. 201110.1111/j.1467-8624.2011.01619.x

Keen R. Representation of Objects and Events Why Do Infants Look So Smart and Toddlers Look So Dumb? Current Directions in Psychological Science. 2003; 12:79–83.

Kwakye LD, Foss-Feig JH, Cascio CJ, Stone WL, Wallace MT. Altered auditory and multisensory temporal processing in autism spectrum disorders. Frontiers in integrative neuroscience. 2010; 4:129. [PubMed: 21258617]

Lewkowicz D, Kraebel K. The value of multisensory redundancy in the development of intersensory perception. The handbook of multisensory processes. 2004:655–678.

Lewkowicz DJ. Bisensory response to temporal frequency in 4-month-old infants. Developmental Psychology. 1985; 21:306–317.

Lewkowicz DJ. Developmental changes in infants’ bisensory response to synchronous durations. Infant Behavior & Development. 1986; 9:335–353.

Lewkowicz DJ. Infants’ responsiveness to the auditory and visual attributes of a sounding/moving stimulus. Perception & Psychophysics. 1992; 52:519–528. [PubMed: 1437484]

Lewkowicz DJ. Perception of auditory-visual temporal synchrony in human infants. Journal of Experimental Psychology: Human Perception & Performance. 1996; 22:1094–1106. [PubMed: 8865617]

Lewkowicz DJ. The development of intersensory temporal perception: An epigenetic systems/limitations view. Psychological Bulletin. 2000a; 126:281–308. [PubMed: 10748644]

Lewkowicz DJ. Infants' perception of the audible, visible and bimodal attributes of multimodal syllables. Child development. 2000b; 71:1241–1257. [PubMed: 11108094]

Lewkowicz DJ. Infant perception of audio-visual speech synchrony. Developmental Psychology. 2010; 46:66–77. [PubMed: 20053007]

Lewkowicz DJ, Ghazanfar AA. The emergence of multisensory systems through perceptual narrowing. Trends in cognitive sciences. 2009; 13:470–478. [PubMed: 19748305]

Lewkowicz DJ, Hansen-Tift AM. Infants deploy selective attention to the mouth of a talking face when learning speech. Proceedings of the National Academy of Sciences. 2012; 109:1431–1436.

Lewkowicz DJ, Leo I, Simion F. Intersensory perception at birth: Newborns match non-human primate faces & voices. Infancy. 2010; 15:46–60.

McGurk H, MacDonald J. Hearing lips and seeing voices. Nature. 1976 Dec.264:229–239.

Navarra J, Vatakis A, Zampini M, Soto-Faraco S, Humphreys W, Spence C. Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. Cognitive Brain Research. 2005; 25:499–507. [PubMed: 16137867]

Neil PA, Chee-Ruiter C, Scheier C, Lewkowicz DJ, Shimojo S. Development of multisensory spatial integration and perception in humans. Developmental Science. 2006; 9:454–464. [PubMed: 16911447]

Partan S, Marler P. Communication goes multimodal. Science. 1999; 283:1272–1273. [PubMed: 10084931]

Paus T, Zijdenbos A, Worsley K, Collins DL, Blumenthal J, Giedd JN, Evans AC. Structural maturation of neural pathways in children and adolescents: in vivo study. Science. 1999; 283:1908–1911. [PubMed: 10082463]

Pons F, Andreu L, Sanz-Torrent M, Buil-Legaz L, Lewkowicz DJ. Perception of audio-visual speech synchrony in Spanish-speaking children with and without specific language impairment. Journal of Child Language. 2012:1–14.10.1017/S0305000912000189 [PubMed: 21418730]

Putzar L, Goerendt I, Lange K, Rösler F, Röder B. Early visual deprivation impairs multisensory interactions in humans. Nature Neuroscience. 2007; 10:1243–1245.

Rosenblum LD. Speech perception as a multimodal phenomenon. Current Directions in Psychological Science. 2008; 17:405. [PubMed: 23914077]

Rosenblum LD, Schmuckler MA, Johnson JA. The McGurk effect in infants. Perception & Psychophysics. 1997; 59:347–357. [PubMed: 9136265]

Rowe C. Receiver psychology and the evolution of multicomponent signals. Animal Behaviour. 1999; 58:921–931. [PubMed: 10564594]

Scheier C, Lewkowicz DJ, Shimojo S. Sound induces perceptual reorganization of an ambiguous motion display in human infants. Developmental Science. 2003; 6:233–244.

Schorr EA, Fox NA, van Wassenhove V, Knudsen EI. Auditory-visual fusion in speech perception in children with cochlear implants. Proc Natl Acad Sci U S A. 2005; 102:18748–18750. 0508862102 [pii]. 10.1073/pnas.0508862102 [PubMed: 16339316]

Schroeder C, Lakatos P, Kajikawa Y, Partan S, Puce A. Neuronal oscillations and visual amplification of speech. Trends in cognitive sciences. 2008; 12:106–113. [PubMed: 18280772]

Smith EG, Bennetto L. Audiovisual speech integration and lipreading in autism. Journal of Child Psychology and Psychiatry. 2007; 48:813–821. [PubMed: 17683453]

Stein, BE.; Meredith, MA. The merging of the senses. Cambridge, MA: The MIT Press; 1993.

Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. Journal of the Acoustical Society of America. 1954; 26:212–215.

Summerfield AQ. Use of visual information in phonetic perception. Phonetica. 1979; 36:314–331. [PubMed: 523520]

Tallal P, Miller S, Fitch RH. Neurobiological basis of speech: a case for the preeminence of temporal processing. Annals of the New York Academy of Sciences. 1993; 682:27–47. [PubMed: 7686725]

Thelen, E.; Smith, LB. A dynamic systems approach to the development of cognition and action. Cambridge, MA: MIT Press; 1994.

van Wassenhove V, Grant KW, Poeppel D. Temporal window of integration in auditory-visual speech perception. Neuropsychologia. 2007; 45:598–607. [PubMed: 16530232]

Vatikiotis-Bateson E, Eigsti IM, Yano S, Munhall KG. Eye movement of perceivers during audiovisual speech perception. Perception & Psychophysics. 1998; 60:926–940. [PubMed: 9718953]

Weikum WM, Vouloumanos A, Navarra J, Soto-Faraco S, Sebastián-Gallés N, Werker JF. Visual language discrimination in infancy. Science. 2007; 316:1159. [PubMed: 17525331]

Yakovlev, P.; Lecours, A. The myelogenetic cycles of regional maturation of the brain. In: Minkowski, A., editor. Regional development of the brain in early life. Philadelphia, PA: Davis; 1967. p. 3-70.

Yehia H, Rubin P, Vatikiotis-Bateson E. Quantitative Association of Vocal-Tract and Facial Behavior. Speech Communication. 1998; 26:23–43.

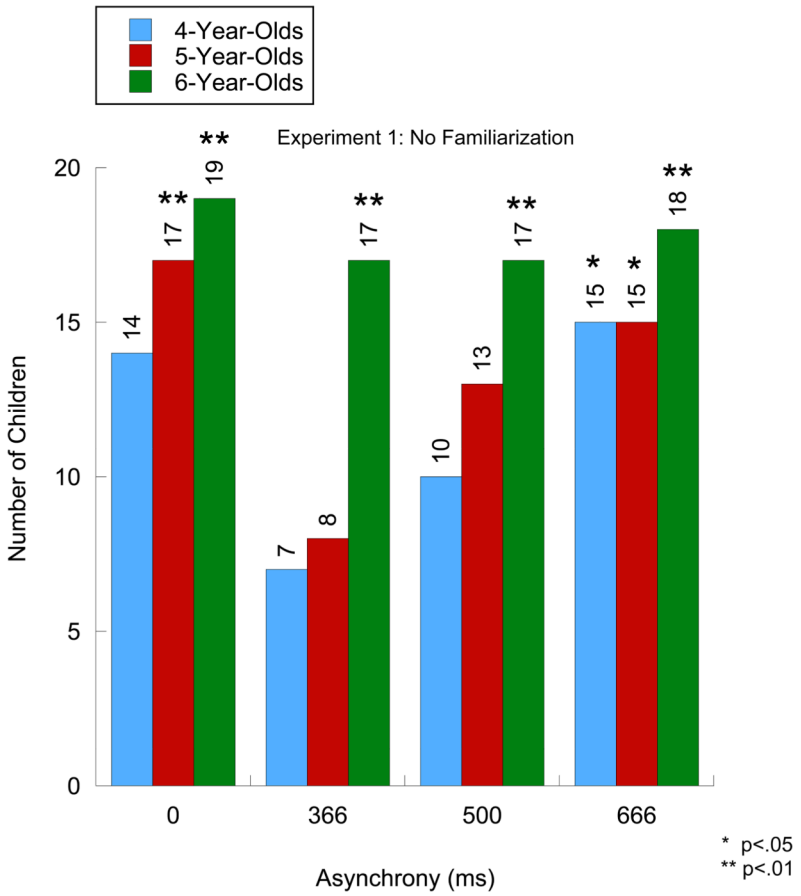

Figure 1.

Number of children correctly identifying both test events as having the voice come at the correct or incorrect time. Asterisks indicate significance level based on a Chi-Square test.

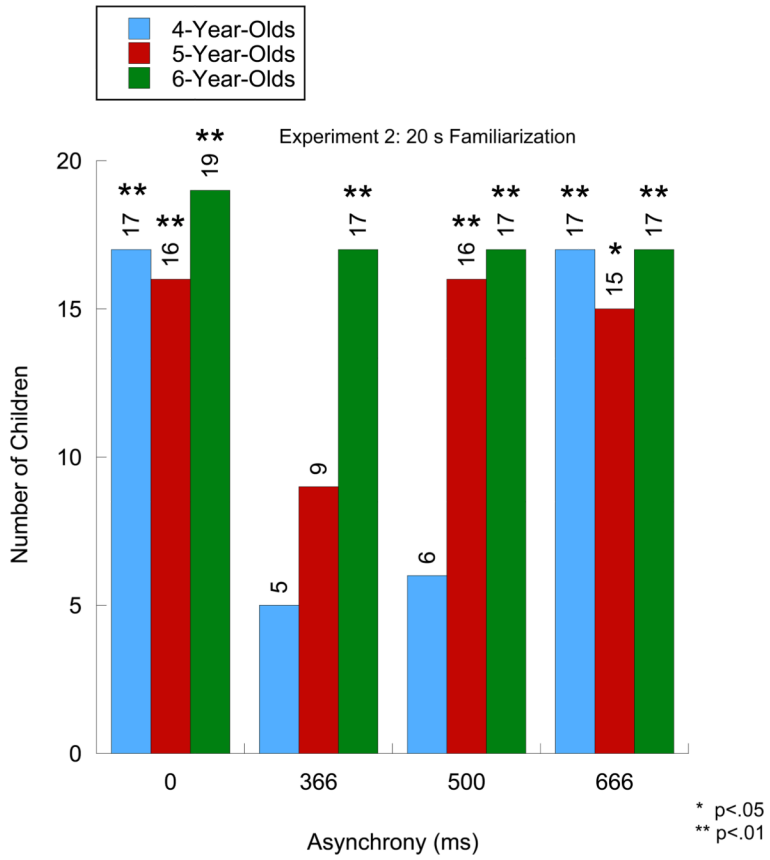

Figure 2.

Number of children correctly identifying both test events as being similar or different from an audiovisually synchronous speech syllable. Asterisks indicate significance level based on a Chi-Square test.