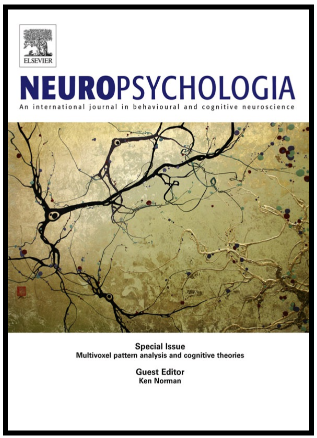
ELSEVIER
NEUROPSYCHOLOGIA
An international journal in behavioural and cognitive neuroscience
Special Issue
Multivoxel pattern analysis and cognitive theories
Guest Editor
Ken Norman

www.elsevier.com/locate/neuropsychologia

The temporal binding window for audiovisual speech: Children *are* like little adults

Andrea Hillock-Dunn, D. Wesley Grantham, Mark T. Wallace

PII: S0028-3932(16)30052-5
DOI: http://dx.doi.org/10.1016/j.neuropsychologia.2016.02.017
Reference: NSY5897

To appear in: *Neuropsychologia*

Received date: 13 July 2015
Revised date: 23 December 2015
Accepted date: 22 February 2016

The temporal binding window for audiovisual speech: Children *are* like little adults

Andrea Hillock-Dunn [a, c], D. Wesley Grantham [a] & Mark T. Wallace [a, b, c]

[a] Dept of Hearing and Speech Sciences, Vanderbilt University, Nashville, TN

[b] Dept of Psychology, Vanderbilt University, Nashville, TN

[c] Vanderbilt University Kennedy Center, Vanderbilt University

Correspondence:

Andrea Hillock-Dunn, Au.D., Ph.D.
1215 21st Avenue South
MCE South Tower, Room 9302
Nashville, TN 37232

Phone: 615-875-0123
Fax: 615-875-5058

andrea.h.dunn@vanderbilt.edu

Abstract

During a typical communication exchange, both auditory and visual cues contribute to speech comprehension. The influence of vision on speech perception can be measured behaviorally using a task where incongruent auditory and visual speech stimuli are paired to induce perception of a novel token reflective of multisensory integration (i.e., the McGurk effect). This effect is temporally constrained in adults, with illusion perception decreasing as the temporal offset between the auditory and visual stimuli increases. Here, we used the McGurk effect to investigate the development of the temporal characteristics of audiovisual speech binding in 7 – 24 year-olds. Surprisingly, results indicated that although older participants perceived the McGurk illusion more frequently, no age-dependent change in the temporal boundaries of audiovisual speech binding was observed.

Keywords: development, maturation, auditory, visual, multisensory, temporal, McGurk effect

Research Highlights

- The temporal profile of audiovisual speech binding is remarkably similar in children, adolescents and adults.
- Children and adolescents integrate audiovisual speech less readily and are more likely to report the auditory cue than adults.
- Mature audiovisual temporal binding windows for speech emerge before adult-like integrative capacity is fully realized.

Introduction

The influence of visual cues on speech perception has been well characterized. This literature has demonstrated that the information conferred by a talker's face and lip movements can markedly increase the intelligibility of the auditory signal (Macleod & Summerfield, 1987; Schwartz, Berthommier, & Savariaux, 2004; Sumby & Pollack, 1954), and improve speech detection thresholds in noise (Grant & Seitz, 2000). Whereas the contribution of vision to speech comprehension is relatively nominal in quiet environments, it is much more robust in noisy and reverberant settings when the auditory signal is masked or distorted (Erber, 1969; Sumby & Pollack, 1954).

While the synergistic use of auditory and visual cues typically provides an adaptive benefit in the realm of speech comprehension, audiovisual interactions can also result in illusory percepts. Although such illusions have less perceptual relevance in real-world environments, they can be used as important tools to index interactive processes taking place within speech networks. One of the most robust and informative of these illusions is the McGurk effect (McGurk & Macdonald, 1976). In this illusion, the presentation of incongruent visual (glottal - /ga/ or /ka/) and auditory (bilabial - /ba/ or /pa/) speech tokens often results in report of an intermediary fused (velor) target such as /ta/ or /da/, reflecting integration of the visual and auditory elements and formation of a new perceptual construct. As with normal speech processes, the strength of the McGurk effect is highly dependent upon the salience of the auditory cues, with the contribution of visual information increasing when the auditory signal is weak, ambiguous and/or degraded (Fixmer & Hawkins, 1998; K. Sekiyama, Kanno, Miura, & Sugita, 2003; Kaoru Sekiyama & Tohkura, 1991).

In addition to being modulated by the efficacy of the paired stimuli, the McGurk effect is also influenced by the temporal relationship between the auditory and visual cues, such that increasing audiovisual asynchrony decreases the strength of the illusion (Jones & Jarick, 2006; D. W. Massaro, Cohen, & Smeele, 1996; Soto-Faraco & Alsius, 2009; van Wassenhove, Grant, & Poeppel, 2007). This effect is asymmetric, such that visual leading presentations are less readily detected than comparable auditory leading asynchronies (e.g., Conrey & Pisoni, 2006). The temporal binding window (TBW) is a construct that indexes the time frame over which manifestations of multisensory integration such as perceptual fusion are likely to be observed (e.g., Colonius & Diederich, 2004; Hairston, Burdette, Flowers, Wood, & Wallace, 2005; Lewkowicz, 2000; Powers, Hillock, & Wallace, 2009; R. A. Stevenson & Wallace, 2013; R. A. Stevenson, Wilson, Powers, & Wallace, 2013; R. A. Stevenson, Zemtsov, & Wallace, 2012; Wallace & Stevenson, 2014). The TBW can be applied to interactive processes at levels ranging from single neurons to behavior and perception. In human perceptual studies it has been measured using stimuli ranging from the highly simple (e.g., flashes and beeps) to the more complex (e.g., speech) (Frens, Van Opstal, & Van der Willigen, 1995; Fujisaki, Shimojo, Kashino, & Nishida, 2004; Hillock, Powers, & Wallace, 2011; Meredith, Nemitz, & Stein, 1987; Powers et al., 2009; Stein, Huneycutt, & Meredith, 1988; Vroomen, Keetels, de Gelder, & Bertelson, 2004; Zampini, Guest, Shore, & Spence, 2005). In addition, the TBW has proven to be a useful tool for identifying multisensory temporal processing abnormalities for basic stimuli (e.g., ring flashes and beeps) in clinical populations including individuals with autism (Foss-Feig et al., 2010; Kwakye, Foss-Feig, Cascio, Stone, & Wallace, 2011) and dyslexia (Hairston et al., 2005), and to speech stimuli in individuals with specific language impairment (Pons, Andreu, Sanz-Torrent, Buil-Legaz, & Lewkowicz, 2013). Finally, the TBW appears to be related to an individual's magnitude of multisensory integration, with the size of this temporal window being strongly correlated with the strength of their integration and binding (R. A. Stevenson et al., 2012). Thus, those with the narrowest binding windows show the largest gains of response, suggesting that better multisensory temporal acuity may serve to more strongly bind signals from multiple modalities.

The breadth of the temporal binding window also varies across stimuli, with audiovisual speech producing larger windows than arbitrary stimulus pairings such as flashes and beeps (e.g., Stevenson & Wallace, 2013; Vatakis & Spence, 2006). Several theories have been advanced to explain the disparity. Some have postulated that the breadth of the window is tied to the stimulus, noting that speech windows approximate the average duration of an English phoneme (van Wassenhove et al., 2007), but also see Munhall et al. (1996) who reported comparable windows for audiovisual speech of varied durations (fast, normal and clear- slow). Others have suggested that complex stimuli like speech might require further unisensory neural processing time before or during multisensory processing, reasoning that wider windows facilitate interactions between stimuli with potentially greater unisensory temporal processing offsets (Wallace & Stevenson, p. 258). Furthermore, it has been speculated that whereas basic signal binding involves analysis of lower-level physical stimulus factors such as spatial and temporal congruence, audiovisual speech binding is also influenced by higher-level factors such as semantics and talker gender (Ryan A. Stevenson, Wallace, & Altieri, 2014). Support for this theory comes from recent research demonstrating the interaction of timing and content (semantic) cues during a syllable perception task (ten Oever, Sack, Wheat, Bien, & van Atteveldt, 2013). Ten Oever and colleagues manipulated visual (place of articulation) and audio (voicing) cues while varying temporal synchrony and reported wider windows for congruent versus disparate audiovisual speech pairings, suggesting that stimuli with inherently stronger, more meaningful relationships may be bound at longer temporal offsets (2013). Similar effects of semantic influences on stimulus binding or “unity” have also been reported by others (e.g., Vatakis & Spence, 2007).

While there is an abundant literature describing stimulus and temporal effects shaping multisensory integration in adults, much less is known about the maturation of multisensory temporal acuity. Previous work in adults using the McGurk illusion found that the temporal constraints for the illusion span asynchronies ranging from approximately -50 to +200 ms, where negative and positive values indicate auditory leading and lagging conditions, respectively (Jones & Callan, 2003; D. W. Massaro et al., 1996; Munhall, Gribble, Sacco, & Ward, 1996; van Wassenhove et al., 2007). Although, to our knowledge, no studies have directly compared the multisensory temporal binding window for speech in younger and older participants, comparisons across work with infants, children and adults suggest that certain audiovisual temporal processing abilities mature over a protracted developmental timeline (e.g., Grant, van Wassenhove, & Poeppel, 2004; Hillock-Dunn & Wallace, 2012; Hillock et al., 2011; Lewkowicz, 1996; Lewkowicz, 2010; Lewkowicz & Flom, 2013; Munhall et al., 1996; van Wassenhove et al., 2007).

Prior work has shown that by four months of age babies can detect tempo, rhythm and synchrony in the auditory and visual modalities and can use these cues for crossmodal matching (Bahrick, 1983; Bahrick, 1987, 1988; Dodd, 1979; Lewkowicz, 1986, 1992, 1996, 2000; Mendelson & Ferland, 1982; Spelke, 1979). Studies also indicate that 4.5 to 5 month old infants can actively integrate cues across the different sensory modalities, and begin to demonstrate behavioral responses suggestive that they are fusing McGurk stimuli (Burnham & Dodd, 2004; Rosenblum, Schmuckler, & Johnson, 1997). For example, Burnahm and Dodd (2004) employed a looking paradigm and habituated 4.5 month old infants to a McGurk stimulus and control infants to an audiovisual /ba/. Unlike the control group, on test trials, the experimental group displayed longer visual fixation to /da/ and /tha/ (common perceptual byproducts arising from integrated McGurk presentations) compared to /ba/. These results suggest that the former stimuli were familiar and the latter was novel. However, despite the presence of multisensory abilities soon after birth (Bahrick & Pickens, 1994), multisensory development is far from complete in infancy. Lewkowicz and colleagues (1996) showed marked differences in the sensitivity of adults compared to infants to temporally asynchronous presentations of a bouncing green disc and a collision sound. Infants habituated to synchronous audiovisual presentations required temporal offsets

4-5 times as large as adult-reported asynchrony detection thresholds to demonstrate release from habituation when the video of the ball bouncing preceded the associated sound.

Recent work from Lewkowicz & Flom (2013) investigating audiovisual asynchrony detection of speech in 4-, 5-, and 6-year old children also provides evidence of ongoing refinement of multisensory processing of complex signals during childhood. They documented responses of 60 pre-school- and kindergarten-age children to a repeating audiovisual speech syllable presented either synchronously or at three temporal lags (366, 500, 666 ms). Children were asked if the face (lips) and voice in the video went together, and if asynchronous versus synchronous audiovisual tracks were the same or different. Results showed an emerging sensitivity to asynchrony with age; only older children were capable of detecting the smallest temporal offset (366 ms). Authors note that this developmental period, around the entrance to school, is associated with significant multisensory temporal processing gains. However, they cautiously conclude that further maturation is needed to reach adult-like competency based on cross-study comparisons including work reporting still smaller asynchrony thresholds to speech in adults (Grant et al., 2004; van Wassenhove et al., 2007). Previous work from our laboratory also showed age-related differences in the temporal processing of basic audiovisual stimuli and protracted maturation. Comparisons of performance between 10 and 11 year-old children and adults on an audiovisual simultaneity judgment task revealed significant differences in the probability of reporting simultaneity at moderate and long asynchronies (150 – 350 ms), with children being more likely to bind low-level stimuli separated by longer temporal delays (Hillock et al., 2011). A later cross-sectional study expanded on this, showing that maturational changes in the size of the multisensory TBW for binding low-level audiovisual stimuli (i.e., flashes and beeps) continue into adolescence (Hillock-Dunn & Wallace, 2012). A recent study from Downing and colleagues measuring responses of children, adolescents and adults on an object recognition task involving complex stimuli (i.e., animal pictures and sounds) presented alone or amidst distractors shows a more dichotomous developmental affect (Downing, Barutchu, & Crewther, 2014). All groups showed decreased reaction times to multisensory presentations, and the reaction times of adolescents were similar to those of adults. However, both children and adolescents exhibited significantly fewer race model violations than adults (greater than expected multisensory response acceleration based on unisensory reaction times), a finding indicative of residual processing immaturity.

These results suggest that the temporal architecture of audiovisual integration is changing during early and middle childhood, but little is known about potential ongoing development in late childhood and adolescence. Furthermore, although prolonged audiovisual temporal processing immaturities have been noted to basic stimuli, it is unknown whether such findings generalize to speech. Adult research has reported differences in asynchrony detection for speech and non-speech stimuli (Dixon & Spitz, 1980), and there has been speculation regarding the role of content-oriented processing of multisensory speech on the temporal window (Ryan A. Stevenson et al., 2014). Higher -level processing may influence the profile and timeline of mature audiovisual temporal perception in children leading to stimulus-specific differences in behavior. Given previous developmental research and stimulus considerations, in the current it was hypothesized that there would be a positive age effect on integrative probability and negative effect on window size. Children were expected to present with wider multisensory temporal speech-binding windows and reduced fusion compared to adults, with adolescents showing intermediary performance. To our knowledge, this study provides the first systematic examination of the temporal architecture of audiovisual binding across a wide range of developmental ages, from early childhood to adulthood using complex speech stimuli.

Methods

Participants

Participants were recruited via institution approved advertising materials and all participants and parents/ guardians of minors were assented and consented prior to study participation in accordance with the regulations of the Vanderbilt Institutional Review Board (IRB). Native English speakers with normal birth and prenatal history, no known medical issues or diagnoses, or learning problems were eligible to participate.

Sixty typically developing individuals (29 males, 31 females) between 6 and 24 years of age were recruited for the study. Participants were screened for hearing loss, uncorrected vision loss and intellectual ability. One child who presented with hearing loss and one adolescent with a below average intelligent quotient were eliminated due to failure to meet screening eligibility criteria. An additional child was excluded because he elected to discontinue participation before completing the experimental task. Of the 57 participants that were screened and tested, a further thirteen participants were excluded. Those excluded comprised 7 participants who had fusion rates below chance level (33 %) for simultaneous McGurk presentations, and 6 whose fusion perception did not change appreciably with increasing stimulus onset asynchrony (SOA). That is, average fusion at one or more SOAs did not decrease to below ¾ maximum as expected based on prior studies. Findings of reduced or atypical fusion patterns have also been reported previously (e.g., van Wassenhove et al., 2007; excluded fusion < 40%).

Data from the 44 participants included in the study (22 males, 22 females) was subdivided into groups to based on age: children (n= 16, 7 – 11.99 years), adolescents (n=14, 12 – 17.99 years), and adults (n=14, 18 - 24.0 years). Individual window size estimates were obtained in all 44 participants via function fitting in Matlab. Only data from individuals with medium to large R^2 values (≥0.3) indicating a modest to strong fit of the function to the data were included in window size comparisons (n = 35; 18 males, 17 females); this comprised 13 children, 10 adolescents and 12 adults.

The cut-offs used to determine study eligibility on initial screening measures were as follows: hearing sensitivity (pure tone thresholds < 25dB HL at octave frequencies from 250 to 8000 Hz), visual acuity (Snellen, 20/20 - 2 or better for each eye), and intellectual ability (i.e., Kaufmann Brief Intelligence, second edition – average or above average composite IQ [standard score >= 85]) (Kaufman & Kaufman, 2004). Screening measures took approximately 45 minutes to complete.

Experimental Conditions

The primary motivation of this study was to evaluate developmental differences in audiovisual integration of temporally synchronous and asynchronous speech syllables in background noise using the McGurk effect. However, to potentially control for individual differences in unisensory performance accuracy and audiovisual speech understanding, unisensory and synchronous multisensory congruent performance was also measured.

An audiovisual speech identification task comprising unisensory and multisensory experimental and control trials was run immediately after completion of an auditory adaptive staircase, which identified the level of noise needed to equate unisensory auditory speech recognition accuracy across participants. For both tasks participants were seated in a quiet, dimly lit room approximately 60 cm from a high refresh-rate computer monitor, which displayed video stimuli (NEC Multisync F3992, 100 Hz refresh rate). Auditory stimuli were presented via Sennheiser (HD 265 linear) supra-aural earphones. Stimulus presentation and data logging was controlled using MATLAB 7.7.0 R2008b software and responses were collected with a response pad (Cedrus RB-530). Stimulus duration and interstimulus delays were externally verified with an oscilloscope within an error tolerance of 5 ms (Hameg Instruments HM507), and auditory stimulus intensity was verified with a sound level meter (Larson Davis LxT2, 375A02 microphone).

Auditory Adaptive Staircase

A three-down, one up adaptive staircase was used to determine the amount of masking noise needed to achieve unisensory auditory accuracy scores of 79% correct on a closed-set, four-alternative forced choice (4AFC) identification task.

Stimuli

The auditory stimuli were comprised of recordings of four voiced consonant-vowel (CV) pairings produced by a native English-speaking male (i.e., /ba/, /ga/, /da/, /tha/, mean duration = 511 ms, range = 463 – 555 ms). The tokens sampled were the same as those used in the audiovisual speech identification task. They included those paired to induce the McGurk illusion, /ba/ and /ga/, as well as the most commonly reported (fused) McGurk percepts, /da/ and /tha/. Auditory tokens were embedded in white noise to degrade the salience of the recordings and thereby increase the frequency of McGurk illusion perception (Soto-Faraco & Alsius, 2009; Sumby & Pollack, 1954). The tokens were presented at a fixed level of 56 dB SPL (+/- 0.5 dB SPL, all tokens), and the level of the broadband noise was individually determined for each participant using an adaptive staircase procedure described below. The staircasing procedure was implemented to control for developmental differences in susceptibility to competing background sound on auditory speech recognition measures (e.g.,Hall, Grose, Buss, & Dev, 2002).

Procedure

Prior to testing, participants were asked to point to the button on the response pad depicting each consonant-vowel (CV) (recited aloud by the experimenter) to verify their ability to correctly associate the visual and auditory representations of the stimuli. All participants were able to successfully identify the tokens. Once the staircase was initiated, participants saw a white crosshair fixation marker on a black screen, which appeared before and after each unisensory auditory presentation. Auditory presentations were paired with a static image of the speaker's face. The speech tokens were played at an average constant level of 56 dB SPL (+/- 0.5) and the intensity of the broadband noise was varied. The starting level of the noise was 50 dB SPL and the initial step size (i.e., amount by which noise level is changed) of 4 dB was decreased to 2 dB after the second reversal. Two independent tracks were run simultaneously (trials interleaved and alternating) and the final noise level was given by the average of the thresholds produced by each track. In instances where only one track terminated normally (8 reversals were not completed, n=2), the noise level was set based on the threshold produced from the terminated track only. Total test time for the staircase was approximately 5 minutes.

Audiovisual Speech Task

The audiovisual speech task was adapted from the seminal study by McGurk and McDonald (1976). It was a four-alternative forced choice task comprised of the following four randomly interleaved stimulus conditions: 1) auditory only, 2) visual only, 3) audiovisual congruent, 4) audiovisual incongruent (McGurk trials). For McGurk stimuli, auditory and video tokens were presented synchronously and asynchronously at a range of audiovisual temporal delays.

Stimuli

The stimuli were comprised of four consonant vowel (CV) pairs (/ba/, /ga/, /da/, /tha/) recorded from a native English-speaking male and played in unisensory and multisensory formats. Video images were recorded at 30 frames per second (fps) with a resolution of 640 x 480 pixels. Auditory stimuli had a 48 kHz sampling rate and 768 kbps bit rate. The auditory speech tokens were presented at 56 dB SPL (+/- 0.5 dB) amidst a background of white noise played which played continuously throughout the test at the pre-determined level established via the auditory adaptive staircase. Unisensory visual presentations showed the male talker mouthing the syllables, whereas unisensory auditory presentations were paired with a static image of the talker's face. Audio and video tracks of multisensory congruent trials (e.g., auditory /ba/, visual /ba/) were presented synchronously. Multisensory incongruent McGurk trials were played at the following visual leading audio stimulus onset asynchronies (SOAs): 0, 50, 100, 150, 200, 250, 300, 400 ms and 500 ms (Figure 1). Observations were restricted to visual leading asynchronies, where fusion has historically been more readily observed across a wider range of delays (e.g., Jones & Jarick, 2006; van Wassenhove et al., 2007). Auditory leading presentations were omitted to reduce task time given the reduced capacity for sustained attention seen in younger children (e.g., Lin, Hsiao, & Chen, 1999). Hence, window estimates comprise only one side of the traditional temporal binding window distribution, reflecting responses to more ethologically valid stimulus combinations (Hillock-Dunn & Wallace, 2012). McGurk videos were created using Adobe Premier Pro CC video editing software. For the synchronous McGurk trials (0 ms SOA), the acoustic burst of the /ba/ audio track was aligned with the burst from the /ga/ soundtrack in the video. For asynchronous McGurk presentations, the desired stimulus onset asynchronies (SOAs) were created by copying and pasting the pre-vocalization recording to the beginning of the audio track.

Procedure

Before beginning the task, all participants viewed a computer-generated story containing instructions for completing the experiment. Pilot testing showed that this promoted understanding in younger participants (Appendix 1). Participants were asked to watch and listen to the recordings and press the button on the response pad corresponding to the CV that they thought the man produced. A practice comprised of seven trials was completed to familiarize participants with the task. Practice could be repeated up to two times before beginning the experiment.

During the experiment, a white crosshair fixation marker (1.9 cm x 1.9 cm) appeared in the center of a black background on the computer screen prior to and following each trial. Trials were initiated 1 second after participants logged their response to the preceding presentation. Unisensory and multisensory congruent and incongruent trials were randomly interleaved and broadband noise was played throughout all presentations. For unisensory conditions, each of the four tokens (/ba/, /ga/, /da/, /tha/] were sampled nine times for a total of 36 presentations/ condition. The total number of multisensory congruent trials (e.g., auditory /ba/, visual /ba/) was also 36 (9 samples/ token). The multisensory incongruent McGurk trials (i.e., auditory /ba/, visual /ga/), which were of greatest interest for the current study, were sampled more frequently. Eighteen samples were obtained at each of the nine stimulus onset asynchronies (SOAs) for a total of 162 McGurk presentations. The experiment was comprised of 270 trials across all conditions and took approximately 30 – 40 minutes to complete. Visual puzzles displaying the percent of finished trials (i.e., 25%, 50%, 75% and 100%) appeared four times during the experiment and became progressively more complete. Participants were given the option of a short break lasting 5 minutes or less at each interval.

Analysis

Accuracy scores on unisensory and congruent multisensory presentations were determined using proportions (correct responses/ total trials). A one-way between-subjects analysis of variance (ANOVA) and follow-up *t*-tests were used to explore the group differences in auditory and audiovisual discrimination ability. A Welch test for equality of means (Welch, 1947) was used to evaluate differences in lipreading proficiency, which corrected for violation of the error variance assumption (Levene, 1960). Follow-up (adjusted) *t*-tests were used to elaborate group differences while accounting for unequal group variance (Satterthwaite, 1946).

Performance on incongruent audiovisual McGurk trials was evaluated in several ways. First, a partial correlation was computed between age and average fusion on McGurk trials across all SOAs, which controlled for visual-only ability. This was used to evaluate whether any significant relationship could be attributed to multisensory processing versus visual perceptual processing differences. Descriptive statistics were used to explore group-average differences in the probability of fusion and of reporting either the auditory or visual token on unfused McGurk trials (averaged SOAs).

Next, developmental changes in the temporal dependency of McGurk illusion perception was investigated by comparing the rate of fusion at each SOA across groups. A corrected mixed-model analysis of variance (ANOVA) with a within-subjects factor of SOA condition (9 levels: 0, 50, 100, 150, 200, 250, 300, 400 ms and 500 ms) and between-subjects factor of age group (3 levels: children, adolescents, adults) was performed (Greenhouse & Geisser, 1959). Follow-up (adjusted) *t*-tests were used to compare average fusion between groups (Satterthwaite, 1946).

To establish a global metric of temporal processing, the multisensory temporal binding window was computed on distributions fitted to individual and group-averaged McGurk responses. To calculate window size, the proportion of fusion responses was plotted for each SOA, and the best fitting 3-parameter sigmoid function was computed over the range 0 to 500 ms using the Matlab function nlinfit. A three parameter sigmoid function was selected to fit the data because it most logically mimicked the nature of multisensory interactions based on the published literature. That is, audiovisual fusion is greatest for shorter SOAs and decreases with a monotonic sigmoidal shape with increasing SOA (e.g., Jones & Jarick, 2006; van Wassenhove et al., 2007). The temporal window was established as the SOA corresponding to ¾ of the maximum fusion value from the fitted function (Powers et al., 2009; Hillock et al., 2011; Hillock-Dunn & Wallace, 2012). A one-way ANOVA was used investigate group differences in window size, and window size estimates were descriptively compared.

Results

The participants included in the analyses (n = 44) ranged in age from 7.1 to 23.6 years. For purposes of comparison, groups were divided into 16 children (7.1 – 11.9 years, *M* = 9.9 years), 14 adolescents (12.3 - 17.4 years, *M* = 15.0 years) and 14 adults (18.1 – 23.6, *M* = 21.2 years). For the analyses of temporal window size (see Methods), only individuals with R^2 values ≥ 0.3 were included. This sample (n = 35) comprised 13 children (*M* = 10.2 years), 10 adolescents (*M* = 15.1 years) and 12 adults (*M* = 21.4 years). Data from over ¾ of children (81%) and adults (86%) met the R^2 criterion for inclusion in the temporal window analysis, compared to roughly 71% of adolescents. However, average R^2 values, indicating the strength of the fit to data were lower for children (*M* = 0.595) compared to other groups (adolescents: *M* = 0.773, adults: *M* = 0.784).

Unisensory and congruent multisensory performance

In order to properly interpret developmental changes in the perception of the McGurk illusion, it was first necessary to examine changes in unisensory and multisensory speech recognition accuracy that could influence McGurk performance. Group-average accuracy scores for the auditory only, visual only and congruent multisensory conditions appear in Figure 2. Since a staircase procedure was used to individually determine the masking level needed to produce equivalent auditory only performance, not surprisingly there were no differences in auditory only accuracy across groups ($p > 0.05$). Mean accuracy scores were 0.732, 0.763, and 0.685 for children, adolescents and adults, respectively. The range of individualized SNRs employed varied from +8.5 to – 12.2 dB signal-to-noise ratio (SNR) (*M* = - 6.1 dB SNR), with younger participants generally requiring more favorable (positive) SNRs. There was also no significant difference in congruent audiovisual speech recognition accuracy scores ($p > 0.05$). All groups performed well above chance levels (children: *M* = 0.771, adolescents: *M* = 0.816, adults: *M* = 0.736) (Figure 2), suggesting that even younger participants were able to successfully complete the task. In contrast to auditory only and audiovisual conditions, a significant group difference was observed for visual only accuracy scores (*Welch's* $F_{2,\,26.253}$ = 4.554, *p* = 0.020). Follow-up tests indicated that the children (*M* = 0.436) performed more poorly than adults (*M* = 0.538) ($t_{23.220}$ = -2.787, *p* = 0.010), but adolescents (*M* = 0.746) did not differ from the other groups (p > 0.05).

Factors influencing McGurk performance

The primary goal of the current work was to investigate age-related behavioral differences in the temporal constraints for integrating audiovisual speech signals, taking advantage of the McGurk illusion. However, it is known that differences in unisensory perception can influence the likelihood of multisensory integration, the magnitude of responses gains, and the probability of reporting the auditory versus visual token on unfused McGurk trials (e.g., Hockley & Polka, 1994). Hence, a partial correlation between age and average fusion that controlled for differences in visual-only ability was performed. As predicted, results showed a significant positive relationship between age and fusion suggesting that differences in multisensory integration cannot be fully attributed to differences seen in visual processing (r_{41} = 0.385, *p* = 0.011) (Figure 3A). Overall adults (*M* = 0.77) perceived the illusion on nearly 20% more trials than children (*M* = 0.58), with adolescents (*M* = 0.64) showing slightly higher fusion than children (Figure 3B). On unfused trials, the younger groups were more than three times as likely to report the auditory (A) token (children: *M* = 0.33, adolescents: *M* = 0.28) than the visual (V) token (children: *M* = 0.09, adolescents: *M* = 0.08), whereas these reports were much more balanced in adults (Vis *M* = 0.10, Aud *M* = 0.13)(Figure 3B).

Age and the temporal profile of McGurk performance

To investigate maturational changes in the temporal dependency of illusion perception, group differences in the probability of fusion on McGurk trials were explored. Data were analyzed using a mixed-model ANOVA with a within-subjects factor of SOA and between-subjects factor of group. Degrees of freedom were Greenhouse Geisser adjusted to correct violation of the sphericity assumption (Mauchley's test). Results revealed significant main effects of group ($F_{2, 41}$ = 6.975, p = 0.002) and SOA ($F_{5.31, 219.1}$ = 54.623, p = 0.000), but no significant interaction (p > 0.05). Follow-up tests comparing average fusion between groups revealed that fusion was significantly reduced in children compared to adults ($t_{27.255}$ = -3.862, p = 0.001), and adolescents relative to adults ($t_{25.683}$ = -2.318, p = 0.029), but there was no difference between the younger groups (p > 0.05). The non-significant Group × SOA interaction effect, however, was quite surprising in that it suggests that audiovisual speech binding is similarly affected by temporal asynchrony in all three age groups.

To further investigate maturational changes in the temporal constraints of audiovisual speech binding, temporal binding window (TBW) size was computed from functions fitted to individual and group-average fusion data in a subset of participants (n = 35; see methods for additional detail). Results of one-way ANOVA showed no significant group difference in window size (p > 0.05). Group-averaged window values representing the breadth of time (ms) over which fusion was likely to be observed was strikingly similar in children (*M* = 315, *SE* = 37.14), adolescents (*M* = 301, *SE* = 27.18), and adults (*M* = 320 ms, *SE* = 35.36). Collectively, these findings converge on the overall conclusion that audiovisual speech perception is similarly affected by temporal asynchrony in each of the three studied age groups, and suggest that the temporal binding window for audiovisual speech stimuli is surprising mature (i.e., adult-like) by the age of 7.

Discussion

There are two key findings to the current study. First, and as has been previously reported (D. W. Massaro, 1984; Dominic W Massaro, Thompson, Barron, & Laren, 1986; McGurk & Macdonald, 1976; Tremblay et al., 2007), perceptual fusions such as those indexed by the McGurk effect become increasingly common as development progresses, and this effect appears to be partly due to the increasing influence of visual stimuli on speech processing. Second, and quite surprising given findings from our earlier work showing protracted development of audiovisual temporal processing for simple non-linguistic stimuli (Hillock-Dunn & Wallace, 2012; Hillock et al., 2011), the temporal constraints for binding auditory and visual speech elements did not change significantly with age. Based on these findings, as well as those from Lewkowiz and Flom (2013) showing the emergence of increased sensitivity to asynchronously presented syllables around 5 years of age, it is theorized that the temporal window for binding audiovisual speech cues contracts rather rapidly in early childhood.

Although all groups performed equally well on congruent audiovisual speech pairings, the younger groups showed significantly less overall perceptual fusion than adults on incongruent synchronous and asynchronous audiovisual McGurk stimuli. These results complement and extend previous findings from McGurk and Macdonald (1976) and others (e.g., D. W. Massaro, 1984; Dominic W Massaro et al., 1986; Tremblay et al., 2007) showing decreased fusion in children for synchronous McGurk presentations, and increased visual report by adults on unfused McGurk trials. This effect is believed to arise from a reduced capacity for multisensory integration in younger participants, as well as reduced visual perceptual accuracy and visual attention. Modeling studies suggest that the final product of a multisensory interaction is based on the statistical weighting of the combined unisensory signals, with these weights being inversely proportional to stimulus reliability (Ernst, 2007; Helbig & Ernst, 2007; Witten & Knudsen, 2005). Generalizing this to the current work, the positive correlation between age and visual-only perception on control trials suggests that the visual signal is potentially less reliable or salient for younger children. This may affect a decrease in the visual contribution to bimodal speech perception and a concomitant increase the relative weight or reliance on the auditory signal in children (Dominic W Massaro et al., 1986).

Age-related improvements in speech reading accuracy have been reported elsewhere in the literature (e.g., Hnath-Chisolm, Laipply, & Boothroyd, 1998; Kyle, Campbell, Mohammed, Coleman, & MacSweeney, 2013; Tye-Murray, Hale, Spehar, Myerson, & Sommers, 2013), which could correspond with changes in cue reliance and attention. In the current study, developmental growth of visual perceptual processing ability was accompanied by a reduction in auditory dominance. Whereas children and adolescents were more apt to report the auditory token on unfused trials and experienced less overall fusion, adults exhibited a roughly equal probability of reporting auditory or visual percepts on unfused presentations. Developmental studies have elaborated changes in auditory versus visual preferential processing and suggest that infants and young children tend to show greater auditory preference even for nonspeech sounds (Lewkowicz, 1988a, 1988b; Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003). For example, infants habituated to a flashing checkboard and pulsing sound detected temporal changes in auditory stimuli at 6- and 10-months of age, but even older infants were unable to consistently detect temporal changes in the visual realm (Lewkowicz, 1988a, 1988b). Similarly, Robinson and Sloutsky (2004) reported that 4 year-olds and adults trained to use auditory and visual stimuli (A1, V1; A2, V2) to determine the location of animated animals showed different patterns of cue reliance during test phases when auditory and visual cue combinations were switched (A1,V2; A2,V1). Whereas children shifted reliance on predictor cues across modalities, adults showed greater visual reliance. Although children demonstrated the capacity to accurately process visual cues, especially when attention was directed toward the non-preferred modality, they were less likely to be encoded when concomitantly presented with auditory information (Robinson & Sloutsky, 2004). It has been theorized that early preferential auditory processing or preference may be due to the enormous importance of auditory word learning in early development (Sloutsky & Napolitano, 2003).

Two interrelated factors also potentially contributing to differences in audiovisual speech binding are top-down attentional processes and memory, which undergo protracted developmental changes. Gathercole (1999) provides an excellent review of developmental increases in verbal (speech based) short-term and working memory capacity between pre-school and adolescence. Similarly, Klingberg and colleagues

(2002) reported increased visual working memory capacity in older children and adolescents (13-18 years) compared to younger children (9-12 years) and found positive correlations between working memory ability and activation in cortical areas (superior frontal and intraparietal cortex) associated with these attentional processes. A recent review suggests that improvements in attentional control during childhood into adolescence may drive or at least partly explain increases in visual short-term and working memory capacity (Astle & Scerif, 2011). In the context of the current work, this literature suggests that group differences may arise from reduced capacity of younger groups to simultaneously or sequentially process and maintain visual and auditory speech representations, or due to alterations in negotiation of discrepancies in multisensory stimulus representations and assignment of stimulus relevancy. Hence, immature higher-level cognitive processes may underlie age-related differences in speech binding and unfused perceptual report.

Despite differences in overall fusion and auditory versus visual perceptual bias, the temporal constraints within which children bind audiovisual speech appears mature by 7 years of age. Although we were unable to establish the specific age at which mature processing emerged in the current study, we can draw upon findings from similar work to posit the maturational time course. Research by Lewkowicz and others suggests that infants and 4 year-olds demonstrate a similar ability to detect asynchrony in auditory leading audiovisual speech presentations, but that 5- and 6 year-olds are more sensitive to shorter interstimulus delays (Lewkowicz, 2010; Lewkowicz & Flom, 2013). Despite improved sensitivity in the older children, continued maturation was theorized given that 6 year-olds still exhibited larger windows compared to those reported in adults (Lewkowicz & Flom, 2013). Based on these findings and results from the current work, it appears that window consolidation for audiovisual speech occurs throughout early childhood reaching full maturity around 7 years of age.

Interestingly, this finding contrasts earlier work in our laboratory showing a more prolonged developmental time course for realization of adult-like multisensory temporal processing of basic stimuli (Hillock-Dunn & Wallace, 2012). Additional research is needed to investigate the basis for apparent differences in processing of linguistic and non-linguistic stimuli, and the emergence of mature audiovisual temporal perception of speech during childhood versus adolescence for low-level stimuli (flashes and tone pips) (Hillock-Dunn & Wallace, 2012; Hillock et al., 2011). The more rapid window consolidation seen for speech may be related to system specialization for language, a process that may be dictated by the greater ethological relevance of speech signals and the accelerated maturation of the neural structures subserving speech processes. In an event-related potentials (ERP) study involving 3-16 year olds, Pang and Taylor (2000) reported that the N1c ERP component to speech was consistent with the adult morphology by 7-8 years of age, but that this same component was not clearly identifiable until 11-12 years for tones, after which time its amplitude continued to increase into adulthood. Speech and nonspeech stimuli have been shown to produce distinct in neural activation patterns in adults (Vouloumanos, Kiehl, Werker, & Liddle, 2001). Perhaps maturational differences in speech- versus nonspeech-evoked ERP responses in children reflect developmental differences in the neural generators for these different stimuli, with the generators for speech stimuli potentially leading in maturation due to the importance of these stimuli for communication.

The maturation of cortical structures and changes in neurocognitive processing during childhood and adolescence can impact integrative probability and have implications for overall functioning. Audiovisual interactions play an important role in role in facilitating stimulus detection, speech recognition, etc. Hence, developmental differences in receptive communication may be particularly apparent in noisy settings such as the classroom, where reductions in the integrity of the auditory signal and reduced audiovisual integration may render children and adolescents less capable of managing the deleterious effects of masking noise on speech understanding. On the other hand, multisensory temporal processing for speech-related signals appears to be relatively mature in early childhood, which may be critically important for reading skill development. Abnormal multisensory temporal processing has been reported in individuals with reading impairment (Hairston et al., 2005; Rose, Feldman, Jankowski, & Futterweit, 1999), and it has been speculated that reductions in multisensory temporal resolution (i.e., increased binding windows) may lead to ambiguities in grapheme-phoneme correspondences. The current study suggests that the basis for the development of normal reading skills are in place at a relatively early age, and that window consolidation may coincide with the onset of formal reading instruction. However, future work should further explore the link between multisensory temporal processing and reading ability, and audiovisual integration and speech recognition ability in younger and older individuals.

Acknowledgments

This work was supported by NIH R21 CA183492, the Vanderbilt Kennedy Center for Research on Human Development and the American Speech Language and Hearing Association. We also wish to thank Matthew Hevey, Nicky Hackett, Olivia Broaddus and Haley Eberly for their assistance with data collection.

References

Astle, D. E., & Scerif, G. (2011). Interactions between attention and visual short-term memory (VSTM): What can be learnt from individual and developmental differences? *Neuropsychologia, 49*(6), 1435-1445.

Bahrick, L. E. (1983). Infants' perception of substance and temporal synchrony in multimodal events. *Infant Behavior & Development, 6*, 429-451.

Bahrick, L. E. (1987). Infants' intermodal perception of two levels of temporal structure in natural events. *Infant Behavior & Development, 10*, 387-416.

Bahrick, L. E. (1988). Intermodal learning in infancy: learning on the basis of two kinds of invariant relations in audible and visible events. *Child Dev, 59*(1), 197-209.

Bahrick, L. E., & Pickens, J. N. (1994). Amodal relations: The basis for intermodal perception and learning in infancy. In D. J. L. R. Lickliter (Ed.), *The development of intersensory perception: Comparative perspectives* (pp. 205-233). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology, 45*(4), 204-220.

Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *J Cogn Neurosci, 16*(6), 1000-1009.

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception, 9*(6), 719-721.

Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology, 11*(4), 478-484.

Downing, H. C., Barutchu, A., & Crewther, S. G. (2014). Developmental trends in the facilitation of multisensory objects with distractors. *Front Psychol, 5*, 1559.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *J Speech Hear Res, 12*(2), 423-425.

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

Fixmer, E., & Hawkins, S. (1998). *The influence of quality of information on the McGurk effect.* Paper presented at the Australian Workshop on Auditory-Visual Speech Processing.

Foss-Feig, J. H., Kwakye, L. D., Cascio, C. J., Burnette, C. P., Kadivar, H., Stone, W. L., et al. (2010). An extended multisensory temporal binding window in autism spectrum disorders. *Exp Brain Res, 203*(2), 381-389.

Frens, M. A., Van Opstal, A. J., & Van der Willigen, R. F. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Percept Psychophys, 57*(6), 802-816.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nat Neurosci, 7*(7), 773-778.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences, 3*(11), 410-419.

Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *J Acoust Soc Am, 108*(3 Pt 1), 1197-1208.

Grant, K. W., van Wassenhove, V., & Poeppel, D. (2004). Detection of auditory (cross-spectral) and auditory–visual (cross-modal) synchrony. *Speech Communication, 44*(1–4), 43-53.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*(2), 95-112.

Hairston, W. D., Burdette, J. H., Flowers, D. L., Wood, F. B., & Wallace, M. T. (2005). Altered temporal profile of visual-auditory multisensory interactions in dyslexia. *Exp Brain Res, 166*(3-4), 474-480.

Hall, J. W., 3rd, Grose, J. H., Buss, E., & Dev, M. B. (2002). Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. *Ear Hear, 23*(2), 159-165.

Helbig, H. B., & Ernst, M. O. (2007). Optimal integration of shape information from vision and touch. *Exp Brain Res, 179*(4), 595-606.

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

Hnath-Chisolm, T. E., Laipply, E., & Boothroyd, A. (1998). Age-Related Changes on a Children's Test of Sensory-Level Speech Perception Capacity. *Journal of Speech, Language, and Hearing Research, 41*(1), 94-106.

Hockley, N. S., & Polka, L. (1994). A developmental study of audiovisual speech perception using the McGurk paradigm. *The Journal of the Acoustical Society of America, 96*(5), 3309-3309.

Jones, J. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *NeuroReport, 14*(8), 1129-1133.

Jones, J. A., & Jarick, M. (2006). Multisensory integration of speech signals: the relationship between space and time. *Exp Brain Res, 174*(3), 588-594.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaugman Brief Intelligence Test, Second Edition*. Minneapolis, MN.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Increased Brain Activity in Frontal and Parietal Cortex Underlies the Development of Visuospatial Working Memory Capacity during Childhood. *Journal of Cognitive Neuroscience, 14*(1), 1-10.

Kwakye, L. D., Foss-Feig, J. H., Cascio, C. J., Stone, W. L., & Wallace, M. T. (2011). Altered auditory and multisensory temporal processing in autism spectrum disorders. *Front Integr Neurosci, 4*, 129.

Kyle, F. E., Campbell, R., Mohammed, T., Coleman, M., & MacSweeney, M. (2013). Speechreading Development in Deaf and Hearing Children: Introducing the Test of Child Speechreading. *Journal of Speech, Language, and Hearing Research, 56*(2), 416-426.

Levene, H. (1960). Robust tests for equality of variances1. *Contributions to probability and statistics: Essays in honor of Harold Hotelling, 2*, 278-292.

Lewkowicz, D. J. (1986). Developmental changes in infants' bisensory response to synchronous durations. *Infant Behavior & Development, 9*(3), 335-353.

Lewkowicz, D. J. (1988a). Sensory dominance in infants: I. Six-month-old infants' response to auditory-visual compounds. *Developmental Psychology, 24*(2), 155-171.

Lewkowicz, D. J. (1988b). Sensory dominance in infants: II. Ten-month-old infants' response to auditory-visual compounds. *Developmental Psychology, 24*(2), 172-182.

Lewkowicz, D. J. (1992). Infants' response to temporally based intersensory equivalence: The effect of synchronous sounds on visual preferences for moving stimuli. *Infant Behavior & Development, 15*(3), 297-324.

Lewkowicz, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception & Performance, 22*(5), 1094-1106.

Lewkowicz, D. J. (2000). Infants' perception of the audible, visible and bimodal attributes of multimodal syllables. *Child Development, 71*(5), 1241-1257.

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

Lewkowicz, D. J., & Flom, R. (2013). The Audiovisual Temporal Binding Window Narrows in Early Childhood. *Child Development*, n/a-n/a.

Lin, C. C., Hsiao, C. K., & Chen, W. J. (1999). Development of sustained attention assessed using the continuous performance test among children 6-15 years of age. *J Abnorm Child Psychol, 27*(5), 403-412.

Macleod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology, 21*(2), 131 - 141.

Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Dev, 55*(5), 1777-1788.

Massaro, D. W., Cohen, M. M., & Smeele, P. M. (1996). Perception of asynchronous and conflicting visual and auditory speech. *J Acoust Soc Am, 100*(3), 1777-1786.

Massaro, D. W., Thompson, L. A., Barron, B., & Laren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. *Journal of Experimental Child Psychology, 41*(1), 93-113.

McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748.

Mendelson, M. J., & Ferland, M. B. (1982). Auditory-Visual Transfer in Four-Month-Old Infants. *Child Development, 53*(4), 1022-1027.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci, 7*(10), 3215-3229.

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Percept Psychophys, 58*(3), 351-362.

Pang, E. W., & Taylor, M. J. (2000). Tracking the development of the N1 from age 3 to adulthood: an examination of speech and non-speech stimuli. *Clin Neurophysiol, 111*(3), 388-397.

Pons, F., Andreu, L., Sanz-Torrent, M., Buil-Legaz, L., & Lewkowicz, D. J. (2013). Perception of audio-visual speech synchrony in Spanish-speaking children with and without specific language impairment. *J Child Lang, 40*(3), 687-700.

Powers, A. R., 3rd, Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *J Neurosci, 29*(39), 12265-12274.

Robinson, C. W., & Sloutsky, V. M. (2004). Auditory Dominance and Its Change in the Course of Development. *Child Development, 75*(5), 1387-1401.

Rose, S. A., Feldman, J. F., Jankowski, J. J., & Futterweit, L. R. (1999). Visual and auditory temporal processing, cross-modal transfer, and reading. *J Learn Disabil, 32*(3), 256-266.

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Percept Psychophys, 59*(3), 347-357.

Satterthwaite, F. E. (1946). An Approximate Distribution of Estimates of Variance Components. *Biometrics Bulletin, 2*(6), 110-114.

Schwartz, J. L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition, 93*(2), B69-78.

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neurosci Res, 47*(3), 277-287.

Sekiyama, K., & Tohkura, Y. i. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *The Journal of the Acoustical Society of America, 90*(4), 1797-1805.

Sloutsky, V. M., & Napolitano, A. C. (2003). Is a Picture Worth a Thousand Words? Preference for Auditory Modality in Young Children. *Child Development, 74*(3), 822-833.

Soto-Faraco, S., & Alsius, A. (2009). Deconstructing the McGurk-MacDonald illusion. *J Exp Psychol Hum Percept Perform, 35*(2), 580-587.

Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology, 15*, 626-636.

Stein, B. E., Huneycutt, W. S., & Meredith, M. A. (1988). Neurons and behavior: the same rules of multisensory integration apply. *Brain Res, 448*(2), 355-358.

Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus dependencies. *Exp Brain Res, 227*(2), 249-261.

Stevenson, R. A., Wallace, M. T., & Altieri, N. (2014). The interaction between stimulus factors and cognitive factors during multisensory integration of audiovisual speech. *Frontiers in Psychology, 5*, 352.

Stevenson, R. A., Wilson, M. M., Powers, A. R., & Wallace, M. T. (2013). The effects of visual training on multisensory temporal processing. *Exp Brain Res, 225*(4), 479-489.

Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *J Exp Psychol Hum Percept Perform, 38*(6), 1517-1529.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America, 26*(2), 212-215.

ten Oever, S., Sack, A. T., Wheat, K. L., Bien, N., & van Atteveldt, N. (2013). Audio-visual onset differences are used to determine syllable identity for ambiguous audio-visual stimulus pairs. *Frontiers in Psychology, 4*, 331.

Tremblay, C., Champoux, F., Voss, P., Bacon, B. A., Lepore, F., & Theoret, H. (2007). Speech and non-speech audio-visual illusions: a developmental study. *PLoS One, 2*(1), e742.

Tye-Murray, N., Hale, S., Spehar, B., Myerson, J., & Sommers, M. S. (2013). Lipreading in School-age Children: The Roles of Age, Hearing Status, and Cognitive Ability. *J Speech Lang Hear Res*.

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., & Spence, C. (2007). Crossmodal binding: Evaluating the “unity assumption” using audiovisual speech stimuli. *Perception & Psychophysics, 69*(5), 744-756.

Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of Sounds in the Auditory Stream: Event-Related fMRI Evidence for Differential Activation to Speech and Nonspeech. *Journal of Cognitive Neuroscience, 13*(7), 994-1005.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Brain Res Cogn Brain Res, 22*(1), 32-35.

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

Welch, B. L. (1947). THE GENERALIZATION OF ‘STUDENT'S’ PROBLEM WHEN SEVERAL DIFFERENT POPULATION VARLANCES ARE INVOLVED. *Biometrika, 34*(1-2), 28-35.

Witten, I. B., & Knudsen, E. I. (2005). Why seeing is believing: merging auditory and visual worlds. *Neuron, 48*(3), 489-496.

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Percept Psychophys, 67*(3), 531-544.

Figures

FIGURE 1.

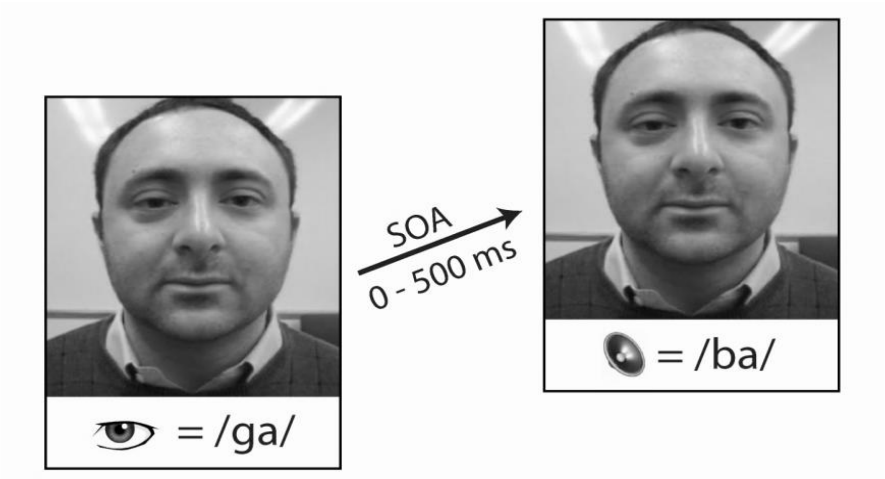

Figure 1. Visual representation of the trial structure for incongruent McGurk stimuli. For McGurk trials, the auditory track was presented either simultaneously with the onset of lip movement, or at the following stimulus onset asynchronies (SOAs): 50, 100, 150, 200, 250, 300, 400, 500 ms.

FIGURE 2.

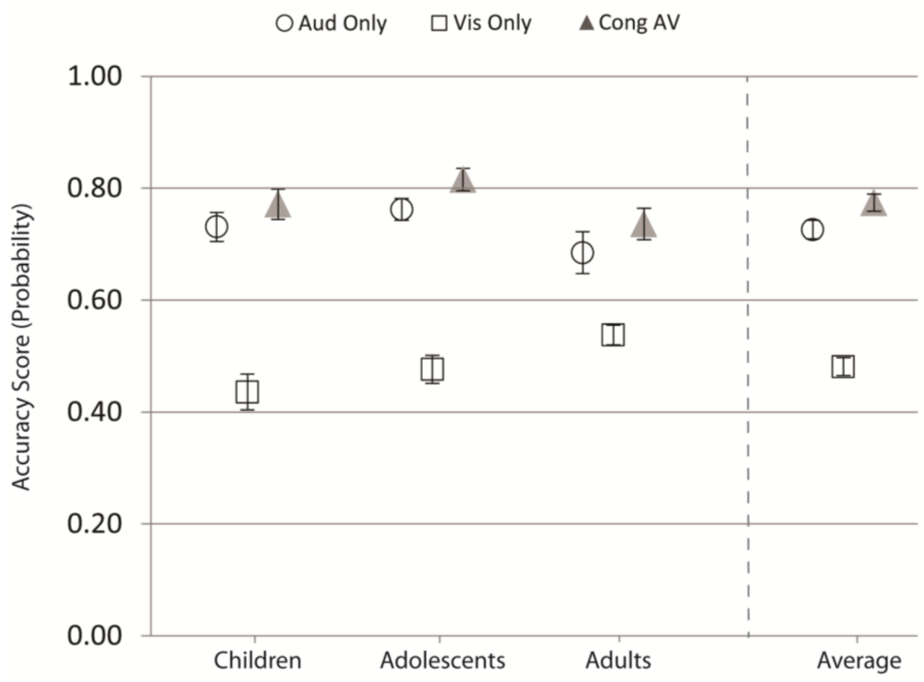

Figure 2. Group-average accuracy scores (calculated using the probability of correct responses) are shown for unisensory and multisensory congruent control trials. Open circles and squares represent the auditory only and visual only conditions, respectively, and filled triangles depict performance on multisensory congruent trials. Error bars represent +/- 1 standard error of the mean (SE).

FIGURE 3.

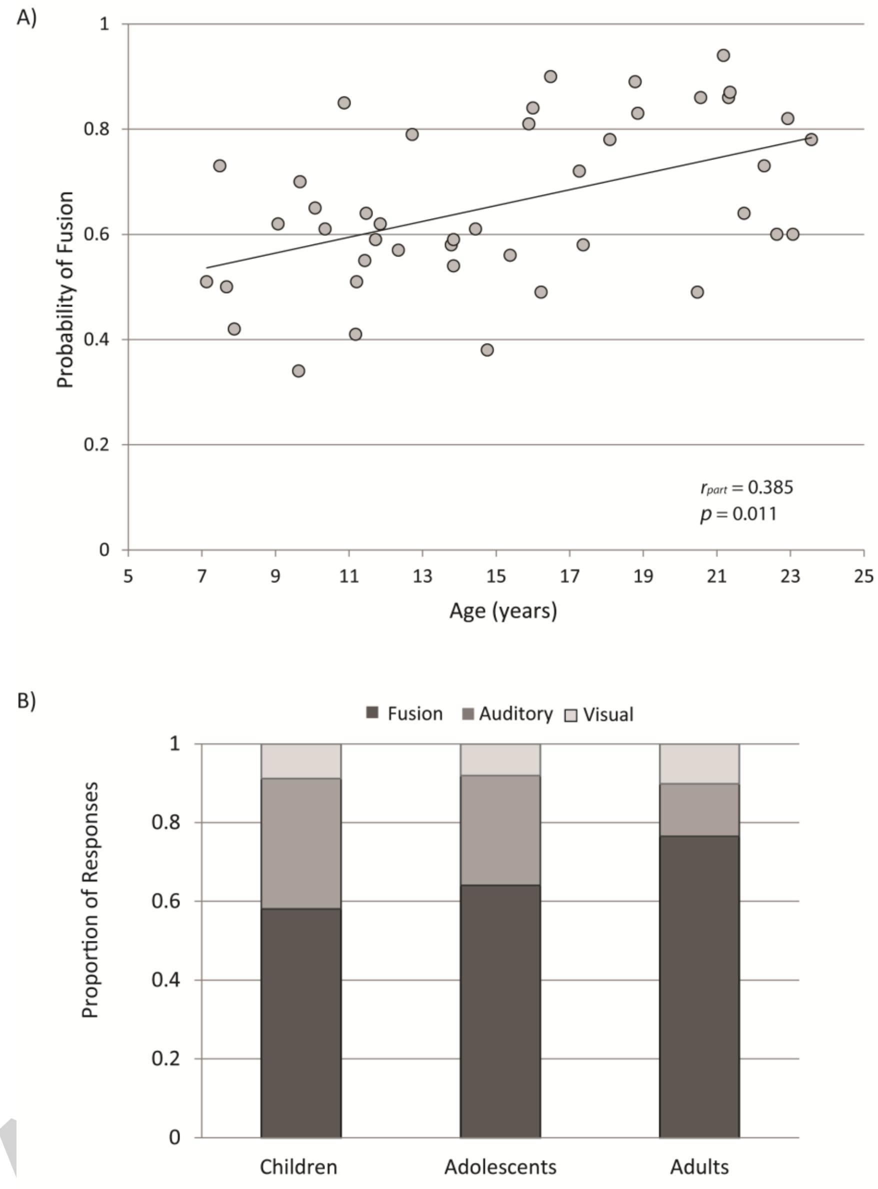

Figure 3 A/B. Older subjects experience the McGurk illusion more frequently than younger participants. A) Scatterplot showing positive relationship between participant age and the likelihood of fusion on McGurk trials (n= 44). **B)** Bar graph displaying the proportion of fused and unfused responses to McGurk trials (dark gray = fusion, medium gray = auditory, light gray = visual). Children and adolescents were much more likely to report the auditory token on unfused McGurk trials whereas unisensory report of adults was nearly balanced.

FIGURE 4.

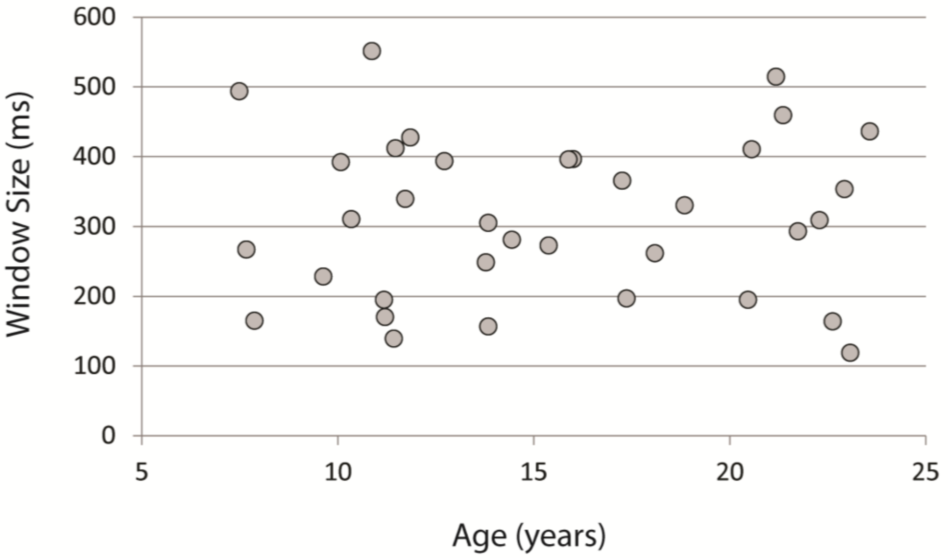

Figure 4. Individual window sizes estimates did not change with participant age (n = 35). Scatterplot of window size as a function of participant age. No age-related group difference was observed in window size. Notably, marked intersubject variability in window size estimates was observed across the age span.

Appendix

Appendix 1. Story script for audiovisual speech task.

Page 1: Billy goes to the movies every Friday. He usually arrives 10 minutes early to get a good seat so that he can see and hear everything.

Page 2: Last Friday Billy arrived a little late and almost all of the seats were taken. The seat he found was in the very back of the theater and it was hard for him to understand from where he was sitting.

Page 3: Please help Billy figure out what speech sounds the man in this movie is saying.

Instructions: You will see a plus sign to remind you to pay attention just before each movie clip starts. Watch the man's lips and listen to the sounds he makes. Sometimes you will hear him talking but his lips won't move. Sometimes you will see his lips moving but not hear his voice. Other times you will hear him and see his lips moving. After each movie clip is finished, press the button that shows what he said. If you aren't sure, take a guess. Do you best! Press any button to start the game.