HHS Public Access

Author manuscript

Brain Lang. Author manuscript; available in PMC 2015 December 01.

Published in final edited form as:

Brain Lang. 2014 December ; 139: 36–48. doi:10.1016/j.bandl.2014.09.011.

Audiovisual integration for speech during mid-childhood: Electrophysiological evidence

Natalya Kaganovich[1,2,*] and **Jennifer Schumaker**[1]

[1]Department of Speech, Language, and Hearing Sciences, Purdue University, 715 Clinic Drive, West Lafayette, IN 47907-2038

[2]Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907-2038

Abstract

Previous studies have demonstrated that the presence of visual speech cues reduces the amplitude and latency of the N1 and P2 event-related potential (ERP) components elicited by speech stimuli. However, the developmental trajectory of this effect is not yet fully mapped. We examined ERP responses to auditory, visual, and audiovisual speech in two groups of school-age children (7–8-year-olds and 10–11-year-olds) and in adults. Audiovisual speech led to the attenuation of the N1 and P2 components in all groups of participants, suggesting that the neural mechanisms underlying these effects are functional by early school years. Additionally, while the reduction in N1 was largest over the right scalp, the P2 attenuation was largest over the left and midline scalp. The difference in the hemispheric distribution of the N1 and P2 attenuation supports the idea that these components index at least somewhat disparate neural processes within the context of audiovisual speech perception.

Keywords

audiovisual speech perception; electrophysiology; child language development

1. Introduction

In the majority of cases, our experience of the world is multisensory in nature, and as children mature, they learn to match, detect various correspondences between, and integrate perception from different senses. Accumulating research suggests that these various sub-components of what is commonly referred to as “multisensory processing” may rely on at least partially disparate brain regions and have different developmental trajectories (e.g., Burr & Gori, 2012; Calvert, 2001; Stevenson, VanDerKlok, Pisoni, & James, 2011).

*Corresponding author: Department of Speech, Language, and Hearing Sciences, Purdue University, Lyles Porter Hall, 715 Clinic Drive, West Lafayette IN 47907-2038, Phone (765)494-4233, Fax (765)494-0771, kaganovi@purdue.edu.

On the one hand, some ability to match and combine different modalities appears to be present during the first year of life (for the purposes of this study, we will review only research on audiovisual processing because it is most relevant to speech perception; however, a number of studies examined development of interaction between other modalities as well (for a review, see Burr & Gori, 2012; Gori, Sandini, & Burr, 2012; Jaime, Longard, & Moore, 2014)). For example, 10–16 week old infants can detect temporal synchrony between lip movements and speech sounds (Dodd, 1979). By 3 months of age, infants can learn arbitrary pairings between faces and voices (Bahrick, Hernandez-Reif, & Flom, 2005; Brookes et al., 2001), and by 4 to 7 months of age, they are able to match faces and voices based on age (Bahrick, Netto, & Hernandez-Reif, 1998). Additionally, multiple studies reported the presence of at least some degree of audiovisual integration in infants as young as 4.5–5 months of age as revealed by their perception of the McGurk illusion (in which, typically, an auditory ‘pa’ or ‘ba’ is dubbed onto the visual articulation of ‘ka’ or ‘ga,’ resulting in the perception of ‘ta’ or ‘da’) (Burnham & Dodd, 2004; Kushnerenko, Teinonen, Volein, & Csibra, 2008; Rosenblum, Schmuckler, & Johnson, 1997). By recording event-related potentials (ERPs) to the auditory pronunciation of a vowel that either matched or mismatched the earlier visual articulation, Bristow and colleagues reported that 10–12 week old infants appear to have a cross-modal representation of phonemes and integrate auditory and visual information during early stages of perception, similar to what had previously been reported for adults (Bristow et al., 2008). Indeed, some visual speech skills of young infants may even surpass those of adults. As an example, Weikum and colleagues have demonstrated that 4- and 6-month-old (but not 8-month-old) infants are able to discriminate between two languages based on visual speech cues alone (Weikum et al., 2007).

On the other hand, a number of studies point to a protracted developmental course of certain aspects of audiovisual processing. For example, Lewkowicz reported that, compared to adults, infants require significantly longer separations between the onsets of auditory and visual stimuli, both in speech and non-speech contexts, in order to detect temporal asynchrony (Lewkowicz, 1996, 2010). In fact, sensitivity to audiovisual temporal offsets remains immature even during mid-childhood (Hillock-Dunn & Wallace, 2012; Hillock, Powers, & Wallace, 2011; Kaganovich, Schumaker, Leonard, Gustafson, & Macias, 2014). Further, although some ability to perceive the McGurk illusion appears to be present early on, multiple studies have documented a reduced susceptibility to the McGurk illusion in children compared to adults, suggesting that the ability to fully integrate auditory and visual speech cues does not mature until late childhood and depends, at least in part, on children’s experience with visual speech (Massaro, 1984; Massaro, Thompson, Barron, & Lauren, 1986; McGurk & MacDonald, 1976; Tye-Murray, Hale, Spehar, Myerson, & Sommers, 2014). Behavioral benefits of audiovisual processing, such as faster reaction time to audiovisual as compared to auditory only or visual only stimuli and better speech-in-noise perception, also either begin to manifest themselves or continue to improve during mid-childhood (Barutchu, Crewther, & Crewther, 2009; Barutchu et al., 2010; Ross et al., 2011).

Although a number of studies have begun charting developmental trajectories of various audiovisual skills beyond infancy, the absolute majority of studies on audiovisual processing have been done with either infants or adults, and relatively little is known about how its

development unfolds during mid-childhood. Additionally, the bulk of studies on audiovisual processing during school years rely on behavioral paradigms, which may be an unreliable measure of audiovisual processing in this population for several reasons. First, even when children's behavioral responses on an audiovisual task are the same as those of adults and suggest the presence of multisensory processing, the neural circuitry engaged by the task may nonetheless be different in children. As an example, the left posterior superior temporal sulcus (STS) has been shown to play an important role in audiovisual speech perception and in the ability to perceive the McGurk illusion in particular (Beauchamp, Nath, & Pasalar, 2010; Calvert, 2001; Nath & Beauchamp, 2012; Nath, Fava, & Beauchamp, 2011). However, a study by Nath and colleagues (Nath et al., 2011) reported that, compared to adults, some children had weak STS responses even when they perceived the McGurk illusion, thus deviating substantially from the adult pattern of neural activity associated with audiovisual processing of speech. Second, behavioral responses are the end result of many sensory, cognitive, and motor processes. Therefore, a lack of multisensory facilitation in reaction time, accuracy, or other behavioral measure may potentially be the result of the immature motor system, overall greater variability of responses in younger research participants (e.g., McIntosh, Kovacevic, & Itier, 2008; Williams, Hultsch, Strauss, Hunter, & Tannock, 2005), or the inability to apply audiovisual skills to a specific task. Lastly, audiovisual integration happens over different stages of speech processing, such as for example acoustic, phonemic, or lexical (Baart, Stekelenburg, & Vroomen, 2014; Hertrich, Mathiak, Lutzenberger, Menning, & Ackermann, 2007; Kaiser, Kirk, Lachs, & Pisoni, 2003). However, behavioral studies typically cannot provide information about the timing and nature of the neural mechanisms engaged by the task. Unlike behavioral measures alone, the ERP method, with its ability to track neural activity on a millisecond basis, allows one to focus on specific stages of audiovisual processing, often without a need for overt behavioral responses.

In the present study, we took advantage of a well-established electrophysiological paradigm in order to examine an early stage of audiovisual integration in school-age children and adults (Besle, Bertrand, & Giard, 2009; Besle et al., 2008; Besle, Fort, Delpuech, & Giard, 2004; Besle, Fort, & Giard, 2004; Giard & Besle, 2010; Knowland, Mercure, Karmiloff-Smith, Dick, & Thomas, 2014). The paradigm is based on the fact that during sensory processing (i.e., within approximately 200 ms post-stimulus onset), ERPs elicited by auditory and visual stimuli sum up linearly. As a result, in the absence of audiovisual integration, the amplitude of the N1 and P2 ERP components (that are typically present during this early time window) elicited by audivisual stimuli (AV condition) is identical to the algebraic sum of the same components elicited by auditory only (A) and visual only (V) stimuli (the A+V condition). Audiovisual integration, on the other hand, leads to the attenuation of the N1 and P2 amplitude and latency in the AV compared to the A+V condition, (Besle, Fort, & Giard, 2004; Giard & Besle, 2010; Stekelenburg & Vroomen, 2007; Van Wassenhove, Grant, & Poeppel, 2005).

Recent research has demonstrated that within the context of this paradigm, changes in the amplitude and latency of the N1 and P2 components to audiovisual stimuli may occur independently of each other and index different aspects of audiovisual processing. Attenuation of the N1 amplitude is thought to depend on how well visual movements can

cue the temporal onset of the auditory signal, with the nature of the audiovisual stimuli – speech or non-speech – being irrelevant. This interpretation agrees with findings showing that N1 attenuation is not sensitive to the degree to which lip movements predict the identity of the phoneme (Van Wassenhove et al., 2005). In fact, N1 reduction was even reported for incongruent audiovisual presentations (Stekelenburg & Vroomen, 2007). However, it is typically absent when visual cues do not precede the auditory signal (e.g., Brandwein et al., 2011) or carry little information about the temporal onset of the latter (e.g., Baart et al., 2014). On the other hand, shortening of the N1 latency is greatest when lip movements are highly predictive of the articulated phoneme (such as in the articulation of bilabial sounds, for example) (Van Wassenhove et al., 2005).

In a recent study, Baart and colleagues (Baart et al., 2014) proposed that changes in the P2 component elicited by AV speech reflect audiovisual phonetic binding (and not just a detection of audiovisual correspondences associated with the N1 component). The authors used sine-wave speech as stimuli, with some study participants perceiving them as speech and some as computer noises. The authors have demonstrated that while changes to N1 in the audiovisual condition were present when participants perceived sine-wave as both speech and non-speech, changes to the P2 component were present only in those research participants who perceived their stimuli as speech. While P2 attenuation is not speech-specific (Vroomen & Stekelenburg, 2010), it appears to reflect the binding of auditory and visual modalities that are perceived as representing a unitary audiovisual event.

Because almost all studies employing the above electrophysiological paradigm have been conducted with adults, very little is known about when in development changes in the N1 and P2 components to audiovisual stimuli can be reliably detected and, consequently, when different aspects of audiovisual processing indexed by such changes reach adult-like levels. A study by Brandwein and colleagues (Brandwein et al., 2011) used simple non-speech stimuli - a pure tone and a red disk – and tested children ranging in age from 7 to 16 as well as adults. They found that the amplitude of the N1 component was larger in the AV compared to the A+V condition in two oldest groups – 13- to 16-year-olds and adults. The direction of the reported effect was, however, opposite to what had previously been reported for adults (namely, the N1 and P2 amplitude elicited by the AV stimuli is typically smaller, rather than larger, compared to that elicited by the A+V stimuli). One reason for this outcome may be that the onset of visual stimuli in this study was temporally aligned with the onset of sounds, instead of preceding it (Stekelenburg & Vroomen, 2007). As a consequence, although generally speaking Brandwein and colleagues' findings are in agreement with earlier reports on the protracted developmental course of audiovisual integration, they may not generalize to more ecologically valid situations, in which the preceding visual information provides cues to the timing and nature of the following auditory signal.

The study by Knowland and colleagues (Knowland et al., 2014) used words as stimuli and compared ERPs elicited by audiovisual presentations with the ERPs elicited by auditory only presentations in children 6 to 12 years of age and in adults. They reported that in children the reduction in the amplitude of the ERP waveform to audiovisual stimuli was present only during the P2 window, while in adults both the N1 and the P2 components were

attenuated. Further, by treating age as a continuous variable and conducting a linear regression on the N1 and P2 amplitudes, the authors found that in children the N1 amplitude to audiovisually presented words did not differ significantly from that to auditorily presented words until approximately 10 years of age, while the P2 component showed such attenuation by approximately 7 years of age. To the best of our knowledge, the study by Knowland and colleagues was the first to measure electrophysiological indices of audiovisual integration for speech in school-age children. Their results are informative because of a broad age range of participants. However, one limitation of the study is that each age group was represented by only 5–8 children, thus limiting its statistical power. Additionally, the amplitudes of the N1 and P2 ERP components were measured as peak-to-peak values, making these measures less independent of each other. Research both in basic electrophysiology (Crowley & Colrain, 2004) and audiovisual processing (Baart et al., 2014) suggests that the P2 component of the ERP waveform indexes neural processes that are at least somewhat independent of those indexed by the N1 component. Under these circumstances, peak-to-peak measurements of N1 and P2 render conclusions about the time window over which audiovisual integration manifests itself less definitive.

In this study, we asked whether the early stages of audiovisual integration are fully mature in children by the time they start school or whether they continue to develop during mid-childhood. To this end, we compared ERP responses elicited by naturally produced audiovisual syllables to the algebraic sum of ERP responses elicited by the same syllables in auditory only and visual only conditions. We compared ERP data collected from two groups of children (7–8-year-olds and 10–11-year-olds) and from adults. If audiovisual integration during sensory processing is mature in children of this age, we expected to see the attenuation of the N1 and P2 components and the shortening of their latency similar in size to that observed in adults. A complete lack of changes in these components or a smaller degree of their attenuation would indicate continuing development.

2. Method

2.1 Participants

Seventeen 7–8-year-old children (8 female, average age 8;2, range 7;3–9;3), seventeen 10–11-year-old children (9 female, average age 11;2, range 10;1–12;2), and seventeen adults (9 female, average age 21 years, range 18–28) participated in the study. All gave their written consent or assent to participate in the experiment. Additionally, at least one parent of each child gave a written consent to enroll their child in the study. The study was approved by the Institutional Review Board of Purdue University, and all study procedures conformed to The Code of Ethics of the World Medical Association (Declaration of Helsinki) (1964).

All children had normal language skills based on the Core Language Score of the Clinical Evaluation of Language Fundamentals – 4th edition (CELF-4; Semel, Wiig, & Secord, 2003) and normal non-verbal intelligence (TONI-4; Brown, Sherbenou, & Johnsen, 2010) (see Table 1). None had the diagnosis of the Attention Deficit (Hyperactivity) Disorder (based on parental report) or showed any signs of autism (CARS-2; Schopler, Van Bourgondien, Wellman, & Love, 2010). The two groups of children did not differ in their socio-economic status (SES), which was measured through the level of mothers' and

fathers' education (group: mother's years of education, $F(1,32)=1.523$, $p=0.226$; father's years of education, $F(1,31)=2.061$, $p=0.161$; data for father's years of education was not available for one child). According to the Laterality Index of the Edinburgh Handedness Questionnaire, one child in the 7–8-year-old group was left-handed; all other participants were right-handed (Cohen, 2008; Oldfield, 1971). All participants were free of neurological disorders, passed a hearing screening at a level of 20 dB HL at 500, 1000, 2000, 3000, and 4000 Hz, reported to have normal or corrected-to-normal vision, and were not taking medications that may affect brain function (such as anti-depressants) at the time of study.

2.2 Stimuli and Experimental Design

The task of this study was similar to the one described by Besle and colleagues (Besle, Fort, Delpuech, et al., 2004), with slight modifications in order to make it appropriate for use with children. Participants watched a female speaker dressed as a clown produce syllables 'ba,' 'da,' and 'ga.' Only the speaker's face and shoulders were visible. Her mouth was brightly colored to attract attention to articulatory movements. Each syllable was presented in three different conditions – audiovisual (AV), visual only (V), and auditory only (A). Audiovisual syllables contained auditory and visual information. Visual only syllables were created by removing the sound track from audiovisual syllables. Lastly, auditory only syllables were created by using the sound track of audiovisual syllables but replacing the video track with a static image of a talker's face with a neutral expression. There were two different tokens of each syllable, and each token was repeated once within each block. All 36 stimuli (3 syllables × 2 tokens × 3 conditions × 2 repetitions) were presented pseudorandomly within each block, with the restriction that no two similar conditions (e.g., auditory only) could occur in a row. Additionally, videos with silly facial expressions – a clown sticking out her tongue or blowing a raspberry – were presented randomly on 25 percent of trials. These "silly" videos were also shown in three conditions – A, V, and AV (see a schematic representation of a block in Figure 1). The task was presented as a game. Participants were told that the clown helped researchers prepare some videos, but occasionally she did something silly. They were asked to assist researchers in removing all the silly occurrences by carefully monitoring presented videos and pressing a button on a response pad every time they either saw or heard something silly. Instructions were kept identical for children and adults. Together with "silly" videos, each block contained 48 trials and lasted approximately 4 minutes. Each participant completed 8 blocks. Hand to response button mapping was counterbalanced across participants. Responses to silly expressions were accepted if they occurred within 3,000 ms following the onset of the video.

A schematic representation of a trial is shown in Figure 2. Each trial started with a blank screen that lasted for 300 ms. It was replaced by a still image of a clown's face with a neutral expression. This still image was taken from the first frame of the following video. The still image remained for 400 to 800 ms, after which a video was played. A video always started and ended with a neutral facial expression and a completely closed mouth. It was then replaced again by a still image of a clown's face, but this time taken from the last frame of the video. This still image remained on the screen for 700–1,000 ms. Similarly to some of the previous studies of audiovisual speech perception, natural articulations of syllables were used (Knowland et al., 2014; Van Wassenhove et al., 2005). As a result, the time between

the first noticeable articulatory movement and sound onset varied from syllable to syllable, with 167 ms delays for each of the 'ba' tokens, 200 and 234 ms delay for the 'da' tokens, and 301 and 367 ms delays for the 'ga' tokens. All syllable videos were 1,400 ms in duration. The time at which the onset of sound occurred in the AV and A trials served as time 0 for all ERP averaging (AV, A, and V trials). The "raspberry" video lasted 2,167 ms, while the "tongue-out" video lasted 2,367 ms.

All videos were recorded with the Canon Vixia HV40 camcorder. In order to achieve a better audio quality, during video recording, the audio was also recorded with a Marantz digital recorder (model PMD661) and an external microphone (Shure Beta 87) at a sampling rate of 44,100 Hz. The camcorder's audio was then replaced with the Marantz version in Adobe Premiere Pro CS5. All sounds were root-mean-square normalized to 70 dB in Praat (Boersma & Weenink, 2011). The video's frame per second rate was 29.97. The video presentation and response recording was controlled by the Presentation program (www.neurobs.com). The refresh rate of the computer running Presentation was set to 75 Hz. Participants were seated in a dimly-lit sound-attenuating booth, approximately 4 feet from a computer monitor. Sounds were presented at 60 dB SPL via a sound bar located directly under the monitor. Before each task, participants were shown the instructions video and then practiced the task until it was clear. Children and adults had a short break after each block and a longer break after half of all blocks were completed. Children played several rounds of the board game of their choice during breaks. Together with breaks, the ERP testing session lasted approximately 2 hours.

2.3 ERP Recordings and Data Analysis

Electroencephalographic (EEG) data were recorded from the scalp at a sampling rate of 512 Hz using 32 active Ag-AgCl electrodes secured in an elastic cap (Electro-Cap International Inc., USA). Electrodes were positioned over homologous locations across the two hemispheres according to the criteria of the International 10–10 System (American Electroencephalographic Society, 1994). The specific locations were as follows: midline sites: Fz, Cz, Pz, and Oz; mid-lateral sites: FP1/FP2, AF3/AF4, F3/F4, FC1/FC2, C3/C4, CP1/CP2, P3/P4, PO3/PO4, and O1/O2; and lateral sites: F7/F8, FC5/FC6, T7/T8, CP5/CP6, and P7/P8; and left and right mastoids. EEG recordings were made with the Active-Two System (BioSemi Instrumentation, Netherlands), in which the Common Mode Sense (CMS) active electrode and the Driven Right Leg (DRL) passive electrode replace the traditional "ground" electrode (Metting van Rijn, Peper, & Grimbergen, 1990). During recording, data were displayed in relationship to the CMS electrode and then referenced offline to the average of the left and right mastoids (Luck, 2005). The Active-Two System allows EEG recording with high impedances by amplifying the signal directly at the electrode (BioSemi, 2013; Metting van Rijn, Kuiper, Dankers, & Grimbergen, 1996). In order to monitor for eye movement, additional electrodes were placed over the right and left outer canthi (horizontal eye movement) and below the left eye (vertical eye movement). Horizontal eye sensors were referenced to each other, while the sensor below the left eye was referenced to FP1 in order to create electro-oculograms. Prior to data analysis, EEG recordings were filtered between 0.1 and 30 Hz. Individual EEG records were visually inspected to exclude trials containing excessive muscular and other non-ocular artifacts.

Ocular artifacts were corrected by applying a spatial filter (EMSE Data Editor, Source Signal Imaging Inc., USA) (Pflieger, 2001). ERPs were epoched starting at 200 ms before the sound onset and ending at 1000 ms post-sound onset. The 200 ms prior to the sound onset served as a baseline.

Only ERPs elicited by the syllables, which were free of response preparation and execution activity, were analyzed. ERPs were averaged across the three syllables ('ba,' 'da,' and 'ga'), separately for A, V, and AV conditions. The mean number of trials averaged for each condition in each group was as follows (out of 96 possible trials): A condition – 84 (range 63–93) for 7–8-year-old children, 88 (range 80–94) for 10–11-year-old children, and 90 (range 63–95) for adults; V condition – 85 (range 63–93) for 7–8-year-old children, 88 (range 79–94) for 10–11-yearold children, and 91 (range 74–96) for adults; AV condition – 84 (range 62–91) for 7–8-year-old children, 88 (range 76–96) for 10–11-year-old children, and 90 (range 68–95) for adults. In order to ascertain that groups did not differ significantly in the number of accepted trials, we conducted a repeated-measures ANOVA with condition (A, V, and AV) and group (7–8 year olds, 10–11 year olds, and adults) as variables. This analysis yielded a significant effect of group ($F(2,48)=4.365$, $p=0.018$, $\eta_p^2=0.154$), with fewer accepted trials in the group of 7–8-year-old children compared to adults ($p=0.017$). However, this group difference never exceeded 5 trials. Importantly, the effect of group did not interact with condition ($F(4,96)<1$), and there was no difference in the number of accepted trials either between 7–8-year-olds and 10–11-year-olds ($p=0.183$) or between 10–11-year olds and adults ($p=0.985$). ERP responses elicited by A and V syllables were algebraically summed resulting in the A+V waveform (EMSE, Source Signal Imaging, USA). The N1 peak amplitude and peak latency were measured between 136 and 190 ms post-stimulus onset while the P2 peak amplitude and peak latency between 200 and 310 ms post-stimulus onset from the waveforms elicited by AV syllables and the waveforms obtained from the A+V summation. The selected N1 and P2 measurement windows were centered on the N1 and P2 peaks and checked against individual records of all participants.

Repeated-measures ANOVAs were used to evaluate behavioral and ERP results. For behavioral measures, we collected accuracy (ACC) and reaction time (RT) of responses to silly facial expressions, which were averaged across “raspberry” and “tongue out” events. The analysis included condition (A, V, and AV) as a within-subjects factor and group (7–8-year-olds, 10–11-year-olds, and adults) as a between-subjects factor. In regard to ERP measures, we examined the N1 and P2 peak amplitude and peak latency in two separate analyses. First, in order to determine whether, in agreement with earlier reports, the AV stimuli elicited components with reduced peak amplitude and peak latency compared to the algebraic sum of the A and V conditions, we conducted a repeated-measures ANOVA with condition (AV, A+V), laterality section (left (FC1, C3, CP1), midline (Fz, Cz, Pz), right (FC2, C4, CP2)), and site (FC1, Fz, FC2; C3, Cz, C4; and CP1, Pz, CP2) as within-subject variables and group (7–8-year-olds, 10–11-year-olds, and adults) as a between-subject variable, separately on the peak amplitude and peak latency measures. Including both conditions as two levels of the same variable in this analysis, rather than focusing on a difference between them, allowed us to examine potential group differences in ERPs elicited in each condition (AV and A+V) and to ascertain whether an attenuation of N1 and P2 to

AV speech was present. However, this analysis could not tell us if the degree of N1 and P2 attenuation was similar across the three groups. Therefore, in order to determine whether groups differed in the amount of the N1 and P2 attenuation, we subtracted the peak amplitude of these components elicited in the AV condition from the peak amplitude in the A+V condition and evaluated these difference scores in a second repeated-measures ANOVA analysis with laterality section (left (FC1, C3, CP1), midline (Fz, Cz, Pz), right (FC2, C4, CP2)) and site (FC1, Fz, FC2; C3, Cz, C4; and CP1, Pz, CP2) as within-subject variables and group (7–8-year-olds, 10–11-year-olds, and adults) as a between-subject variable. The selection of sites for ERP analyses was based on a typical distribution of the auditory N1 and P2 components (Crowley & Colrain, 2004; Näätänen & Picton, 1987) and a visual inspection of the grand average waveforms.

In all statistical analyses, significant main effects with more than two levels were evaluated with a Bonferroni post-hoc test. In such cases, the reported *p* value indicates the significance of the Bonferroni test, rather than the adjusted alpha level. When omnibus analysis produced a significant interaction, it was further analyzed with step-down ANOVAs, with factors specific to any given interaction. For these follow-up ANOVAs, significance level was corrected for multiple comparisons by dividing the alpha value of 0.05 by the number of follow-up tests. Only results with *p* values below this more stringent cut-off were reported as significant. Mauchly's test of sphericity was used to check for the violation of sphericity assumption in all repeated-measures tests that included factors with more than two levels. When the assumption of sphericity was violated, we used the Greenhouse-Geisser adjusted *p*-values to determine significance. Accordingly, in all such cases, adjusted degrees of freedom and the epsilon value (ε) are reported. Effect sizes, indexed by the partial eta squared statistic (η_p^2), are reported for all significant repeated-measures ANOVA results.

3. Results

3.1. Behavioral Results

We measured ACC and RT of detecting silly facial expressions when the latter were presented in each of the three conditions – A, V, and AV. Analysis of ACC showed only the effect of group ($F(2,48)=5.527$, $p=0.007$, $\eta_p^2=0.187$). Although all groups detected silly faces with very high accuracy (group accuracies ranged from 93.9 to 99.5 percent correct), the youngest group of children was significantly less accurate than adults ($p=0.005$), but did not differ from older children ($p=0.276$). The effect of group did not interact with the factor of condition ($F(4,96)<1$, and there was no overall effect of condition ($F(1.501,72.052)=1.966$, $p=0.158$, $\eta_p^2=0.039$, $\varepsilon=0.751$).

The effect of group was also present in the analysis of RT ($F(2,48)=16.962$, $p<0.001$, $\eta_p^2=0.414$), with faster RT in adults compared to either group of children (adults vs. 7–8-yearolds, $p<0.001$; adults vs. 10–11-year-olds, $p<0.001$; 7–8-year olds vs. 10–11-year-olds, $p=0.878$). RT was also longer in the A condition compared to either the V or the AV condition (condition, $F(2,96)=35.717$, $p<0.001$, $\eta_p^2=0.427$; A vs. V, $p<0.001$; A vs. AV, $p<0.001$; V vs. AV, $p=1$). Longer RT in the A condition is likely due to the fact that facial changes indicating the onset of silly facial expressions (that were present in the V and AV, but not in the A conditions) could be detected prior to the onset of silly sounds in the A

condition. This effect was further defined by a group by condition interaction ($F(4,96)=3.159$, $p=0.017$, $\eta_p^2=0.116$). Follow up tests revealed that while in absolute terms RT in the A condition was always longer than that in either the V or the AV condition, in the group of 10–11-year old children, the difference between the A and AV conditions failed to reach significance (condition: 7–8-year olds, $F(2,32)=15.398$, $p<0.001$, $\eta_p^2=0.49$; A vs. AV, $p<0.001$; A vs. V, $p=0.033$; AV vs. V, $p=0.132$; 10–11-year olds, $F(2,32)=6.034$, $p=0.007$, $\eta_p^2=0.274$; A vs. AV, $p=0.272$; A vs. V, $p=0.02$; AV vs. V, $p=0.218$; adults, $F(1.22,19.525)=49.215$, $p<0.001$, $\eta_p^2=0.755$, $\varepsilon=0.61$; A vs. AV, $p<0.001$; A vs. V, $p<0.001$; AV vs. V, $p=1$).

3.2 ERP Results

ERPs elicited in the three groups of participants in the AV and A+V conditions are shown in Figure 3, and ERPs elicited by A and V syllables alone are shown in Figure 4. Lastly, Figure 5 shows a direct overlay of ERPs elicited by AV syllables across the three groups in order to better reveal developmental changes in brain responses to audiovisual speech. As can be seen from these figures, the two groups of children differed substantially from adults in the amplitude and morphology of the ERP waveforms. This finding is in agreement with earlier reports of immature auditory and visual ERPs in this age group (e.g., Bishop, Hardiman, Uwer, & von Suchodoletz, 2007; De Haan, 2008; Ponton, Eggermont, Kwong, & Don, 2000; M. J. Taylor & Baldeweg, 2002). Despite such differences, however, all groups showed clear N1 and P2 components to A and AV syllables. No clear ERPs were elicited by V syllables. This finding is in agreement with earlier reports (e.g., Van Wassenhove et al., 2005) and is due to the fact that articulation was already in progress during sound onset (which served as time 0 for all ERP averaging). Because within the context of the additive ERP paradigm employed in this study the linear summation of auditory and visual ERP signals can only be assumed during sensory processing (Besle, Fort, Delpuech, et al., 2004; Besle, Fort, & Giard, 2004; Giard & Besle, 2010), we focused our statistical analyses on the peak amplitude and peak latency of the N1 and P2 components in the AV and A+V conditions (see Figure 3).

N1 component—Analysis of the N1 peak amplitude revealed a significant effect of condition ($F(1,48)=20.463$, $p<0.001$, $\eta_p^2=0.299$), with a smaller N1 peak in the AV compared to the A+V condition. This effect did not interact with group (group by condition, $F(2,48)<1$). The effect of laterality was also significant ($F(1.666,79.957)=11.87$, $p<0.001$, $\eta_p^2=0.198$, $\varepsilon=0.833$), with the N1 peak being larger over the right than either over the left or the midline scalp (right vs. left, $p=0.006$; right vs. midline, $p<0.001$; left vs. midline, $p=1$). The laterality effect was further defined by the laterality by group ($F(4,96)=5.943$, $p=0.001$, $\eta_p^2=0.198$) and the laterality by condition ($F(1.698,81.524)=8.162$, $p=0.001$, $\eta_p^2=0.145$, $\varepsilon=0.849$) interactions. Follow-up analyses revealed that the effect of laterality was present only in children (laterality: 7–8-year-olds, $F(2,32)=7.829$, $p=0.004$, $\eta_p^2=0.329$; 10–11-year-olds, $F(2,32)=9.598$, $p=0.001$, $\eta_p^2=0.375$), with larger N1 over the right than either over the left or the midline scalp. In adults, the amplitude of N1 tended to be larger over the left scalp, but this effect did not reach the Bonferroni-corrected alpha level of 0.017 (laterality, $F(2,32)=4.508$, $p=0.019$, $\eta_p^2=0.22$). Additionally, the effect of laterality was present only in the A+V condition (laterality: AV condition, $F(1.608,77.177)=1.785$, $p=0.181$, $\varepsilon=0.804$; A

+V condition, $F(1.7, 81.605)=21.081$, $p<0.001$, $\eta_p^2=0.305$, $\varepsilon=0.85$). None of the other effects were significant (group, $F(2,48)=2.916$, $p=0.064$, $\eta_p^2=0.108$; condition by group, $F(2,48)<1$; site by group, $F(4,96)=1.567$, $p=0.189$; condition by laterality by group, $F(4,96)=1.35$, $p=0.257$; condition by site, $F(1.264,60.665)=2.254$, $p=0.133$, $\varepsilon=0.632$; condition by site by group, $F(4,96)<1$; condition by laterality by site, $F(2.572, 123.469)=2.144$, $p=0.101$, $\varepsilon=0.643$).

We also examined the degree of the N1 peak amplitude attenuation (measured as a difference between the A+V and AV conditions) among the three groups of participants. This analysis showed no effect of group ($F(2,48)<1$), but revealed that the N1 attenuation was larger over the right than either over the left or the midline scalp (laterality, $F(1.698,81.526)=8.162$, $p=0.001$, $\eta_p^2=0.145$, $\varepsilon=0.849$; right vs. left, $p=0.013$; right vs. midline, $p=0.002$; left vs. midline, $p=1$). All other effects were non-significant (laterality by group, $F(4,96)=1.35$, $p=0.257$; site by group, $F(4,96)<1$; laterality by site, $F(2.572,123.467)=2.144$, $p=0.108$, $\varepsilon=0.643$; laterality by site by group, $F(8,192)<1$).

Lastly, analysis of the N1 peak latency yielded two significant interactions – group by laterality ($F(4,96)=3.911$, $p=0.006$, $\eta_p^2=0.14$) and condition by site ($F(1.676,80.429)=5.994$, $p=0.006$, $\eta_p^2=0.111$, $\varepsilon=0.838$). However, none of the follow-up analyses reached significance at the Bonferroni-corrected alpha level of 0.017 (group over the left and midline scalp, $F(2,48)<1$; group over the right scalp, $F(2,48)=1.56$, $p=0.221$; laterality in 7–8-year olds, $F(2,32)=4.229$, $p=0.035$, $\eta_p^2=0.209$; laterality in 10–11-year olds, $F(2,32)=3.664$, $p=0.04$[1], $\eta_p^2=0.186$; laterality in adults, $F(1.486,23.768)= 2.487$, $p=0.117$, $\varepsilon=0.743$; condition over F and FC sites, $F(1,48)<1$; condition over C sites, $F(1,48)=1.048$, $p=0.311$; condition over CP and P sites, $F(1,48)=4.589$, $p=0.037$). A three-way interaction between condition, laterality, and site was also significant ($F(3.245,155.754)=2.947$, $p=0.031$, $\eta_p^2=0.058$, $\varepsilon=0.811$). However, similarly to the other two interactions, none of the follow up tests reached the Bonferroni-corrected alpha level of 0.006- 0.008 (p values ranged from 0.019 to 0.749). No other results were significant (group, $F(2,48)<1$; condition, $F(1,48)=1.126$, $p=0.294$; condition by group, $F(2,48)=1.424$, $p=0.251$; laterality, $F(2,96)=2.594$, $p=0.08$; site by group, $F(4,96)<1$; condition by laterality, $F(2,96)=2.04$, $p=0.136$; condition by laterality by group, $F(4,96)<1$; condition by site by group, $F(4,96)=1.992$, $p=0.102$; laterality by site, $F(3.275, 157.206)=1.845$, $p=0.136$, $\varepsilon=0.819$; laterality by site by group, $F(8,192)=1.267$, $p=0.263$; condition by laterality by site by group, $F(8,192)=1.004$, $p=0.435$).

In sum, in children only, the amplitude of N1 was largest over the right scalp in the A+V condition and was of similar amplitude over all laterality segments in the AV condition. The peak amplitude of N1 was significantly reduced in the AV compared to the A+V condition in all groups. The degree of such N1 attenuation in the AV condition did not differ among groups, but was larger over the right than over either the left or the midline scalp. No significant N1 peak latency effects were found.

[1]Although p values for the laterality effect were significant at the uncorrected alpha level in both groups of children, this effect was driven by a longer N1 peak latency over the right as compared to over the midline scalp. There was no difference in the N1 peak latency over the left and the right scalp.

P2 component—Similarly to the analysis of the N1 peak amplitude, the P2 peak amplitude was smaller in the AV compared to the A+V condition (condition, $F(1,48)=6.735$, $p=0.013$, $\eta_p^2=0.123$). This effect did not interact with group (condition by group, $F(2,48)<1$), but did interact with laterality (condition by laterality, $F(1.773,85.105)=8.674$, $p=0.001$, $\eta_p^2=0.153$, $\varepsilon=0.887$). Follow-up analyses revealed that the effect of condition was present over the left ($F(1,48)=8.207$, $p=0.006$, $\eta_p^2=0.146$) and midline ($F(1,48)=8.797$, $p=0.005$, $\eta_p^2=0.155$) scalp, but failed to reach significance over the right scalp ($F(1,48)=3.285$, $p=0.076$). The effect of group ($F(2,48)=4.553$, $p=0.015$, $\eta_p^2=0.159$) and the group by site interaction ($F(4,96)=5.174$, $p=0.004$, $\eta_p^2=0.177$) were also significant, with a smaller P2 in 7–8-year-old children compared to adults over frontal/fronto-central sites (group, $F(2,48)=6.293$, $p=0.004$, $\eta_p^2=0.208$; 7–8-year-olds vs. adults, $p=0.003$; 7–8-year-olds vs. 10–11-year-olds, $p=0.356$; 10–11 years olds vs. adults, $p=0.17$) and compared to both 10–11-year-olds and adults over centro-parietal/parietal sites (group, $F(2,48)=4.466$, $p=0.017$, $\eta_p^2=0.157$; 7–8-year-olds vs. adults, $p=0.048$; 7–8-year-olds vs. 10–11-year-olds, $p=0.031$; 10–11-year-olds vs. adults, $p=1$). All other results were non-significant (condition by group, $F(2,48)<1$; laterality by group, $F(4,96)=2.219$, $p=0.073$; condition by laterality by group, $F(4,96)=1.454$, $p=0.222$; condition by site, $F(2,96)<1$; laterality by site by group, $F(8,192)=1.744$, $p=0.091$; condition by laterality by site by group, $F(8,192)<1$).

Analysis of the degree of the P2 peak amplitude attenuation (measured as a difference between the A+V and AV conditions) among the three groups of participants showed that over the central scalp, the P2 attenuation was larger over the midline site than either over the left or the right site and was also larger over the left than over the right site (laterality, $F(1.773,85.105)=8.674$, $p=0.001$, $\eta_p^2=0.153$, $\varepsilon=0.887$; laterality by site, $F(3.248, 155.922)=22.258$, $p<0.001$, $\eta_p^2=0.317$, $\varepsilon=0.689$; laterality over C sites, $F(2,96)=32.892$, $p<0.001$, $\eta_p^2=0.407$; Cz vs. C3, $p<0.001$; Cz vs. C4, $p<0.001$; C3 vs. C4, $p=0.003$; laterality over F and FC sites, $F(2,96)=2.326$, $p=0.103$; laterality over P and CP sites, $F(1.589, 76.28)=4.495$, $p=0.021$[2], $\eta_p^2=0.086$, $\varepsilon=0.795$). Although the group by site interaction was also significant (group by site, $F(4,96)=4.132$, $p=0.014$, $\eta_p^2=0.147$), follow up tests failed to reach significance (group over F and FC sites, $F(2,48)<1$; group over C sites, $F(2,48)<1$; group over CP and P sites, $F(2,48)=1.065$, $p=0.353$, $\eta_p^2=0.042$). None of the other results were significant (group, $F(2,48)<1$; laterality by group, $F(4,96)=1.454$, $p=0.228$; laterality by site by group, $F(8,192)<1$).

The P2 peak latency was shorter in the AV compared to the A+V condition ($F(1,48)=4.76$, $p=0.034$, $\eta_p^2=0.09$). This effect was modified by a significant group by condition interaction ($F(2,48)=3.206$, $p=0.049$, $\eta_p^2=0.118$); however, none of the follow-up analyses reached significance at the Bonferroni-corrected alpha level of 0.017(condition in 7–8-year olds, $F(1,16)=5.071$, $p=0.039$, $\eta_p^2=0.241$; condition in 10–11-year olds, $F(1,16)=5.177$, $p=0.037$[3]; condition in adults, $F(1,16)<1$). The effect of group was also significant ($F(2,48)=8.035$, $p=0.001$, $\eta_p^2=0.251$), with longer latency in adults compared to either group

[2]Although below 0.05, this *p* value does not reach the Bonferroni-corrected alpha level of 0.017. Significance at the uncorrected level was due to greater P2 attenuation over the left (CP1) as compared to the midline (Pz) site (left vs. midline, $p=0.035$; left vs. right, $p=0.371$; right vs. midline, $p=0.297$).

[3]Although the condition effect fell short of significance at the Bonferroni-corrected alpha level in the two groups of children, the tendency was for shorter P2 latency in the AV as compared to the A+V condition.

of children (adults vs. 7–8-year-olds, p=0.001; adults vs. 10–11-year-olds, p=0.029; 7–8-year-olds vs. 10–11-year-olds, p=0.675). Lastly, the P2 peaked earlier over the left scalp than over the midline scalp (laterality, F(1.724,82.752)=3.754, p=0.033, η_p^2=0.073, ε=0.862; left vs. midline, p=0.02); however, the latency of P2 did not differ either between the left and the right scalp (p=0.998) or between the midline and the right scalp (p=0.195). None of the other results reached significance (laterality by group, F(4,96)<1; site by group, F(4,96)=1.132, p=0.34; condition by laterality, F(2,96)<1; condition by laterality by group, F(4,96)=1.642, p=0.17; condition by site, F(1.35, 64.803)=1.074, p=0.324, ε=0.675; condition by site by group, F(4,96)<1; laterality by site, F(2.758,132.365)=1.531, p=0.212, ε=0.689; laterality by site by group, F(8,192)=1.738, p=0.123; condition by laterality by site, F(4,192)<1; condition by laterality by site by group, F(8,192)<1).

In sum, 7–8-year-old children had a significantly smaller P2 compared to both 10–11-year-old children and adults, indicating that this component is not yet fully developed during early school years. Despite this overall amplitude difference, all groups showed a reduction in the P2 peak amplitude and peak latency in the AV compared to the A+V condition. However, unlike the N1 peak amplitude reduction described above, the P2 attenuation was largest over the midline and left scalp. This effect was limited to the central sites.

Finally, we examined whether a degree of N1 and P2 attenuation to AV speech depended on children's age. To do so, we conducted a linear regression analysis between, on the one hand, the average of N1 and P2 attenuation over the 9 sites used for ERP analyses (FC1, C3, CP1, Fz, Cz, Pz, FC2, C4, CP2) and, on the other hand, children's age. Neither regression produced significant results (N1, R=0.092, F(1,33)<1; P2, R=0.158, F(1,33)<1).

4. Discussion

We examined electrophysiological indices of audiovisual integration during sensory processing in two groups of school-age children and in adults. We found that, despite differences in the overall amplitude and morphology of ERP waveforms between children and adults, by and large, changes in the N1 and P2 ERP components elicited in the AV compared to the A+V condition were very similar in all age groups. More specifically, both children and adults showed a clear reduction in the amplitude of N1 and P2 in the AV condition. These findings suggest that early stages of audiovisual processing as indexed by these components are functional by early school years (7–8 years of age). In addition, we showed that N1 and P2 amplitude attenuations have a different hemispheric distribution – with larger N1 reduction in the AV condition over the right scalp and larger P2 reduction over the left and midline scalp. This finding supports earlier work indicating that changes in these two components index at least somewhat different neural processes within the context of audiovisual perception.

We did not replicate the earlier reports of shorter N1 latency in the AV compared to the A +V condition (e.g., Van Wassenhove et al., 2005). Previous studies showed that the degree of N1 latency shortening depends on the saliency with which lip movements provide cues to phoneme's identity. In order to obtain a sufficient number of trials from young children, we averaged ERP data across syllables with different places of articulation (bilabial 'ba,'

alveolar ‘da,’ and velar ‘ga’), which provide different degrees of certainty as to the nature of the articulated sound, with bilabial articulations being most informative and alveolar and velar articulations significantly less so. As a result, the shortening of the N1 latency in the AV condition has likely been lost. Unlike the N1 peak latency, the P2 peak latency was in fact shorter in the AV condition. This effect interacted with the factor of group, and even though follow up tests that examined the presence of the P2 peak latency shortening in each group failed to reach the Bonferroni-corrected alpha level, only in the two groups of children did it reach significance at the uncorrected alpha level of 0.05. This pattern of results suggests that the effect was likely carried by children and thus may indicate a developmental trend. However, the immature state of P2 in children (as is evidenced by its much smaller amplitude compared to adults) and the lack of significance at the more stringent alpha level during follow up tests require a replication of this result before firmer conclusions can be drawn.

Our findings only partially replicated the results of Knowland and colleagues (Knowland et al., 2014). As discussed in the Introduction, these authors reported that while P2 attenuation to AV speech in their group of children was present already at approximately 7 years of age, the N1 attenuation did not develop until closer to 10 years of age. In contrast, we found adult-like attenuation of both components by 7–8 years of age. In fact, N1 attenuation to AV syllables was stronger than the corresponding attenuation of P2, at least based on the difference in the effect sizes measured by partial eta squared (0.299 for N1 attenuation vs. 0.123 for the P2 attenuation). It is possible that, at least in part, differences in the studies’ designs could have contributed to differences in the outcomes. For example, the task performed by children in Knowland et al.’s study was semantic in nature. Children were presented with real words (rather than syllables as in the current study) and had to press a button when they heard animal names. One might hypothesize that when processing longer linguistic units, such as words, that typically contain a series of lip movements (and thus multiple visual cues to a word’s identity) and in which articulation of the last phoneme may be just as informative as that of the initial phoneme, differences in ERP responses to AV presentations may become evident over a slightly later temporal window. In contrast, syllables presented in the current study differed only in the initial consonant and yielded ERP differences between the AV and A+V condition during the N1 time period. Future studies examining audiovisual processing of both syllables and words in school-age children should provide better understanding of this issue.

The study by Knowland and colleagues has also reported a significant correlation between the degree of N1 and P2 attenuation to the AV as compared to the A speech and children’s age. We conducted a similar regression between N1 and P2 attenuation to the AV as compared to the A+V speech and children’s age. Neither analysis yielded significant results. Somewhat surprisingly, in an earlier study from our laboratory, which compared N1 and P2 attenuation to audiovisual speech in school-age children with a history of specific language impairment and their typically developing peers (Kaganovich, Schumaker, Macias, & Anderson, accepted), we did see a significant relationship between age and N1 amplitude reduction in the AV condition. However, the presence of a clinical population and an overall smaller number of children representing each age group in our earlier report makes a direct comparison difficult. Nevertheless, the inconsistent pattern of findings regarding the

relationship between the attenuation of N1 and P2 to audiovisual speech and children's age may suggest that age is only one out of a number of factors that can influence the degree of attenuation of early sensory components to audiovisual speech. Given that cross-sectional studies show significant individual variability in N1 and P2 attenuation even at an early age, in the future, a longitudinal approach to this issue may better clarify a relationship between age and N1 and P2 sensitivity to audiovisual speech.

Because a number of studies have shown a protracted developmental course of audiovisual speech perception in school-age children (e.g., Massaro, 1984; Massaro et al., 1986; McGurk & MacDonald, 1976; Tye-Murray et al., 2014), our results of adult-like N1 and P2 attenuation to audiovisual speech in 7- and 8-year olds may appear to contradict earlier reports. However, the degree to which early sensory encoding of audiovisual information can contribute to later cognitive processing of such information in children and adults remains to be understood. Audiovisual speech perception relies on a complex network of neural structures, which include both primary sensory and association brain areas (Alais, Newell, & Mamassian, 2010; Calvert, 2001; Senkowski, Schneider, Foxe, & Engel, 2008) and that have different developmental trajectories (Gogtay et al., 2004; Huttenlocker & Dabholkar, 1997). Adult-like audiovisual function may require not only that all components of the network are fully developed but also that adult-like patterns of connectivity between them are also in place. Therefore, the seemingly mature state of early audiovisual processing cannot by itself guarantee efficient audiovisual integration. Neuroimaging studies of audiovisual processing in children are very scarce. However, at least one study reported that interconnectivity between different components of the audiovisual processing network is not yet mature during mid-childhood. More specifically, Dick and colleagues (Dick, Solodkin, & Small, 2010) recorded brain activity to auditory only and audiovisual speech in 8–11-year old children and in adults and reported that while in both groups the same areas of the brain were activated during perception, the pattern of their interaction was different in children when listening to audiovisual speech – namely, children showed a reduced influence of the posterior inferior frontal gyrus and ventral premotor cortex on activity in the supramarginal gyrus in the left hemisphere. The authors suggest that this pathway may be engaged when using sensory and motor information for identifying speech sounds. Its immaturity in school-age children would therefore indicate that they cannot use experience with sensory and motor aspects of speech events for audiovisual speech perception as well as adults can.

Analysis of N1 and P2 attenuation to AV speech syllables revealed an intriguing hemispheric difference, with N1 attenuation being largest over the right scalp and the P2 attenuation being largest over the left and midline scalp. Previous studies did not report such hemispheric effects, possibly because their statistical analyses of ERP data did not include a factor of hemisphere or laterality (see, for example, Baart et al., 2014; Knowland et al., 2014; Van Wassenhove et al., 2005). It has been suggested that the attenuation of the N1 peak amplitude indexes how well lip movements can cue the onset of the auditory signal (Baart et al., 2014; Klucharev, Möttönen, & Sams, 2003; Van Wassenhove et al., 2005). In our stimuli, the onset of sound followed the first noticeable articulation-related facial movement by 167–367 milliseconds depending on the syllable. Assuming that processing temporal relationships within auditory only and audiovisual contexts taps into similar neural networks, at least hypothetically, the greater attenuation of N1 over the right scalp could be

driven by greater sensitivity of the right hemisphere to events that unfold over a time window of several hundred milliseconds as proposed by the 'asymmetrical sampling in time' hypothesis of Poeppel and colleagues (Boemio, Fromm, Braun, & Poeppel, 2005; Poeppel, 2003). Although evidence for the 'asymmetric sampling in time' hypothesis comes from work with auditory stimuli, the fact that silent lip reading can activate some of the same areas of the auditory cortices as heard speech (Bernstein et al., 2002; Calvert et al., 1997; Pekkola et al., 2005; Sams et al., 1991) lends some support to the proposition that similar neural structures may be engaged by processing both auditory and audiovisual temporal information as long as it unfolds over similar temporal intervals. However, a replication of our finding in both children and adults is needed before stronger conclusions can be drawn.

Contrary to N1, P2 attenuation was largest over the left and midline scalp. Reduction in the amplitude of this relatively later ERP component to audiovisual speech is thought to index phonetic binding (Baart et al., 2014) and is present when auditory and visual signals are thought to be part of the same event. Some of the previous electrophysiological and neuroimaging studies showed that phonological processing may be left-hemisphere lateralized under certain circumstances – such as during a phonological categorization task – but is typically bilateral during passive listening (e.g., Poeppel et al., 1996; Zatorre, Evans, Meyer, & Gjedde, 1992) (although see also work by Shtyrov and colleagues for an alternative view on hemispheric specialization of phonemic processing (Shtyrov, Pihko, & Pulvermüller, 2005). Greater activation of the left hemisphere during phonological categorization is thought to stem at least in part from the need to access articulatory representations for specific phonemes (e.g., Zatorre et al., 1992). This hypothesis is supported by more recent findings of greater left-hemisphere activation during lip-reading (e.g., Pekkola et al., 2005). While our task did not require phonemic categorization – in fact, participants were asked to screen presented stimuli for non-linguistic expressions – it required greater attention to facial movements and to lip movements in particular than what could be expected during passive listening. Additionally, because research participants knew that the speaker always pronounced one of three possible syllables – 'ba,' 'da,' or 'ga' – it is very likely that they could often identify the pronounced phoneme even in the visual only condition. Therefore, our task might have led to the activation of the articulatory representations for the three syllables, which could be reflected in the greater P2 attenuation over the left and midline electrode sites.

In sum, neural mechanisms underlying the sensory stage of audiovisual speech processing as indexed by the attenuation of the N1 and P2 ERP components appear to be active by early school years. Previous studies have already begun to tease apart functional significance of N1 and P2 attenuation to audiovisual stimuli by showing that different factors influence their characteristics in an ERP waveform and that their amplitude and latency can be modified independently of each other. This study further contributes to this literature by showing that N1 and P2 attenuation can also have different hemispheric distribution, lending further support to the proposition that they reflect at least somewhat disparate neural processes within the context of audiovisual perception. Understanding the maturational trajectory of specific audiovisual integration processes in typically developing children provides an important benchmark for evaluating audiovisual processing in a number of developmental

disorders for which at least some form of impairment in audiovisual function has been indicated, such as autism (Foxe et al., in press; Guiraud et al., 2012; Saalasti et al., 2012; Stevenson, Siemann, Schneider, et al., 2014; Stevenson, Siemann, Woynaroski, et al., 2014; N. Taylor, Isaac, & Milne, 2010; Woynaroski et al., 2013), dyslexia (Bastien-Toniazzo, Stroumza, & Cavé, 2010), specific language impairment (Boliek, Keintz, Norrix, & Obrzut, 2010; Hayes, Tiippana, Nicol, Sams, & Kraus, 2003; Kaganovich et al., 2014; Leybaert et al., 2014; Meronen, Tiippana, Westerholm, & Ahonen, 2013; Norrix, Plante, & Vance, 2006; Norrix, Plante, Vance, & Boliek, 2007), and phonological disorders (Dodd, McIntosh, Erdener, & Burnham, 2008).

Acknowledgement

This research was supported in part by grants P30DC010745 and R03DC013151 from the National Institute on Deafness and Other Communicative Disorders, National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official view of the National Institute on Deafness and Other Communicative Disorders or the National Institutes of Health. We are thankful to Daniel Noland for assistance with programming and to Dana Anderson, Camille Hagedorn, Danielle Macias, Courtney Rowland, and Casey Spelman for their assistance with different stages of data collection and processing. Bobbie Sue Ferrel-Brey's acting skills were invaluable during audiovisual recordings. Last, but not least, we are immensely grateful to children and their families for participation.

References

Alais D, Newell FN, Mamassian P. Multisensory processing in review: From physiology to behavior. Seeing and Perceiving. 2010; 23:3–38. [PubMed: 20507725]

American Electroencephalographic Society. Guideline thirteen: Guidelines for standard electrode placement nomenclature. Journal of Clinical Neurophysiology. 1994; 11:111–113. [PubMed: 8195414]

Baart M, Stekelenburg JJ, Vroomen J. Electrophysiological evidence for speech-specific audiovisual integration. Neuropsychologia. 2014; 53:115–121. [PubMed: 24291340]

Bahrick LE, Hernandez-Reif M, Flom R. The development of infant learning about specific face-voice relations. Developmental Psychology. 2005; 41(3):541–552. [PubMed: 15910161]

Bahrick LE, Netto D, Hernandez-Reif M. Intermodal perception of adult and child faces and voices by infants. Child Development. 1998; 69(5):1263–1275. [PubMed: 9839414]

Barutchu A, Crewther DP, Crewther SG. The race that precedes coactivation: Development of multisensory facilitation in children. Developmental Science. 2009; 12(3):464–473. [PubMed: 19371371]

Barutchu A, Danaher J, Crewther SG, Innes-Brown H, Shivdasani MN, Paolini AG. Audiovisual integration in noise by children and adults. Journal of Experimental Child Psychology. 2010; 105:38–50. [PubMed: 19822327]

Bastien-Toniazzo M, Stroumza A, Cavé C. Audio-visual perception and integration in developmental dyslexia: An exploratory study using the McGurk effect. Current Psychology Letters: Behaviour, Brain and Cognition. 2010; 25(3):1–15.

Beauchamp MS, Nath AR, Pasalar S. fMRI-guided Transcranial Magnetic Stimulation reveals that the Superior Temporal Sulcus is a cortical locus of the McGurk effect. The Journal of Neuroscience. 2010; 30(7):2414–2417. [PubMed: 20164324]

Bernstein LE, Auer ET, Moore JK, Ponton CW, Don M, Singh M. Visual speech perception without primary auditory cortex activation. NeuroReport. 2002; 13(3):311–315. [PubMed: 11930129]

Besle J, Bertrand O, Giard M-H. Electrophysiological (EEG, sEEG, MEG) evidence for multiple audiovisual interactions in the human auditory cortex. Hearing Research. 2009; 258:2225–2234.

Besle J, Fischer C, Bidet-Caulet A, Lecaignard F, Bertrand O, Giard M-H. Visual activation and audivisual interaction in the auditory cortex during speech perception: Intracranial recordings in humans. The Journal of Neuroscience. 2008; 28(52):14301–14310. [PubMed: 19109511]

Besle J, Fort A, Delpuech C, Giard M-H. Bimodal speech: early suppressive visual effects in human auditory cortex. European Journal of Neuroscience. 2004; 20:2225–2234. [PubMed: 15450102]

Besle J, Fort A, Giard M-H. Interest and validity of the additive model in electrophysiological studies of multisensory interactions. Cognitive Processing. 2004; 5:189–192.

BioSemi. Active Electrodes. 2013 2013, from http://www.biosemi.com/active_electrode.htm.

Bishop DVM, Hardiman M, Uwer R, von Suchodoletz W. Maturation of the long-latency auditory ERP: step function changes at the start and end of adolescence. Developmental Science. 2007; 10(5):565–575. [PubMed: 17683343]

Boemio A, Fromm S, Braun A, Poeppel D. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nature Neuroscience. 2005; 8(3):389–395.

Boersma P, Weenink D. Praat: doing phonetics by computer (version 5.3) [Computer program]. 2011 Retrieved from http://www.praat.org (Version 5.1).

Boliek CA, Keintz C, Norrix LW, Obrzut J. Auditory-visual perception of speech in children with leaning disabilities: The McGurk effect. Canadian Journal of Speech-Language Pathology and Audiology. 2010; 34(2):124–131.

Brandwein AB, Foxe JJ, Russo NN, Altschuler TS, Gomes H, Molholm S. The development of audiovisual multisensory integration across childhood and early adolescence: A high-density electrical mapping study. Cerebral Cortex. 2011; 21(5):1042–1055. [PubMed: 20847153]

Bristow D, Dehaene-Lambertz G, Mattout J, Soares G, Glida T, Baillet S, Mangin J-F. Hearing faces: How the infant brain matches the face it sees with the speech it hears. Journal of Cognitive Neuroscience. 2008; 21(5):905–921. [PubMed: 18702595]

Brookes H, Slater A, Quinn PC, Lewkowicz DJ, Hayes R, Brown E. Three-month-old infants learn arbitrary auditory-visual pairings between voices and faces. Infant and Child Development. 2001; 10:75–82.

Brown, L.; Sherbenou, RJ.; Johnsen, SK. Test of Nonverbal Intelligence. 4th ed.. Austin, Texas: Pro-Ed: An International Pubilsher; 2010.

Burnham D, Dodd B. Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. Developmental Psychobiology. 2004; 45(4):204–220. [PubMed: 15549685]

Burr, D.; Gori, M. Multisensory integration develops late in humans. In: Murray, MM.; Wallace, MT., editors. The Neural Bases of Multisensory Processes. New York: CRC Press; 2012.

Calvert GA. Crossmodal processing in the human brain: Insights from functional neuroimaging studies. Cerebral Cortex. 2001; 11:1110–1123. [PubMed: 11709482]

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SCR, McGuire PK, David AS. Activation of auditory cortex during silent lipreading. Science. 1997; 276:593–596. [PubMed: 9110978]

Cohen MS. Handedness Questionnaire. 2008 Retrieved 05/27/2013, 2013, from http://www.brainmapping.org/shared/Edinburgh.php#.

Crowley KE, Colrain IM. A review of the evidence for P2 being an independent component process: age, sleep and modality. Clinical Neurophysiology. 2004; 115:732–744. [PubMed: 15003751]

De Haan, M. Neurocognitive mechanisms for the development of face processing. In: Nelson, CA.; Luciana, M., editors. Handbook of Developmental Cognitive Neuroscience. Cambridge, Massachusetts: The MIT Press; 2008. p. 509-520.

Dick AS, Solodkin A, Small SL. Neural development of networks for audiovisual speech comprehension. Brain and Language. 2010; 114:101–114. [PubMed: 19781755]

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

Dodd B, McIntosh B, Erdener D, Burnham D. Perception of the auditory-visual illusion in speech perception by children with phonological disorders. Clinical Linguistics and Phonetics. 2008; 22(1):69–82. [PubMed: 18092221]

Foxe JJ, Molholm S, Del Bene VA, Frey H-P, Russo NN, Blanco D, Ross LA. Severe multisensory speech integration deficits in high-functioning school-aged children with Autism Spectrum Disorder (ASD) and their resolution during adolescence. Cerebral Cortex. (in press).

Giard, M-H.; Besle, J. Methodological considerations: Electrophysiology of multisensory interactions in humans. In: Naumer, MJ., editor. Multisensory Object Perception in the Primate Brain. New York: Springer; 2010. p. 55-70.

Gogtay N, Giedd JN, Lusk L, Hayashi KM, Greenstein D, Vaituzis AC, Thompson PM. Dynamic mapping of human cortical development during childhood through early adulthood. Proceedings of the National Academy of Sciences. 2004; 101(21):8174–8179.

Gori M, Sandini G, Burr D. Development of visuo-auditory integration in space and time. Frontiers in Integrative Neuroscience. 2012; 6

Guiraud JA, Tomalski P, Kushnerenko E, Ribeiro H, Davies K, Charman T, Team tB. Atypical audiovisual speech integration in infants at risk for autism. PLOS ONE. 2012; 7(5):e36428. [PubMed: 22615768]

Hayes EA, Tiippana K, Nicol TG, Sams M, Kraus N. Integration of heard and seen speech: a factor in learning disabilities in children. Neuroscience Letters. 2003; 351:46–50. [PubMed: 14550910]

Hertrich I, Mathiak K, Lutzenberger W, Menning H, Ackermann H. Sequential audiovisual interactions during speech perception: A whole-head MEG study. Neuropsychologia. 2007; 45:1342–1354. [PubMed: 17067640]

Hillock-Dunn A, Wallace MT. Developmental changes in the multisensory temporal binding window persist into adolescence. Developmental Science. 2012; 15(5):688–696. [PubMed: 22925516]

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]

Huttenlocker PR, Dabholkar AS. Regional differences in synaptogenesis in human cerebral cortex. The Journal of Comparative Neurology. 1997; 387:167–178. [PubMed: 9336221]

Jaime M, Longard J, Moore C. Developmental changes in the visual proprioceptive integration threshold of children. Journal of Experimental Child Psychology. 2014; 125:1–12. [PubMed: 24814203]

Kaganovich N, Schumaker J, Leonard LB, Gustafson D, Macias D. Children with a history of SLI show reduced sensitivity to audiovisual temporal asynchrony: An ERP study. Journal of Speech, Language, and Hearing Research. 2014; 57:1480–1502.

Kaganovich N, Schumaker J, Macias D, Anderson D. Processing of audiovisually congruent and incongruent speech in school-age children with a history of Specific Language Impairment: a behavioral and event-related potentials study. Developmental Science. (accepted).

Kaiser AR, Kirk KI, Lachs L, Pisoni DB. Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. Journal of Speech, Language, and Hearing Research. 2003; 46(2):390–404.

Klucharev V, Möttönen R, Sams M. Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. Cognitive Brain Research. 2003; 18(1):65–75. [PubMed: 14659498]

Knowland VCP, Mercure E, Karmiloff-Smith A, Dick F, Thomas MSC. Audio-visual speech perception: A developmental ERP investigation. Developmental Science. 2014; 17(1):110–124. [PubMed: 24176002]

Kushnerenko E, Teinonen T, Volein A, Csibra G. Electrophysiological evidence of illusory audiovisual speech percept in human infants. Proceedings of the National Academy of Sciences. 2008; 105(32):11442–11445.

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

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

Leybaert J, Macchi L, Huyse A, Champoux F, Bayard C, Colin C, Berthommier F. Atypical audio-visual speech perception and McGurk effects in children with specific language impairment. Frontiers in Psychology. 2014; 5

Luck, SJ. An Introduction to the Event-Related Potential Technique. Cambridge, MA: The MIT Press; 2005.

Massaro DW. Children's perception of visual and auditory speech. Child Development. 1984; 55:1777–1788. [PubMed: 6510054]

Massaro DW, Thompson LA, Barron B, Lauren E. Developmental changes in visual and auditory contributions to speech perception. Journal of Experimental Child Psychology. 1986; 41:93–113. [PubMed: 3950540]

McGurk H, MacDonald J. Hearing lips and seeing voices. Nature. 1976; 264:746–748. [PubMed: 1012311]

McIntosh AR, Kovacevic N, Itier RJ. Increased brain signal variability accompanies lower behavioral variabiilty in development. PLOS ONE. 2008; 4(7):e1000106.

Meronen A, Tiippana K, Westerholm J, Ahonen T. Audiovisual speech perception in children with developmental language disorder in degraded listening conditions. Journal of Speech, Language, and Hearing Research. 2013; 56:211–221.

Metting van Rijn, AC.; Kuiper, AP.; Dankers, TE.; Grimbergen, CA. Low-cost active electrode improves the resolution in biopotential recordings. Paper presented at the 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Amsterdam, The Netherlands. 1996.

Metting van Rijn AC, Peper A, Grimbergen CA. High-quality recording of bioelectric events. Part 1: Interference reduction, theory and practice. Medical and Biological Engineering and Computing. 1990; 28:389–397. [PubMed: 2277538]

Näätänen R, Picton T. The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. Psychophysiology. 1987; 24(4):375–425. [PubMed: 3615753]

Nath AR, Beauchamp MS. A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. NeuroImage. 2012; 59:781–787. [PubMed: 21787869]

Nath AR, Fava EE, Beauchamp MS. Neural correlates of interindividual differences in children's audiovisual speech perception. The Journal of Neuroscience. 2011; 31(39):13963–13971. [PubMed: 21957257]

Norrix LW, Plante E, Vance R. Auditory-visual speech integration by adults with and without language-learning disabilities. Journal of Communication Disorders. 2006; 39:22–36. [PubMed: 15950983]

Norrix LW, Plante E, Vance R, Boliek CA. Auditory-visual integration for speech by children with and without specific language impairment. Journal of Speech, Language, and Hearing Research. 2007; 50:1639–1651.

Oldfield RC. The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia. 1971; 9:97–113. [PubMed: 5146491]

Pekkola J, Ojanen V, Autti T, Jääskeläinen IP, Möttönen R, Tarkiainen A, Sams M. Primary auditory cortex activation by visual speech: an fMRI study at 3T. NeuroReport. 2005; 16(2):125–128. [PubMed: 15671860]

Pflieger, ME. Theory of a spatial filter for removing ocular artifacts with preservation of EEG. Paper presented at the EMSE Workshop, Princeton University; 2001. http://www.sourcesignal.com/SpFilt_Ocular_Artifact.pdf

Poeppel D. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Communication. 2003; 41:245–255.

Poeppel D, Yellin E, Phillips C, Roberts TPL, Rowley HA, Wexler K, Marantz A. Task-induced asymmetry of teh auditory evoked M100 neuromagnetic field elicited by speech sounds. Cognitive Brain Research. 1996; 4:231–242. [PubMed: 8957564]

Ponton CW, Eggermont JJ, Kwong B, Don M. Maturation of human central auditory system activity: evidence from multi-channel evoked potentials. Clinical Neurophysiology. 2000; 111:220–236. [PubMed: 10680557]

Rosenblum LD, Schmuckler MA, Johnson JA. The McGurk effect in infants. Perception and Psychophysics. 1997; 59(3):347–357. [PubMed: 9136265]

Ross LA, Molholm S, Blanco D, Gomez-Ramirez M, Saint-Amour D, Foxe JJ. The development of multisensory speech perception continues into the late childhood years. European Journal of Neuroscience. 2011; 33(12):2329–2337. [PubMed: 21615556]

Saalasti S, Kätsyri J, Tiippana K, Laine-Hernandez M, von Wendt L, Sams M. Audiovisual speech perception and eye gaze behavior of adults with Asperger syndrome. Journal of Autism and Developmental Disorders. 2012; 42:1606–1615. [PubMed: 22068821]

Sams M, Aulanko R, Hämäläinen M, Hari R, Lounasmaa OV, Lu S-T, Simola J. Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. Neuroscience Letters. 1991; 127:141–145. [PubMed: 1881611]

Schopler, E.; Van Bourgondien, ME.; Wellman, GJ.; Love, SR. Childhood Autism Rating Scale. 2nd ed.. Western Psychological Services; 2010.

Semel, E.; Wiig, EH.; Secord, WA. CELF4: Clinical Evaluation of Language Fundamentals. 4th ed.. San Antonio, TX: Pearson Clinical Assessment; 2003.

Senkowski D, Schneider TR, Foxe JJ, Engel AK. Crossmodal binding through neural coherence: Implications for multisensory processing. Trends in Neurosciences. 2008; 31(8):401–409. [PubMed: 18602171]

Shtyrov Y, Pihko E, Pulvermüller F. Determinants of dominance: Is language laterality explained by physical or linguistic features of speech? NeuroImage. 2005; 27:37–47. [PubMed: 16023039]

Stekelenburg JJ, Vroomen J. Neural correlates of multisensory integration of ecologically valid audiovisual events. Journal of Cognitive Neuroscience. 2007; 19(12):1964–1973. [PubMed: 17892381]

Stevenson RA, Siemann JK, Schneider BC, Eberly HE, Woynaroski TG, Camarata SM, Wallace MT. Multisensory temporal integration in autism spectrum disorders. The Journal of Neuroscience. 2014; 34(3):691–697. [PubMed: 24431427]

Stevenson RA, Siemann JK, Woynaroski TG, Schneider BC, Eberly HE, Camarata SM, Wallace MT. Brief report: Arrested development of audiovisual speech perception in autism spectrum disorders. Journal of Autism and Developmental Disorders. 2014; 44(6):1470–1477. [PubMed: 24218241]

Stevenson RA, VanDerKlok RM, Pisoni DB, James TW. Discrete neural substrates underlie complementary audiovisual speech integration processes. NeuroImage. 2011; 55:1339–1345. [PubMed: 21195198]

Taylor MJ, Baldeweg T. Application of EEG, ERP and intracranial recordings to the investigtaion of cognitive functions in children. Developmental Science. 2002; 5(3):318–334.

Taylor N, Isaac C, Milne E. A comparison of the development of audiovisual integration in children with Autism Spectrum Disorders and typically developing children. Journal of Autism and Developmental Disorders. 2010; 40:1403–1411. [PubMed: 20354776]

Tye-Murray N, Hale S, Spehar B, Myerson J, Sommers MS. Lipreading in school-age children: The roles of age, hearing status, and cognitive ability. Journal of Speech, Language, and Hearing Research. 2014; 57:556–565.

Van Wassenhove V, Grant KW, Poeppel D. Visual speech speeds up the neural processing of auditory speech. Proceedings of the National Academy of Sciences. 2005; 102(4):1181–1186.

Vroomen J, Stekelenburg JJ. Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. Journal of Cognitive Neuroscience. 2010; 22(7):1583–1596. [PubMed: 19583474]

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]

Williams BR, Hultsch DF, Strauss EH, Hunter MA, Tannock R. Inconsistency in reaction time across the life span. Neuropsychology. 2005; 19(1):88–96. [PubMed: 15656766]

Woynaroski TG, Kwakye LD, Foss-Feig JH, Stevenson RA, Stone WL, Wallace MT. Multisensory speech perception in children with autism spectrum disorders. Journal of Autism and Developmental Disorders. 2013; 43(12):2891–2902. [PubMed: 23624833]

Zatorre RJ, Evans AC, Meyer E, Gjedde A. Lateralization of phonetic and pitch discrimination in speech processing. Science. 1992; 256:846–849. [PubMed: 1589767]

Highlights

- We recorded ERPs to auditory, visual, and audiovisual speech in children and adults
- N1 and P2 components were smaller to audiovisual speech in all groups
- N1 attenuation to audiovisual speech was largest over the right scalp
- P2 attenuation to audiovisual speech was largest over the left and midline scalp
- Audiovisual integration at the sensory stage is functional by early school years

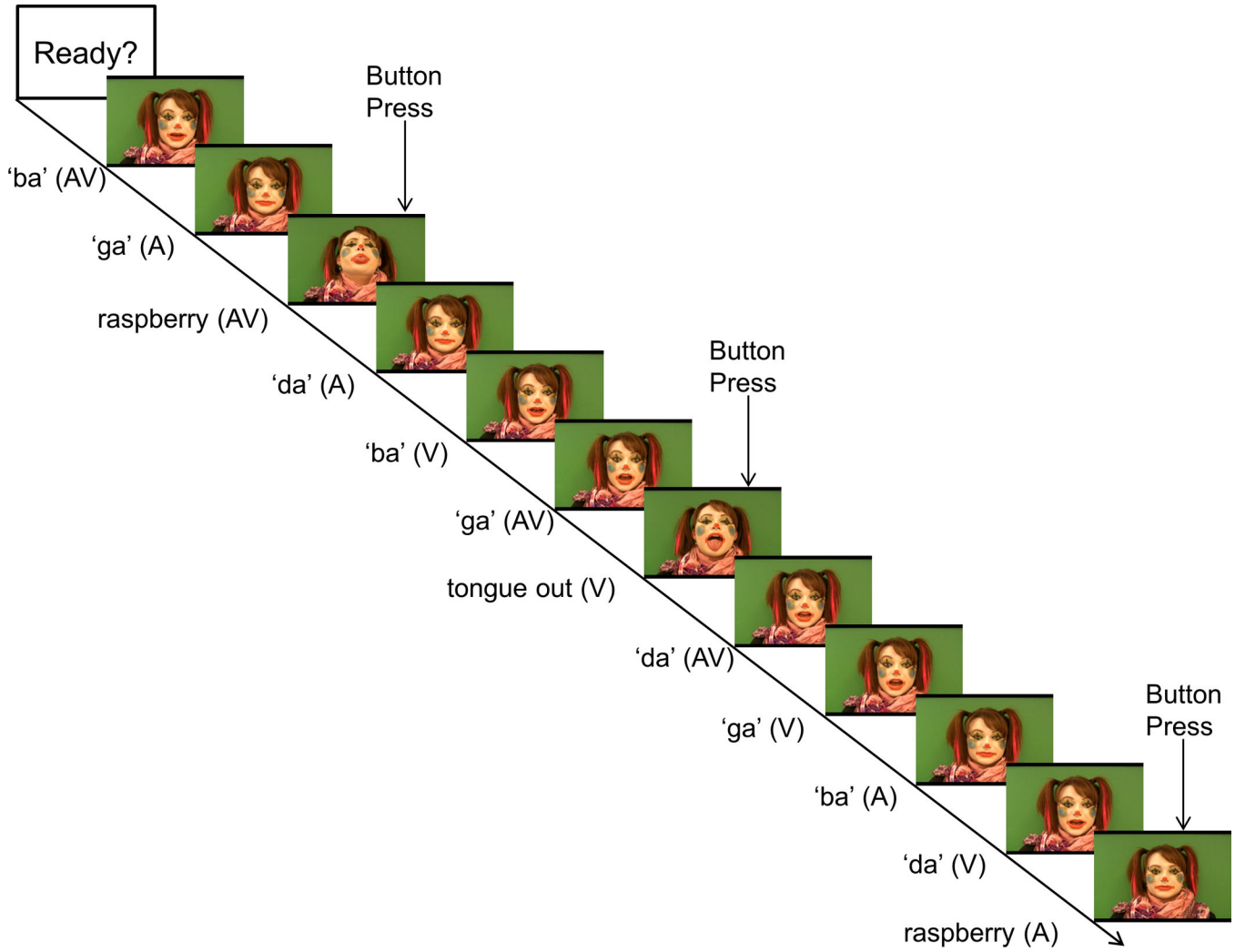

Figure 1. Schematic Representation of a Block

The timeline shows a succession of trials from top left to bottom right. AV = audiovisual trials, A = auditory only trials, V = visual only trials. This figure was originally published in the following study: Kaganovich, N., Schumaker, J., Macias, D., & Anderson, D. (accepted). Processing of audiovisually congruent and incongruent speech in school-age children with a history of Specific Language Impairment: A behavioral and event-related potentials study. Developmental Science.

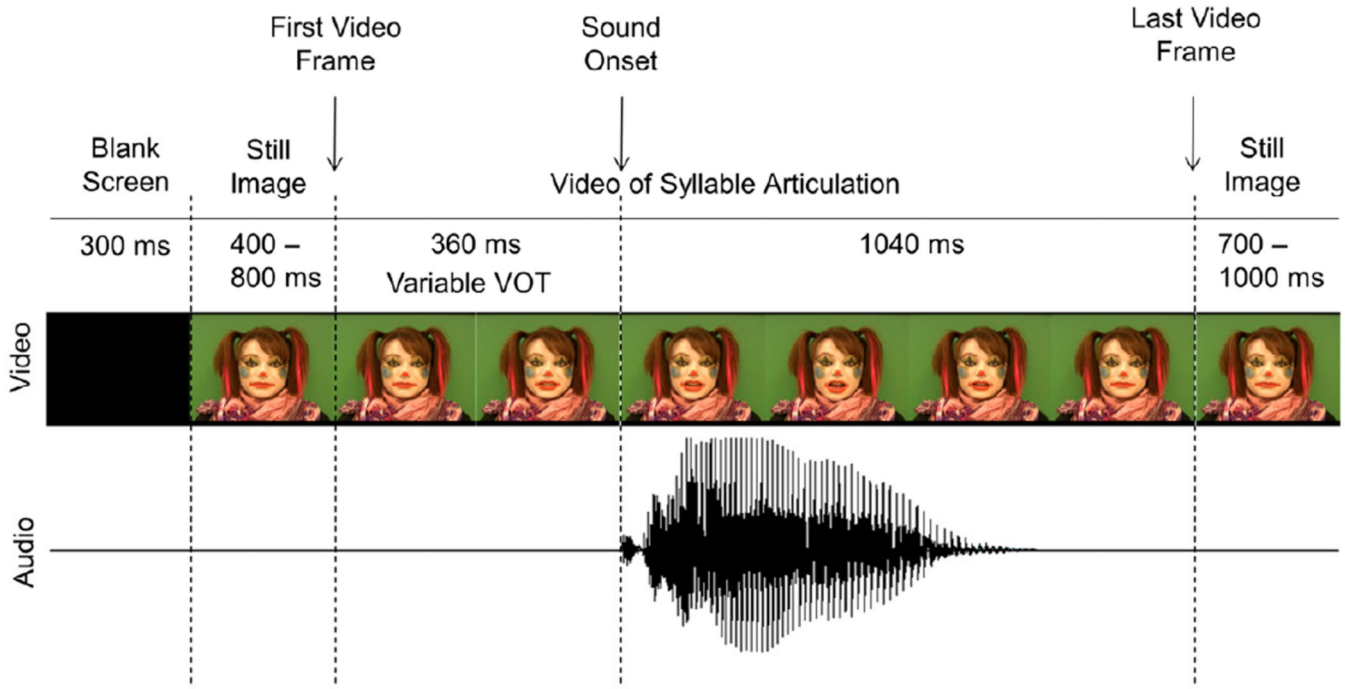

Figure 2. Schematic Representation of a Trial

Main events within a single trial are shown from left to right. Note that only representative frames are displayed. The sound onset was used as time 0 for all ERP averaging. The time between the first noticeable facial movement and sound onset (VOT) varied between 167 and 367 milliseconds depending on the syllable. This figure was originally published in the following study: Kaganovich, N., Schumaker, J., Macias, D., & Anderson, D. (accepted). Processing of audiovisually congruent and incongruent speech in school-age children with a history of Specific Language Impairment: A behavioral and event-related potentials study. Developmental Science.

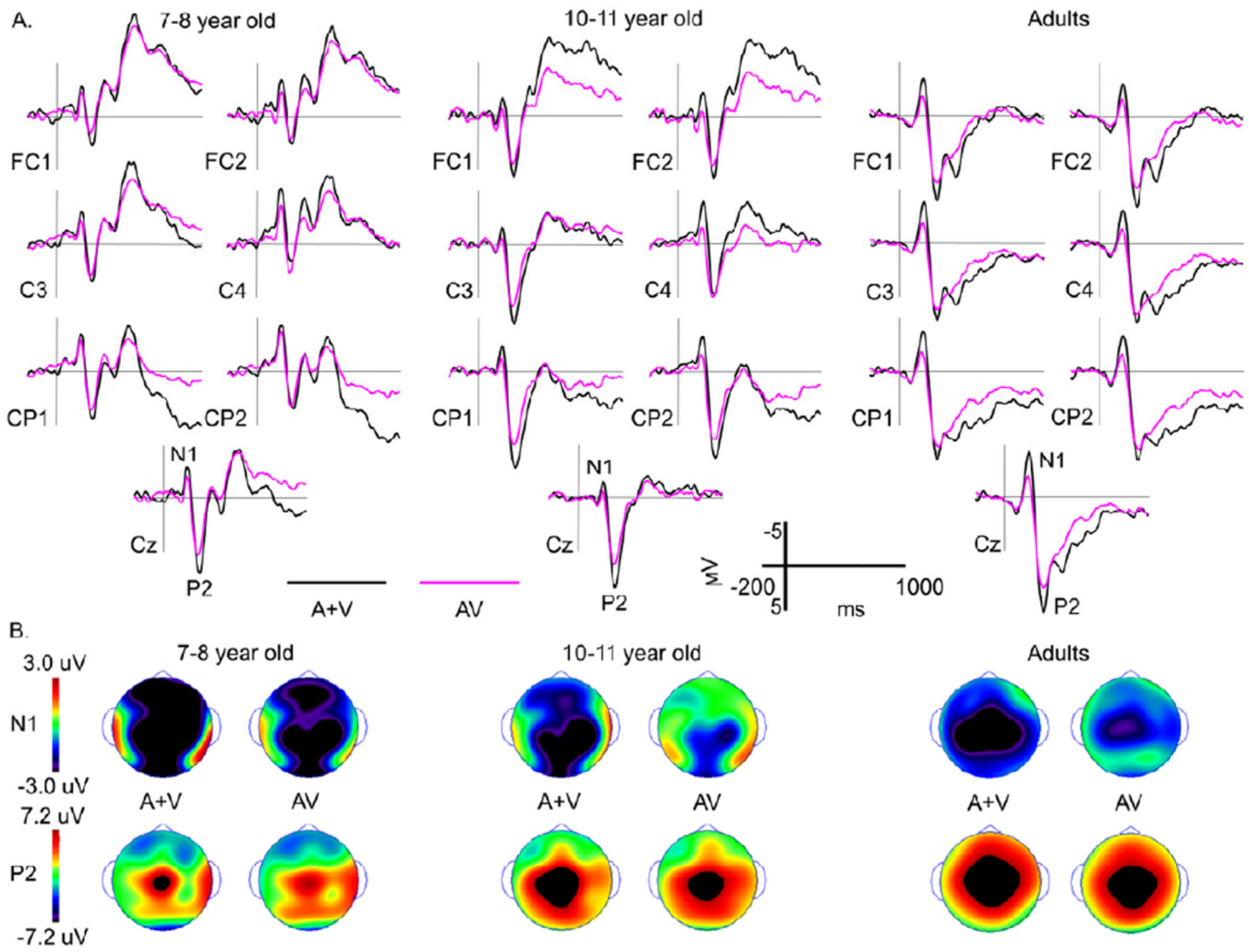

Figure 3. ERPs Elicited in the AV and A+V Conditions

A. Grand average waveforms elicited by audiovisual syllables (AV) and by the sum of auditory only and visual only (A+V) syllables are overlaid separately for 7–8-year-old children, 10–11-year-old children, and adults. Six mid-lateral sites and one midline site are shown for each group. All three groups displayed clear N1 and P2 components, which are marked on the Cz site. Negative is plotted up.

B. Scalp topographies of voltage at the peak of the N1 and P2 components are displayed for each group of participants in the AV and A+V conditions. Note a significant reduction in negativity for the N1 component and a significant reduction in positivity for the P2 component in the AV condition in each group.

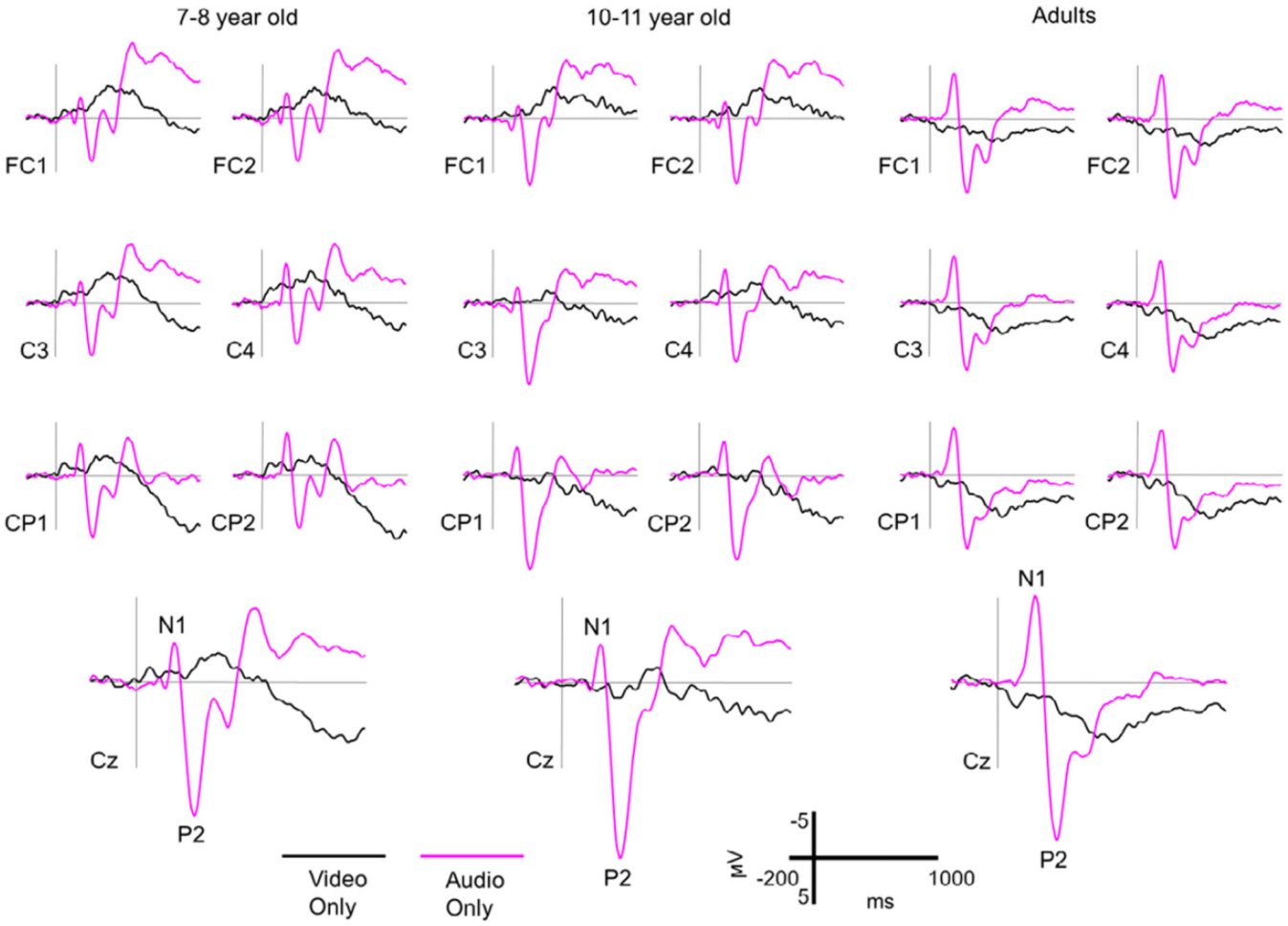

Figure 4. ERPs Elicited in the A and V Conditions

Grand average waveforms elicited by auditory only (A) and visual only (V) syllables are shown separately for 7–8-year old children, 10–11-year old children, and adults. Note that a lack of clear visual components is in agreement with earlier reports and is due to the fact that articulation was in progress during sound onset (which served as time 0 for all ERP averaging). Six mid-lateral sites and one midline site are shown for each group. Negative is plotted up.

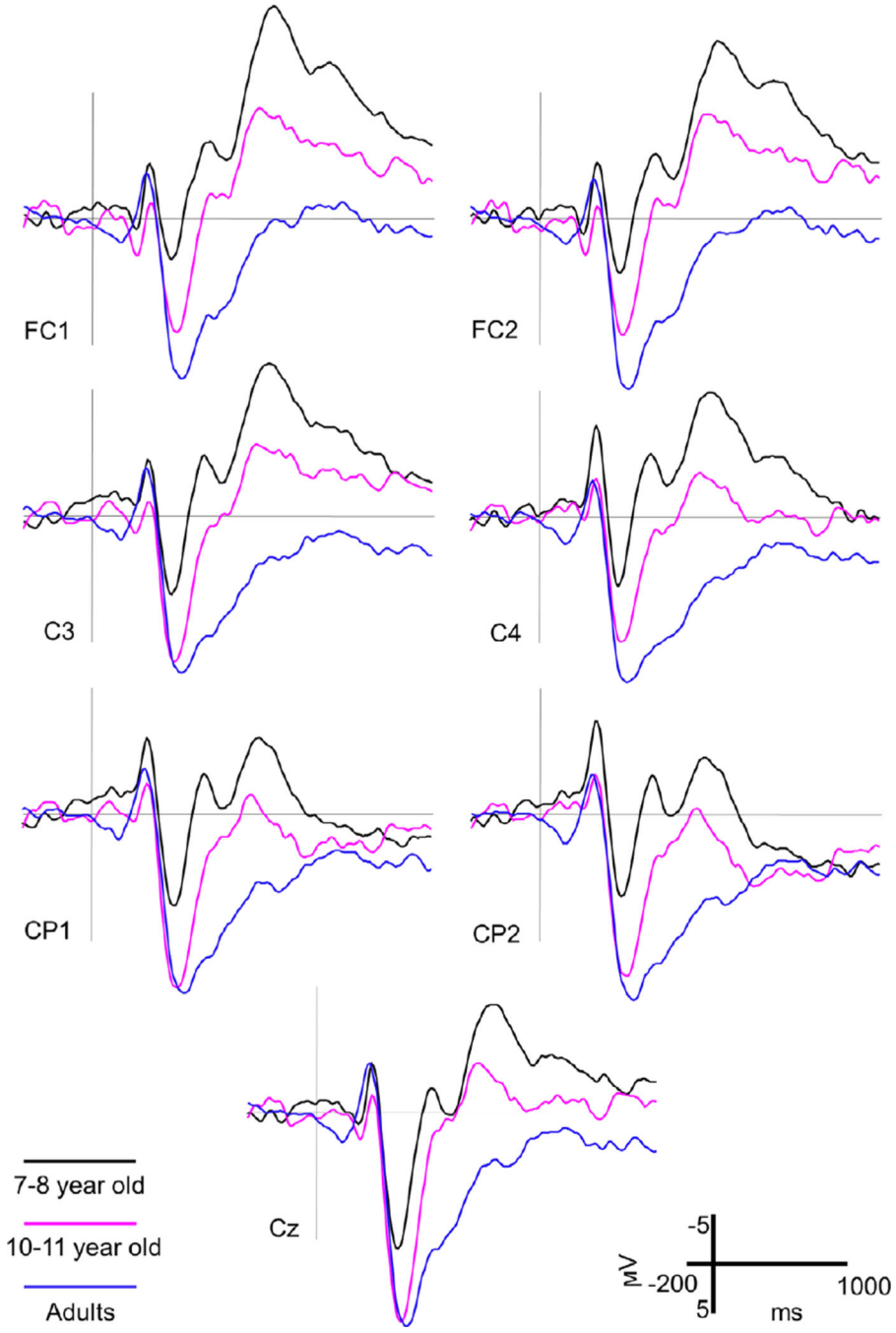

Figure 5. Group Comparison of ERPs elicited in the AV Condition

Grand average waveforms elicited by AV syllables are overlaid for the three groups of participants. Note that ERPs elicited in both groups of children differ significantly from those elicited in adults, with the waveform elicited in 10–11-year olds more closely resembling the adult waveform. Six mid-lateral sites and one midline site are shown. Negative is plotted up.

Brain Lang. Author manuscript; available in PMC 2015 December 01.

Table 1

TONI and CELF-4 scores for 7–8-year olds and 10–11-year olds

	TONI		C&FD		RS		FS		WS		WC						CLS	
Sub	**SS**	**%ile**	**SS**	**%ile**	**SS**	**%ile**	**SS**	**%ile**	**SS**	**%ile**	**R-SS**	**R-%ile**	**E-SS**	**E-%ile**	**T-SS**	**T-%ile**	**SS**	**%ile**
1	123	94	14	91	12	75	15	95	12	75	-	-	-	-	-	-	120	91
2	109	73	12	75	12	75	11	63	9	37	-	-	-	-	-	-	106	66
3	96	39	11	63	14	91	12	75	12	75	-	-	-	-	-	-	114	82
4	118	88	11	63	9	37	14	91	12	75	-	-	-	-	-	-	109	73
5	106	66	14	91	9	37	14	91	10	50	-	-	-	-	-	-	111	77
6	101	52	13	84	11	63	13	84	10	50	-	-	-	-	-	-	111	77
7	114	83	16	98	14	91	15	95	11	63	-	-	-	-	-	-	124	95
8	104	61	12	75	10	50	12	75	10	50	-	-	-	-	-	-	106	66
9	123	94	14	91	11	63	14	91	13	84	-	-	-	-	-	-	118	88
10	100	50	13	84	11	63	12	75	13	84	-	-	-	-	-	-	114	82
11	113	81	12	75	13	71	10	50	11	63	-	-	-	-	-	-	109	73
12	101	52	12	75	9	37	9	37	10	50	-	-	-	-	-	-	99	47
13	110	74	12	75	10	50	11	63	12	75	-	-	-	-	-	-	108	70
14	113	81	15	95	10	50	15	95	10	50	-	-	-	-	-	-	115	84
15	113	81	11	63	9	37	11	63	11	63	-	-	-	-	-	-	102	55
16	112	79	12	75	12	75	12	75	10	50	-	-	-	-	-	-	109	73
17	113	81	14	91	13	84	11	63	9	37	-	-	-	-	-	-	111	77
Mean	109.94		12.82		11.12		12.41		10.88								110.94	
1	105	63	12	75	16	98	13	84	-	-	14	91	12	81	13	84	121	92
2	96	39	7	16	8	25	14	91	-	-	10	50	13	84	12	75	100	50
3	100	50	10	50	10	50	11	63	-	-	14	91	10	50	12	75	106	66
4	102	55	11	63	10	50	14	91	-	-	13	84	9	37	11	63	109	73
5	114	83	10	50	11	63	14	91	-	-	16	98	14	91	15	95	115	84
6	101	52	12	75	12	75	12	75	-	-	14	91	9	37	12	75	112	79
7	130	98	12	75	13	84	12	75	-	-	16	98	13	84	15	95	118	88
8	102	55	13	84	12	75	13	84	-	-	10	50	8	25	9	37	111	77

	TONI		C&FD		RS		FS		WS		WC						CLS	
Sub	**SS**	**%ile**	**SS**	**%ile**	**SS**	**%ile**	**SS**	**%ile**	**SS**	**%ile**	**R-SS**	**R-%ile**	**E-SS**	**E-%ile**	**T-SS**	**T-%ile**	**SS**	**%ile**
9	105	63	13	84	17	99	11	63	-	-	16	98	15	95	16	98	126	96
10	98	45	10	50	10	50	10	50	-	-	15	95	13	84	14	91	106	66
11	103	58	12	75	9	37	8	25	-	-	11	63	11	63	11	63	99	47
12	94	34	11	63	10	50	13	84	-	-	11	63	8	25	9	37	104	61
13	96	39	8	25	11	63	10	50	-	-	11	63	10	50	10	50	98	45
14	127	96	12	75	12	75	11	63	-	-	16	98	12	75	14	91	114	82
15	110	74	12	75	15	95	11	63	-	-	16	98	15	95	16	98	121	92
16	108	70	13	84	11	63	14	91	-	-	12	75	12	75	12	75	115	84
17	99	48	12	75	13	84	12	75	-	-	14	91	11	63	13	84	115	84
Mean	105.29		11.18		11.76		11.94				13.47		11.47		12.59		111.18	

Scores for 7–8-year-olds are shown in the top half of the table while scores for 10–11-year-olds are shown in the bottom half of the table. TONI = Test of Non-Verbal Intelligence, C&FD = Concepts and Following Directions, RS = Recalling Sentences, FS = Formulated Sentences, WS = Word Structure; WC = Word Classes; R = Receptive; E = expressive; T = Total; CLS = Core Language Score, SS = standardized score, %ile = percentile.