



PAPER

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

Natalya Kaganovich,^{1,2} Jennifer Schumaker,¹ Danielle Macias¹ and Dana Gustafson¹

1. Department of Speech, Language, and Hearing Sciences, Purdue University, USA

2. Department of Psychological Sciences, Purdue University, USA

Abstract

Previous studies indicate that at least some aspects of audiovisual speech perception are impaired in children with specific language impairment (SLI). However, whether audiovisual processing difficulties are also present in older children with a history of this disorder is unknown. By combining electrophysiological and behavioral measures, we examined perception of both audiovisually congruent and audiovisually incongruent speech in school-age children with a history of SLI (H-SLI), their typically developing (TD) peers, and adults. In the first experiment, all participants watched videos of a talker articulating syllables 'ba', 'da', and 'ga' under three conditions – audiovisual (AV), auditory only (A), and visual only (V). The amplitude of the N1 (but not of the P2) event-related component elicited in the AV condition was significantly reduced compared to the N1 amplitude measured from the sum of the A and V conditions in all groups of participants. Because N1 attenuation to AV speech is thought to index the degree to which facial movements predict the onset of the auditory signal, our findings suggest that this aspect of audiovisual speech perception is mature by mid-childhood and is normal in the H-SLI children. In the second experiment, participants watched videos of audiovisually incongruent syllables created to elicit the so-called McGurk illusion (with an auditory 'pa' dubbed onto a visual articulation of 'ka', and the expectant perception being that of 'ta' if audiovisual integration took place). As a group, H-SLI children were significantly more likely than either TD children or adults to hear the McGurk syllable as 'pa' (in agreement with its auditory component) than as 'ka' (in agreement with its visual component), suggesting that susceptibility to the McGurk illusion is reduced in at least some children with a history of SLI. Taken together, the results of the two experiments argue against global audiovisual integration impairment in children with a history of SLI and suggest that, when present, audiovisual integration difficulties in this population likely stem from a later (non-sensory) stage of processing.

Research highlights

- Previous research shows that children with specific language impairment (SLI) are less able to combine auditory and visual information during speech perception. However, little is known about how audiovisual speech perception develops in this group when they reach school age and no longer show significant language impairment.
- We report that children with a history of SLI (H-SLI) show the same reduction in the amplitude of the N1 event-related potential component to audiovisual speech as their typically developing (TD) peers and adults, suggesting that early sensory processing of audiovisual speech is normal in this group.
- However, a large proportion of H-SLI children fail to perceive the McGurk illusion, suggesting a weaker reliance on visual speech cues compared to TD children and adults at a later stage of processing.
- Our findings indicate that when present, audiovisual integration difficulties in the H-SLI children likely stem from a later (non-sensory) stage of processing.

Address for correspondence: Natalya Kaganovich, Department of Speech, Language, and Hearing Sciences, Purdue University, Lyles Porter Hall, 715 Clinic Drive, West Lafayette, IN 47907-2038, USA; e-mail: kaganovi@purdue.edu

Introduction

A large and rapidly growing body of literature shows that speech perception is audiovisual in nature. For example, the presence of a talker's face has been shown to facilitate the perception of speech when it is embedded in noise (Barutchu, Danaher, Crewther, Innes-Brown, Shivdasani *et al.*, 2010; Ross, Saint-Amour, Leavitt, Javitt & Foxe, 2007; Sumbly & Pollack, 1954), has ambiguous content, or is spoken in a non-native language (Navarra & Soto-Faraco, 2007; Reisberg, McLean & Goldfield, 1987). In fact, visual speech cues can not only make us hear sounds that are not present in the auditory input (as in the McGurk illusion; McGurk & MacDonald, 1976) but also perceive sounds as coming from a location that is significantly different from the sound's true source (as in the ventriloquism illusion; e.g. Magosso, Cuppini & Ursino, 2012).

Audiovisual integration is a complex phenomenon. Accumulating research suggests that its various aspects, such as audiovisual matching, audiovisual identification, or audiovisual learning, to name a few, 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). While the full mastery of combining auditory and visual speech cues does not develop until adolescence (Brandwein, Foxe, Russo, Altschuler, Gomes *et al.*, 2011; Dick, Solodkin & Small, 2010; Knowland, Mercure, Karmiloff-Smith, Dick & Thomas, 2014; Lewkowicz, 1996, 2010; Massaro, 1984; Massaro, Thompson, Barron & Lauren, 1986; McGurk & MacDonald, 1976), children begin to display sensitivity to audiovisual correspondences in their environment during the first year of life (Bahrnick, Hernandez-Reif & Flom, 2005; Bahrnick, Netto & Hernandez-Reif, 1998; Bristow, Dehaene-Lambertz, Mattout, Soares, Glida, Baillet *et al.*, 2008; Brookes, Slater, Quinn, Lewkowicz, Hayes & Brown, 2001; Burnham & Dodd, 2004; Dodd, 1979; Kushnerenko, Teinonen, Volein & Csibra, 2008; Rosenblum, Schmuckler & Johnson, 1997). Importantly, such sensitivity appears to facilitate certain aspects of language acquisition, such as the learning of phonemes, phonotactic constraints (Kushnerenko *et al.*, 2008; Teinonen, Aslin, Alku & Csibra, 2008), and new words (Hollich, Newman & Jusczyk, 2005). Indeed, how infants fixate on a talker's face during audiovisual speech perception at the age of 6–9 months has been shown to predict receptive and expressive language skills during their second year of life (Kushnerenko, Tomalski, Ballieux, Potton, Birtles *et al.*, 2013; Young, Merin, Rogers & Ozonoff, 2009). Such findings suggest that we may need to reconsider the

factors that shape successful language acquisition. More specifically, if visual speech cues facilitate language learning, then we would expect that those children who are good audiovisual integrators should also be effective language learners. On the other hand, a breakdown in audiovisual processing may be expected to delay or disrupt at least some aspects of language acquisition and thus, hypothetically, contribute to the development of language disorders.

To date, only a few studies have examined audiovisual processing in children with language learning difficulties. These studies have focused on children with specific language impairment (SLI) (Boliek, Keintz, Norrix & Obrzut, 2010; Hayes, Tiippana, Nicol, Sams & Kraus, 2003; Meronen, Tiippana, Westerholm & Ahonen, 2013; Norrix, Plante & Vance, 2006; Norrix, Plante, Vance & Boliek, 2007). SLI is a language disorder that is associated with significant linguistic difficulties in the absence of hearing impairment, frank neurological disorders, or low non-verbal intelligence (Leonard, 2014). At 7% prevalence (Tomblin, Records, Buckwalter, Zhang, Smith *et al.*, 1997), it affects more children than either autism (approximately 1.5%, Center for Disease Control and Prevention, 2014) or stuttering (approximately 5%, Yairi & Ambrose, 2004).

To the best of our knowledge, almost all previous studies of audiovisual speech perception in SLI employed the McGurk paradigm (in which, typically, an auditory 'pa' or 'ba' is dubbed onto a visual articulation of 'ka' or 'ga', with the resultant perception being that of 'ta' or 'da' if audiovisual integration took place) (Boliek *et al.*, 2010; Hayes *et al.*, 2003; Meronen *et al.*, 2013; Norrix *et al.*, 2006; Norrix *et al.*, 2007). The exact measures used to evaluate the susceptibility to this illusion varied, with some studies focusing on the number of fused perceptions (e.g. Boliek *et al.*, 2010; Norrix *et al.*, 2007) and others examining the influence of visual speech cues more broadly, by including in their analyses the number of children's responses to incongruent audiovisual speech that were based on visual speech cues alone (so-called visual capture) (e.g. Meronen *et al.*, 2013). The degree of susceptibility to the illusion in both typically developing (TD) children and in children with SLI also varied. For example, Norrix and colleagues (Norrix *et al.*, 2007) reported that when presented with an incongruent audiovisual syllable (auditory 'bi' dubbed onto visual 'gi'), TD children heard 'bi' on 65.4% of trials, while children with SLI heard 'bi' on 88.6% of trials. On the other hand, the study by Meronen and colleagues reported that both TD and SLI children perceived the fused McGurk syllable at approximately similar levels, both in quiet and in noise

(22–38%), but that only TD children were more likely to base their responses on visual modality when either sound intensity or signal-to-noise ratio decreased. Regardless of the measures used, most authors conclude that individuals with SLI show reduced susceptibility to the McGurk illusion, suggesting that they are less able to use visual speech cues during audiovisual speech perception.

The majority of children are diagnosed with SLI during pre-school years (Leonard, 2014), and, in many cases, their language difficulties continue during school years and into adulthood (Conti-Ramsden, Botting, Simkin & Knox, 2001; Hesketh & Conti-Ramsden, 2013; Law, Rush, Schoon & Parsons, 2009; Rice, Hoffman & Wexler, 2009; Stark, Bernstein, Condino, Bender, Tallal *et al.*, 1984). Even when H-SLI children appear to catch up with their TD peers in linguistic proficiency, their apparent recovery may be only ‘illusory’ as suggested by Scarborough and Dobrich, with a high risk for such children to fall behind again (Scarborough & Dobrich, 1990). In some cases, linguistic difficulties become more subtle with age, which makes their detection more challenging. Standardized tests currently available to clinicians and researchers vary substantially in the ability to reliably identify language impairment in older children and adults (Conti-Ramsden, Botting & Faragher, 2001a; Fidler, Plante & Vance, 2011; Spaulding, Plante & Farinella, 2006). Importantly, normal scores on such tests may be misleading because they can hide atypical cognitive strategies employed by H-SLI children during test taking (Karmiloff-Smith, 2009). Furthermore, when H-SLI children’s scores do reach the normal range on standardized tests, they often fall within what can be called the low normal range. A recent study by Hampton Wray and Weber-Fox has shown that even in TD children low normal and high normal scores on standardized language tests were associated with distinct patterns of brain activity to semantic and syntactic violations (Hampton Wray & Weber-Fox, 2013), suggesting that achieving scores within the normal range on language tests may not be sufficient for effective linguistic processing. During school years, language skills are called upon to build knowledge in various academic domains, and even a subtle weakness in linguistic competence may put an H-SLI child at a disadvantage compared to his or her TD peers. And indeed previous research shows that H-SLI children are at high risk for learning disorders, most notably reading (Catts, Fey, Tomblin & Zhang, 2002; Hesketh & Conti-Ramsden, 2013; Scarborough & Dobrich, 1990; Stark *et al.*, 1984). Social skills may also suffer, with poor language proficiency being a strong predictor of behavioral problems in children

(Fujiki, Brinton, Morgan & Hart, 1999; Fujiki, Brinton & Todd, 1996; Peterson, Bates, D’Onofrio, Coyne, Lansford *et al.*, 2013).

Given the continuing vulnerability of language skills in many H-SLI children, the ability to exploit speech-related visual information may be very beneficial for this population, especially if we consider that in many schools the background noise can reach 65 dB (Shield & Dockrell, 2003). Yet, while growing evidence shows that children with a current diagnosis of SLI show audiovisual deficits, the state of audiovisual skills in school-age H-SLI children is largely unknown. In an earlier study, we reported that, despite scoring within the normal range on standardized tests of linguistic competence, children with a history of SLI had significantly more difficulty in differentiating synchronous and asynchronous presentations of simple audiovisual stimuli (Kaganovich, Schumaker, Leonard, Gustafson & Macias, 2014), suggesting that at least some aspects of audiovisual processing are atypical in this group. However, no study to date has examined audiovisual speech perception in this population.

In this study, we evaluated audiovisual speech perception in children with a history of SLI, their TD peers, and college-age adults by using both electrophysiological and behavioral measures. All of our H-SLI children were first diagnosed with SLI when they were 4–5 years old and were tested in the current study when they were 7–11 years old. Each participant took part in two experiments. In the first, we took advantage of a well-established electrophysiological paradigm that permits the evaluation of an early stage of audiovisual speech perception on the basis of naturally produced and audiovisually congruent speech (Besle, Bertrand & Giard, 2009; Besle, Fischer, Bidet-Caulet, Lecaigard, Bertrand *et al.*, 2008; Besle, Fort, Delpuech & Giard, 2004a; Besle, Fort & Giard, 2004b; Giard & Besle, 2010; Knowland *et al.*, 2014). This paradigm compares event-related potentials (ERPs) elicited by audiovisual (AV) speech with the sum of ERPs elicited by auditory only (A) and visual only (V) speech. During the sensory stage of encoding – approximately the first 200 milliseconds (ms) post-stimulus onset – ERPs elicited by auditory and visual stimulation are assumed to sum up linearly (Besle *et al.*, 2004b; Giard & Besle, 2010). Therefore, 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 AV speech is identical to the algebraic sum of the same components elicited by A and V speech (the A+V condition). Audiovisual integration, on the other hand, leads to the attenuation of the N1 and P2 amplitude (and sometimes latency) in the AV stimuli as compared to

the A+V condition (Besle *et al.*, 2004b; Giard & Besle, 2010; Van Wassenhove, Grant & Poeppel, 2005).

Within the context of this paradigm, changes in the N1 and P2 amplitude and latency to AV stimuli can occur independently of each other and thus are thought to index different aspects of audiovisual processing. The N1 amplitude is sensitive to how well facial movements can cue the temporal onset of the auditory signal (Baart, Stekelenburg & Vroomen, 2014; Van Wassenhove *et al.*, 2005). Its attenuation does not depend on the nature of the stimuli (speech or non-speech) and may even be observed in incongruent audiovisual presentations (Stekelenburg & Vroomen, 2007). This hypothetical functional role of N1 attenuation to AV stimuli is supported by the fact that it is typically absent when the visual stimuli are presented simultaneously with the auditory signals (e.g. Brandwein *et al.*, 2011) or when they carry little information about the timing of the sound onset (e.g. Baart *et al.*, 2014). The shortening of the N1 latency, on the other hand, is influenced by the degree to which lip movements predict the identity of the articulated phoneme (Van Wassenhove *et al.*, 2005). Accordingly, articulation of bilabial phonemes may lead to a greater shortening of the N1 latency than the articulation of velar phonemes.

The functional role of the P2 attenuation to AV stimuli has recently been examined by Baart and colleagues (Baart *et al.*, 2014). These authors presented two groups of participants with sine-wave speech. One group was trained to hear it as speech while another perceived it as computer noise. It was observed that while N1 attenuation to the AV condition was evident in both groups, the P2 attenuation was present only in the ERP records of those individuals who perceived their stimuli as speech. The authors proposed that P2 attenuation in their study indexed audiovisual phonetic binding, in which auditory and visual modalities are perceived as representing a unitary audiovisual event.

In the first experiment of our study, which was based on the additive ERP paradigm discussed above, participants watched a speaker articulate syllables 'ba', 'da', and 'ga' presented in an auditory only, visual only, or audiovisual manner. Their task was to press a button every time they saw the speaker do something silly. Such silly events were intermixed with syllable videos. The use of this ERP paradigm has, therefore, allowed us to zoom in on a relatively early stage of audiovisual processing and examine its status without the need for overt response to or identification of specific phonemes. If audiovisual integration during sensory processing is normal in H-SLI children, 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 TD children. A complete lack of changes in these components or a smaller degree of their attenuation would indicate age-inappropriate development. In addition, by comparing the degree of N1 and P2 attenuation to AV speech in children and adults, we aimed to understand whether neural mechanisms underlying such attenuation are mature by mid-childhood years.

Because research on audiovisual processing in SLI is based almost entirely on the McGurk paradigm, in the second experiment of this study we administered the McGurk test to our groups of participants. The goal of this experiment was to understand whether susceptibility to this illusion is reduced in H-SLI children in a fashion similar to that reported earlier for children with a current diagnosis of SLI. Based on earlier studies of McGurk perception in children with SLI, we hypothesized that H-SLI children may have weaker susceptibility to the McGurk illusion compared to their TD peers; however, the degree of such difference remained to be determined. While the McGurk paradigm has been used extensively in research on audiovisual processing, recent findings suggest that the neural mechanisms underlying it may be at least partially different from those engaged during the perception of congruent audiovisual speech (Erickson, Zielinski, Zielinski, Liu, Turkeltaub *et al.*, 2014). By testing H-SLI children on both the additive electrophysiological paradigm and the McGurk illusion paradigm, we were able to examine the processing of both audiovisually congruent and audiovisually incongruent speech in this group.

Method

Participants

Sixteen children with a history of SLI (H-SLI) (3 female; average age 9;5; range 7;3–11;5), 16 typically developing (TD) children (4 female; average age 9;7; range 7;9–11;6), and 16 adults (8 female; average age 22; range 18–28) participated in the study. All but one of the H-SLI children and all but two TD children also participated in an earlier study by Kaganovich and colleagues (Kaganovich *et al.*, 2014). All gave their written consent or assent to participate in the experiment. In addition, at least one parent of each child gave 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).

Information on children's linguistic ability and verbal working memory was collected with the following

standardized tests: the Concepts and Following Directions, Recalling Sentences, Formulated Sentences, Word Structure (7- and 8-year-olds only), and Word Classes-2 Total (9–11-year-olds only) sub-tests of the Clinical Evaluation of Language Fundamentals – 4th edition (CELF-4; Semel, Wiig & Secord, 2003); non-word repetition test (Dollaghan & Campbell, 1998); Number Memory Forward and Number Memory Reversed sub-tests of the Test of Auditory Processing Skills – 3rd edition (TAPS-3; Martin & Brownell, 2005). In addition, in order to rule out the influence of attentional skills on possible group differences between TD and H-SLI children, we administered the Sky Search, Sky Search Dual Task, and Map Mission sub-tests of the Test of Everyday Attention for Children (TEA-Ch; Manly, Robertson, Anderson & Nimmo-Smith, 1999). In the Sky Search sub-test of TEA-Ch, children circled matching pairs of spaceships, which appeared among similar looking distractor ships. In the Dual Task version of this task, children also counted concurrently presented sound pips. Lastly, in Map Mission, children were given 1 minute to circle as many ‘knives-and-forks’ symbols on a city map as possible. TEA-Ch tasks measured children’s ability to selectively focus, sustain, or divide their attention. In addition, all children were administered the Test of Nonverbal Intelligence – 4th edition (TONI-4; Brown, Sherbenou & Johnsen, 2010) to rule out mental retardation and the Childhood Autism Rating Scale – 2nd edition (Schopler, Van Bourgondien, Wellman & Love, 2010) to rule out the presence of autism spectrum disorders. The level of mothers’ and fathers’ education was measured as an indicator of children’s socioeconomic status (SES). In all participants, handedness was assessed with an augmented version of the Edinburgh Handedness Questionnaire (Cohen, 2008; Oldfield, 1971). Tables 1 and 2 show

group means and standard errors (*SE*) for all of the aforementioned measures.

H-SLI children were diagnosed as having SLI during pre-school years based on either the Structured Photographic Expressive Language Test – 2nd Edition (SPELT-II; Werner & Kresheck, 1983) or the Structured Photographic Expressive Language Test – Preschool 2 (Dawson, Eyer & Fonkalsrud, 2005). Both tests have shown good sensitivity and specificity (Greenslade, Plante & Vance, 2009; Plante & Vance, 1994). During diagnosis, all children fell below the 10th percentile on the above tests, with the majority falling below the 2nd percentile (range 1st–8th percentile), thus revealing severe impairment in linguistic skills. At the time of the present testing, all but two of these children scored within the normal range of language abilities based on the Composite Language Score (CLS) of CELF-4. However, as a group, their CLS scores were nonetheless significantly lower than those of the TD children (see Tables 1 and 2 for statistics on all group comparisons). More specifically, out of the four sub-tests of CELF-4 that collectively comprise the composite CLS, the H-SLI children performed significantly worse on two – Recalling Sentences and Formulated Sentences. Children in the H-SLI group also had significantly more difficulty repeating 3-syllable and 4-syllable nonwords, with a similar trend for 2-syllable nonwords. Importantly, the Recalling Sentences sub-test of CELF-4 and the non-word repetition test, both of which tap into short-term working memory, have been shown to have high sensitivity and specificity when identifying children with language disorders (Archibald, 2008; Archibald & Gathercole, 2006; Archibald & Joanisse, 2009; Conti-Ramsden, Botting & Faragher, 2001b; Dollaghan & Campbell, 1998; Ellis Weismer, Tomblin, Zhang, Buckwalter, Chynoweth *et al.*, 2000; Graf Estes, Evans & Else-Quest,

Table 1 Group means for age, non-verbal intelligence (TONI-4), presence of autism (CARS-2), socioeconomic status (parents’ education level) and linguistic ability (CELF-4)

	Age (years, months)	TONI-4	CARS	Mother’s education (years)	Father’s education (years)	CELF-4					
						CFD	RS	FS	WS	WC R/E/T	CLS
H-SLI	9;5(0.4)	107.9(2.3)	15.4(0.1)	16.3(0.8)	14.8(0.4)	10.8(0.6)	7.7(0.7)	10.4(0.7)	10.3(1.0)	10.6(0.9)/9.6(1.4)/10.1(1.1)	98.8(3.2)
TD	9;7(0.3)	108.4(2.4)	15.3(0.2)	16.4(0.8)	17.1(0.8)	12.1(0.4)	11.4(0.5)	12.6(0.4)	11.3(0.6)	12.7(0.8)/11.6(1.0)/12.3(0.9)	111.9(1.9)
<i>F</i>	<1	<1	<1	<1	6.189	3.384	19.998	6.779	<1	3.014/1.315/2.442	12.661
<i>p</i>	.820	.898	.592	.956	.021	.076	<.001	.014	.407	.103/0.270/0.139	.001

Note: Numbers for TONI-4, CARS-2, and the CELF-4 subtests represent standardized scores. Numbers in parentheses are standard errors of the mean. *P* and *F* values reflect a group comparison. For age, TONI-4, CARS, mother’s education, father’s education, and the CFD, RS, and FS sub-tests of CELF-4, as well as for the CLS index, the degrees of freedom were (1,31). For the WS sub-test of CELF-4, the degrees of freedom were (1,14), and for the WC sub-tests of CELF-4, the degrees of freedom were (1,16). CFD = 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.

Table 2 Group means for nonword repetition, auditory processing (TAPS-3), and attention (TEA-Ch) skills

	Nonword Repetition					TAPS-3		TEA-Ch		
	Syllables					Number memory				
	1	2	3	4	Total	Forward	Reversed	Sky Search	Sky Search Dual Task	Map Mission
H-SLI	91.1(1.8)	95.3(1.1)	87.5(3.0)	62.7(2.9)	80.3(1.7)	7.9(0.4)	9.3(0.6)	9.4(0.7)	7.1(0.7)	7.5(0.9)
TD	92.2(1.2)	98.1(1.0)	96.9(0.9)	81.4(2.9)	90.8(1.2)	10.4(0.6)	11.8(0.6)	9.8(0.5)	8.6(0.9)	9.0(0.7)
<i>F</i>	<1	3.671	8.813	20.512	24.992	11.952	8.727	<1	1.771	1.901
<i>p</i>	.618	.065	.008	<.001	<.001	.002	.006	.723	.193	.178

Note: Numbers for TAPS-3 and TEA-Ch represent standardized scores. TEA-Ch tasks measured children's ability to selectively focus, sustain, or divide their attention. Numbers for Nonword Repetition are percent correct of repeated phonemes. Numbers in parentheses are standard errors of the mean. *P* and *F* values reflect a group comparison. For all group comparisons, the degrees of freedom were (1,31).

2007; Hesketh & Conti-Ramsden, 2013). Furthermore, parents of 10 H-SLI children noted the presence of at least some language-related difficulties in their children during the time of testing, and two of the H-SLI children had stopped speech therapy just a year earlier. We take these results to indicate that, despite few overt grammatical errors produced by our cohort of H-SLI children, they continued to lag behind their TD peers in linguistic competence.

In addition to the CELF-4 and nonword repetition group differences described above, the H-SLI children also performed worse than their TD peers on the Number Memory Forward and the Number Memory Reversed sub-tests of TAPS-3. The two groups of children did not differ in age, non-verbal intelligence, or SES as measured by mothers' years of education. Although fathers of the H-SLI children had, on average, fewer years of education than fathers of the TD children, this group difference was less than 3 years. None of the children had any signs of autism spectrum disorders. Two children in the H-SLI group had a diagnosis of attention deficit/hyperactivity disorder (ADD/ADHD), but were not taking medication at the time of the study. One child in the H-SLI group had a diagnosis of dyslexia, based on parental report.

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 having normal or corrected-to-normal vision, and were not taking medication that may affect brain function (such as anti-depressants) at the time of study. All children in the TD group had language skills within the normal range and were free of ADD/ADHD. None of the adult participants had a history of SLI, based on self-report. According to the Laterality Index of the Edinburgh Handedness Questionnaire, one child in the H-SLI group was left-handed. All other participants were right-handed.

Stimuli and experimental design

All participants performed two tasks. The main task was similar to the one described by Besle and colleagues (Besle *et al.*, 2004a), with slight modifications in order to make it appropriate for use with children. This task is referred to hereafter as the Congruent Syllables task. 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 – auditory only (A), visual only (V), and audiovisual (AV). 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. In addition, videos with silly facial expressions – a clown sticking out her tongue or blowing a raspberry – were presented randomly on 25% 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 the videos presented 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

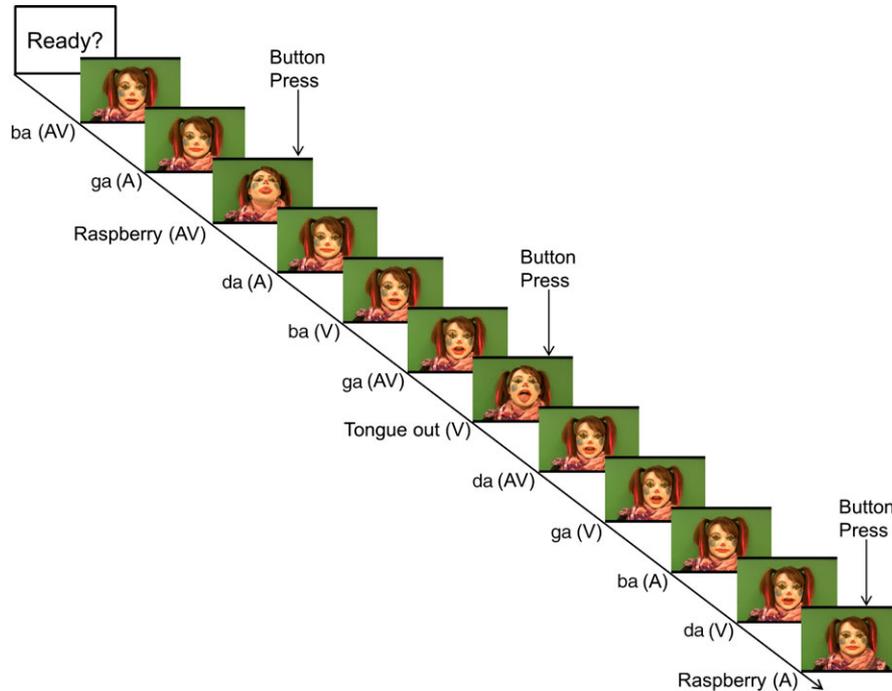


Figure 1 Schematic representation of a block in the Congruent Syllables task. This timeline shows a succession of trials from top left to bottom right. AV = audiovisual trials, A = auditory only trials, V = visual only trials.

block contained 48 trials and lasted approximately 4 minutes. Each participant completed eight blocks. Hand to response button mapping was counterbalanced across participants. Responses to silly expressions were accepted if they occurred within 3000 ms following the onset of the video. This task was combined with EEG recordings (see below).

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 700–1000 ms. All syllable videos were 1400 ms in duration. The onset of sound in the A and AV videos always occurred at 360 ms post video onset. This point served as time 0 for all ERP averaging. 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 for the 'ga' tokens. The 'raspberry' video lasted 2167 ms, while the 'tongue-out' video lasted 2367 ms.

In the second experiment (referred to hereafter as the McGurk task), participants watched short videos, presented one at a time and in a random order, in which the same female clown as described above articulated one of three syllables. Two of these syllables had congruent auditory and visual components – one was a 'pa' and one was a 'ka'. The third was the McGurk syllable, in which the auditory 'pa' was dubbed onto the visual articulation of a 'ka', with the resulting percept being that of a 'ta' (hereafter we refer to the McGurk syllable as 'ta'^{McGurk}). The McGurk syllable was created by aligning the burst onset of the 'pa' audio with the burst onset of the 'ka' audio in the 'ka' video clip and then deleting the original 'ka' audio (Adobe Premiere Pro CS5, Adobe Systems Incorporated, USA). All syllable videos were 1634 ms in duration. On each trial, participants watched one of three videos, which was followed by a screen with syllables 'pa', 'ta', and 'ka' written across it. This screen signaled the onset of a response window, which lasted for 2000 ms. Participants indicated which syllable they heard by pressing a response pad button corresponding to the position of the syllable on the screen (RB-530, Cedrus Corporation). The order of written syllables on the screen was counterbalanced across participants. Prior to testing, all children were asked to read syllables 'pa', 'ta', and 'ka' presented on a screen to ensure that reading difficulties did not affect the results of the study. The inter-stimulus interval (ISI)

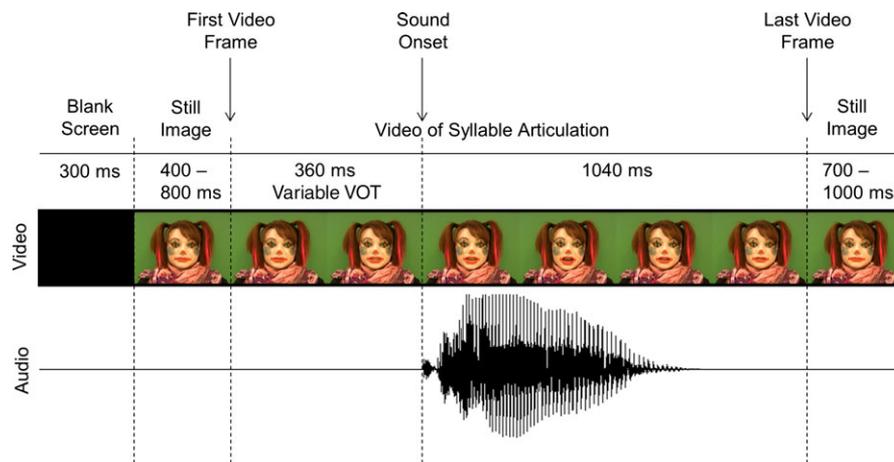


Figure 2 Schematic representation of a trial in the Congruent Syllables task. Main events within a single trial are shown from left to right. Note that only representative frames are displayed. The sound onset was used as time 0 for all ERP averaging. The time between the first noticeable facial movement and sound onset (VOT) varied between 167 and 367 ms depending on the syllable (see text for details).

varied randomly between 250, 500, 750 and 1000 ms. Each participant was administered four blocks with 15 trials each (five presentations of each of the ‘pa’, ‘ka’, and ‘ta’^{McGurk} syllables), yielding 20 responses to each type of syllable from each participant. No electroencephalographic (EEG) data were collected during this part.

All videos for both tasks 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 (Borsma & 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.

Behavioral and ERP testing took place over three non-consecutive days. All language, attention, and non-verbal intelligence tests were administered during two separate sessions in order to keep each session’s duration to less than 1.5 hours. Both the Congruent Syllable task and the McGurk task were administered during the third session, with the McGurk task always being first. 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 in each task. Children played several rounds of the board game of their choice during breaks. Participants were fitted with an EEG cap during a longer break between the two tasks. Together with breaks, the ERP testing session lasted approximately 2 hours.

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, The 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 pre-stimulus and ending at 1000 ms post-stimulus onset. The 200 ms prior to the stimulus 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 – 83 (range 62–94) for H-SLI children, 85 (range 68–94) for TD children, and 90 (range 63–95) for adults; V condition – 86 (range 65–93) for H-SLI children, 85 (range 72–94) for TD children, and 91 (range 74–96) for adults; AV condition – 82 (range 63–93) for H-SLI children, 84 (range 71–96) for TD children, and 90 (range 68–95) for adults. Adults had significantly more clean trials than the H-SLI group of children (group, $F(1, 45) = 4.275$, $p = .02$, $\eta_p^2 = 0.16$; adults vs. H-SLI, $p = .025$; adults vs. TD, $p = .096$, TD vs. H-SLI, $p = 1$). However, although significant, this difference was only seven trials. 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 and the P2 peak amplitude and peak latency were measured between 198 and 298 ms post-stimulus onset from the ERP waveforms elicited by AV syllables and the ERP waveforms obtained from the A+V summation. Peaks were defined as the largest points within a specified window that were both preceded and followed by smaller values. 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 the Congruent Syllables Task, 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 (H-SLI, TD, 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 the N1 and P2 components with reduced peak amplitude 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 (H-SLI, TD, adults) as a between-subjects variable, separately on the N1 and P2 components. Second, 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 repeated-measures ANOVA 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 (H-SLI, TD, and adults) as a between-subjects variable, again separately for the N1 and P2 components. 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.

We recorded the following measures during the McGurk study: the accuracy of responses to the 'pa', 'ta'^{McGurk}, and 'ka' syllables; the type and number of errors made when responding to the 'ta'^{McGurk} syllable (namely, how often it was perceived as 'pa' (in agreement with its auditory component) or as 'ka' (in agreement with its visual component)); reaction time, and the number of misses for each of the three syllables. Analyses of ACC, RT, and misses included syllable type as a within-subject variable ('pa', 'ta'^{McGurk}, and 'ka') and group as a between-subjects variable (H-SLI, TD, and adults). Analysis of error types included a within-subject factor of error type ('pa', 'ka') and a between-subjects factor of group (H-SLI, TD, and adults). One child in the H-SLI group failed to provide any accurate responses to the congruent 'pa' syllable of the McGurk task, suggesting that he was not following the experimenter's instructions. His data were excluded from the McGurk study analysis.

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 are 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.

Results

Congruent syllables study

Behavioral results

As expected, all three groups were able to detect 'raspberry' and 'tongue-out' events with high accuracy (see Figure 3). The H-SLI children performed overall less accurately than adults (group, $F(2, 45) = 4.429$, $p = .018$, $\eta_p^2 = 0.164$; H-SLI vs. adults, $p = .014$; TD vs. adults, $p = .368$). However, the two groups of children did not differ from each other (H-SLI vs. TD, $p = .504$), and there was no group by condition interaction ($F(4, 90) = 1.553$, $p = .194$). Adults were also overall faster than either group of children (group, $F(2, 45) = 18.053$, $p < .001$, $\eta_p^2 = 0.445$; adults vs. H-SLI, $p < .001$; adults vs. TD, $p < .001$; H-SLI vs. TD, $p = 1$). All participants were slower to respond to A silly events as compared to V and AV silly events (condition, $F(2, 90) = 26.165$, $p < .001$, $\eta_p^2 = 0.368$; A vs. V, $p < .001$; A vs. AV, $p < .001$; V vs. AV, $p = 1$). Shorter RTs in the V and AV conditions were most likely due to the fact that facial changes indicating the onset of silly events could be detected prior to the onset of the accompanying silly sound.

ERP results

ERPs elicited by A and V conditions are shown in Figure 4, while Figure 5 displays ERPs from the AV and A+V conditions. As can be easily seen from these figures, although children's and adults' waveforms differed

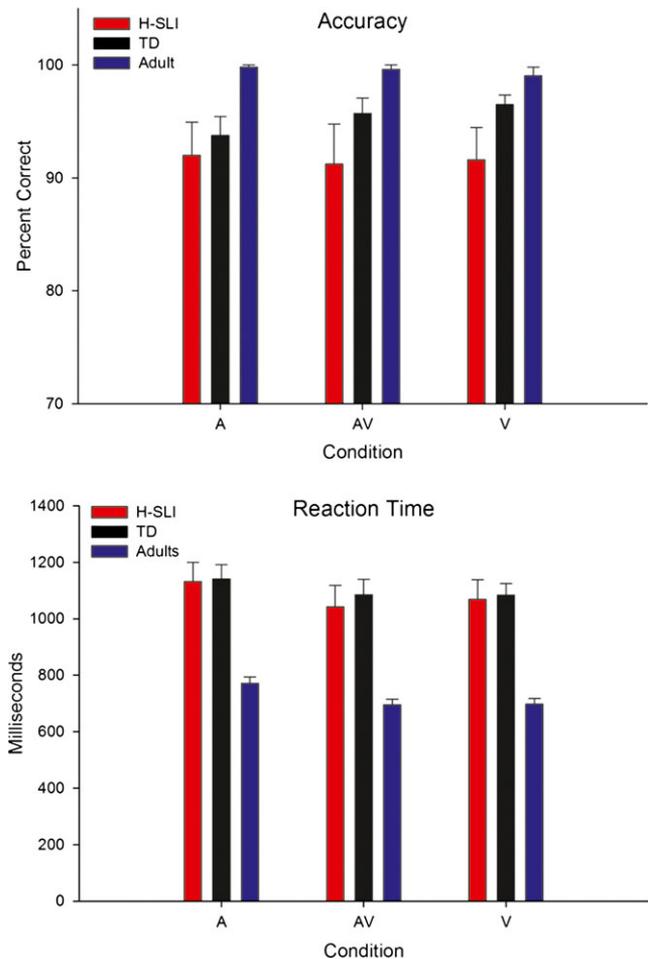


Figure 3 Performance on the Congruent Syllables task. Accuracy and reaction time for the task of detecting silly facial expressions are shown for each group and each condition. Error bars represent standard errors of the mean.

vastly, both in the size and the nature of observed components, clear N1 and P2 were present in all groups. It is important to note that despite significant ERP differences between children and adults during the later portion of the 1000 ms epoch, our analyses focus on just the N1 and P2 ERP components. As mentioned in the Introduction, electrical signals elicited by auditory and visual modalities can be assumed to sum up linearly only during early sensory processing (Besle *et al.*, 2004b; Giard & Besle, 2010). Therefore, comparing AV responses with the sum of A and V responses during a later time window becomes invalid.

N1 analysis. The N1 peak amplitude was smaller in the AV compared to the A+V condition (see panels (a) and (b) of Figure 5) (condition: $F(1, 45) = 34.224$, $p < .001$, $\eta_p^2 = 0.432$). This effect did not interact with the factor of

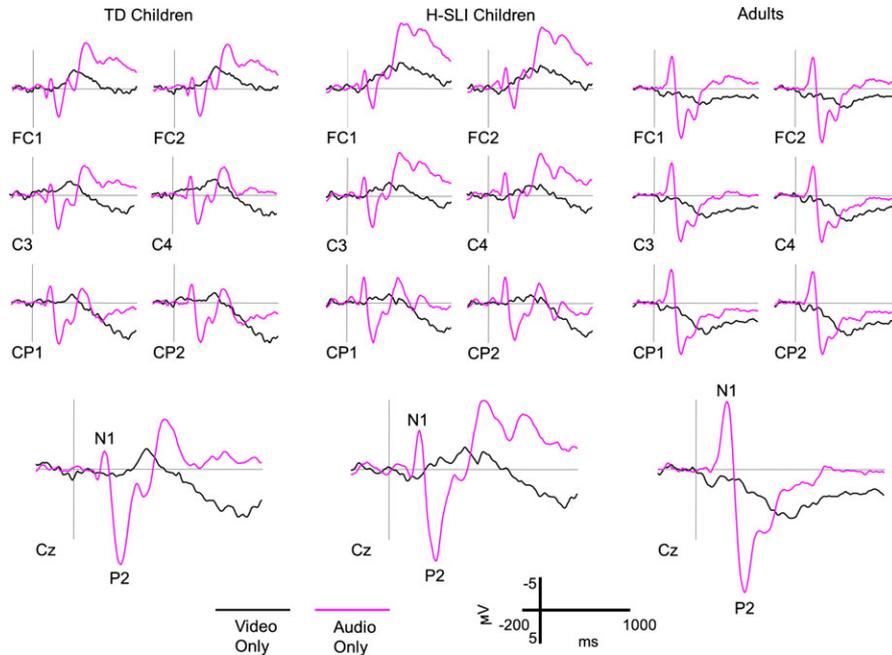


Figure 4 ERP components in A and V conditions. Grand average waveforms elicited by auditory only (A) and visual only (V) syllables are shown for three groups of participants. Six mid-lateral sites and one midline site are included for each group. All groups displayed clear N1 and P2 components, which are marked on the Cz site. Negative is plotted up. Visual syllables did not elicit clear peaks because visual articulation was in progress at the time of the sound onset. This observation is in agreement with earlier reports.

group ($F(2, 45) = 1.241, p = .299$), and there was no overall group effect ($F(2, 45) < 1$). Analysis of the N1 peak amplitude difference between the A+V and AV conditions showed the main effect of laterality section ($F(1.439, 64.77) = 6.834, p = .005, \eta_p^2 = 0.132, \varepsilon = 0.72$), with smaller N1 reduction over the left sites as compared to over the midline and right sites (left vs. midline, $p = .035$, left vs. right, $p = .016$; midline vs. right, $p = .189$). Although, overall, groups did not differ from each other in the amount of the N1 peak amplitude attenuation ($F(2, 45) = 1.241, p = .299$), the group by site interaction was significant ($F(2.714, 61.057) = 7.14, p = .001, \eta_p^2 = 0.241, \varepsilon = 0.678$). Follow-up analyses that aimed to determine whether groups differed at each level of site (FC1, Fz, FC2 vs. C3, Cz, C4 vs. CP1, Pz, CP2) revealed a significant group effect only over frontal and fronto-central sites ($F(2, 45) = 4.486, p = .017, \eta_p^2 = 0.166$), with larger N1 peak amplitude reduction in the H-SLI group compared to their TD peers (H-SLI vs. TD, $p = .013$; H-SLI vs. adults, $p = .465$, TD vs. adults, $p = .386$). Analysis of the N1 peak latency produced no significant findings. In addition, because at least one previous report (Knowland *et al.*, 2014) suggested that N1 attenuation to AV speech increases with age, we conducted a linear regression analysis between the average of the N1 attenuation over all nine sites of our ERP analysis and children's age. Because the degree of

N1 attenuation did not differ between groups, we combined TD and H-SLI children for this regression analysis. We found that, indeed, N1 attenuation was significantly correlated with age, with older children having on average greater N1 attenuation ($R = 0.385, F(1, 31) = 5.224, p = .03$).

In sum, the N1 peak amplitude was significantly reduced in the AV as compared to the A+V condition in all three groups. The N1 attenuation was largest over the right and midline electrode sites and was positively correlated with age in children. The distribution of the N1 attenuation differed somewhat between the two groups of children, with H-SLI children showing significantly larger N1 peak amplitude reduction over frontal and fronto-central sites compared to their TD peers. No significant latency effects have been found.

P2 analysis. The amplitude of P2 differed only marginally between the AV and A+V conditions ($F(1, 45) = 2.732, p = .105, \eta_p^2 = 0.057$); however, it differed significantly between groups (group, $F(2, 45) = 4.666, p = .014, \eta_p^2 = 0.172$; group by site ($F(2.592, 58.309) = 3.594, p = .024, \eta_p^2 = 0.138, \varepsilon = 0.648$). Follow-up analyses examined the effect of group at each level of site (FC1, Fz, FC2 vs. C3, Cz, C4 vs. CP1, Pz, CP2) and found that adults had a significantly larger P2 peak amplitude over frontal and fronto-central sites compared

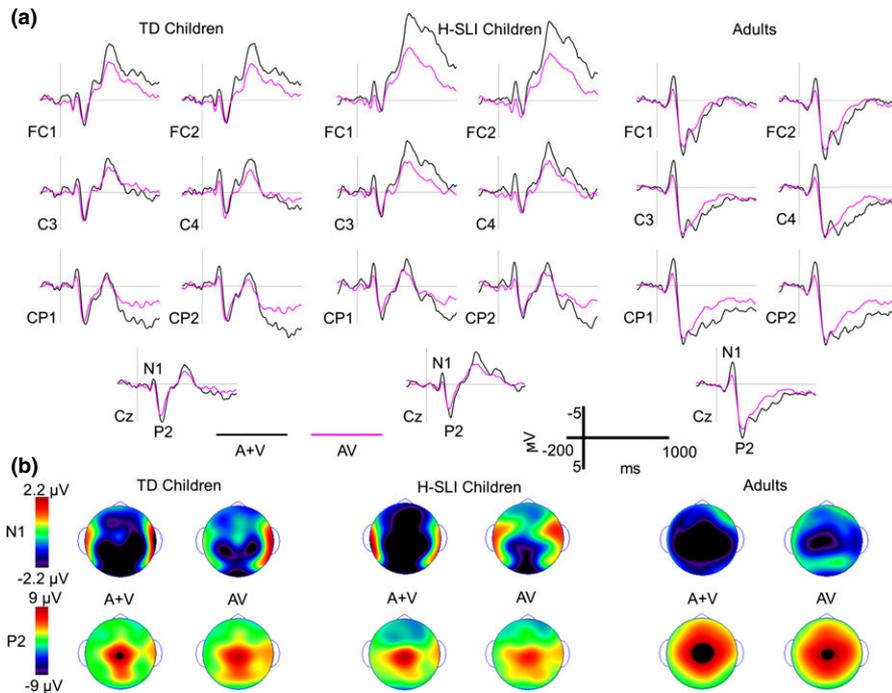


Figure 5 ERP components in AV and A+V conditions. (a) Grand average waveforms elicited by audiovisual syllables (AV) and by the sum of auditory only and visual only (A+V) syllables are overlaid separately for TD children, H-SLI children, and adults. Six mid-lateral sites and one midline site are shown for each group. All 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 peaks of the N1 and P2 components are displayed for each group of participants in the AV and A+V condition. Note a significant reduction in negativity at the N1 peak in the AV compared to the A+V condition in each group. Although noticeable, a reduction in P2 in the AV condition failed to reach significance.

to either group of children (adults vs. H-SLI, $p = .001$; adults vs. TD, $p = .047$; H-SLI vs. TD, $p = .432$). A similar trend was also observed over central sites ($F(2, 45) = 3.013$, $p = .059$, $\eta_p^2 = 0.118$). Greater amplitude of P2 in adults was also accompanied by a longer latency of this component (group, $F(2, 45) = 13.489$, $p < .001$, $\eta_p^2 = 0.375$; adults vs. H-SLI, $p < .001$; adults vs. TD, $p < .001$; H-SLI vs. TD, $p = 1$). Analysis of the P2 peak amplitude difference between the AV and A+V conditions yielded no significant results.

In sum, the P2 amplitude was marginally smaller in the AV compared to the A+V condition. In regard to group differences, the P2 amplitude was larger and its latency longer in adults compared to both groups of children. All other analyses yielded insignificant results.

McGurk study

The results of the McGurk study are reported in Table 3. Overall, all groups performed the McGurk task with high accuracy (see Figure 6, top panel, for overall results and Figure 7 for individual differences in each group), but were less accurate when identifying the 'ta'^{McGurk}

syllable compared to when identifying either 'pa' or 'ka'. Accuracies for 'pa' and 'ka' did not differ from each other. The H-SLI group performed less accurately than adults but similarly to their TD peers. This accuracy group effect did not interact with syllable type. A lower accuracy of responses to the 'ta'^{McGurk} syllable was accompanied by a prolonged RT in all groups (see Figure 6, bottom panel). Importantly, groups differed in the types of errors made when responding to the 'ta'^{McGurk} syllable, with the H-SLI group perceiving significantly more instances of 'pa' than 'ka' when viewing the 'ta'^{McGurk} syllable. This relationship was absent in both the TD and the adult groups. Lastly, the analysis of misses revealed that the H-SLI children had more misses than adults when responding to the 'ka' syllable. Notably, groups did not differ in the number of misses to the 'ta'^{McGurk} syllable.

Discussion

We examined the processing of audiovisually congruent and incongruent speech in children with a history of SLI,

Table 3 Results of the McGurk study

	ACC	RT	Misses	Error type
Factors/Interactions				
Syllable Type	$F(2, 88)=9.368$, $p = .003$, $\eta_p^2=0.176$	$F(2, 88)=10.527$, $p < .001$, $\eta_p^2=0.193$		
‘ta’ vs. ‘pa’	$p = .004$, ‘ta’<‘pa’	$p = .001$, ‘ta’>‘pa’		
‘ta’ vs. ‘ka’	$p = .025$, ‘ta’<‘ka’	$p = .002$, ‘ta’>‘ka’		
‘pa’ vs. ‘ka’	$p = .184$	$p = 1$		
Group	$F(2, 44)=4.758$, $p = .013$, $\eta_p^2=0.178$			
H-SLI vs. Adults	$p = .014$, H-SLI<Adults			
H-SLI vs. TD	$p = .104$			
TD vs. Adults	$p = 1$			
Group × Syllable Type	$F(4, 88)=2.021$, $p = .138$		$F(4, 88)=2.582$, $p = .043$, $\eta_p^2=0.105$	
group for ‘ta’			$F(2, 44)=1.554$, $p = .223$	
group for ‘pa’			$F(2, 44)=2.628$, $p = .083$	
group for ‘ka’			$F(2, 44)=5.017$, $p = .011$	
H-SLI vs. Adults			$p = .009$, H-SLI > Adults	
H-SLI vs. TD			$p = .094$	
TD vs. Adults			$p = .592$	
Group × Error Type				$F(2, 44)=2.657$, $p = .081^1$, $\eta_p^2=0.108$
H-SLI				$F(1, 14)=7.999$, $p = .013$, $\eta_p^2=0.364$, ‘pa’>‘ka’
TD				$F(1, 15)=1.154$, $p = .3$
Adults				$F(1, 15)=1.15$, $p = .3$

¹Although the error type by group interaction fell somewhat short of significance, the medium effect size of this interaction warranted follow-up analyses.

their TD peers, and college-age adults. In the first experiment, we evaluated the N1 and P2 ERP components elicited by audiovisually congruent syllables in the AV compared to the A+V condition as a measure of audiovisual integration during the first 200 milliseconds of processing. In agreement with previous research, we found that the N1 amplitude elicited by AV syllables was significantly smaller compared to the N1 amplitude obtained from the A+V syllables. This effect was present in all groups of participants. Children with a history of SLI showed greater N1 attenuation than their TD peers over frontal and fronto-central sites, but did not differ from other groups over either central or parietal sites. This finding might indicate a difference in the orientation of the neural structures generating the N1 component in H-SLI children, or, assuming that a broader distribution of an electrophysiological effect implies the involvement of a larger amount of neural tissue, this group difference may also suggest an increased demand on neural resources during audiovisual processing in the H-SLI group. However, because this group difference was limited to a relatively small set of electrodes, a replication of this effect is needed before firmer conclusions can be drawn.

We also found that N1 attenuation in the AV condition was significantly larger over the right and midline sites (see Figure 5). Previous studies did not report such

hemispheric asymmetry, possibly because their statistical analyses of ERP data did not include a factor of hemisphere or laterality (see, for example, Knowland *et al.*, 2014; Van Wassenhove *et al.*, 2005). It has been suggested that the attenuation of the N1 peak amplitude is sensitive to how well lip movements can cue the temporal onset of the auditory signal (Baart *et al.*, 2014; Klucharev, Möttönen & Sams, 2003; Van Wassenhove *et al.*, 2005) and as such is an index of the temporal relationship between visual and auditory modalities during speech perception. In our stimuli, the onset of sound followed the first noticeable articulation-related facial movement by 167–367 milliseconds. A plausible reason behind the stronger attenuation of N1 over the right scalp could be the 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). This proposition of course rests on the assumption that at least partially similar neural networks are involved in the processing of temporal relationships in auditory and audiovisual contexts – an assumption that requires further study. Overall, however, ERP results suggest that the neural mechanisms underlying the N1 attenuation during audiovisual speech perception were largely similar in both groups of children and in adults.

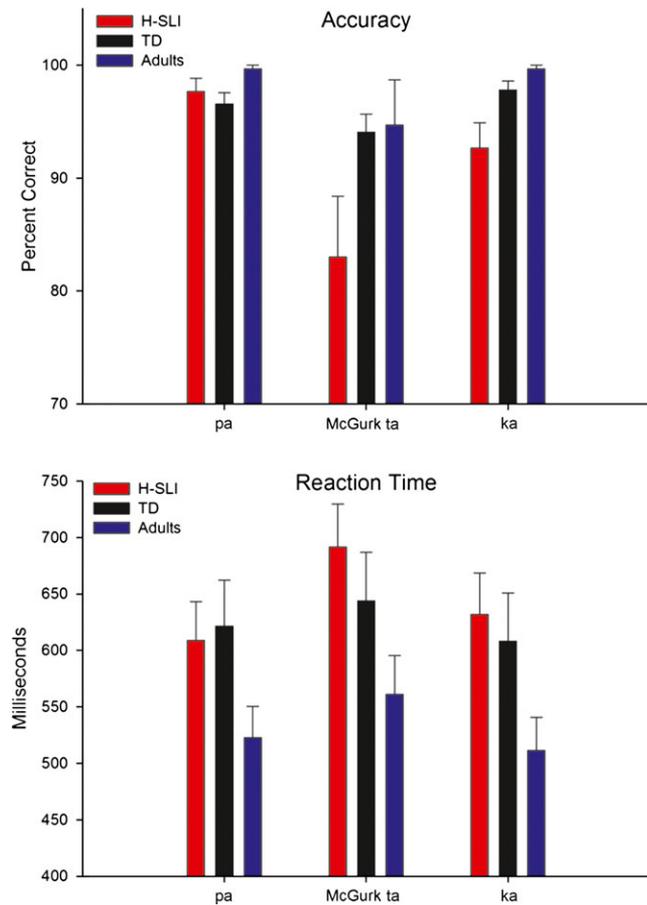


Figure 6 Performance on the McGurk task. Error bars represent standard errors of the mean.

Contrary to our expectations, we did not observe a statistically significant reduction in the P2 amplitude to AV speech. Although in absolute terms the amplitude of P2 tended to be smaller in the AV condition compared to the A+V condition (as can be seen in Panel B of Figure 5), this effect was overall weak. Of note here is the fact that both groups of children had a markedly smaller P2 component compared to adults. Thus, P2 appears to have a protracted developmental course, at least within the context of audiovisual speech. Given the immature state of P2 in our two groups of children, significant inter-group variability in its properties has likely contributed to this negative finding.

Some studies reported the shortening of the N1 and/or P2 peak latency in the AV as compared to the A conditions (e.g. Knowland *et al.*, 2014; Van Wassenhove *et al.*, 2005). In the study by Van Wassenhove and colleagues, such speeding up of the neural responses in the AV condition was dependent on the predictability value of articulatory movements, such that the greatest shortening of latency was observed for the /p/ phoneme

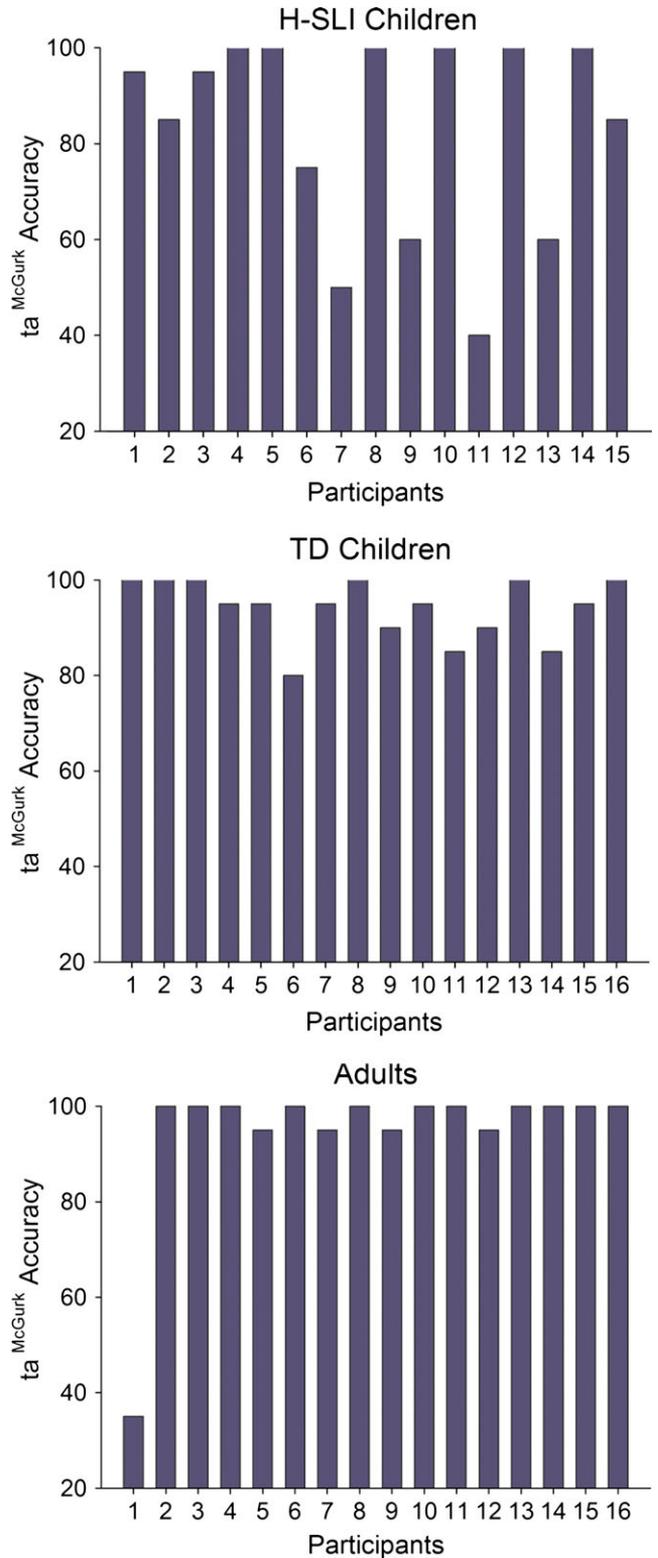


Figure 7 'ta' McGurk accuracy: individual differences. Y-axis shows percent of 'ta' responses to the 'ta' McGurk syllable.

due to its clear bilabial articulation, with smaller shortening for /t/ and /k/ phonemes, both of which have less salient visual cues. We did not replicate this finding in our study. One reason for this may be the fact that we had to keep the overall number of trials relatively low to make sure that all children were able to complete the ERP session without becoming excessively tired. As a result, we did not have enough trials to analyze ERP components elicited by separate phonemes. Without such an analysis, the effect might have failed to reach significance due to latency variability within and across groups.

Although the N1 attenuation to AV as compared to A+V speech in adults has been replicated in multiple studies, relatively little is known about the developmental course of this effect. To the best of our knowledge, to date only the study by Knowland and colleagues (Knowland *et al.*, 2014) has addressed this issue. These authors compared N1 and P2 ERP responses to auditory and audiovisual words in children who ranged in age from 6 years to 11 years and 10 months. They found that significant attenuation of N1 to AV speech was not present in children until approximately 10 years of age, while the amplitude of P2 was reduced already at the age of 7 years and 4 months. Our findings are in partial agreement with this earlier study. Our group of TD children was relatively small compared to that of Knowland and colleagues. As a result, analyses of the N1 attenuation over smaller age ranges (such as 1–2 years) were not feasible. It is, therefore, possible that in our results older children contributed more to the N1 attenuation in response to AV speech at the group level than younger children. A significant positive correlation between N1 attenuation and children's age provides some support to this hypothesis. However, many of the youngest children in our study still had N1 attenuation to AV speech, suggesting that early stages of audiovisual integration may develop early at least in some children. Alternatively, a difference in the timing of the main effect (N1 in our study vs. P2 in Knowland *et al.*'s study) could be due to differences in the stimuli and the main task of the two studies. Knowland and colleagues used real words as stimuli and asked participants to press a button every time they heard an animal word. This task required semantic processing and might have been more demanding than the task of monitoring facial changes for silly events employed in our study, potentially delaying the audiovisual integration effect until the P2 time window, at least in younger children.

In the second experiment, we have examined susceptibility to the McGurk illusion in the three groups of participants. We found that compared to their TD peers and adults, H-SLI children were much more likely to

misidentify the 'ta'^{McGurk} syllable as 'pa' (based on its auditory component) than as 'ka' (based on its visual component). This relationship was absent in TD children and adults. Although the bias toward hearing 'pa' was significant at the group level, examination of individual data, as shown in Figure 7, suggests that only approximately one-third of H-SLI children had a genuine difficulty perceiving the McGurk illusion, with the rest performing similarly to their TD peers. Furthermore, even the H-SLI children with the fewest McGurk percepts did report hearing the 'ta'^{McGurk} syllable on at least some trials, indicating that the ability to combine incongruent auditory and visual modality into a coherent speech percept is not completely absent in these children. Previous studies of the McGurk illusion in individuals with SLI differed in the ages of participants, the analyses used for determining the presence of the McGurk perception, and the strength of the McGurk illusion in control populations, which makes comparing susceptibility rates across studies somewhat difficult. However, results of at least some previous investigations also suggest that difficulties with McGurk perception are not an all-or-nothing phenomenon in individuals with language impairment. For example, in the study by Norrix and colleagues (Norrix *et al.*, 2006), only half of adults with language-learning disabilities showed significant impairment on the McGurk task. In a different study by the same group (Boliek *et al.*, 2010), which included SLI children in the age range comparable to ours (6–12 years of age), the reported size of standard deviation suggested that at least some of the SLI children performed within the normal range. Our results are in agreement with these earlier studies and extend them by showing that low susceptibility to the McGurk illusion is also present in some children with a history of SLI, even when such children's scores on standardized tests of language ability fall within the normal range.

Overall, however, a tendency for the H-SLI children to perceive the McGurk stimulus as 'pa' rather than as 'ka' presents a paradox that requires further study. Given the documented weakness of phonemic processing in this population (e.g. Hesketh & Conti-Ramsden, 2013), one could expect that within the context of the McGurk stimulus, in which auditory and visual modalities provide conflicting cues to the phoneme identity, the H-SLI children would rate the auditory modality as less reliable than the visual one and thus would be more likely to report hearing 'ka' (in agreement with the visual component of the syllable). The lack of this tendency is especially surprising given that the typically very strong visual cues for the /p/ phoneme are absent in the McGurk stimulus. The bias toward hearing 'pa' in

the H-SLI children would, therefore, seem to indicate that, for reasons yet unknown, a sizeable sub-group of these children are more likely than their TD peers to disregard visual speech cues and rely on auditory information alone, at least when the two modalities are incongruent.

Successful performance on any given task depends on a whole set of cognitive skills (Barry, Weiss & Sabisch, 2013; Moore, Ferguson, Edmondson-Jones, Ratib & Riley, 2010). It is therefore possible that H-SLI children's difficulty with the McGurk paradigm might have been influenced by deficits in one or more of these higher cognitive abilities required by the task, rather than by deficits in audiovisual processing *per se*. Within this context, we examined whether group differences in overall processing speed or attention might have contributed to lower susceptibility to the McGurk illusion in the H-SLI children. Given the lack of ACC and RT differences between the two groups of children when classifying congruent syllables of the McGurk task, the tendency to rely on auditory modality alone in the H-SLI group is difficult to ascribe to a general limitation in the processing speed or efficiency. Furthermore, although two children in the H-SLI group had a diagnosis of ADD/ADHD, the two groups of children did not differ in their ability to selectively focus, sustain, or divide attention based on the results of the TEA-Ch tests (see Table 2), making it less likely that group differences in McGurk susceptibility were due to differences in attentional skills.

Based on the CELF-4 test of language skills, most H-SLI children who participated in our study fell within the normal range of linguistic abilities, despite performing overall significantly worse than their TD peers. It is possible, therefore, that age-appropriate N1 attenuation to AV speech in this population is a sign of recovery. All of the H-SLI children we tested were diagnosed as severely impaired during pre-school years, with most falling below the 2nd percentile on standardized language tests at the time of the diagnosis. It is not, therefore, the case that our group of children might have started out as less impaired on language skills than what may be typical for SLI children in general. In addition, as discussed above, we know that children with a history of SLI are at an increased risk for developing other learning disorders, suggesting that their performance on standardized language tests may hide continuing language problems. Nevertheless, future studies that compare children with a current diagnosis of SLI and children with a history of this disorder are needed in order to determine whether normal N1 attenuation to AV speech is a sign of (at least partial) recovery or is present in children with a current diagnosis of SLI as well.

Lastly, this study focused on just two relatively early stages of audiovisual speech perception. Visual speech cues available to listeners during word and sentence processing may differ significantly from those available during early acoustic and phonemic processing analyzed here. For example, the rhythmic opening of the jaw may index the syllabic structure of the incoming signal in longer utterances. An important empirical question to be addressed in future research is whether audiovisual integration in children with either a current or a former diagnosis of SLI is normal beyond the phonemic level of linguistic analysis. In addition, 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). 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 (Dick *et al.*, 2010). Therefore, the seemingly mature state of the early stage of audiovisual processing in TD and H-SLI children reported in this study reflects only one component within a complex network of sensory and cognitive processes that are engaged during audiovisual processing and cannot by itself guarantee efficient audiovisual integration.

In conclusion, as a group, the H-SLI children showed normal sensory processing of audiovisually congruent speech but reduced ability to perceive the McGurk illusion. This finding strongly argues against global audiovisual integration impairment in children with a history of SLI and suggests that, when present, audiovisual integration difficulties in the H-SLI group likely stem from a later (non-sensory) stage of processing.

Acknowledgements

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 Dr. Patricia Deevy, Daniel Noland, Camille Hagedorn, Courtney Rowland, and Casey Spelman for their assistance with different stages of this project. 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, F.N., & Mamassian, P. (2010). Multisensory processing in review: from physiology to behavior. *Seeing and Perceiving*, **23**, 3–38.
- American Electroencephalographic Society (1994). Guideline thirteen: guidelines for standard electrode placement nomenclature. *Journal of Clinical Neurophysiology*, **11**, 111–113.
- Archibald, L.M.D. (2008). The promise of nonword repetition as a clinical tool. *Canadian Journal of Speech-Language Pathology and Audiology*, **32** (1), 21–28.
- Archibald, L.M.D., & Gathercole, S.E. (2006). Nonword repetition: a comparison of tests. *Journal of Speech, Language, and Hearing Research*, **49**, 970–983.
- Archibald, L.M.D., & Joanisse, M.F. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research*, **52**, 899–914.
- Baart, M., Stekelenburg, J.J., & Vroomen, J. (2014). Electrophysiological evidence for speech-specific audiovisual integration. *Neuropsychologia*, **53**, 115–121.
- Bahrnick, L.E., Hernandez-Reif, M., & Flom, R. (2005). The development of infant learning about specific face–voice relations. *Developmental Psychology*, **41** (3), 541–552.
- Bahrnick, L.E., Netto, D., & Hernandez-Reif, M. (1998). Intermodal perception of adult and child faces and voices by infants. *Child Development*, **69** (5), 1263–1275.
- Barry, J.G., Weiss, B., & Sabisch, B. (2013). Psychophysical estimates of frequency discrimination: more than just limitations of auditory processing. *Brain Sciences*, **3** (3), 1023–1042.
- Barutchu, A., Danaher, J., Crewther, S.G., Innes-Brown, H., Shivdasani, M.N. *et al.* (2010). Audiovisual integration in noise by children and adults. *Journal of Experimental Child Psychology*, **105**, 38–50.
- Besle, J., Bertrand, O., & Giard, M.-H. (2009). Electrophysiological (EEG, sEEG, MEG) evidence for multiple audiovisual interactions in the human auditory cortex. *Hearing Research*, **258**, 2225–2234.
- Besle, J., Fischer, C., Bidet-Caulet, A., Lecaigard, F., Bertrand, O. *et al.* (2008). Visual activation and audiovisual interaction in the auditory cortex during speech perception: intracranial recordings in humans. *Journal of Neuroscience*, **28** (52), 14301–14310.
- Besle, J., Fort, A., Delpuech, C., & Giard, M.-H. (2004a). Bimodal speech: early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, **20**, 2225–2234.
- Besle, J., Fort, A., & Giard, M.-H. (2004b). Interest and validity of the additive model in electrophysiological studies of multisensory interactions. *Cognitive Processing*, **5**, 189–192.
- BioSemi (2013). Active electrodes. Available from http://www.biosemi.com/active_electrode.htm
- Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, **8** (3), 389–395.
- Boersma, P., & Weenink, D. (2011). Praat: doing phonetics by computer (version 5.3) [Computer program]. Retrieved from: <http://www.praat.org>
- Boliek, C.A., Keintz, C., Norrix, L.W., & Obrzut, J. (2010). Auditory-visual perception of speech in children with learning disabilities: the McGurk effect. *Canadian Journal of Speech-Language Pathology and Audiology*, **34** (2), 124–131.
- Brandwein, A.B., Foxe, J.J., Russo, N.N., Altschuler, T.S., Gomes, H. *et al.* (2011). The development of audiovisual multisensory integration across childhood and early adolescence: a high-density electrical mapping study. *Cerebral Cortex*, **21** (5), 1042–1055.
- Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, G., Glida, T. *et al.* (2008). Hearing faces: How the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience*, **21** (5), 905–921.
- Brookes, H., Slater, A., Quinn, P.C., Lewkowicz, D., Hayes, R. *et al.* (2001). Three-month-old infants learn arbitrary auditory-visual pairings between voices and faces. *Infant and Child Development*, **10**, 75–82.
- Brown, L., Sherbenou, R.J., & Johnsen, S.K. (2010). *Test of Nonverbal Intelligence* (4th edn.). Austin, TX: Pro-Ed.
- 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.
- Burr, D., & Gori, M. (2012). Multisensory integration develops late in humans. In M.M. Murray & M.T. Wallace (Eds.), *The neural bases of multisensory processes (ch. 18)*. New York: CRC Press.
- Calvert, G.A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral Cortex*, **11**, 1110–1123.
- Catts, H., Fey, M.E., Tomblin, J.B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research*, **45**, 1142–1157.
- Center for Disease Control and Prevention (2014). *CDC estimates 1 in 68 children has been identified with autism spectrum disorder*. Atlanta, GA: Department of Health and Human Services.
- Cohen, M.S. (2008). Handedness Questionnaire. Retrieved 27 May 2013, from: <http://www.brainmapping.org/shared/Edinburgh.php#>
- Conti-Ramsden, G., Botting, N., & Faragher, B. (2001a). Psycholinguistic markers for specific language impairment. *Journal of Child Psychology and Psychiatry*, **42** (6), 741–748.
- Conti-Ramsden, G., Botting, N., & Faragher, B. (2001c). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*, **42** (6), 741–748.
- Conti-Ramsden, G., Botting, N., Simkin, Z., & Knox, E. (2001b). Follow-up of children attending infant language units: outcomes at 11 years of age. *International Journal of Language and Communication Disorders*, **36** (2), 207–219.

- Crowley, K.E., & Colrain, I.M. (2004). A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clinical Neurophysiology*, **115**, 732–744.
- Dawson, J., Eyer, J., & Fonkalsrud, J. (2005). *Structured Photographic Expressive Language Test - Preschool* (2nd edn.). DeKalb, IL: Janelle Publications.
- Dick, A.S., Solodkin, A., & Small, S.L. (2010). Neural development of networks for audiovisual speech comprehension. *Brain and Language*, **114**, 101–114.
- Dodd, B. (1979). Lip reading in infants: attention to speech presented in- and out-of-synchrony. *Cognitive Psychology*, **11**, 478–484.
- Dollaghan, C., & Campbell, T.F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, **41**, 1136–1146.
- Ellis Weismer, S., Tomblin, J.B., Zhang, X., Buckwalter, P., Chynoweth, J.G. *et al.* (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research*, **43**, 865–878.
- Erickson, L.C., Zielinski, B.A., Zielinski, J.E.V., Liu, G., Turkeltaub, P.E. *et al.* (2014). Distinct cortical locations for integration of audiovisual speech and the McGurk effect. *Frontiers in Psychology*, **5**, 534.
- Fidler, L.J., Plante, E., & Vance, R. (2011). Identification of adults with developmental language impairments. *American Journal of Speech-Language Pathology*, **20**, 2–13.
- Fujiki, M., Brinton, B., Morgan, M., & Hart, C. (1999). Withdrawn and sociable behavior of children with language impairment. *Language, Speech, and Hearing Services in Schools*, **30**, 183–195.
- Fujiki, M., Brinton, B., & Todd, C. (1996). Social skills with specific language impairment. *Language, Speech, and Hearing Services in Schools*, **27**, 195–202.
- Giard, M.-H., & Besle, J. (2010). Methodological considerations: electrophysiology of multisensory interactions in humans. In M.J. Naumer (Ed.), *Multisensory object perception in the primate brain* (pp. 55–70). New York: Springer.
- Graf Estes, K., Evans, J.L., & Else-Quest, N.M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: a meta-analysis. *Journal of Speech, Language, and Hearing Research*, **50**, 177–195.
- Greenslade, K., Plante, E., & Vance, R. (2009). The diagnostic accuracy and construct validity of the Structured Photographic Expressive Language Test – Preschool: Second Edition (SPELT-P2). *Language, Speech, and Hearing Services in Schools*, **40**, 150–160.
- Hampton Wray, A., & Weber-Fox, C. (2013). Specific aspects of cognitive and language proficiency account for variability in neural indices of semantic and syntactic processing in children. *Developmental Cognitive Neuroscience*, **5**, 149–171.
- Hayes, E.A., Tiippana, K., Nicol, T.G., Sams, M., & Kraus, N. (2003). Integration of heard and seen speech: a factor in learning disabilities in children. *Neuroscience Letters*, **351**, 46–50.
- Hesketh, A., & Conti-Ramsden, G. (2013). Memory and language in middle childhood in individuals with a history of specific language impairment. *PLOS ONE*, **8** (2), doi:10.1371/journal.pone.0056314
- Hollich, G., Newman, R.S., & Jusczyk, P.W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Development*, **76** (3), 598–613.
- Kaganovich, N., Schumaker, J., Leonard, L.B., Gustafson, D., & Macias, D. (2014). Children with a history of SLI show reduced sensitivity to audiovisual temporal asynchrony: an ERP study. *Journal of Speech, Language, and Hearing Research*, **57**, 1480–1502.
- Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: rethinking the study of developmental disorders. *Developmental Psychology*, **45** (1), 56–63.
- Klucharev, V., Möttönen, R., & Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cognitive Brain Research*, **18** (1), 65–75.
- Knowland, V.C.P., Mercure, E., Karmiloff-Smith, A., Dick, F., & Thomas, M.S.C. (2014). Audio-visual speech perception: a developmental ERP investigation. *Developmental Science*, **17** (1), 110–124.
- Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences, USA*, **105** (32), 11442–11445.
- Kushnerenko, E., Tomalski, P., Ballieux, H., Potton, A., Birtles, D. *et al.* (2013). Brain responses and looking behavior during audiovisual speech integration in infants predict auditory speech comprehension in the second year of life. *Frontiers in Psychology*, **4**, 432. doi:10.3389/fpsyg.2013.00432
- Law, J., Rush, R., Schoon, I., & Parsons, S. (2009). Modeling developmental language difficulties from school entry into adulthood: literacy, mental health, and employment outcomes. *Journal of Speech, Language, and Hearing Research*, **52**, 1401–1416.
- Leonard, L. (2014). *Children with specific language impairment* (2nd edn.). Cambridge, MA: The MIT Press.
- Lewkowicz, D.J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance*, **22**, 1094–1106.
- Lewkowicz, D.J. (2010). Infant perception of audio-visual speech synchrony. *Developmental Psychology*, **46**, 66–77.
- Luck, S.J. (2005). *An Introduction to the event-related potential technique*. Cambridge, MA: The MIT Press.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, **264**, 746–748.
- Magosso, E., Cuppini, C., & Ursino, M. (2012). A neural network model of ventriloquism effect and aftereffect. *PLOS ONE*, **7** (8), e42503. doi:10.1371/journal.pone.0042503
- Manly, T., Robertson, I.H., Anderson, V., & Nimmo-Smith, I. (1999). *Test of everyday attention for children*. London: Pearson Assessment.
- Martin, N., & Brownell, R. (2005). *Test of auditory processing skills* (3rd edn.). Novato, CA: Academic Therapy Publications.
- Massaro, D.W. (1984). Children's perception of visual and auditory speech. *Child Development*, **55**, 1777–1788.

- Massaro, D.W., Thompson, L.A., Barron, B., & Lauren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. *Journal of Experimental Child Psychology*, **41**, 93–113.
- Meronen, A., Tiippana, K., Westerholm, J., & Ahonen, T. (2013). Audiovisual speech perception in children with developmental language disorder in degraded listening conditions. *Journal of Speech, Language, and Hearing Research*, **56**, 211–221.
- vanMetting Rijn, A.C., Kuiper, A.P., Dankers, T.E., & Grimbergen, C.A. (1996). 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.
- Metting van Rijn, A.C., Peper, A., & Grimbergen, C.A. (1990). High-quality recording of bioelectric events. Part 1: Interference reduction, theory and practice. *Medical and Biological Engineering and Computing*, **28**, 389–397.
- Moore, D.R., Ferguson, M.A., Edmondson-Jones, A.M., Ratib, S., & Riley, A. (2010). Nature of auditory processing disorder in children. *Pediatrics*, **126** (2), e382–e390.
- Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, **24** (4), 375–425.
- Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: visual articulatory information enables the perception of second language sounds. *Psychological Research*, **71**, 4–12.
- Norrix, L.W., Plante, E., & Vance, R. (2006). Auditory-visual speech integration by adults with and without language-learning disabilities. *Journal of Communication Disorders*, **39**, 22–36.
- Norrix, L.W., Plante, E., Vance, R., & Boliek, C.A. (2007). Auditory-visual integration for speech by children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, **50**, 1639–1651.
- Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, **9**, 97–113.
- Peterson, I.T., Bates, J.E., D'Onofrio, B.M., Coyne, C.A., Lansford, J.E. *et al.* (2013). Language ability predicts the development of behavior problems in children. *Journal of Abnormal Psychology*, **122** (2), 542–557.
- Pflieger, M.E. (2001). Theory of a spatial filter for removing ocular artifacts with preservation of EEG. Paper presented at the EMSE Workshop, Princeton University. Available at: http://www.sourcesignal.com/SpFilt_Ocular_Artifact.pdf
- Plante, E., & Vance, R. (1994). Selection of preschool language tests: a data-based approach. *Language, Speech, and Hearing Services in the Schools*, **25**, 15–24.
- Poeppl, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*, **41**, 245–255.
- Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: a lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 97–113). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rice, M.L., Hoffman, L., & Wexler, K. (2009). Judgments of omitted BE and DO in questions as extended finiteness clinical markers of specific language impairment (SLI) to 15 years: a study of growth and asymptote. *Journal of Speech, Language, and Hearing Research*, **52**, 1417–1433.
- Rosenblum, L.D., Schmuckler, M.A., & Johnson, J.A. (1997). The McGurk effect in infants. *Perception & Psychophysics*, **59** (3), 347–357.
- Ross, L.A., Saint-Amour, D., Leavitt, V.M., Javitt, D.C., & Foxe, J.J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, **17**, 1147–1153.
- Scarborough, H.S., & Dobrich, W. (1990). Development of children with early language delay. *Journal of Speech and Hearing Research*, **33**, 70–83.
- Schopler, E., Van Bourgondien, M.E., Wellman, G.J., & Love, S.R. (2010). *Childhood Autism Rating Scale* (2nd edn.). Los Angeles, CA: Western Psychological Services.
- Semel, E., Wiig, E.H., & Secord, W.A. (2003). *CELF4: Clinical Evaluation of Language Fundamentals* (4th edn.). San Antonio, TX: Pearson Clinical Assessment.
- Senkowski, D., Schneider, T.R., Foxe, J.J., & Engel, A.K. (2008). Crossmodal binding through neural coherence: implications for multisensory processing. *Trends in Neurosciences*, **31** (8), 401–409.
- Shield, B.M., & Dockrell, J.E. (2003). The effects of noise on children at school: a review. *Journal of Building Acoustics*, **10** (2), 97–106.
- Spaulding, T.J., Plante, E., & Farinella, K.A. (2006). Eligibility criteria for language impairment: is the low end of normal always appropriate? *Language, Speech, and Hearing Services in Schools*, **37**, 61–72.
- Stark, R.E., Bernstein, L.E., Condino, R., Bender, M., Tallal, P. *et al.* (1984). Four-year follow-up study of language impaired children. *Annals of Dyslexia*, **34**, 49–68.
- Stekelenburg, J.J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience*, **19** (12), 1964–1973.
- Stevenson, R.A., VanDerKlok, R.M., Pisoni, D.B., & James, T.W. (2011). Discrete neural substrates underlie complementary audiovisual speech integration processes. *NeuroImage*, **55**, 1339–1345.
- Sumbly, W.H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, **26** (2), 212–215.
- Teinonen, T., Aslin, R.N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, **108**, 850–855.
- Tomblin, J.B., Records, N.L., Buckwalter, P., Zhang, X., Smith, E. *et al.* (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, **40**, 1245–1260.

- Van Wassenhove, V., Grant, K.W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences, USA*, **102** (4), 1181–1186.
- Werner, E.O., & Kresheck, J.D. (1983). *Structured Photographic Expressive Language Test – II*. DeKalb, IL: Janelle Publications.
- Yairi, E., & Ambrose, N. (2004). Stuttering: recent developments and future directions. *The ASHA Leader* (5 October).

- Young, G.S., Merin, N., Rogers, S.J., & Ozonoff, S. (2009). Gaze behavior and affect at 6 months: predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Developmental Science*, **12** (5), 798–814.

Received: 15 October 2013

Accepted: 7 September 2014