

BRIEF RESEARCH REPORT

Perception of audio-visual speech synchrony in Spanish-speaking children with and without specific language impairment*

FERRAN PONS

*Departament de Psicologia Bàsica, Facultat de Psicologia,
Universitat de Barcelona and Institute for Brain, Cognition and Behaviour
(IR3C), Barcelona, Spain*

LLORENÇ ANDREU

*Departament de Psicologia Bàsica, Facultat de Psicologia, Universitat de
Barcelona and Cognitive Neuroscience and Information Technologies Research
Program, IN3, Universitat Oberta de Catalunya*

MONICA SANZ-TORRENT, LUCÍA BUIL-LEGAZ

*Departament de Psicologia Bàsica, Facultat de Psicologia,
Universitat de Barcelona*

AND

DAVID J. LEWKOWICZ

*Department of Psychology and Center for Complex Systems & Brain Sciences,
Florida Atlantic University*

*(Received 4 March 2011 – Revised 4 January 2012 – Accepted 25 April 2012 –
First published online 9 July 2012)*

ABSTRACT

Speech perception involves the integration of auditory and visual articulatory information, and thus requires the perception of temporal synchrony between this information. There is evidence that children with specific language impairment (SLI) have difficulty with auditory

[*] This work was supported by the Spanish Ministerio de Ciencia e Innovación (SEJ2007-62743 to M. S., and PSI2010-20294 to F. P.) and by the National Science Foundation (grants BCS-0751888 to D. J. L.). Special thanks to all the children who participated and to CEIP Els Pins (Barcelona), CREDA Narcís Massó (Girona), and the School of Educational Psychology Services – SPE (Castelló). Address for correspondence: Ferran Pons, Departament de Psicologia Bàsica, Facultat de Psicologia, Universitat de Barcelona, Pg. de la Vall d'Hebrón, 171, 08035 – Barcelona, Spain. tel: +34 933125144; fax: +34934021363; e-mail: ferran.pons@ub.edu

speech perception but it is not known if this is also true for the integration of auditory and visual speech. Twenty Spanish-speaking children with SLI, twenty typically developing age-matched Spanish-speaking children, and twenty Spanish-speaking children matched for MLU-w participated in an eye-tracking study to investigate the perception of audiovisual speech synchrony. Results revealed that children with typical language development perceived an audiovisual asynchrony of 666 ms regardless of whether the auditory or visual speech attribute led the other one. Children with SLI only detected the 666 ms asynchrony when the auditory component followed the visual component. None of the groups perceived an audiovisual asynchrony of 366 ms. These results suggest that the difficulty of speech processing by children with SLI would also involve difficulties in integrating auditory and visual aspects of speech perception.

INTRODUCTION

Whenever we interact with other people, we can usually see as well as hear them talking. As a result, everyday speech is audiovisual rather than auditory in nature. Normally, because of our ability to integrate the auditory and visual streams of speech information, we perceive them as part of a unified multisensory event (Alsius, Navarra, Campbell & Soto-Faraco, 2005; Lewkowicz, 2010; Munhall & Vatikiotis-Bateson, 2004). Integration is facilitated by the fact that the dynamic auditory and visual signals that specify audiovisual speech are temporally coupled, and thus highly redundant (Chandrasekaran, Trubanova, Stillitano, Caplier & Ghazanfar, 2009; Yehia, Rubin & Vatikiotis-Bateson, 1998). Once the streams of auditory and visual information are integrated, the speech becomes more salient (Sumby & Pollack, 1954; Summerfield, 1979) and more intelligible (Munhall, Gribble, Sacco & Ward, 1996) and, as evidence of this, both infant (Lewkowicz & Hansen-Tift, 2012) and adult listeners (Sumby & Pollack, 1954; Summerfield, 1979) take advantage of the greater intelligibility of audiovisual as opposed to auditory speech.

Integration of auditory and visual speech and, by extension, its intelligibility, is affected by the specific temporal relationship between the auditory and visual streams of speech information (van Wassenhove, Grant & Poeppel, 2007). When the auditory and visual streams of information are delayed with respect to one another, intelligibility is adversely affected in an asymmetrical fashion: when auditory speech leads visual speech intelligibility declines much more than when auditory speech follows visual speech. Importantly, however, as long as the auditory and visual components fall into what is known as the intersensory temporal contiguity window (ITCW: Lewkowicz, 1996), perceivers experience those components as part of

a unitary event. The size of the ITCW differs as a function of whether the auditory component follows the visual component (V–A asynchrony) or whether it precedes it (A–V asynchrony). In adults, the ITCW is approximately 180–240 ms for a V–A asynchrony but only 60–120 ms for an A–V asynchrony (Dixon & Spitz, 1980; Munhall *et al.*, 1996; van Wassenhove *et al.*, 2007). Developmental studies have found similar differences in infants, although the ITCW in infancy is considerably larger (Lewkowicz, 1996; 2000; 2010). For example, when infants are habituated to a speech syllable and then tested for detection of an A–V asynchrony between its audible and visible attributes, they detect the asynchrony only when it reaches 633–666 ms. This finding indicates that the ITCW narrows with development. The developmental narrowing of the ITCW is also evident in findings from studies of responsiveness to non-speech events. Thus, infants can only detect a V–A asynchrony between a bouncing object and its impact sound when the asynchrony reaches 450 ms, and an A–V asynchrony between the visible and audible bounce when it reaches 350 ms (Lewkowicz, 1996). Consistent with the developmental narrowing of the ITCW, adults can detect lower asynchronies for non-speech events (Dixon & Spitz, 1980).

Why might the ITCW be larger for V–A than for A–V asynchrony? One reason is that whenever people speak, the motion of their lips can be seen before their vocalizations can be heard (Chandrasekaran *et al.*, 2009). As a result, the perceptual system expects a delay. In addition, because lip motion is of a continuous nature, it is more difficult to determine the precise point when the vocalization begins in relation to lip motion. In contrast, when a vocalization begins first, the perceptual system does not expect a delay and the punctate onset of the vocalization makes it easier to determine that this is the point when lip motion corresponds.

If perception of the unitary nature of speech is critical for adaptive functioning, and if this depends on the ability to perceive the temporal relationship between the auditory and visual attributes of speech, then impaired detection of audiovisual temporal synchrony would be maladaptive. One specific disorder where this may be the case is specific language impairment (SLI). In general, children with SLI are characterized by developmental delays in a number of different language domains, including semantic, morphosyntactic, pragmatic, and discourse skills in oral and/or written language (Leonard, 1998). In addition, studies have found that children with SLI perform poorly on tasks requiring the processing of relatively brief (250 ms) synthetic CV syllables for which the critical formant transition was short in duration (40 ms), as well as stimuli presented in rapid succession (Tallal & Piercy, 1973; 1975). Studies also have found that children with SLI have more difficulty in tasks requiring identification of brief stimuli than do age-matched peers (Elliott & Hammer, 1988; Tallal, Stark & Mellits, 1985; Wright, Lombardino, King, Puranik, Leonard

& Merzenich, 1997). These findings have led some researchers to argue that the deficits in processing brief sounds underlie SLI (AUDITORY TEMPORAL PROCESSING HYPOTHESIS: Tallal, 1984). However, a number of concerns have been raised about this conclusion. For example, the tasks used by Tallal and colleagues require substantial attention and memory skills, suggesting that task-specific effects might account for the Tallal *et al.* (1985) results (see Elliott & Hammer, 1988). Furthermore, some children with SLI have difficulty with particular processing tasks, but not with the processing of rapid auditory transitions (Stark & Heinz, 1996).

Regardless of the specific auditory processing difficulties that children with SLI may have, it is likely that visual speech information might facilitate auditory processing in these children, especially when auditory perception alone is less than optimal. An examination of the empirical literature in audiovisual (AV) speech perception provides mixed clues about the ability of children with speech and language disorders to lip-read and whether they can take advantage of visual information to compensate for auditory processing difficulties. On the one hand, some studies of AV speech perception have indicated that preschool children who make developmental speech errors perform differently from their controls on lip-reading tasks (Desjardins, Rogers & Werker, 1997). On the other hand, a study of children's response to the McGurk effect¹ has found that children with speech disorders do not differ from matched controls in their perception of the illusion, or in their favored strategy in response to incongruent AV speech sounds (Dodd, McIntosh, Erdener & Burnham, 2008). However, it has also been reported that children with language disorders show a diminished McGurk effect relative to their peers (Boliek, Keintz, Norrix & Obrzut, 2010). In particular, Norrix and colleagues (Norrix, Plante, Vance & Boliek, 2007) found that children with SLI are less influenced by the visual information in a McGurk task than their peers, and concluded that children with SLI may differ both from adults and from their normal peers in the degree to which the visual dimensions of articulated speech affect their response to audiovisual speech. If that is the case then these findings suggest that speech perception difficulties in children with SLI may not be specific to the auditory modality but may reflect an inability to respond to the combination of auditory and visual information.

Given the potential benefit of audiovisual, as opposed to auditory only, articulatory information (Sumby & Pollack, 1954; Summerfield, 1979), it would be beneficial for children with SLI to respond to the combination of

[1] The McGurk effect (McGurk & MacDonald, 1976) is a powerful illustration that speech perception is, by default, a multisensory process where the auditory and visual information is integrated into a novel percept. For example, when a listener is presented with the sound 'ba-ba' while the lips of a speaker are silently mouthing 'ga-ga', the listener perceives 'da-da'.

auditory and visual speech information in a manner similar to that found in their typically developing peers. One way to determine whether they do is to test their ability to perceive the temporal relationship between the audible and visible attributes of speech and to investigate whether their ITCW is similar to the ITCW in typically developing peers. Norrix *et al.* (2007) suggested that audiovisual integration skills in children with SLI should be investigated but, to date, no studies of this ability have appeared. As a result, in the current study we investigated the perception of audio-visual temporal synchrony in fluent speech in children with SLI and compared their performance to that of typically developing children.

Bebko, Weiss, Denmark and Gómez (2006) examined responsiveness to audiovisual temporal synchrony in speech and non-speech events in young children with autism spectrum disorder, children with other forms of developmental disability but no autism, and typically developing children. Findings indicated that typically developing children as well as children with a developmental disability but no autism preferred looking at synchronous rather than asynchronous events regardless of whether they were speech or non-speech events. In contrast, children with autism did not exhibit a preference for synchronous speech events. Critically, these children failed to detect the difference between synchrony and asynchrony even though the asynchrony was as large as three seconds, and, even though the asynchrony was far larger (i.e. 3 s) than the asynchrony that infants can detect (Lewkowicz, 1996; 2000; 2010).

To determine whether children with SLI are impaired in their ability to perceive combined auditory and visual speech, we investigated their preference for one of two audiovisual speech events. One of these events showed a talker's face whose visible speech was synchronized with a concurrently presented soundtrack while the other event showed the same talker's face whose visible speech was desynchronized with respect to the soundtrack by 666 or 366 ms. To determine whether the children with SLI are impaired, we compared their performance to typically developing children. Based on Norrix *et al.*'s (2007) findings, we expected that children with SLI would exhibit impaired detection of audiovisual synchrony relations compared to typically developing children.

METHOD

Participants

All participants were native Spanish speakers selected from state schools in Catalonia and Valencia (Spain) and did not need eyeglasses to see the computer screen. Three groups took part in this study, a group of twenty children with SLI (aged 4;04–7;02), a group of twenty typically developing age-matched children (aged 4;04–6;10), and a group of twenty children

matched for mean length of utterance (MLU-w; aged 3;04–6;02). The parents of each child gave their written informed consent prior to their child's participation in the study.

The children with SLI were diagnosed with specific language impairment by speech and language therapists from school educational psychology services and were receiving language intervention. They were selected according to standard criteria for diagnosing SLI (Leonard, 1998; Stark & Tallal, 1981). Specifically, children with SLI were tested to assess their non-verbal intelligence and level of language development. Tests included the Wechsler Intelligence Scale for Children (WISC-R; Spanish version: Wechsler, Cordero & de la Cruz, 1993) or the Kaufman Brief Intelligence Test (KBIT; Spanish version: Kaufman & Kaufman, 2004). Every child with SLI obtained a non-verbal IQ standard score above 85. Language ability was assessed by language profiles following the Spanish protocol for evaluation of language delay, the *Análisis del Retraso del Lenguaje (AREL)* (Pérez & Serra, 1998), the Peabody Picture Vocabulary Test III (PPVT-III, Spanish version: Dunn, Dunn & Arribas, 2006) and the ELI (Early Language Inventory) child language scale (Saborit & Julián, 2003) for children younger than six years. The ELI scale includes several subtests for phonetics, lexical reception, lexical production and pragmatics. Children with SLI had scores of at least a -1.25 standard deviation below the mean, both on the Peabody III and the ELI. Language profiles based on transcripts of spontaneous conversations provided further information about the characteristics of the language production of the children. These analyses showed that these children had a delay of at least one year in language production, based on MLU-w values. Children were excluded if they had difficulty hearing pure tones in normal frequency ranges, or had neurological dysfunction, oral or motor dysfunction, or impaired social functioning. A summary of the descriptive data for the three groups of children can be found in Table 1.

The second group consisted of twenty children matched on age ($+/-2$ months) and gender with the children with SLI. Children were not selected if they had a history of speech therapy or psychological therapy. Teachers confirmed that the control participants' language development was typical for their age. Finally, the third group consisted of twenty children matched with the children with SLI on MLU in terms of words ($+/-0.6$ words) and gender. In addition, non-verbal intelligence and language ability was assessed in all children selected in both the age-control and MLU-w groups using the same tests and protocols applied to children in the SLI group (seven children were not tested with the ELI scale given that they were not under six years). The socioeconomic background of the children based on occupational status and educational degree of parents was established as a middle socioeconomic status.

TABLE I. *Group age, cognitive measures and language performance*

	SLI group Mean (SD)	Age control Mean (SD)	MLUw control Mean (SD)
Age (years)	6.69 (0.90)	6.72 (0.92)	5.51 (1.05)
NVIQ	96.1 (7.9)	106.3 (6.0)	93.13 (9.32)
PPVT-III	77.45 (8.96)	112.07 (14.37)	92 (12.87)
ELI-Receptive vocabulary*	36.27 (18.84)	73.07 (17.97)	67.85 (26.13)
ELI-Expressive vocabulary*	8.62 (1.8)	60.38 (15.06)	52.27 (28.84)
MLUw	3.95 (1.39)	6.86 (1.76)	3.46 (1.55)

NOTES: Chronological age in years; NVIQ (Non-verbal Intelligence Quotient) in standard score; PPVT-III (Peabody Picture Vocabulary Test III. Spanish version) in standard score; ELI (Evaluación del Lenguaje Infantil); ELI-Phonetics in mean number of errors; ELI-Receptive vocabulary. ELI-Expressive vocabulary and ELI-Pragmatics in percentiles; MLU-w (Mean Length of Utterance by words). * Values only calculated for children younger than six years old.

Apparatus

A Tobii T120 Eye-tracker was used to collect and store eye-tracking data. These data consisted of the participants' eye position sampled at 120 Hz (approximately 8 ms intervals). The Tobii T120 Eye-tracker is integrated together with a 17" TFT monitor, and thus the visual stimuli were presented on this monitor and the soundtrack was presented via a built-in speaker.

Stimuli

The stimuli consisted of multimedia movies which were constructed with Premiere 6.0 (Adobe Corporation) and consisted of two side-by-side video clips of the same female speaker looking directly at the camera and uttering a prepared script (see Appendix). The movies were presented at 30 frames/sec and had a resolution of 1024 × 480 pixels. The sound track portion of the movie was made with an audio sampling rate of 1024 kbps. Across all the movies, one of the faces (counterbalanced for side across trials) was always synchronized with the soundtrack while the other one was not. In two of the movies, the soundtrack for the desynchronized face preceded the visual speech by 366 ms (A-V 366) or by 666 ms (A-V 666). In the other two movies, the soundtrack for the desynchronized face followed the visual speech by 366 ms (V-A 366) or by 666 ms (V-A 666).

Procedure

Children were tested individually at their school. They were seated approximately 22" in front of the Tobii T120 Eye-tracker. A nine-point

calibration was carried out at the beginning of the experiment. The Tobii Studio Software automatically validates calibrations and the experimenter could, if required, repeat the calibration process if validation was poor.

The experiment consisted of four trials during which different video clips were presented. Each trial consisted of the presentation of a synchronous clip paired with one of the following clips where the auditory and visual speech streams were desynchronized (A–V 666, A–V 366, V–A 666 and V–A 366). All clips had a duration of 30 s. An attention-getter (a cross-hair) was presented in the middle of the screen between trials to center the children's attention prior to the next trial. Side of presentation and trial order were counterbalanced across children. Children were seated in front of the monitor and told that there were two faces talking and that one of them corresponded to the voice that they were hearing. They were given no explicit task to perform.

To acquaint the child with the procedure, a familiarization/baseline trial was presented prior to the start of the experiment. The same two faces were presented during this trial with one face in synchrony and the other desynchronized with respect to the audio by one second (the audio preceded the video). All the children easily identified the synchronous face, indicating that they were able to solve the task. Once this baseline trial ended, the experiment began. As soon as the child fixated the attention-getter, the test movies started to play.

RESULTS

To measure preferences, we divided the screen into a left and a right area of interest (AOI) and calculated the duration of fixation that each participant directed at each AOI during each trial. To determine whether children perceived AV speech asynchrony, and at what degree of asynchrony they did so, for each trial we computed the total time children spent looking at the synchronized face versus the total time they spent looking at the desynchronized face. Based on the unity assumption (Vatakis & Spence, 2007; Welch & Warren, 1980), according to which observers prefer concordant versus discordant multisensory events, we expected that children would be able to identify the talking face that was synchronized with the sound track and that this preference would be evident in greater looking at the synchronized face.

As indicated in the 'Introduction', the size of the ITCW is smaller for A–V asynchrony than for V–A asynchrony. Consequently, we analyzed responsiveness to these two types of asynchrony separately. In the first analysis, we examined responsiveness to A–V asynchrony and, to do so, we submitted the duration of looking scores to a repeated-measures analysis of variance (ANOVA), with trial (2) and synchrony (2) as the within-subjects

factors and group (3) as the between-subjects factor. The ANOVA yielded a main effect of group ($F(2, 57) = 5.993$, $p = 0.004$, $\eta_p^2 = 0.179$), which was due to less overall looking in children with SLI than in children in the two control groups. Despite the absence of any interaction involving trial as a factor, we felt that the theoretical predictions offered in the 'Introduction' provide a strong, empirically based, a priori justification for examining the data separately for each trial (we expected that children with SLI would exhibit difficulties in detecting the specific asynchronies presented). Tests of these a priori hypotheses were conducted using Bonferroni adjusted alpha levels of 0.016 per test (0.05/3).

Separate two-tailed t -tests revealed that the A-V 666 asynchrony was perceived by all three groups; the children looked longer at the synchronized than at the desynchronized face (Age-Control: $t(15) = 2.669$, $p = 0.002$; MLU-w Control: $t(16) = 3.118$, $p = 0.007$; SLI: $t(19) = 2.402$, $p = 0.003$). Looking times of each group are shown in Table 2 and illustrated in Figure 1. For the A-V 366 trial, two-tailed t -tests indicated that, given the Bonferroni adjustment, the A-V 366 asynchrony was not detected by any of the groups (Age-Control: $t(19) = 0.963$, $p = 0.347$; MLU-w Control: $t(19) = 3.338$, $p = 0.02$, Cohen's $d = 0.60$, medium effect size; SLI: $t(18) = 1.255$, $p = 0.225$).

The same repeated-measures analysis of variance on the data from the V-A asynchrony trials also revealed only a main effect of group ($F(2, 57) = 4.219$, $p = 0.02$, $\eta_p^2 = 0.129$). As before, despite the absence of a significant interaction, we used t -tests to explore our a priori hypothesis (alpha level of 0.016). The analyses of the V-A 666 trial indicated that this degree of asynchrony was detected by the two control groups but not by the SLI group (Age-Control: $t(19) = 2.594$, $p = 0.008$; MLU-w Control: $t(19) = 3.216$, $p = 0.01$; SLI: $t(19) = 1.623$, $p = 0.121$). The t -tests also indicated that none of the groups detected the V-A 366 asynchrony (Age-Control: $t(16) = 1.594$, $p = 0.130$; MLU-Control: $t(16) = 1.482$, $p = 0.158$; SLI: $t(19) = 1.177$, $p = 0.254$).

The three groups detected an asynchrony of 666 ms when the voice led lip motion. Furthermore, only the two control groups detected an asynchrony of 666 ms when lip motion led the voice. Finally, none of the groups were able to detect an asynchrony of 366 ms.

DISCUSSION

The purpose of the current study was to investigate the detection of audiovisual fluent speech asynchrony in children with and without SLI. To do so, children watched side-by-side faces of the same person mouthing a short and identical passage and heard the person talking at the same time. The person's vocalizations were synchronized with one of the two faces and desynchronized with the other face either by 366 or 666 ms. As predicted, we

TABLE 2. Mean and standard deviations of the looking time children spent looking at the synchronized face versus the total time they spent looking at the desynchronized face (Bold numbers indicate significant effects; * $p \leq 0.01$)

	A-V design		V-A design	
	A-V 666	A-V 366	V-A 666	V-A 366
Children with SLI	16.3 (6.2)/ 10.2 (5.8)*	11.1 (6.7)/12.3 (7.3)	13.4 (7.0)/11.7 (6.8)	14.0 (6.1)/13.2 (6.1)
Age-Control	19.1 (5.8)/ 11.7 (6.0)*	15.3 (6.5)/14.6 (6.8)	17.3 (5.4)/ 12.5 (5.1)*	14.3 (7.3)/15.5 (5.9)
MLU-w Control	18.3 (6.1)/ 11.3 (5.7)*	15.1 (4.9)/11.0 (5.6)	18.6 (4.9)/ 13.3 (5.9)*	14.3 (6.3)/14.1 (6.8)

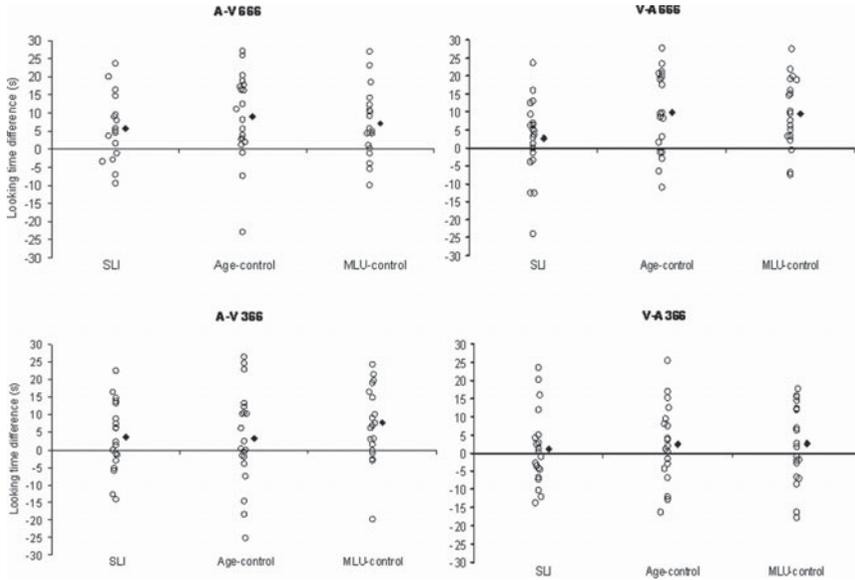


Fig. 1. Distribution of looking time difference scores (in seconds) to the synchronized face for each type of trial and children group. Open circles represent each child's score. Filled diamonds represent mean difference score for each group.

found that children with SLI exhibited poorer detection of audiovisual asynchrony than did children without SLI. That is, whereas the children in the two control groups preferred the synchronized face and voice in the 666 A-V and V-A asynchrony conditions, children with SLI only preferred the synchronized face and voice in the 666 ms A-V asynchrony condition. None of the children preferred the synchronized face and voice in the A-V and V-A 366 ms condition, indicating that none of them detected an asynchrony of 366 ms regardless of whether the auditory or visual speech attribute led the other one.

The fact that children with SLI could not detect the difference between two identical talking faces based on whether the concurrent voice they were hearing corresponded to one of them or not indicates impaired perception of audiovisual temporal relations. This impairment may be due to difficulties in auditory processing, an inability to speech-read, and/or attentional control problems. With specific regard to speech perception difficulties, the present results suggest that, at a minimum, the typical kinds of speech perception difficulties that children with SLI have may not be solely due to problems in auditory processing. In other words, in addition to difficulties in auditory processing (Tallal, 1984) there may be other factors that may be responsible for speech perception difficulties in children with SLI.

The current findings suggest that children with SLI are not only impaired in their speech perception abilities but also in their processing of the temporal coherence of auditory and visual speech information. One possible reason for this may be because these children are poorer at speech-reading. Recent studies of selective attention to audiovisual speech in infancy have shown that at about six months of age, when infants babble they shift their attention to the lips of their interlocutors and continue to focus on the speaker's lips until nearly twelve months of age (Lewkowicz & Hansen-Tift, 2012). The correlation between the sustained period of focusing on the speaker's lips between six and twelve months of age and the development of speech production capacity during that time surely contributes in important ways to the development multisensory perception. That is, they increase the opportunity to experience the temporal coherence of audiovisual speech. This fact raises the possibility that children with SLI may be impaired in the detection of audiovisual synchrony because they may not have attended to the mouth of their interlocutors during infancy as much as children without SLI. This, in turn, may make it difficult for children with SLI to integrate the articulatory code associated with visual and heard speech (Desjardins *et al.*, 1997; Siva, Stevens, Kuhl & Meltzoff, 1995). Further studies should explore this possibility.

Finally, it may be that the attentional control difficulties of children with SLI (Noterdaeme, Amorosa, Mildenerger, Sitter & Minow, 2001) make it difficult for them to detect audiovisual synchrony relations. That is, although children with SLI are able to attend to the stimuli, they may not be able to divide their attention between simultaneously occurring congruent and incongruent visual and auditory stimuli as efficiently as typically developing children do. It should be noted, however, that this explanation does not account for the fact that these children can detect the A-V 666 asynchrony. Thus, although attentional problems may have contributed in some subtle way to the deficit, they do not appear to play a major role in it.

In sum, the current study has found that children with SLI are poorer than their typically developing peers in their ability to distinguish between temporally coherent and incoherent auditory and visual attributes of speech. The reasons for this impairment are currently not clear; further research is required.

REFERENCES

- Alsius, A., Navarra, J., Campbell, R. & Soto-Faraco, S. (2005). AV integration of speech falters under high attention demands. *Current Biology* **15**, 839–43.
- Bebko, J. M., Weiss, J. A., Denmark, J. L. & Gómez, P. (2006). Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism. *Journal of Child Psychology and Psychiatry* **47**(1), 88–98.

- Boliek, C., Keintz, C. K., 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**(6), 124–31.
- Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A. & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology* **5**, e1000436.
- Desjardins, R., Rogers, J. & Werker, J. F. (1997). An exploration of why preschoolers perform differently than do adults in AV perception tasks. *Journal of Experimental Child Psychology* **66**, 85–110.
- Dixon, N. F. & Spitz, L. T. (1980). The detection of auditory visual desynchrony. *Perception* **9**, 719–21.
- Dodd, B., McIntosh, B., Erdener, D. & Burnham, D. (2008). Perception of the auditory-visual illusion in speech perception by children with phonological disorders. *Clinical Phonetics & Linguistics* **22**, 69–82.
- Dunn, L. M., Dunn, L. M. & Arribas, D. (2006). *PPVT-III. Peabody. Test de vocabulario en imágenes*. Madrid: TEA Ediciones.
- Elliott, L. L. & Hammer, M. A. (1988). Longitudinal changes in auditory discrimination in normal children and children with language-learning problems. *Journal of Speech and Hearing Disorders* **53**, 467–74.
- Kaufman, A. S. & Kaufman, N. L. (2004). *KBIT: Kaufman Brief Intelligence Test (KBIT, Spanish version)*. Madrid: TEA Editions.
- Leonard, L. (1998). *Specific Language Impairment*. Cambridge, MA: MIT Press.
- Lewkowicz, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception & Performance* **22**(5), 1094–1106.
- Lewkowicz, D. J. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin* **126**(2), 281–308.
- Lewkowicz, D. J. (2010). Infant perception of audio-visual speech synchrony. *Developmental Psychology* **46**(1), 66–77.
- Lewkowicz, D. J. & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences* **109**(5), 1431–36.
- McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* **264**, 746–48.
- Munhall, K., Gribble, P., Sacco, L. & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception and Psychophysics* **58**, 351–62.
- Munhall, K. G. & Vatikiotis-Bateson, E. (2004). Spatial and temporal constraints on audiovisual speech perception. In G. A. Calvert, C. Spence & B. E. Stein (eds), *The handbook of multisensory processes*, 177–88. Cambridge, MA: MIT Press.
- 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**, 1639–51.
- Noterdaeme, M., Amorosa, H., Mildenerger, K., Sitter, S. & Minow, F. (2001). Evaluation of attention problems in children with autism and in children with specific language disorder. *European Child and Adolescent Psychiatry* **10**, 58–66.
- Pérez, E. & Serra, M. (1998). *Análisis del retraso del lenguaje (AREL)*. Barcelona: Ariel.
- Saborit, C. & Julián, J. P. (2003). *ELI – La evaluación del lenguaje infantil*. Castelló de la Plana: Universitat Jaume I.
- Siva, N., Stevens, E., Kuhl, P. & Meltzoff, A. (1995). A comparison between cerebral-palsied and normal adults in the perception of auditory-visual illusions. *Journal of the Acoustical Society of America* **98**, 2983.
- Stark, R. & Heinz, J. M. (1996). Vowel perception in children with and without language impairment. *Journal of Speech and Hearing Research* **39**, 860–69.
- Stark, R. E. & Tallal, P. (1981). Selection of children with specific language deficits. *Journal of Speech and Hearing Disorders* **46**, 114–22.
- Sumby, W. H. & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America* **26**, 212–15.

- Summerfield, A. Q. (1979). The use of visual information in phonetic perception. *Phonetica* **36**, 314–31.
- Tallal, P. (1984). Temporal or phonetic processing deficit in dyslexia? That is the question. *Applied Psycholinguistics* **5**, 167–69.
- Tallal, P. & Piercy, M. (1973). Developmental aphasia: Impaired rate of non-verbal processing as a function of sensory modality. *Neuropsychologia* **11**, 389–98.
- Tallal, P. & Piercy, M. (1975). Developmental aphasia: The perception of brief vowels and extended stop consonants. *Neuropsychologia* **13**, 69–74.
- Tallal, P., Stark, R. & Mellits, E. (1985). Identification of language-impaired children on the basis of rapid perception and production skills. *Brain and Language* **25**, 314–22.
- van Wassenhove, V., Grant, K. W. & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia* **45**(3), 598–607.
- Vatakis, A. & Spence, C. (2007). Crossmodal binding: Evaluating the ‘unity assumption’ using audiovisual speech stimuli. *Perception & Psychophysics* **69**, 744–56.
- Wechsler, D., Cordero, A. & de la Cruz, M. V. (1993). *WISC-R: escala de inteligencia de Wechsler para niños – revisada: manual*. Madrid: TEA Ediciones.
- Welch, R. B. & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin* **88**, 638–67.
- Wright, B. A., Lombardino, L. J., King, W. M., Puranik, C. S., Leonard, C. M. & Merzenich, M. M. (1997). Deficits in auditory temporal and spectral resolution in language-impaired children. *Nature* **387**, 176–78.
- Yehia, H., Rubin, P. & Vatikiotis-Bateson, E. (1998). Quantitative association of vocal-tract and facial behavior. *Speech Communication* **26**(1/2), 23–43.

APPENDIX

SCRIPT: *¡Buenos días, despiértate ya! ¡Si te levantas ahora tendremos una hora entera para jugar! Me encantan estas mañanas largas, ¿y a ti? Ojala no se acabaran nunca. Bueno, por lo menos es viernes y tenemos todo el sábado para descansar, excepto por lo de la fiesta. Me vas a ayudar a arreglar la casa, ¿sí? Tenemos que comprar flores, preparar la comida, sacar el polvo, aspirar la casa y limpiar los discos.*

ENGLISH TRANSLATION: *Good morning! Get up. Come on now. If you get up right away, we have a whole hour to putter around the house. I love these long mornings, don't you? I wish that they could last all day. Well, at least it's Friday and we can loaf around all day Saturday, except of course, for the party. Are you going to help me fix up the house? We have to buy flowers, prepare the food, vacuum the house, dust everything and clean the records.*