Published in final edited form as:
Child Dev. 2011 September ; 82(5): 1397–1403. doi:10.1111/j.1467-8624.2011.01619.x.

Can children with autism spectrum disorders "hear" a speaking face?

Julia R. Irwin[1,2], **Lauren A. Tornatore**[1,3], **Lawrence Brancazio**[1,2], and **D.H. Whalen**[1]

[1]Haskins Laboratories

[2]Southern Connecticut State University

[3]University of Massachusetts, Amherst

Abstract

This study used eye-tracking methodology to assess audiovisual (AV) speech perception in 26 children ranging in age from 5-15 years, half with autism spectrum disorders (ASD) and half with typical development (TD). Given the characteristic reduction in gaze to the faces of others in children with ASD, it was hypothesized that they would show reduced influence of visual information on heard speech. Responses were compared on a set of auditory, visual and audiovisual speech perception tasks. Even when fixated on the face of the speaker, children with ASD were less visually influenced than TD controls. This indicates fundamental differences in the processing of AV speech in children with ASD, which may contribute to their language and communication impairments.

Individuals diagnosed with autism spectrum disorders (ASD) show marked deficits in social and communicative functioning (American Psychiatric Association, 2000). Children with ASD often exhibit significant delays in the development of language (LeCouteur et al,1989; Lord & Paul, 1997). A lack of attention to *visual* speech information, which is known to facilitate language processing, may be a source of these communicative deficits. Sensitivity to audiovisual (AV) speech appears to be present in very early development (Meltzoff & Kuhl, 1994; Lewkowicz, 1996; Rosenblum, Schmuckler & Johnson, 1997). This implies that early sensitivity to AV speech is crucial in native language acquisition (Legerstee, 1990). Given the potentially important role of both visual and auditory speech in language development, a deficit in AV speech processing may contribute to language impairment.

One demonstration of visual influence on heard speech is the McGurk effect (McGurk & MacDonald, 1976), in this effect, synchronized but mismatching audio and video consonant-vowel tokens elicit a percept influenced by visual information (e.g., a visual /ga/ and an auditory /ma/ are "heard" as /na/). Children with ASD are less influenced by visual speech than those with typical development (TD) as evidenced by reduced visual influence in a McGurk-type paradigm (Massaro & Bosseler, 2003; Mongillo et al, 2008). Further, Smith & Bennetto (2007) report both weaker lipreading and reduced integration of matched AV speech in the context of auditory noise for individuals with ASD compared with TD controls.

Although several studies have shown a reduced role of visible speech information in children with ASD (de Gelder, Vroomen & van der Heide, 1991; Massaro & Bosseler, 2003; Mongillo et al, 2008; Williams, Massaro, Peel, Bosseler & Suddendorf, 2004), complicating this evidence is the characteristic reduction in gaze to the faces of others in these children

julia.irwin@haskins.yale.edu .

(Hobson, Ouston & Lee, 1988). Thus, it is difficult to determine whether children with ASD detect or integrate visual and auditory information less than TD controls, or whether they neglect visual information because they are not fixated on the speaker's face. If children with ASD exhibit poorer AV speech perception than TD children even when fixated on the face of a speaker, such deficits may reflect difficulty in extracting phonetic information from the face *or* in processing AV phonetic information.

The current study uses eye-tracking methodology paired with audiovisual speech perception tasks to examine the relationship between fixation on the face of the speaker and speech perception in children with ASD and TD controls. The use of eye-tracking allows us to discriminate between possible underlying causes of atypical AV perception of speech in ASD, that 1) children with ASD show reduced visual influence because they are not gazing at the face of a speaker, 2) children with ASD have an underlying weakness in the processing of AV speech, or 3) children with ASD have a general deficit in processing of AV (both speech and non-speech) stimuli.

Method

Participants

Participants were 30 native English speaking monolingual children, 15 with ASD diagnosed prior to the study using DSM-IV criteria by a licensed clinician, and 15 TD controls. Pairwise matching of a subgroup of TD controls taken from a larger group (N=80) of typical participants was done with the ASD participants based on cognitive and language functioning. To participate in the study, participants had to be able to identify /ma/ and /na/ in an auditory-only pretest at a rate of 80% or greater. Two TD participants and two with ASD did not correctly identify the syllables, and were excluded from analyses. The final sample of participants was 26 children, 13 with ASD (9 boys, mean age 9.08 years, age range 5-15 years) and 13 with TD (9 boys, mean age 9.16, age range 7-12 years). The mean age of ASD and TD participants did not differ significantly (t=.07, ns). The groups also did not significantly differ in identification of the auditory only tokens t (24) = 1.09, ns (ASD: M=96.46% correct, SD=6.14; TD: M=98.85% correct, SD=21.93).

Six participants had a clinical diagnosis of autism, three had Asperger Syndrome, and four had pervasive developmental disorder not otherwise specified (PDD-NOS); these diagnoses fall within the classification of ASD. In addition to the clinical diagnosis, participants with ASD were assessed with the Autism Diagnostic Observation Schedule Generic (ADOS-G; Lord et al, 2000; Lord, Rutter, DiLavore & Risi, 2002), a semi-structured standardized assessment of communication, social interaction, and play/imaginative use of materials for individuals suspected of having ASD. All participants with ASD met or exceeded cutoff scores for autism spectrum or autism proper on the ADOS algorithm.

Caregivers (N=13 females) of the children with ASD were interviewed with the Autism Diagnostic Interview-Revised (ADI-R; Lord, Rutter & LeCouteur, 1994). The ADI-R is a standardized, semi-structured interview for caregivers of individuals with ASD. Scores obtained from caregivers showed that the diagnosed children met or exceeded cutoff criteria on the language/communication and social interaction domain on the ADI-R. All but one of the children also met the criteria on the restricted, repetitive and stereotyped behaviors domain. All participants were reported by parents to have normal or corrected-to-normal hearing and vision. The TD controls had no history of developmental delays or speech or language problems by parent report.

Materials

Stimuli consisted of consonant-vowel (CV) syllables /ma/, /na/, /ga/ and the consonant-vowel-consonant-vowel (CVCV) syllables /bada/ recorded to create video clips. A male, monolingual native speaker of American English produced the stimuli in a recording booth.

Visual only (speechreading) stimuli

The visual only stimuli were silent versions of the speaker producing /ma/ and /na/. In this condition, participants were told that they would see a man saying some sounds that they would not be able to hear, and then asked to report what they thought the man was saying, for a total of 20 trials.

Speech in noise stimuli

Noise was added to the 60 dBA /ma/ and /na/ tokens to create a range of signal-to-noise levels at 5, 0, −5, −10, −15 and −20 dB, from less to more noisy. The audiovisual stimuli were the same auditory tokens with video of the speaker producing the same CV syllables. For both auditory and audiovisual stimuli, there were 24 trials.

AV match and mismatch (McGurk) stimuli

The mismatch stimuli were dubbed by placing the audio track such that the point of consonant release at the beginning of the vowel for a new auditory token matched the point of release for the original token, at the resolution of a single video frame, for a total of 12 trials. Mismatched stimuli were always a visual /ga/ token paired with an auditory /ma/. Matched stimuli replaced the audio from tokens of the same CV (e.g., a /ma/ visual token paired with a different auditory /ma/), for a total of 16 trials.

For the speech in noise and the AV match-mismatch conditions, participants were instructed to watch and listen to the video display. They were then told that they would hear a man saying some sounds that were not words and to say out loud what they heard. Results reported for the AV in noise, visual only (speechreading), and match and mismatch (McGurk) trials include only those trials where the participant was fixated on the face of the speaker within a time window crucial for phonetic judgment with these stimuli: the transition into the consonantal closure, during closure and through to the beginning of the release.

AV asynchronous stimuli

Audiovisual /bada/ tokens were edited such that the auditory and visual signals were separated and realigned at various temporal offsets. In addition to one synchrony condition, there were four asynchrony conditions: auditory leading visual (auditory lead) of 250 ms and 550 ms, visual leading auditory (visual lead) of 250 ms and 550 ms. Participants were told to watch and listen to the video and report whether the speaker's face and voice "talked" at the same time, a match; or at different times, a mismatch, for a total of 20 trials.

AV non-speech stimuli

The audiovisual non-speech stimuli consisted of a set of figure-eight shapes that increased and decreased in size, paired with sine-wave tones that varied in frequency and amplitude. These stimuli were modeled on the speaker's productions of /ma/ and /na/ to retain the temporal characteristics of speech, but did not look or sound like speech. To create the visual stimulus, the lip aperture was measured in every video frame of the /ma/ and /na/ syllables. These aperture values were the used to drive the size of the figure: when the lips closed the figure was small, upon consonant release into the vowel the figure expanded (insert Figure 1 about here). The auditory stimuli were created by converting the auditory /

ma/ and /na/ syllables into sine-wave analogs, which consist of three or four time-varying sinusoids, following the center-frequency and amplitude pattern of the spectral peaks of an utterance (Remez, Rubin, Pisoni & Carrell, 1981). These sine-wave analogs sound like chirps or tones. Thus, the audiovisual non-speech stimuli retained the temporal dynamics of speech, without looking or sounding like a speaking face.

Participants were told that they would see two shapes that would open and close and report whether they opened and closed in the same way (the shapes modeled on /ma/-/ma/ or /na/-/na/) or if the way that they closed was different (those modeled on /ma/-/na/ or /na/-/ma/), for a total of 28 trials.

Assessment—Language ability was assessed with the Clinical Evaluation of Language Fundamentals 4th Edition (CELF-4; Semel, Wiig & Secord, 2003, 5-21 years). The CELF-4 provides a core language index (CLI), which quantifies overall language ability. There were no significant group differences in language functioning, and scores for both groups fell within the typical range. Mean CLI for the TD group was 93.92 (SD=13.9) and for the ASD group was 91.38 (SD=13.3), t (24)=.71, ns.

Cognitive ability was assessed using the Differential Ability Scales School Age Cognitive Battery (DAS; Elliott, 1991). The DAS provides a General Conceptual Ability (GCA) score, which assesses Verbal Ability, Nonverbal Reasoning Ability, and Spatial Ability. A t-test showed that there were no significant group differences in cognitive functioning. Further, the mean standard score for the GCA for the TD group was 96.62 (SD=12.0) and for the ASD group 93.00 (SD=13.8), t (24)=.47, ns, both within the typical range.

Visual Tracking Methodology—Visual tracking was assessed with an ASL Model 504 pan/tilt remote tracking system. To optimize the accuracy of the pupil coordinates, this model has a magnetic head tracking unit that tracks the position of a small magnetic sensor attached above the left eye of the participant.

Procedure

After parental consent and child assent were obtained in accordance with the (name withheld to retain anonymity) Institutional Review Board, participants completed the experimental tasks in the eye-tracker laboratory at (name withheld to retain anonymity). Calibration of fixation points in the eye-tracker was completed first. Prior to stimulus presentation, directions appeared on the monitor and were read aloud by a researcher to ensure that the child understood the task. In addition, two practice items for each condition were completed with the researcher present to confirm that the child understood and could complete the task.

After every five trials, participants saw a video of animated shapes, to maintain attention to the task. Tasks were blocked, with stimuli presented in random order within a block. The inter-stimulus interval for all trials within the blocks was 3 seconds. The blocks were presented in a pseudo-random order; all participants were presented with the auditory-only stimuli first to ensure reliable discrimination between /ma/ and /na/. Audio stimuli were presented at a comfortable listening level (60 dBA) from a centrally located speaker under the eye-tracker.

Coding of behavioral responses

Responses were coded from videotapes of the experimental session. Each session was independently evaluated three times by trained coders blind to the participant's group membership. For the speechreading, speech in noise and match mismatch condition, coding was done by viseme class; that is, it was scored for correct place of articulation but not

manner or voicing. Specifically, for a /ma/ syllable, the responses /ma/, /ba/ and /pa/ were all accepted; for a /na/ syllable, the responses /na/, /da/, /ta/ and /la/ were accepted (this was done because some participants made systematic and consistent voicing or manner errors that preserved viseme class). In speechreading, speech in noise and the match component of the match mismatch task, accuracy reflects correct viseme class. For the mismatch condition, visually influenced responses include the dominant McGurk percept /na/ and the visemically equivalent /da/ and /la/. (Pilot testing with healthy child participants indicated that the dominant McGurk response for the mismatched AV stimuli was /na/. This reflects a percept with the auditory manner and a visual place of articulation that is intermediate between the auditory and visual signal. Other common responses had the same place, but different manner, namely /da/ and /la/.)

Coder responses for the non-speech condition were at 100% agreement. For the remaining conditions, Cohen's Kappa (Cohen, 1988) values were within the moderate to strong range (speechreading =.60, .93 for within viseme class responses, auditory and audiovisual speech in noise, both=.93, mismatched AV speech= .94, AV asynchrony=.96).

Results

Speech in noise and speechreading

There was a significantly greater number of dropped trials for the ASD than the TD group in the AV speech in noise condition because of lack of fixation on the face of the speaker during consonantal closure, t (24)=−2.15, p<05 (ASD: M=5.4, SD=3.6, 22.5% of trials; TD: M=, 3.2, SD 3.6, 13% of trials).

There were no significant group differences in the auditory speech in noise condition, indicating that both children with ASD and their TD peers were able to identify syllables in the context of auditory noise to a similar degree, $t(24)$=.52, ns, (ASD: M=56.8% correct place of articulation, SD=25.2; TD: M=61.3% correct, SD=18.8). (There was also no group difference when all noise levels were included.)

The AV speech in noise condition allows us to measure an increase in identification of the CV syllable in the presence of the face scaled to performance with auditory alone. To remove ceiling effects in the auditory condition, only the data from the three highest levels of noise **was included (**−10, −15 and −20 S/N ratio). To increase statistical power, mean accuracy of place of articulation across the noise levels was calculated. Thus, AV gain was the improvement in accuracy from A to AV relative to the maximum possible gain using the formula [(AV-A)/(100-A)]. Importantly, for trials in which children fixated on the face of the speaker, children with ASD showed significantly less visual gain compared to the TD controls. A group comparison revealed a significant difference in visual gain, t (24)= 2.71 p<.01 (ASD: M=57.5%, SD=32.9; TD M=88.9%, SD=25.8), Cohen's d =1.06. This suggests that even when visible articulatory information is available and they are fixated on it, children with ASD do not benefit from this information as much as the TD controls.

In the speechreading condition, there were also significantly more trials that had to be dropped for the ASD than the TD group because of lack of fixation on the face of the speaker during consonantal closure, t (24)=2.17, p<05 (ASD: M=8.2, SD=3.8, 41.0% of trials; M=5.7, SD=1.7, 28.0%). However, the comparison of interest, where participants were fixated on the face of the speaker revealed that participants with ASD were significantly less accurate speechreaders than TD controls, t (24)=2.50, p<.05 (ASD: M=87.9% correct place of articulation, SD=13.3; TD: M=97.6%, SD=3.9), Cohen's d = .98. Notably, the performance for both groups was relatively good, suggesting that there may be even larger differences between the two groups for a more difficult speechreading task.

AV matched and mismatched (McGurk)

As in the speech in noise and speechreading conditions, significantly more trials were dropped for the ASD than the TD group for lack of fixation on the face of the speaker during consonantal closure for the match-mismatch AV condition, t (24)=−5.88, p<.001 (ASD: M=5.35, SD=2.7, 19.1% of trials; TD: M=.92, SD=.27, 3.2% of trials). For the matched AV syllables, both groups were close to ceiling in place of articulation accuracy, and there is no between-group difference t (24)=1.3, ns (ASD: M=95.3, SD=11.4;TD:M=99.5, SD=1.73). In the mismatched auditory and visual condition (auditory /ma/ and visual /ga/), the groups were compared on percent of visually influenced responses. Children with ASD were significantly less visually influenced for the mismatched condition, even when fixating on the face, t (24)=2.74, p<.01 (ASD: M=55.7%, SD=33.5; TD: M=87.6%, SD=24.8), Cohen's d = 1.0.

AV asynchrony

To compare sensitivity to timing in speech perception in children with ASD and their TD controls, A', a non-parametric signal detection measure of perceptual sensitivity to differences between stimuli was employed. A' was calculated for an asynchrony condition by comparing it to the synchronous condition. Thus, a "mismatch" response to an asynchronous stimulus was coded as a "hit", and a "match" response to a synchronous stimulus was coded as a "correct rejection." The A' measure ranges from 1.0 (perfect performance) to 0 (consistently incorrect) with an A' of .5 corresponding to chance responding (Pollack & Norman, 1964).

Typically developing adult perceivers are more accurate at detecting larger asynchronies, and also show an asymmetry in ability to detect asynchrony: stimuli with a visual lead are more difficult to detect as asynchronous than those with auditory leads (Conrey & Pisoni, 2006). Overall, both groups in the current study performed better with large (550 ms) than small (250 ms) asynchronies. Further, at the small asynchrony, both groups performed worse with visual than auditory lead. A 2 (Group, ASD vs. TD) × 2 (Asynchrony, auditory vs. visual) × 2 (Timing, 250 vs. 550 ms) mixed analysis of variance (ANOVA) showed the expected effect of Timing F(1,22)=20.6, p<001 and an interaction of Timing and Asynchrony, F (1,22)=27.0, p<.001 (see Table 1). Critically, there was no interaction with timing or asynchrony for Group, (group × timing F(1,22)=1.78, ns; group × asynchrony F<1, ns) (see Table 1). This indicates that both the ASD and TD children show the characteristic asymmetry to asynchrony in AV speech stimuli. Further, to determine whether sensitivity with A' was above chance responding, one-sample t-tests were run for the ASD and TD groups. There were significant differences for both groups in a comparison of the A' value to .5 or chance responding at each level of timing × asynchrony (p<.01 or less for all comparisons, see Table 1).

AV non-speech

To compare performance in children with ASD and their TD controls in detecting non-speech cross-modal matching, A' was employed again. Thus, a "same" response to two AV shapes modeled on the same syllable was coded as a "hit", and a "different" response to two AV shapes, one modeled on /na/, the other on /ma/ was coded as a "correct rejection." The groups did not differ on ability to detect whether the non-speech AV tokens were same or different t (24)=.52, ns (mean A' ASD: .67, SD= .27;TD: M=.72, SD=.19). A comparison of the A' value to .05 (chance) responding indicated significant differences for both groups by comparing the A' value to .5 or chance responding, with t (12)=2.37, p<.05 for the ASD group and t (12)=4.13, p<.001 for the TD group. Thus, the groups did not differ in sensitivity to AV non-speech tasks modeled on the dynamics of speech.

Discussion

This study used visual tracking methodology to assess visual influence on heard speech in children with ASD. Even when fixated on the face of the speaker, children with ASD were less visually influenced than TD controls for tasks that involved phonetic processing of visual speech. Children with ASD were significantly weaker at speechreading than TD controls and showed reduced visual influence for the mismatched auditory and visual (McGurk) and AV speech in noise stimuli, where they reported auditory-only percepts significantly more often than the TD controls (Magnee, de Gelder, van Engeland & Kemner, 2008; Smith & Bennetto, 2007). Insufficient speechreading skill in children with ASD may account for impaired performance in AV tasks, or could be due to a unique deficit in AV integration.

While the children with ASD gazed less at the face of the speaker than TD controls, differences between children with ASD and TD in perception of AV speech stimuli were not due to lack of gaze to a speaker's face, because only responses during fixation on the face were analyzed. Children with ASD performed similarly to TD children on the AV asynchrony and the non-speech tasks, suggesting that the impairment in processing of AV stimuli is speech-specific. Children with ASD exhibited particular difficulty with processing of AV *phonetic* information, including speechreading, AV speech in noise, and AV matched and mismatched speech. However, a difference in the task (producing what the speaker said versus identifying a match or mismatch) could also contribute to observed group differences.

The current data suggest that children with ASD were not globally impaired in perception of audiovisual information. While there is evidence that they prefer synchronous AV stimuli (Klin, Lin, Gorrindo, Ramsay & Jones, 2009), they are able to detect temporal offset in AV stimuli (Grossman, Schneps & Tager-Flusberg, 2009). In this study children with ASD showed an asymmetric pattern in detection of this AV asynchrony, similar to TD children; however, the temporally asynchronous speech stimuli did not require phonetic processing. Because there is evidence of impaired performance of ASD participants in tasks involving audiovisual phonetic information (i.e., AV speech in noise and mismatched McGurk stimuli), there may be an important distinction between audiovisual *processing* that involves phonetic perception as opposed to those involving audiovisual *timing* perception. Children with ASD were impaired in using audiovisual information in phonetic perception but not in non-phonetic judgments in the asynchrony task. Further, they showed no differences in comparison to TD children in their sensitivity to non-speech (and non-face) cross-modal inconsistencies. Thus, the current study reveals a potential mechanism that underlies the speech and language difficulties in children with ASD, a deficit in phonetic processing of AV speech.

These findings inform us about the significant developmental consequences of a lack of gaze to the face of a speaker. Beginning early in development, young children with ASD likely look less at a speaking face than their typically developing peers. This behavior could lead to weaker AV speech perception, which may have cascading effects on language development. In this manner, fundamental differences in attention during social interactions may influence the development of language perception and use. Even in the current sample of children with ASD who fell within the typical range on standardized tests of language skill, evidence of deficits in the perception of AV articulatory information was found. This raises the possibility that children with ASD with more significant language impairments have even greater deficits in speech perception, and furthermore, that these difficulties in speech processing underlie their language impairments. Continuing to pursue the etiology of the deficits in phonetic perception in children with ASD using both auditory and audiovisual

speech stimuli will lead to a better understanding of both typical and atypical language development.

Acknowledgments

This work was supported by NIH grants R03-007339, R01-000403, and HD001994. Thanks to Catherine Best, Alice Carter, Mike D'Angelo, Carol Fowler, Mark Gancsos, Tiffany Gooding, Jessica Grittner, Dominic Massaro, Karen Mulak, and Mark Tiede.

References

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed.. Washington, DC: 2000. Text Revision

Cohen, J. Statistical power analysis for the behavioral sciences. 2nd ed.. Lawrence Erlbaum Associates; Hillsdale, NJ: 1988.

Conrey BL, Pisoni DB. Auditory-visual speech perception and synchrony detection for speech and non-speech signals. Journal of the Acoustical Society of America. 2006; 119:4065–4073. [PubMed: 16838548]

De Gelder B, Vroomen J, van der Heide L. Face recognition and lipreading in autism. European Journal of Cognitive Psychology. 1991; 3:69–86.

Elliot, CD. Differential Ability Scales: Introductory and Technical Handbook. The Psychological Corporation; San Antonio, TX: 1991.

Grossman RB, Schneps MH, Tager-Flusberg H. Slipped lips: onset asynchrony detection of auditory-visual language in autism. Journal of Child Psychology and Psychiatry. 2009; 50:491–497. [PubMed: 19207623]

Hobson RP, Ouston J, Lee A. What's in a face? The case of autism. British Journal of Psychology. 1988; 79(4):441–453. [PubMed: 3208000]

Klin A, Lin DJ, Gorrindo P, Ramsay G, Jones W. Two-year-olds with autism orient to non-social contingencies rather than biological motion. Nature. 2009; 459:257–261. [PubMed: 19329996]

LeCouteur A, Rutter M, Lord C, Rios P, Robertson S, Holdgrafer M, McLennan J. Autism diagnostic interview: A standardized investigator based instrument. Journal of Autism and Developmental Disorders. 1989; 19(3):363–387. [PubMed: 2793783]

Legerstee M. Infants use multimodal information to imitate speech sounds. Infant Behavior and Development. 1990; 13:343–354.

Lewkowicz DJ. Perception of auditory-visual temporal synchrony in human infants. Journal of Experimental Psychology. 1996; 22(5):1094–1106. [PubMed: 8865617]

Lord, C.; Paul, R. Language and communication in autism. In: Cohen, DJ.; Volkmar, FR., editors. Handboook of Autism and Pervasive Developmental Disorders. Wiley; New York, NY: 1997. p. 195-225.

Lord C, Risi S, Lambrecht L, Cook EH Jr. Leventhal BL, DiLavore PC, Pickles A, Rutter M. The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism. Journal of Autism & Developmental Disorders. 2000; 30(3):205–223. [PubMed: 11055457]

Lord, C.; Rutter, M.; DiLavore, PC.; Risi, S. Autism Diagnostic Observation Schedule: Manual. Western Psychological Services; Los Angeles, CA: 2002.

Lord C, Rutter M, LeCouteur A. Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. Journal of Autism and Developmental Disorders. 1994; 24:659–685. [PubMed: 7814313]

Magnee MJCM, de Gelder B, van Engeland H, Kemner C. Audiovisual speech integration in Pervasive Developmental Disorder: Evidence from Event Related Potentials. Journal of Child Psychology and Psychiatry. 2008; 49:995–1000. [PubMed: 18492039]

Massaro DW, Bosseler A. Perceiving speech by ear and eye: Multimodal integration by children with autism. The Journal of Developmental and Learning Disorders. 2003; 7:111–146.

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

Meltzoff, AN.; Kuhl, PK. Faces and speech: intermodal processing of biologically relevant signals in infants and adults. In: Lewkowicz, DJ.; Lickliter, R., editors. The development of intersensory perception: comparative perspectives. 1994. p. 335-398.

Mongillo EA, Irwin JR, Whalen DH, Klaiman C, Carter AS, Schultz RT. Audiovisual processing in children with and without autism spectrum disorders. Journal of Autism and Developmental Disorders. 2008; 38(7):1349–1358. [PubMed: 18307027]

Pollack I, Norman DA. A non-parametric analysis of recognition experiments. Psychonomic Science. 1964; 1(5):125–126.

Remez RE, Rubin PE, Pisoni DB, Carrell TD. Speech perception without traditional speech cues. Science. 1981; 212(4497):947–949. [PubMed: 7233191]

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

Semel, E.; Wiig, E.; Secord, W. Clinical evaluation of language fundamentals (4th ed.): Examiner's manual. Harcourt Assessment; San Antonio TX: 2003.

Smith EG, Bennetto L. Audiovisual speech integration and lipreading in autism. Journal of Child Psychology and Psychiatry. 2007; 48(8):813–821. [PubMed: 17683453]

Williams JHG, Massaro DW, Peel NJ, Bosseler A, Suddendorf T. Visual-auditory integration during speech imitation in autism. Research in Developmental Disabilities. 2004; 25(6):559–575. [PubMed: 15541632]

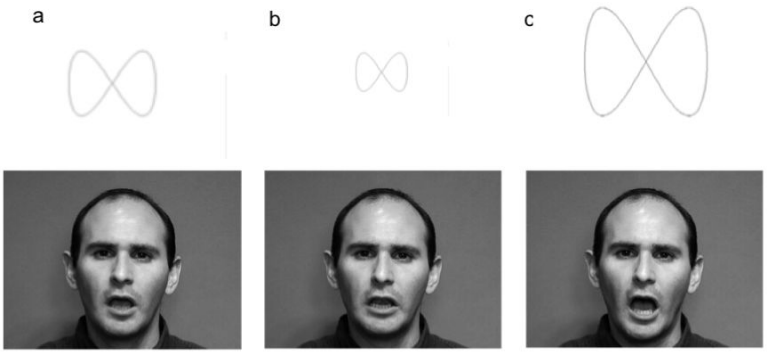

Figure 1.
Selected video frames of the non-speech figure driven by lip aperture from a video /na/ token. The images correspond to a) opening prior to consonantal closure, b) consonantal closure, and c) maximum opening for the vowel.

Table 1

Mean A' by group for detection of audiovisual asynchrony

	ASD		TD	
	250 ms Mean (SD)	**550 ms Mean (SD)**	**250 ms Mean (SD)**	**550 ms Mean (SD)**
Auditory lead	.87 (.08)	.90 (.07)	.79 (.07)	.83 (.07)
Video lead	.47 (.05)	.76 (.06)	.53 (.05)	.80 (.06)