Published in final edited form as:

Res Autism Spectr Disord. 2017 July ; 39: 11–19. doi:10.1016/j.rasd.2017.04.001.

Temporal synchrony and audiovisual integration of speech and object stimuli in autism

Elizabeth Smith^{a,1}, Shouling Zhang^a, and Loisa Bennetto^a

^aDepartment of Clinical and Social Sciences in Psychology, University of Rochester, Rochester, NY USA

Abstract

Background—Individuals with Autism Spectrum Disorders (ASD) have been shown to have multisensory integration deficits, which may lead to problems perceiving complex, multisensory environments. For example, understanding audiovisual speech requires integration of visual information from the lips and face with auditory information from the voice, and audiovisual speech integration deficits can lead to impaired understanding and comprehension. While there is strong evidence for an audiovisual speech integration impairment in ASD, it is unclear whether this impairment is due to low level perceptual processes that affect all types of audiovisual integration or if it is specific to speech processing.

Method—Here, we measure audiovisual integration of basic speech (i.e., consonant-vowel utterances) and object stimuli (i.e., a bouncing ball) in adolescents with ASD and well-matched controls. We calculate a temporal window of integration (TWI) using each individual's ability to identify which of two videos (one temporally aligned and one misaligned) matches auditory stimuli. The TWI measures tolerance for temporal asynchrony between the auditory and visual streams, and is an important feature of audiovisual perception.

Results—While controls showed similar tolerance of asynchrony for the simple speech and object stimuli, individuals with ASD did not. Specifically, individuals with ASD showed less tolerance of asynchrony for speech stimuli compared to object stimuli. In individuals with ASD, decreased tolerance for asynchrony in speech stimuli was associated with higher ratings of autism symptom severity.

Conclusions—These results suggest that audiovisual perception in ASD may vary for speech and object stimuli beyond what can be accounted for by stimulus complexity.

Keywords

speech processing; audiovisual integration; temporal perception

Present Address: Elizabeth Smith, The Child Development Lab, Department of Human Development, 3304 Benjamin Building, University of Maryland, College Park, MD 20742-1131

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Multisensory integration is a ubiquitous process by which information combined from multiple sensory streams leads to faster, more robust, and more accurate perception of the world around us. For example, when listening to a person speak, watching their mouth move leads to a faster auditory brain response (van Wassenhove, Grant, & Poeppel, 2005), louder-sounding speech (MacLeod & Summerfield, 1987), and increased understanding of both words (Schwartz, Berthommier, & Savariaux, 2004) and ideas (Arnold & Hill, 2001; Reisberg, McLean, & Goldfield, 1987). Some aspects of audiovisual integration are already developed and robust within the first year of life both at neural (Hyde, Jones, Porter, & Flom, 2010) and behavioral levels (Lewkowicz, 2010). In fact, audiovisual integration in infancy predicts later language and communication abilities (Kushnerenko et al., 2013), which is evidence for the importance of audiovisual integration in early language development. While audiovisual integration is present and important early in life, it continues to mature through childhood and adolescence (Dick, Solodkin, & Small, 2010; Knowland, Mercure, Karmiloff-Smith, Dick, & Thomas, 2014; Lewkowicz & Flom, 2014).

Mounting evidence supports an audiovisual integration impairment in ASD, although signs of this impairment may be subtle. For example, individuals with autism show a reduced McGurk effect (Mongillo et al., 2008; Woynaroski et al., 2013), which persists even when their looking time to faces is the same as in controls (Irwin, Tornatore, Brancazio, & Whalen, 2011). Individuals with autism also show reduced audiovisual benefit when listening to speech in background noise (Smith & Bennetto, 2007), although this may improve with age (Foxy et al., 2013). These differences are reflected at the neurophysiological level as shown in studies using both electroencephalography (A. Brandwein et al., 2014; A. B. Brandwein et al., 2013; Megnin et al., 2012) and fMRI (Hubbard et al., 2012). Notably, studies using the non-social “sound induced flash illusion” (Shams, Kamitani, & Shimojo, 2000) have shown intact integration of nonsocial stimuli in autism (i.e., presence of the illusion, Foss-Feig et al., 2010; van der Smagt, van Engeland, & Kemner, 2007).

Temporal synchronization of auditory and visual cues is a feature of audiovisual integration that plays a significant role in the degree of audiovisual benefit. Reducing audiovisual temporal synchrony can reduce both illusory audiovisual effects (Ryan A. Stevenson, Zemtsov, & Wallace, 2012) and audiovisual reaction time benefits (Diederich & Colonius, 2004). For any construct capturing audiovisual integration, there is thought to be a window of audiovisual asynchronies within which the dependent variable increases probabilistically, frequently referred to as the Temporal Window of Integration (TWI, Spence & Squire, 2003; van Wassenhove, Grant, & Poeppel, 2007). Recently, the effects of temporal synchrony on audiovisual integration in autism have been measured across multiple studies (de Boer-Schellekens, Eussen, & Vroomen, 2013; de Boer-Schellekens, Keetels, Eussen, & Vroomen, 2013; Foss-Feig et al., 2010; Grossman, Schneps, & Tager-Flusberg, 2009; Kwakye, Foss-Feig, Cascio, Stone, & Wallace, 2011; R. A. Stevenson et al., 2014; Woynaroski et al., 2013). Findings have included a larger temporal window of integration in autism across multiple stimulus types (de Boer-Schellekens, Eussen, et al., 2013; de Boer-Schellekens, Keetels, et al., 2013; Foss-Feig et al., 2010; Kwakye et al., 2011; Woynaroski et al., 2013),

differentially larger temporal windows for speech stimuli only (Stevenson et al., 2014), or no difference in temporal windows (Grossman et al., 2009). Importantly, TWI can be affected by stimulus type (e.g., speech, object), stimulus complexity (e.g. complex human speech vs consonant-vowel utterances), and choice of dependent variable (e.g., presence of illusions, temporal order judgment, or simultaneity judgment), which may explain these discrepancies (R. A. Stevenson & Wallace, 2013; van Eijk, Kohlrausch, Juola, & van de Par, 2008; Vatakis, Navarra, Soto-Faraco, & Spence, 2008; Vatakis & Spence, 2006, 2010). While it is known that social and nonsocial stimuli are associated with different TWIs and that individuals with autism tend to show less impairment in nonsocial audiovisual integration paradigms, only one study directly compares social and nonsocial temporal synchrony perception in autism (Stevenson et al., 2014). Thus, while differences in multisensory integration for social versus nonsocial stimuli are a hallmark of the audiovisual integration literature in autism, we know little about the role of temporal synchrony detection in this pattern.

Most audiovisual integration tasks by their nature require various degrees of verbal processing. Specifically, they may require the subject to report on their perception, stating 1) whether sound or video appeared first (temporal order judgment) 2) whether the audio and video were in synch (simultaneity judgment) 3) what speech sound they heard (McGurk effect) or 4) how many times they heard a beep (flash-beep illusion). While responding with single words is not likely to be particularly taxing for the individuals with autism participating in these studies (i.e. those without intellectual disability), reduction of verbal task demands is preferable, particularly if those studies are to be extended to individuals with more significant verbal deficits.

While there is evidence that minimizing verbal demands still leads to group differences in audiovisual integration tasks in autism (Collignon et al., 2013), it is unknown whether this is the case for temporal synchrony tasks. Beyond verbal processes related to task demands, many investigations of temporal synchrony detection and its role in audiovisual integration in autism use speech stimuli that are both more visually and auditorily complex than those used for nonspeech conditions. Because more complex stimuli are generally associated with larger temporal windows of integration (Vatakis & Spence, 2006, 2010), this could lead to finding larger TWIs in autism that are related to the complexity of the stimuli rather than its social nature (D. L. Williams, Minshew, & Goldstein, 2015). Here, we use speech and object-based audiovisual stimuli designed to minimize differences in complexities (i.e., in timing, pitch, etc) paired with a nonverbal temporal synchrony task to determine the role of stimulus type in audiovisual integration in autism. Specifically, we used a paradigm based on detection versus discrimination of temporal synchrony, and used speech and object stimuli shown in a pilot study to have similar TWIs. The task in the present study requires participants to use a physical response (i.e., pushing arrow buttons vs. verbal response) to select which of two videos matches auditory stimuli in both speech and object conditions. Selecting the matching video parallels the essential process of using visual cues to locate a speaker in a noisy room (i.e., the Audiovisual Cocktail Party Effect, (Zion Golumbic, Cogan, Schroeder, & Poeppel, 2013). While similar designs have shown sensitivity to audiovisual temporal synchrony in autism and typical development, (Baart, Vroomen, Shaw, & Bortfeld, 2014; Bebko, Weiss, Demark, & Gomez, 2006; Patten, Watson, & Baranek, 2014), the

present design incorporates subtly varying audiovisual asynchronies in order to measure the TWI for responses to both speech and object stimuli.

Methods

This study was approved by the University of Rochester Institutional Review Board, and all subjects and parents of subjects 17 years of age and younger completed informed consent.

Participants

Participants were 40 children and adolescents (ages 11-17, 20 with autism and 20 typically developing controls) matched on chronological age, sex, verbal and performance IQ (Table 1). Diagnoses of autism were confirmed in the autism group using both the Autism Diagnostic Observation Schedule (Lord, Rutter, Di Lavore, & Risi, 2000) and the Autism Diagnostic Interview- Revised (Rutter, LeCouteur, & Lord, 2003). Only participants who met criteria for autism on both instruments were included in the autism group. Autism spectrum disorder was ruled out in the control group with a combination of the Social Communication Questionnaire (Rutter, Bailey, & Lord, 2004) and the ADOS. Participants were only included in the control group if they failed to meet diagnostic criteria for any pervasive developmental disorder. Parents of participants in both groups also completed the Social Responsiveness Scale (Constantino & Gruber, 2005) which is a 65-item, parent-report scale measuring autism symptomatology across five scales (i.e., social awareness, social cognition, social communication, social motivation, and autistic mannerisms). It provides standardized, continuous variables that can be used to assess symptoms associated with autism.

Exclusion criteria for all participants included: 1) prior or concurrent diagnosis of a genetic disorder, neurological disorder, or neurological abnormalities other than autism 2) history of head trauma with loss of consciousness. Additional exclusion criteria for the control group included: 1) a first or second degree relative with an autism spectrum disorder or 2) evidence of psychiatric or behavioral impairment, as indicated by parent report of DSM-IV Axis I diagnosis or a factor T-score in the clinical range on any of the DSM-oriented scales of the Child Behavior Checklist (CBCL, Achenbach et al., 2001). Individuals in the autism group were not excluded based on CBCL symptoms or concurrent Axis I diagnoses with the exception of exclusion for concurrent thought disorder.

All participants were native English speakers with normal hearing and normal or corrected to normal vision. Auditory acuity was measured using a Maico MA32 advanced diagnostic audiometer. Subjects with hearing loss greater than 20 decibels within the frequency range of 1000-4000 Hz were excluded based on the American Speech and Hearing Association's criteria for determining hearing impairment in children 5-18 years old. Visual acuity was measured for each eye using a Snellen chart; subjects with corrected visual acuity worse than 20/25 in either eye were excluded.

Stimuli

Speech stimuli—A 20-year old female speaker was recorded with a digital video camera equipped with an external microphone while saying multiple tokens of the consonant-vowel

utterances “ba,” “ga,” and “da”. The speaker was instructed to maintain flat prosody across utterances. Stimuli were then edited using Praat (Boersma & Weenink, 2009) to remove prominent pitch fluctuations and give all stimulus tokens the same average pitch. Stimuli were edited into 1 second segments that began and ended with the mouth closed, with the mouth open for approximately 800ms. Segments including blinking, glottal stops or starts, or other anomalies were not used.

Object stimuli—A black rubber ball with a diameter of 1.5 cm was video-recorded while being dropped through a tube and landing at a 90 degree angle on a wooden platform. Sound was recorded with a microphone embedded in the platform. The 1 second visual stimuli began showing the platform only, then showed a ball appearing at the top of the screen, falling to a platform, and bouncing up and out of the screen, with the ball present on the screen for approximately 200 ms. Bouncing ball stimuli were edited in Praat to have the same mean frequency and peak intensity as stimuli from the speech condition. Frequencies of all sounds from the speech stimuli and bouncing ball stimuli were within those measured by the hearing test utilized for study inclusion.

The speech and object stimuli were edited using Finalcut Pro so that each stimulus presentation included 5 repetitions of the original stimulus, lasting for a total of 5 seconds (see Figure 1). Further, stimuli were edited to create 16 variations of each stimulus, each with varying degrees of audiovisual temporal asynchrony. These included both auditory leads (i.e., for an audiolead of 100ms, the audio track is advanced 100 ms relative to the video track) and visual leads (i.e., for a 100 ms video lead, the video starts 100 ms prior to the audio track).

Final stimuli presented to research subjects show two visually identical videos presented side by side via DirectRT stimulus presentation software. One of the two videos (either on the left or right, determined randomly) always matched the sound that was played and one video either started earlier or later than the sound such that it was off by 33 to 433 milliseconds either in the audio or visual lead direction. The first frames of the 2 videos were always identical, to prevent participants from inferring the synchronous video based on initial characteristics. Stimuli began 1 second after participants responded to the previous trial, and were presented on an LCD monitor with an integrated sound bar in front of participants in a sound attenuated room.

Participants were told that they would see two videos and hear one sound, and that their job was to select which video on the screen matched the sound using the arrow keys on the keyboard (i.e., left, right, similar to (Baart et al., 2014)). We then presented four blocks of stimuli using an adaptive paradigm to present audiovisual stimuli at difficulty levels dependent on individual performance. Examples of stimuli at the easiest level (i.e., highest asynchrony between the audio and the nonmatching video) were presented as practice. Then, for each condition, participants began at the easiest level (i.e., highest asynchrony) and selected which video (left,right) matched the sound. If they selected the correct response, the level of asynchrony was decreased, whereas if they selected an incorrect response, the level increased (i.e., 1 up, 1 down rule), completing 60 trials per condition. Participants completed separate auditory lead and visual lead conditions, with the adaptive paradigm completed

separately for auditory and visual leads. Then, using each participant's number of presented and number of correct trials at each level of asynchrony for each condition, we fit two psychometric sigmoid functions to each individual's data to account for the audio leads and visual leads separately (e.g., R. A. Stevenson et al., 2014). We used the glm package in Matlab, which accounts for both accuracy and number of trials, to find the best fitting sigmoid functions, then used those functions to calculate the TWI with the window set at a correct discrimination rate of 75%. This method of fitting sigmoid curves to the data was associated with high goodness of fit, with likelihood ratio chi-square values ranging from 34.9 to 258.6 when compared to the intercept-only model (for goodness-of-fit estimates for each group and condition, see Supplemental Table 1.)

Pilot Study

Because TWI can vary depending on place of articulation for speech (Vatakis, Maragos, Rodomagoulakis, & Spence, 2012), we piloted the object stimuli (bouncing balls) and three consonant-vowel utterances (“ba”, “da” and “ga”) with 16 typically developing young adults (9 female, ages 18-30) to determine which of the three CV utterances was most similar to the object stimulus. All pilot subjects passed vision and hearing screening using a Snellen Chart and a Maico Audiometer.

Our pilot study showed that the TWIs for the bouncing ball stimuli were most similar to those of the “ba” CV utterances (see Figure 2.) Specifically, the TWI for the rubber ball and the “ba” stimuli did not vary ($t(15)=.02$, $p=.99$, $\text{cohen's } d=.005$), while the TWI for the rubber ball and the “da” stimulus did vary ($t(15)=-2.17$, $p=.05$, $\text{Cohen's } d=.55$). The TWI for the rubber ball and the “ga” stimulus did not vary in this sample, but showed an effect size indicating that it might vary in a larger sample ($t(15)=-1.3$, $p=.2$, $\text{cohen's } d=.54$). We therefore selected the “ba” CV utterances for use in the speech condition for the study in individuals with autism and controls.

Procedure

Participants in the autism and control groups completed this session over a period of 2 hours. Their session began with a hearing and vision screen, as well as a brief medical history. Then, participants completed the temporal asynchrony task as part of a larger battery of experimental tasks. Participants completed the audiovisual temporal asynchrony task, using identical stimuli and procedure as described above for the pilot participants, except that child and adolescent participants saw only the “ba” tokens for the speech condition combined with the bouncing ball used for the object condition.

Analysis

TWIs were calculated as described above for each participant for each condition. All variables evaluated met the assumptions of normality. Effects of group and condition were then evaluated using a repeated-measures ANOVA with group as the between subjects factor. Significant results led to follow up with paired-samples t-tests and ANOVAs.

Results

There was a significant interaction between condition (Speech, Object), and Group (Autism, Control) ($F=6.7$, $p=.01$, $\eta^2_{\text{partial}}=.15$). This interaction was characterized by a differentially smaller TWI for speech compared to object stimuli in autism in comparison to the control group (see Figure 3). Pairwise comparisons showed a difference in TWI for speech and object stimuli in autism ($t=-2.68$, $p=.02$, $d=.67$) but not in controls ($t=1.11$, $p=.28$, $d=.32$). Independent samples t-test showed a trend toward a smaller TWI for the speech condition in autism versus controls ($t=1.95$, $p=.06$, $d=.63$) but no difference in the TWI for the object condition ($t=1.45$, $p=.15$, $d=.47$). TWIs for the speech and object stimuli were not correlated (Pearson's $r=.13$, $p=.41$). The condition by group interaction was replicated when looking in males only ($F=8.02$, $p=.008$, $\eta^2_{\text{partial}}=.19$).

Because TWI for the speech condition was also associated with FSIQ (Pearson $r=-.324$, $r^2=.10$, $p=.04$), we determined effects of this association on the group interaction, and found that the interaction continued to be significant after controlling for the effects of FSIQ ($F=6.7$, $p=.01$, $\eta^2_{\text{partial}}=.154$). The TWI for the object condition was not associated with FSIQ (Pearson $r=-.15$, $r^2=.02$, $p=.37$).

In the autism group, there was a significant correlation between the TWI for the speech condition and the overall T-score for the SRS (Pearson $r=-.64$, $r^2=.4$, $p=.003$). Specifically, having a narrower speech TWI (i.e., being more sensitive to audiovisual asynchronies) was associated with an increase in autism symptom severity (see Figure 4). This correlation remained significant after accounting for the effects of FSIQ (Pearson $r=-.73$, $r^2=.53$, $p<.001$). There was no relation between the width of the object TWI and symptom severity (Pearson $r=-.17$, $r^2=.03$, $p=.48$).

In the control group, there was no correlation between the TWI for either the speech or object conditions and the SRS overall T-scores ($r=.02$, $r^2=.0004$, $p=.93$; $r=-.36$, $r^2=.13$, $p=.12$).

Discussion

In this study, we investigate temporal synchrony detection in a sample of children and adolescents with autism and well-matched controls using both speech and object-based stimuli in a task designed to simulate the process of locating a talking person. Individuals with autism showed a differentially smaller temporal window of integration (TWI) for the speech vs. object condition in comparison to typical controls. In spite of creating speech and object stimuli designed to minimize differences in complexity, and demonstrating no difference in TWI for the speech and object conditions in a pilot group of adults and a well-matched control group, individuals in the autism group showed a smaller TWI for the speech vs. object stimuli. Finally, having a narrow TWI for the speech stimuli was associated with parent-reported autism symptom severity in the group of individuals with autism.

This study supports findings of differential audiovisual integration in autism for social versus nonsocial tasks (Bebko et al., 2006; Mongillo et al., 2008; R. A. Stevenson et al., 2014). However, finding a smaller TWI for speech versus object stimuli in autism differs

from the pattern shown in Stevenson et al., 2014, where individuals with autism had wider TWI's on a simultaneity judgment task for McGurk syllables compared to controls. Potential methodological differences that might contribute to these discrepancies include uses of both “ba” and “ga” in Stevenson et al. (vs. “ba” only in the present study), group differences at baseline, and the length of the study session. Specifically, all consonant-vowel stimuli presented to individuals with autism and controls in the present study were bilabial (“ba”), whereas stimuli from the McGurk task include both bilabial and velar places of articulation (“ba” and “ga”, respectively). While it is known that place of articulation alters the TWI (Vatakis et al., 2012) it is unknown whether this effect varies in individuals with autism. It is therefore possible that group differences in TWI were differentially affected by perceived place of articulation. In addition, the autism group showed a lower mean VIQ compared to controls in the Stevenson et al. study. In the present study we found that higher IQ was associated with a narrower TWI for the speech condition in our VIQ-matched autism and control groups, a pattern that may contribute to different findings between studies. Finally, wider TWIs in the autism group during speech stimuli in the Stevenson et al. study could also be affected by group differences in temporal recalibration. Specifically, while repeated presentation of asynchronous audiovisual stimuli leads to widening of the TWI, this appears to affect speech stimuli to a greater degree in autism (Noel, De Nier, Stevenson, Alais, & Wallace, 2016). As a result, individuals with autism may show smaller TWIs to speech stimuli as reported here, but when exposed to a greater number of trials, as occurs in the method of constant stimuli compared to the present adaptive paradigm, would appear to have a wider window due to unequal recalibration.

Aside from discrepancies related to methodological differences, it is also possible for individuals with autism to show markedly different patterns of temporal sensitivity across tasks even for very similar stimuli. Specifically, in simultaneity and temporal order judgment tasks, participants are being asked to describe “what” they are perceiving (i.e., if it is “audio first”, “video first”, or “matched”), whereas the present task refocuses the question on “where” simultaneity is perceived. Importantly, TWIs for temporal and spatial binding processes are not correlated within typically developing individuals (B. Odegaard & L. Shams, 2016), emphasizing potential independence of the mechanisms for temporal and spatial audiovisual integration processes. In fact, previous research suggests that we ought to expect smaller TWIs in autism for social stimuli specifically during spatial localization tasks (i.e., the present study). Specifically, visual search is best facilitated by linked auditory cues when attention is distributed as opposed to captured (Van der Burg, Olivers, & Theeuwes, 2012). In this way, individuals with autism could benefit from decreased attentional engagement with faces (Kikuchi et al., 2011), orienting more quickly than controls to a matching audiovisual face, allowing them superiority in audiovisual localization of speaking faces but not in tasks related to awareness of audiovisual temporal synchrony (e.g., Stevenson et al., 2014). In short, the TWI in the present study is meant to capture the window within which participants can correctly locate a matching video, which may not parallel the TWI within which identification of sounds is altered.

It is unclear, based on the extant literature, whether a wider TWI for audiovisual stimuli affords an individual perceptual benefit, harm, or even a combination of the two. It has generally been assumed that a smaller TWI indicates better use of temporal cues during the

process of audiovisual integration. However, it could also be that having a smaller TWI leads to reduced integration, because integration is limited to that smaller window. This implies that individuals with smaller TWIs would only receive the benefits of audiovisual integration (i.e., improved hearing and comprehension) within a more narrow set of circumstances. Research supporting this possibility shows that individuals with smaller TWIs for simple nonsocial stimuli showed decreased integration of those same types of stimuli (Ryan A. Stevenson et al., 2012). It will be important for future work to clarify the process by which a smaller window of integration relates to perceptual integration, and how this process might be different in individuals with autism.

This study was designed to clarify the existing literature on audiovisual temporal synchrony detection in autism by reducing potential confounds related to social and nonsocial stimulus features. Persisting differences in the processing of speech and object stimuli reported here provide support for differences in processing of audiovisual temporal synchrony for social stimuli in autism. Sensory integration, rather than being a unitary process in the brain, likely emerges from multiple processes (Brian Odegaard & Ladan Shams, 2016). These processes may be driven by other cues seemingly unrelated to audiovisual integration that vary between social and nonsocial stimuli. For example, the presence of an upright (versus inverted) faces leads to differences in both the strength of the McGurk effect (Eskelund, MacDonald, & Andersen, 2015) and temporal synchrony detection (Vatakis & Spence, 2008). Upright faces and other cues intrinsic to social stimuli (e.g., biological motion, direct gaze) may influence audiovisual integration more globally, leading to differential effects in autism that cannot be accounted for by a central deficit in temporal synchrony detection.

This study has several important limitations. First, the two conditions capture perception of simple speech and object stimuli, but did not include an intermediary social/non-speech stimulus (e.g., hand clapping). In order to clarify group differences seen here, it will be important to include several subtly varying versions of social stimuli in future studies. As such, it is not possible to say whether group differences observed are related to social stimuli generally or speech sounds specifically. Second, participants in this study had average to above average cognitive abilities, and their performance cannot be assumed to reflect abilities of all persons with autism. In fact, significant correlations here between the TWI for the speech condition and FSIQ indicate that the TWI might vary differently in individuals with autism and intellectual disability.

In this study, we did not assess for effects of adaptation. Evidence is emerging that for typical development, TWIs for multiple types of tasks can be affected by repeated exposure to asynchronous stimuli (Fujisaki, Shimojo, Kashino, & Nishida, 2004; J. Navarra et al., 2005; Vatakis, Navarra, Soto-Faraco, & Spence, 2007; Vatakis et al., 2008), and that this type of adaptation may be different or absent in individuals with autism (Noel et al., 2016). As such, it will be important for future studies to determine if adaptation to repeated asynchronous stimuli varies for speech and object stimuli in autism or in typical development.

Finally, this study used an adaptive paradigm as opposed to the method of constant stimuli used frequently in the temporal synchrony literature. The main advantage of an adaptive

paradigm is that it can be completed much more quickly, which is especially useful for tasks being used in a clinical setting, with young children, or with children with developmental disabilities. It also may reduce the effects of stimulus adaptation, which increase with repeated exposure to asynchrony (Jordi Navarra, García-Morera, & Spence, 2012). However, this method does result in increasing the trials given at or around the inflection point for each individual's sigmoid function relating accuracy to temporal asynchrony and in fewer trials at the extreme ends of the function. Given that this paradigm varies from the standard method of constant stimuli, it will be important for future studies to investigate any effects of adaptive versus constant stimuli methods on temporal synchrony detection both in typical and atypical development.

Implications

Audiovisual speech integration deficits in autism likely lead to real-life impairments in hearing and understanding human speech. There are studies showing that training in particular aspects of audiovisual integration (e.g., lipreading (J. H. Williams, Massaro, Peel, Bosseler, & Suddendorf, 2004), temporal synchrony detection (Powers, Hillock, & Wallace, 2009)) might improve audiovisual speech integration for individuals with ASD. The present study implies that features of audiovisual speech integration specific to social stimuli, such as biological motion perception and attention to faces, may prove to be useful targets for training.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was funded by an NIDCD Individual Predoctoral NRSA (F31 DC010769), with additional funding through the following grants: NIDCD R01 DC009439, NIDCD R21 DC011094. We thank Rafael Klorman for support with statistical analyses and Ashley Wilson for her assistance in stimulus development. We are also grateful to all of the families who invested their time to contribute to this research.

References

- Achenbach T, Rescorla L, McConaughy S, Pecora P, Wetherbee K, Ruffle T. Child behavior Checklist for Ages 6-18. Burlington, VT: 2001.
- Arnold P, Hill F. Bisenory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology*. 2001; 92(2):339–355. DOI: 10.1348/000712601162220
- Baart M, Vroomen J, Shaw K, Bortfeld H. Degrading phonetic information affects matching of audiovisual speech in adults, but not in infants. *Cognition*. 2014; 130(1):31–43. DOI: 10.1016/j.cognition.2013.09.006 [PubMed: 24141035]
- Bebko JM, Weiss JA, Demark JL, Gomez P. Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. 2006; 47(1):88–98. DOI: 10.1111/j.1469-7610.2005.01443.x
- Boersma P, Weenink D. Praat: doing phonetics by computer. 2009. Retrived from <http://www.praat.org/>: Glot International
- Brandwein A, Foxe J, Butler J, Frey HP, Bates J, Shulman L, Molholm S. Neurophysiological Indices of Atypical Auditory Processing and Multisensory Integration are Associated with Symptom

- Severity in Autism. *Journal of Autism and Developmental Disorders*. 2014; :1–15. DOI: 10.1007/s10803-014-2212-9
- Brandwein AB, Foxe JJ, Butler JS, Russo NN, Altschuler TS, Gomes H, Molholm S. The development of multisensory integration in high-functioning autism: High-density electrical mapping and psychophysical measures reveal impairments in the processing of audiovisual inputs. 2013;6.
- Collignon O, Charbonneau G, Peters F, Nassim M, Lassonde M, Lepore F, et al. Bertone A. Reduced multisensory facilitation in persons with autism. *Cortex*. 2013; 49(6):1704–1710. DOI: 10.1016/j.cortex.2012.06.001 [PubMed: 22818902]
- Constantino J, Gruber CP. *Social Responsiveness Scale (SRS)*. Los Angeles: Western Psychological Services; 2005.
- de Boer-Schellekens L, Eussen M, Vroomen J. Diminished sensitivity of audiovisual temporal order in autism spectrum disorder. *Frontiers in Integrative Neuroscience*. 2013; 7
- de Boer-Schellekens L, Keetels M, Eussen M, Vroomen J. No evidence for impaired multisensory integration of low-level audiovisual stimuli in adolescents and young adults with autism spectrum disorders. *Neuropsychologia*. 2013; 51(14):3004–3013. DOI: 10.1016/j.neuropsychologia.2013.10.005 [PubMed: 24157536]
- Dick AS, Solodkin A, Small SL. Neural development of networks for audiovisual speech comprehension. *Brain and Language*. 2010; 114(2):101–114. DOI: 10.1016/j.bandl.2009.08.005 [PubMed: 19781755]
- Diederich A, Colonius H. Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception and Psychophysics*. 2004; 66(8):1388–1404. DOI: 10.3758/bf03195006 [PubMed: 15813202]
- Eskelund K, MacDonald EN, Andersen TS. Face configuration affects speech perception: Evidence from a McGurk mismatch negativity study. *Neuropsychologia*. 2015; 66(0):48–54. DOI: 10.1016/j.neuropsychologia.2014.10.021 [PubMed: 25447378]
- Foss-Feig J, Kwakye LD, Cascio CJ, Burnette CP, Kadivar H, Stone WL, Wallace MT. An extended multisensory temporal binding window in autism spectrum disorders. *Experimental Brain Research*. 2010; 203(2):381–389. [PubMed: 20390256]
- Foxe JJ, Molholm S, Del Bene VA, Frey HP, Russo NN, Blanco D, et al. Ross LA. Severe Multisensory Speech Integration Deficits in High-Functioning School-Aged Children with Autism Spectrum Disorder (ASD) and Their Resolution During Early Adolescence. *Cerebral Cortex*. 2013; doi: 10.1093/cercor/bht213
- Fujisaki W, Shimojo S, Kashino M, Nishida S. Recalibration of audiovisual simultaneity. *Nature Neuroscience*. 2004; 7(7):773–778. [PubMed: 15195098]
- 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(4):491–497. [PubMed: 19207623]
- Hubbard AL, McNealy K, Scott-Van Zeeland AA, Callan DE, Bookheimer SY, Dapretto M. Altered integration of speech and gesture in children with autism spectrum disorders. *Brain Behav*. 2012; 2(5):606–619. DOI: 10.1002/brb3.81 [PubMed: 23139906]
- Hyde DC, Jones BL, Porter CL, Flom R. Visual stimulation enhances auditory processing in 3-month-old infants and adults. *Developmental Psychobiology*. 2010; 52(2):181–189. DOI: 10.1002/dev.20417 [PubMed: 20014224]
- Irwin JR, Tornatore LA, Brancazio L, Whalen DH. Can children with autism spectrum disorders “hear” a speaking face? *Child Development*. 2011; 82(5):1397–1403. [PubMed: 21790542]
- Kikuchi Y, Senju A, Akechi H, Tojo Y, Osanai H, Hasegawa T. Atypical disengagement from faces and its modulation by the control of eye fixation in children with autism spectrum disorder. *J Autism Dev Disord*. 2011; 41(5):629–645. DOI: 10.1007/s10803-010-1082-z [PubMed: 20714799]
- Knowland VCP, Mercure E, Karmiloff-Smith A, Dick F, Thomas MSC. Audio-visual speech perception: a developmental ERP investigation. *Developmental Science*. 2014; 17(1):110–124. DOI: 10.1111/desc.12098 [PubMed: 24176002]
- Kushnerenko E, Tomalski P, Ballieux H, Potton A, Birtles D, Frostick C, Moore DG. Brain responses and looking behavior during audiovisual speech integration in infants predict auditory speech comprehension in the second year of life. *Frontiers in Psychology*. 2013; 4

- Kwakye LD, Foss-Feig JH, Cascio CJ, Stone WL, Wallace MT. Altered auditory and multisensory temporal processing in autism spectrum disorders. *Frontiers in Integrative Neuroscience*. 2011; 4:129. [PubMed: 21258617]
- Lewkowicz DJ. Infant Perception of Audio-Visual Speech Synchrony. *Developmental Psychology*. 2010; 46(1):66–77. DOI: 10.1037/a0015579 [PubMed: 20053007]
- Lewkowicz DJ, Flom R. The audiovisual temporal binding window narrows in early childhood. *Child Development*. 2014; 85(2):685–694. DOI: 10.1111/cdev.12142 [PubMed: 23888869]
- Lord C, Rutter M, Di Lavore PC, Risi S. Autism diagnostic observation schedule (ADOS). Los Angeles: Western Psychological Services; 2000.
- MacLeod A, Summerfield Q. Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*. 1987; 21(2):131–141. [PubMed: 3594015]
- Megnin O, Flitton A, Jones CR, de Haan M, Baldeweg T, Charman T. Audiovisual speech integration in autism spectrum disorders: ERP evidence for atypicalities in lexical-semantic processing. *Autism Res*. 2012; 5(1):39–48. DOI: 10.1002/aur.231 [PubMed: 22162387]
- Mongillo E, Irwin J, Whalen DH, Klaiman C, Carter A, Schultz R. Audiovisual Processing in Children with and without Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*. 2008; 38(7):1349–1358. DOI: 10.1007/s10803-007-0521-y [PubMed: 18307027]
- Navarra J, García-Morera J, Spence C. Temporal adaptation to audiovisual asynchrony generalizes across different sound frequencies. *Frontiers in Psychology*. 2012; 3doi: 10.3389/fpsyg.2012.00152
- Navarra J, Vatakis A, Zampini M, Soto-Faraco S, Humphreys W, Spence C. Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Cognitive Brain Research*. 2005; 25(2):499–507. DOI: 10.1016/j.cogbrainres.2005.07.009 [PubMed: 16137867]
- Noel JP, De Nier MA, Stevenson R, Alais D, Wallace MT. Atypical rapid audio-visual temporal recalibration in autism spectrum disorders. *Autism Research*. 2016; :n/a–n/a. DOI: 10.1002/aur.1633
- Odegaard B, Shams L. The Brain's Tendency to Bind Audiovisual Signals Is Stable but Not General. *Psychol Sci*. 2016; 27(4):583–591. DOI: 10.1177/0956797616628860 [PubMed: 26944861]
- Odegaard B, Shams L. The Brain's Tendency to Bind Audiovisual Signals Is Stable but Not General. *Psychological Science*. 2016; 27(4):583–591. DOI: 10.1177/0956797616628860 [PubMed: 26944861]
- Patten E, Watson LR, Baranek GT. Temporal Synchrony Detection and Associations with Language in Young Children with ASD. *Autism Research and Treatment*. 2014.
- Powers AR, Hillock AR, Wallace MT. Perceptual Training Narrows the Temporal Window of Multisensory Binding. *Journal of Neuroscience*. 2009; 29(39):12265–12274. DOI: 10.1523/jneurosci.3501-09.2009 [PubMed: 19793985]
- Reisberg D, McLean J, Goldfield A. Easy to hear but hard to understand: A speechreading advantage with intact auditory stimuli. In: Dodd B, Campbell R, editors *Hearing by eye: the psychology of lipreading*. London: Erlbaum; 1987. 97–113.
- Rutter M, Bailey AJ, Lord C. Social Communication Questionnaire. Los Angeles, CA: Western Psychological Services; 2004.
- Rutter M, LeCouteur A, Lord C. Autism diagnostic Interview-Revised (ADI-R). Los Angeles: Western Psychological Services; 2003.
- Schwartz JL, Berthommier F, Savariaux C. Seeing to hear better: evidence for early audio visual interactions in speech identification. *Cognition*. 2004; 93(2):B69–78. DOI: 10.1016/j.cognition.2004.01.006 [PubMed: 15147940]
- Shams L, Kamitani Y, Shimojo S. Illusions: What you see is what you hear. *Nature*. 2000; 408(6814):788–788. [PubMed: 11130706]
- Smith EG, Bennetto L. Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry*. 2007; 48(8):813–821. DOI: 10.1111/j.1469-7610.2007.01766.x [PubMed: 17683453]
- Spence C, Squire S. Multisensory integration: Maintaining the perception of synchrony. *Current Biology*. 2003; 13(13):R519–R521. [PubMed: 12842029]

- Stevenson RA, Siemann JK, Schneider BC, Eberly HE, Woynaroski TG, Camarata SM, Wallace MT. Multisensory temporal integration in autism spectrum disorders. *Journal of Neuroscience*. 2014; 34(3):691–697. [PubMed: 24431427]
- Stevenson RA, Wallace MT. Multisensory temporal integration: task and stimulus dependencies. *Experimental Brain Research*. 2013; 227(2):249–261. [PubMed: 23604624]
- Stevenson RA, Zemtsov RK, Wallace MT. Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology: Human Perception and Performance*. 2012; 38(6):1517–1529. [PubMed: 22390292]
- Van der Burg E, Olivers CN, Theeuwes J. The attentional window modulates capture by audiovisual events. *PLoS One*. 2012; 7(7):e39137.doi: 10.1371/journal.pone.0039137 [PubMed: 22808027]
- van der Smagt MJ, van Engeland H, Kemner C. Brief report: can you see what is not there? low-level auditory-visual integration in autism spectrum disorder. *Journal of Autism and Developmental Disorders*. 2007; 37(10):2014–2019. [PubMed: 17273934]
- van Eijk RLJ, Kohlrausch A, Juola JF, van de Par S. Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Perception and Psychophysics*. 2008; 70(6):955–968. DOI: 10.3758/pp.70.6.955 [PubMed: 18717383]
- van Wassenhove V, Grant KW, Poeppel D. Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102(4):1181–1186. DOI: 10.1073/pnas.0408949102 [PubMed: 15647358]
- van Wassenhove V, Grant KW, Poeppel D. Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*. 2007; 45(3):598–607. [PubMed: 16530232]
- Vatakis A, Maragos P, Rodomagoulakis I, Spence C. Assessing the effect of physical differences in the articulation of consonants and vowels on audiovisual temporal perception. *Frontiers in integrative neuroscience*. 2012; 6:71–71. [PubMed: 23060756]
- Vatakis A, Navarra J, Soto-Faraco S, Spence C. Temporal recalibration during asynchronous audiovisual speech perception. *Experimental Brain Research*. 2007; 181(1):173–181. DOI: 10.1007/s00221-007-0918-z [PubMed: 17431598]
- Vatakis A, Navarra J, Soto-Faraco S, Spence C. Audiovisual temporal adaptation of speech: temporal order versus simultaneity judgments. *Experimental Brain Research*. 2008; 185(3):521–529. DOI: 10.1007/s00221-007-1168-9 [PubMed: 17962929]
- Vatakis A, Spence C. Audiovisual synchrony perception for music, speech, and object actions. *Brain Research*. 2006; 1111:134–142. DOI: 10.1016/j.brainres.2006.05.078 [PubMed: 16876772]
- Vatakis A, Spence C. Investigating the effects of inversion on configural processing with an audiovisual temporal-order judgment task. *Perception*. 2008; 37(1):143–160. DOI: 10.1068/p5648 [PubMed: 18399253]
- Vatakis A, Spence C. Audiovisual Temporal Integration for Complex Speech, Object-Action, Animal Call, and Musical Stimuli. In: Kaiser J, Naumer MJ, editors *Multisensory Object Perception in the Primate Brain*. Springer; New York: 2010. 95–121.
- Williams DL, Minshew NJ, Goldstein G. Further understanding of complex information processing in verbal adolescents and adults with autism spectrum disorders. *Autism*. 2015; 19(7):859–867. DOI: 10.1177/1362361315586171 [PubMed: 26019307]
- Williams JH, 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. DOI: 10.1016/j.ridd.2004.01.008 [PubMed: 15541632]
- Woynaroski TG, Kwakye LD, Foss-Feig JH, Stevenson RA, Stone WL, Wallace MT. Multisensory speech perception in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2013; 43(12):2891–2902. [PubMed: 23624833]
- Zion Golumbic E, Cogan GB, Schroeder CE, Poeppel D. Visual input enhances selective speech envelope tracking in auditory cortex at a “cocktail party”. *J Neurosci*. 2013; 33(4):1417–1426. DOI: 10.1523/jneurosci.3675-12.2013 [PubMed: 23345218]

Highlights

- We measured sensitivity to audiovisual asynchrony for speech and object stimuli
- Controls showed similar sensitivity to the two, carefully matched stimulus types
- Adolescents with autism showed higher sensitivity to asynchrony in audiovisual speech
- Higher sensitivity to asynchronous speech was associated with autism severity

Figure 1. Video presentation for the speech and object conditions

Figure 2. Temporal window of Integration in milliseconds for the object (rubber ball) condition and three versions of the speech condition in young adult pilot subjects. Error bars show Standard Error of the Mean.

Figure 3.
Temporal window of integration in milliseconds by condition (speech, object) and group (autism, control). Error bars show Standard Error of the Mean.

Figure 4. Scatterplot relating temporal window of integration for the social condition to autism symptom severity (T-score on the Social Responsiveness Scale).

Table 1
Characteristics of the autism and comparison groups

	Autism Group <i>M (SD)</i> [range]	Comparison Group <i>M (SD)</i> [range]	<i>F</i>	<i>p</i>
<i>n</i>	20 (18 male)	20 (18 male)		
Age (years)	14.55 (2.18) [11.25-17.99]	13.96 (1.94) [11.1-17.41]	.83	.37
Verbal IQ ^a	113.6 (16.04) [87-146]	117.55 (7.42) [103-130]	1.0	.32
Performance IQ ^a	111.45 (15.57) [80-148]	109.25 (10.94) [91-131]	.27	.61

^aEstimated from a short form of the Wechsler Intelligence Scale for Children, 4th ed., or the Wechsler Adult Intelligence Scale, 4th Ed.