International Journal of Engineering Sciences Paradigms and Researches (IJESPR)
Volume 47, Issue 01, Quarter 01 (January to March 2018)
An Indexed and Referred Journal with Impact Factor: 2.80
ISSN (Online): 2319-6564
www.ijesonline.com

Audiovisual Speech Integration and lip Reading in Autism

Dr. Satyabrata Dash[1], Sunil Panigrahi[2] and Archana Panda[3]

[1,3]Associate Professor, Department of Computer Science Engineering, Gandhi Institute For Technology (GIFT), Bhubaneswar
[2]Assistant Professor, Department of Computer Science Engineering, Gandhi Engineering College, Bhubaneswar

Publishing Date: 2nd March, 2018

Abstract

Background: During speech perception, the ability to integrate auditory and visual information causes speech to sound louder and be more intelligible, and leads to quicker processing. This integration is important in early language development, and also continues to affect speech comprehension throughout the lifespan. Previous research shows that individuals with autism have difficulty integrating information, especially across multiple sensory domains. Methods: In the present study, audio-visual speech integration was investigated in 18 adolescents with high-functioning autism and 19 well-matched adolescents with typical development using a speech in noise paradigm. Speech reception thresholds were calculated for auditory only and audiovisual matched speech, and lip-reading ability was measured. Results: Compared to individuals with typical development, individuals with autism showed less benefit from the addition of visual information in audiovisual speech perception. We also found that individuals with autism were significantly worse than those in the comparison group at lip-reading. Hierarchical regression demonstrated that group differences in the audiovisual condition, while influenced by auditory perception and especially by lip-reading, were also attributable to a unique factor, which may reflect a specific deficit in audiovisual integration. Conclusions: Combined deficits in audiovisual speech integration and lip-reading in individuals with autism are likely to contribute to ongoing difficulties in speech comprehension, and may also be related to delays in early language development. Keywords: Speech reception threshold, speech in noise, audiovisual speech integration, autism. Abbreviations: SNR: speech to noise ratio; SRT: speech reception threshold.

Keywords: *Audiovisual Speech Integration and lip Reading in Autism.*

Introduction

One of the hallmarks of autism is impairment in communication, which can range from severe delays in language development to relatively intact language accompanied by problems with functional communication (Tager Flusberg, Paul, & Lord, 2005). Perception of speech is a particular aspect of communication that may be altered in autism, and further investigation of this domain may shed light on the development of communication deficits. For example, understanding a person's speech often requires that listeners integrate information from the speaker's voice, lips, face, and body. This audio-visual speech integration increases identification and comprehension of the information being communicated (Calvert, Brammer, & Iverson, 1998). However, individuals with autism often show deficits in crossmodal integration (Iarocci & McDonald, 2006), which might put them at a disadvantage during speech perception.

Audiovisual speech integration in typical development

Audiovisual speech perception has primarily been investigated in typical development using the 'McGurk effect' paradigm (McGurk & Macdonald, 1976). In this paradigm, unisyllabic or disyllabic, non-word utterances are presented either visually (i.e., individual sees model's lips move without sound), auditory (i.e., individual hears utterance without visual

information), or audio visually (i.e., individual hears utterance and sees model's lips move). Results from McGurk's initial work showed an interesting effect when mismatching auditory and visual stimuli were presented together: the reported percept sometimes represented a fusion between the auditory and visual modes (e.g., auditory/ba/and visual/ga/are perceived as/da/). Studies using the McGurk effect have shown that multisensory speech integration is mandatory and unmodulated by attention (Soto-Faraco, Navarra, & Alsius, 2004). In addition, Driver (1996) used the ventriloquist effect to show that audiovisual information guides atten-tion and also that it is processed prior to attentional modulation. Studies using the McGurk effect have also shown that audiovisual speech perception is present in very young infants and plays an important role in speech production (Desjardins, Rogers, & Werker, 1997; Patterson & Werker, 1999). Audio-visual speech continues to assist older child and adult listeners in comprehension of speech in daily social situations.

Audiovisual integration in autism

There is evidence that individuals with autism have difficulty integrating information across auditory and visual modes (see Iarocci & McDonald, 2006 for a review), including matching voices to faces (Boucher, Lewis, & Collis, 1998; Loveland et al., 1995), forming cross modal associations between sound beeps and light flashes (Martineau et al., 1992), discriminating temporal synchrony of audiovisual speech (Bebko, Weiss, Demark, & Gomez, 2006), and blending auditory and visual speech (Williams, Massaro, Peel, Bosseler, & Suddendorf, 2004). However, deficits in either auditory or visual per-ception alone might account for differences in multimodal integration in autism (Ceponiene et al., 2003; Williams et al., 2004).

Until recently, relatively little research has explored audiovisual integration of speech in autism. Investigation of the McGurk effect has shown that children and adolescents with autism report fewer fusions than typical children, reflecting that children with autism are less likely to take the non-matching, visual syllable into account during speech perception (de Gelder, Vroomen, & van der Heide, 1991). Williams and colleagues (2004) replicated this find-ing, and examined the contributions of unisensory components to this deficit. When visual accuracy (i.e., lipreading) was controlled for, individuals with autism were no longer significantly worse than those in the comparison group in audiovisual integration. Thus, the essential step to be taken in this field re-quires characterization of both unisensory and multisensory speech perception in autism using ecologically valid stimuli.

Speech in noise paradigm

Ecological validity can be improved by employing a paradigm that is correlated with everyday perceptual challenges, such as the speech in noise paradigm. Speech is often heard in varying levels of background noise (e.g., at a loud party), requiring listeners to filter the background noise out of the speech. Individuals with autism appear to have a relative weakness in understanding speech presented in background noise compared to individuals with typical develop-ment (Alcantara, Weisblatt, Moore, & Bolton, 2004).

While typical individuals are better at under-standing speech in noise, they are also able to use visual information to enhance the auditory signal of speech presented in noise (Schwartz, Berthommier, & Savariaux, 2004). The addition of visual information does not simply add unimodal information; the visual information actually enhances the individual's ability to perceive the auditory information. Thus, the ability to use visual information when listening to speech (i.e., audiovisual speech perception) is linked to the ability to perceive and comprehend speech in noise (Rudmann, McCarley, & Kramer, 2003). If, indeed, individuals with autism experience deficits in both audiovisual speech perception and speech in noise perception, these deficits could produce an additive, deleterious effect on comprehension in everyday situations.

In the current study, we investigated whether individuals with autism were capable of

using visual information to enhance an auditory signal embedded in background noise. Based on previous research, we predicted that individuals with autism would be worse at processing audiovisual speech in back-ground noise, and that these deficits in integration would not be explained by auditory or visual processing deficits alone.

Methods

Participants

Participants were 18 adolescents with autism and 19 adolescents with typical development matched by group on chronological age, gender, Full Scale IQ, and the Receptive Language Index (RLI) from the Clinical Evalu-ation of Language Essentials, 4th Ed. (CELF-4; see Table 1). Since there is a small verbal memory compon ent in the speech in noise paradigm, we also adminis- tered the Recalling Sentences subtest from the CELF-4 to ensure that all individuals were able to recall sentences at least as long as those presented in the stimuli.

Diagnoses of autism were confirmed in the autism group and ruled out in the comparison group with a combination of the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 1999) and the Autism Diagnostic Interview-Revised (ADI-R; Rutter, Le Couteur, & Lord, 2003). Individuals with autism were excluded if they had a diagnosis of a gen-etic syndrome (e.g., fragile X syndrome) or a definable postnatal etiology for their developmental symptoms (e.g., head trauma). Individuals in the comparison group were excluded if there were concerns about learning disabilities, mental retardation, language de-lays, head trauma, or other psychiatric conditions, or if there were concerns about autism spectrum disorders in their first- or second-degree relatives.

All participants were native English speakers, and had normal or corrected vision and hearing acuity within the range required for speech perception. Visual acuity was evaluated for each eye using the Snellen chart. Auditory acuity was measured separately for each ear with an audiometer (Maico MA32 Advanced Diagnostic Audiometer). All participants were capable of hearing tones represented at frequencies of 1000– 4000 hertz and loudness of 25 decibels.

This research was approved by the University of Rochester's Research Subjects Review Board. Prior to testing, written informed consent was obtained from parents and from individuals aged 18 or 19; written assent was also obtained from younger adolescents.

Measures

Stimuli were short sentences (5–7 words long, 4 sec-onds in duration) containing three key words (e.g., The cat jumped over the fence). A person's response was considered correct when he or she was able to correctly report all three key words for a trial. Sentences were presented in Auditory Only and Audiovisual conditions, and were presented in a speech in noise paradigm. Manipulating the loudness of speech relative to back-ground noise in a systematic way reveals a particular speech to noise ratio (SNR) for each condition. Differ-ences in SNR across conditions reveal the benefit pro-vided by the addition of visual information. In addition, a lipreading condition evaluated perception of visual information presented alone.

We used 48 sentences from Rosenblum, Johnson, and Saldana's (1996) list, which was composed for American English listeners and designed to provide a balance of lexical and semantic clarity across the sentences. Similar sentence lists have been used to measure speech in noise perception in children, including those with specific language impairments (Stollman, van Velzen, Simkens, Snik, & van den Broek, 2003, 2004).

The speakers for the target sentences included five females between the ages of 23 and 28 years. Previous research has shown that younger women are the easiest to lipread and understand auditorily (Bench, Daly, Doyle, & Lind, 1995). We used several speakers to minimize learning across the session. Each of the speakers was recorded speaking 9 or 10 sentences in a sound- attenuated chamber

with a professional-quality digital video camera equipped with a unidirectional microphone. Movements of the neck and throat can aid audiovisual speech perception (Thomas & Jordan, 2004), so our speakers wore a black turtleneck to cover these areas.

Background noise

Four additional females were re-corded reading excerpts from children's books. All articulatory sounds were low-pass filtered out from the speech by removing frequencies containing segmental content with Praat, a computer phonetics program (Boersma & Weenink, 2006). The filtered streams were then overlapped in Final Cut Pro and divided into 48, 4-second blocks. Since the type of information typically available in background noise (e.g., temporal and spectral dips) differentially affects speech intelligibility for individuals with autism (Alcantara et al., 2004), combining multiple speech streams and removing seg-mental content from articulatory sounds yielded back-ground noise that was equally difficult for both groups. Background noise was presented at 70 decibels across all stimuli and conditions.

Pilot study to determine final sentence lists

We piloted all 48 sentences with 16 young adults to deter-mine the average auditory SNR for each sentence when presented without visual information. Each sentence was presented first at the quietest, or most difficult level (i.e., 40 dB), resulting in an SNR of)30. The speech stream volume was then raised by 2 dB steps until the individual could accurately report all three key words. This allowed us to determine the average SNR at which each of the 48 sentences was intelligible across subjects.

To determine lipreadability, the 48 sentences were shown to a different group of 15 young adults, who were asked to report (or guess) any words they could lipread. The average number of correctly identified key words yielded a mean lipreading accuracy score for each sentence.

We used our piloting data to balance our five speakers across conditions and also to ensure homogeneity of our lists (see Table 2). We first constructed the list for the Lipreading condition from the 12 sentences with the highest average lipreadability scores. This resulted in a list with a high probability of evoking lipreading abilities for all individuals and a low probability of a floor effect. The remaining three lists of 12 sentences each were used in the Auditory Only and Audiovisual conditions, and were matched in terms of average SNR and lipreading accuracy scores. Three presentation sets were constructed, distributing the three balanced lists across the two conditions, so that each person received a different list for each condition. The presentation sets were distributed equally across groups; later analyses revealed no effects of presentation set on performance.

Discussions

The results of this study provide evidence of an audiovisual integration impairment in autism. While the comprehension of speech in noise of both groups improved with the addition of visual information, this improvement was markedly stronger for indi-viduals with typical development compared to those with autism. We also found that individuals with autism were significantly less skilled on a lipreading task that was closely matched to the audiovisual paradigm. Regression analyses showed that even after accounting for the variance due to unisensory factors, the group differences in audiovisual speech remained.

References

[1] Alcantara, J.I., Weisblatt, E.J.L., Moore, B.C.J., & Bolton, P.F. (2004). Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome. Journal of Child Psychology and Psychiatry, 45, 1107–1114.

[2] Bebko, J.M., Weiss, J.A., Demark J.L., & Gomez, 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, 88–98.

[3] Bench, J., Daly, N., Doyle, J., & Lind, C. (1995). Choosing talkers for the BKB/A Speechreading test: A procedure with observations on talker age and gender. British Journal of Audiology, 29, 172–187.

[4] Boddaert, N., Chabane, N., Belin, P., Bourgeois, M., Royer, V., Barthelemy, C., Mouren-Simeoni, M.-C., Philippe, A., Brunelle, F., Samson, Y., & Zilbovicius, M. (2004). Perception of complex sounds in autism: Abnormal auditory cortical processing in children. American Journal of Psychiatry, 161, 2117–2120.

[5] Boersma, P., & Weenink, D. (2006). Praat: Doing phonetics by computer (Version 4.4.30).

[6] Boucher, J., Lewis, V., & Collis, G. (1998). Familiar face and voice matching and recognition in children with autism. Journal of Child Psychology and Psychiatry, 39, 171–181.

[7] Bradlow, A.R., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning disabilities: Sentence perception in noise. Journal of Speech, Language, and Hearing Research, 46, 80–97.

[8] Brambilla, P., Hardan, A.Y., di Nemi, S.U., Caverzasi, E., Soares, J.C., Perez, J., & Barale, F. (2004). The functional neuroanatomical of autism. Functional Neurology, 19, 9–17.

[9] Calvert, G.A. (2001). Cross modal processing in the human brain: Insights from functional neuroimaging studies. Cerebral Cortex, 11, 1110–1123.

[10] Calvert, G.A., Brammer, M.J., & Iverson, S.D. (1998). Cross modal identification. Trends in Cognitive Sciences, 2, 247–253.