5-2015

Speech Reading Training and Audio-Visual Integration in a Child with Autism Spectrum Disorder

Kara L. Saulsbury
University of Arkansas, Fayetteville

Speech Reading Training and Audio-Visual Integration in a Child with Autism Spectrum Disorder

Kara Saulsbury

University of Arkansas

Abstract

Children with Autism Spectrum Disorder (ASD) typically have deficits in communication abilities. Deficits include social, linguistic, and pragmatic difficulties and difficulties in the ability to perceive and integrate audiovisual (AV) stimuli. It is also common for those with ASD to have weaker speech reading skills as compared to typically developing, age-matched peers. Speech reading skills are known to enhance speech perception in naturalistic, noisy environments. In children with ASD, the combination of both poor AV integration and poor speech reading is thought to have significant effects on vocabulary acquisition. Studies have demonstrated that speech reading training can significantly enhance syllable discrimination in noise. However, it has not been investigated whether speech reading training could be generalized to more naturalistic stimuli such as words and in noisy environments.

The purpose of the current study was to implement speech training in a child with ASD at the word level using a multiple baseline, changing criterion design. The child identified words in increasingly higher levels of background noise. During the baseline measures, AV speech was presented at a Signal to Noise Ratio, SNR, of 0dB. At the SNR of 0dB, both the speech and noise signals were equal. Speech reading training was implemented at the SNR of +4dB. At +4dB, the speech signal was louder than noise signal, making the task less challenging. The child was asked to watch and listen to the AV speech and choose what word he heard from a four choice list.

The participant showed increases in receptive language processing over the course of the four training sessions when compared to the multiple baseline measures. Speech reading training enhanced receptive language processing for words in the SNR of 0dB from the initial pre-training baseline to post-training measure. The results from the study

are consistent with previous findings that demonstrate increases in syllable identification after speech training using AV speech and suggest that such gains may also be trained for words in noisy environments.

Table of Contents

List of Figures

Speech Reading Training and Audiovisual Integration in a Child with Autism Spectrum Disorder

Autism Spectrum Disorder (ASD) include a range of developmental disorders that can cause deficits in social interaction, communication, and behavior. Deficits in social communication and social interactions across multiple contexts are present in a diagnosis of ASD (DSMV-V). Deficits in communication could be in the perception of audiovisual (AV) speech integration. There are many benefits that come about from the integration of audiovisual information that strengthen both language skills and development. The integration of audiovisual information has been found to make speech appear louder and be more intelligible (Sumby & Pollack, 1954). Electrophysiological studies have also shown that visual mouth movements can increase the speed of speech processing (Wassenhove, Grant, & Poeppel, 2005). In noisy environments the use of AV integration can remove the equivalent of up to 20 decibels of environmental noise (Sumby & Pollack, 1954). However, it is unclear how children with ASD make use of combined auditory and visual cues in naturalistic, noisy environments.

Children with ASD have been found to have difficulty integrating auditory and visual stimuli as opposed to typically developing children who are perceptually influenced by the integration of both auditory and visual stimuli (Mongillo, Irwin, Whalen, Klaiman, Carter, Schultz, 2008). This deficit may lead to a language delay in children with ASD (Guiraud, Tomalski, Kushnereko, Ribeiro, Davies, Charman, Elsabbagh, & Johnson, 2012). It has also been found that children with ASD have weaker speech reading skills when compared to age- matched peers (Smith & Bennetto, 2007). The combination of both poor AV and poor speech reading skills may have significant effects on language development (Guiraud, et al., 2012). Current theories on autism suggest children with ASD have multisensory integration deficits (Iarocii and McDonald, 2006). For example the “weak central coherence” theory suggests that children with ASD have impaired abilities in integrating higher level processing of information across a variety of contexts (Frith, 1989). This could be a possible explanation as to why children with ASD have trouble integrating AV stimuli and have poor speech reading abilities. Other theories suggest abnormalities in the brain that cause an impairment in the attention system, specifically in the ability to switch attention between both audio and visual modalities (Ciesielski, Knight, Prince, Harris, & Handmaker, 1995). Whatever the cause, the inability to integrate both audio and visual information could lead to receptive language deficits and limited vocabulary knowledge.

The McGurk Effect, discovered by Harry McGurk and John MacDonald, demonstrates the relationship between hearing and vision in speech. When individuals are shown a clip of a face uttering the syllable [ba] with the speech stimuli of the syllable [ga], typical adults will hear the sound [da] (MacDonald & McGurk, 1976). Children with ASD have difficulty in integrating auditory and visual speech cues and it has been observed that children with ASD do not perceive the McGurk Effect (Carter, Irwin, Klaiman, Mongillo, Schultz, & Whalen, 2008). Carter, et al, (2008) observed that children with ASD integrate both auditory and visual stimuli differently from typically developing peers. In this study, the researchers compared AV processing in children with ASD and without ASD in order to observe how each group used visual stimuli. They found that children with ASD may use the visual information they receive differently than those without ASD. Charman, Davies, Elsabbagh, Guiraud, and Johnson (2012), researched both infants at low risk for ASD and those at high risk for ASD for AV integration. It was found that infants who were at low risk of developing ASD could integrate AV speech and also perceive mismatched auditory and visual stimuli, while infants who were at a high risk had difficulty with this task. The researchers concluded that this inability to accurately integrate AV speech could have negative impacts on the development of vocabulary.

While possible audiovisual deficits may be present, Massaro and Bossler (2003), found children with ASD could be trained in speech reading abilities to improve multimodal integration. Through the use of a computer animated talking head, "Baldi", children with ASD showed improvements in identification of the consonant in consonant-vowel (CV) syllables after being trained in speech reading. This study showed that children with ASD are influenced by speech information in the face and can integrate across modalities. While studies have found that it is possible to train speech reading skills in ideal listening environments, no studies have investigated whether speech reading skills can improve receptive language comprehension in noisy, naturalistic listening environments. Also, little research has been done to examine whether children with ASD would benefit from being taught speech reading skills in therapy to improve AV integration in noisy environments. A preliminary single subject study, conducted at the University of Arkansas, recently demonstrated that training speech reading skills can enhance audio-visual discrimination of syllables in background noise (Parkridge & Bowers, 2014). However, that study did not investigate whether the same procedure might generalize to more naturalistic speech at the word level. As such, the specific aims of the current study were to: 1) enhance receptive language processing for words through the use of speech reading training in background noise and 2) enhance single-word receptive language processing in multitalker background noise (MBN) in a multiple baseline, pretest/posttest design. It was predicted that the child's single receptive language processing skills in noise will increase on a measure of percentage of correct words with the implementation of speech reading skills in noisy environments.

Methods

Participant

One eight year old child with a diagnosis of Autism Spectrum Disorder participated in the study. The child was diagnosed with ASD with symptoms falling in the moderate range through a multidisciplinary team in Northwest Arkansas. The child attended school in the Fayetteville Public School District and spoke American English. The child's mother reported that he had no other diagnosis of a disorder or other condition. Written informed consent, approved by the University of Arkansas Institutional Review Board (IRB), was obtained from his mother.

Stimuli

The stimuli used were 19, one and two syllable, high frequency, nouns. Examples include: "student," "line," "world," "father." The stimuli were recorded on a Cannon HD Vixia M32 video camera with a female speaker in a sound-treated booth. The stimuli were low-pass filtered with a cutoff at 5,000 Hz and were root-mean-square normalized (RMS) at 70dB. Multi-talker babble noise was excised from a standard recording used in audiological testing, the Hearing-in-noise test (Nillson, Soil, & Sullivan, 1994). The multi-talker babble noise was filtered the same way as the speech stimuli with the cutoff at 5,000 Hz and RMS normalized to the same intensity as the speech stimuli. Both the speech stimuli and noise recordings were combined using PRAAT (www.prat.org) on a Dell desktop computer. To train speech reading in a variety of levels of background noise, the speech stimuli were generated using multiple signal-to-noise ratios (SNRs). The SNRs were 0dB and +4dB because those levels have been found to elicit speech discrimination at chance and at levels of ~70% (Binder, Liebenthal, Possing, Medler, & Ward, 2004).

Procedure

The speech training occurred under the supervision of trained speech-language pathologists at the University of Arkansas Speech and Hearing clinic. The speech reading training involved the audio-visual presentation of each word. The baseline was collected at the SNR of 0dB. For training, the SNR of +4dB was used in order to increase the number of correct responses. To assess whether or not training at +4dB generalized to the 0dB SNR, posttesting was conducted using the 0dB SNR stimuli. Words were randomized for each trial block consisting of 19 trials. The child was asked to watch and listen to the recording, then indicate the word being spoken by pointing to the corresponding picture from colored line drawings in a four choice picture array. Foils included other high frequency words. Pictures were chosen because the participant's mother reported that using orthographic representations of the spoken word would likely lead to failure on the task. A time-line representing one trial is displayed in Figure 1. Verbal praise and a token economy system were used as reinforcement. The participant also received gift cards for his participation in the experiment.

Analysis

After each session the accuracy data were charted and analyzed using single subject statistical procedures. The changes in accuracy over the sessions were analyzed using traditional visual inspection and also statistical process control (SPC). First, the data were represented visually via line-graphs. Following visual inspection of the results, the mean and standard deviation (SD) were computed and represented on the line chart. With the SPC analysis, it was predicted that variation would fall within a specific range for a given subject (i.e., 3SD) and data outside of the range of predictable variability would indicate a change that is related to treatment (>3SD). Finally, a measure of percentage change (y2-y1)/y1)*100 was used to assess differences between pre and post-test measures using an SNR of 0dB.

Results

Percentage of correct words identified (% CW) for the pretraining and posttraining for the audio-visual (AV) modality are displayed in Figure 2. Percentage correct responses for perceptual training on using an SNR of +4dB over four training sessions are displayed in Figure 3. Pretraining baseline measures were 42%, 36%, and 32% respectively and were relatively stable with a standard deviation (SD) of 5.5% over the course of three sessions. SPC analysis showed that the participant increased on the measure of %CW in training session three (T3) and training session four (T4) more than 3SD from the mean over the three baselines. As shown in Figure 2, the participant increased the %CW from 36% in the initial training with successive increases reaching a maximum of 89% in the final training session. Percentage change from pretesting to postesting using a SNR of 0dB was 111.90% and the percentage value fell more than +3SD from the mean.

Discussion

The study evaluated the effectiveness of speech reading training with the following specific aims and hypotheses. First, it was expected that training speech reading for common (i.e., high-frequency) nouns would increase the %CW identified in a four forced choice task. Second, it was predicted that receptive language processing in a highly noisy environment in which speech and noise levels were equal would increase from pretest to posttest measures of %CW following speech reading training in MTB noise. As expected, training a participant with ASD was associated with significant increases in %CW over the course of four training sessions over four weeks that were more than 3SD from the multiple baseline measures taken prior to training. Further, in accordance with initial expectations, speech reading training enhanced receptive language processing for words from the initial pre-training baseline to post-training measures in 0dB background noise, suggesting that training resulted in significant increases in AV identification of words. Overall, results are consistent with previous findings that increases in syllable identification after training using AV speech and suggest that such gains may also be trained for words.

Consistent with previous studies investigating AV integration (Mongillo, Irwin, Whalen, Klaiman, Carter, Schultz, 2008), and theories predicting AV integration deficits in children with autism (e.g., weak central coherence), the participant in this study showed low word identification performance throughout multiple baseline sessions and early in the training sessions. This is expected in low SNRs as the acoustic cues associated with segmental units are obscured by the MTB noise. The results from the current study are also consistent with results from previous studies using a training procedure to increase AV integration. Massaro and Bossler (2003) found that children with ASD could be trained in speech reading abilities in order to improve multimodal integration in ideal listening environments for consonant-vowel syllables. A similar single subject study using syllables in multitalker babble noise for speech reading training was found to show that training can significantly enhance audiovisual discrimination of syllables in background noise for a single subject with high functioning autism (Parkridge and Bowers, 2014). The findings in the current study suggest that AV speech perception can be also be trained for words in noisy listening environments for a single subject with high-functioning autism. This is of value because natural listening environments often have varying levels of background noise. The use of word stimuli is of importance as perceptual deficits in integrating multimodal stimuli in those with ASD may have escalating effects on the later development of vocabulary (Charman et al., 2012; Iarocci & McDonald, 2006). This idea suggests that individuals with ASD could miss fundamental perceptual cues in background noise that decrease the acquisition and comprehension of new vocabulary. The findings from both Parkridge and Bowers (2014), and the current study demonstrate that the speech reading training approach used may be implemented into therapy sessions to train for speech perception in noise.

Limitations

Despite the results that support the previously mentioned specific aims, there are several limitations to this single subject case study. This particular study was a single subject design, which limits the generalizability beyond the experimental procedures used in this study and to larger population of high-functioning children with ASD. In regards to the participant, there are several factors that limit generalizability. First, the child was diagnosed with ASD with the level of symptoms falling in the moderate range, which implies that children with more severe symptoms may not benefit from training. Second, the child was motivated and enjoyed the training tasks, which may be a significant aspect of the training results. Third, the child was also engaged in other speech-language therapy at his school during the training period. Consequently, multiple treatment interferences cannot be ruled out as factors influencing the results. Lastly, maturation is a limitation that is common to time-series treatment designs. However, in the current study, with the use of multiple baseline testing and statistical process control, maturation alone cannot explain increases in %CW.

Future Directions

Despite the limitations, findings from this single subject design suggest that AV training can enhance perceptual performance of single words in noisy background environments for a child with ASD. The use of this simple and brief training procedure may easily be implemented into speech-language therapy. Various aspects of the current study could be altered in future studies involving speech reading training and audiovisual integration. Group studies could be tested using the training protocol to better assess generalizability in the population of children ASD. The task could also be altered to involve phrase or sentence contexts as opposed to single words. Finally, as it is always challenging to demonstrate generalization from training to more naturalistic contexts, measurements involving receptive speech processing in conversational contexts should be devised.

References

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th edition). Washingon, DC, American Psychiatric Association.

Binder, J., Liebenthal, E., Possing, E., Medler, D., & Ward B. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience,* 7, 295-301.

Carter, A., Irwin, J., Klaiman, C., Mongillo, E., Schultz, R., & Whalen, D. (2008). Audiovisual processing in children with and without autism spectrum disorders. *Journal of Autism & Developmental Disorders,*38, 1349-1358.

Ciesielski, K., Knight, J., Prince, R., Harris, R., & Handmaker, S. (1995). Event-related cross modal divided attention in autism. *Neuropsychologia.* 33, 225-246.

Frith, U. (1989). Autism: Explaining the enigma. Oxford: Basil Blackwell.

Gibson, E. (1969). Principles of perceptual learning and development. New York: Appleton-Century-Crofts. 537.

Guiraud, J., Tomalski, P., Kushnerenko, E., Ribeiro, H., Davies, K., Charman, T., Elsabbagh, M., & Johnson, M. (2012). Atypical audiovisual speech integration in infants at risk for autism. *PLosOne*, 7(5).

Iarocci, G., & McDonald, J. (2006). Sensory integration and the perceptual experience of persons with autism. *Journal of Autism and Developmental Disorders,* 36, 77-90.

MacDonald, J. & McGurk, H. (1976). Hearing lips and seeing voices. *Nature,* 264, 746-748

Massaro, D., & Bosselerm A., (2003). Perceiving speech by ear and eye: Multimodal integration by children with autism. *Journal on Developmental Learning Disorders,* 7, 111-114.

Mongillo, E., Irwin, J., Whalen, D., Klaiman, C., Carter, A., & Schultz, R. (2008). Audiovisual processing in children with and without autism spectrum disorders. *Journal of Autism and Developmental Disorders.* 38, 1349-1358.

Nillson, M, Soil, S., Sullivan, J. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and noise. *Journal of Accostical Society of America,* 95(2), 1085-1099.

Smith, E., & Bennetto, L. (2007). Audiovisual speech integration and lip-reading in autism. *Journal of Child Psychology and Psychiatry.* 48, 813-821.

Sumby, W., & Pollack, I. (1954). Visual contribution of speech intelligibility in noise. *J. Acoust. Soc. America*, 26, 212-215.

Wassenhove, V., Grant, K., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci*, 102, 1181-1186.

Appendix

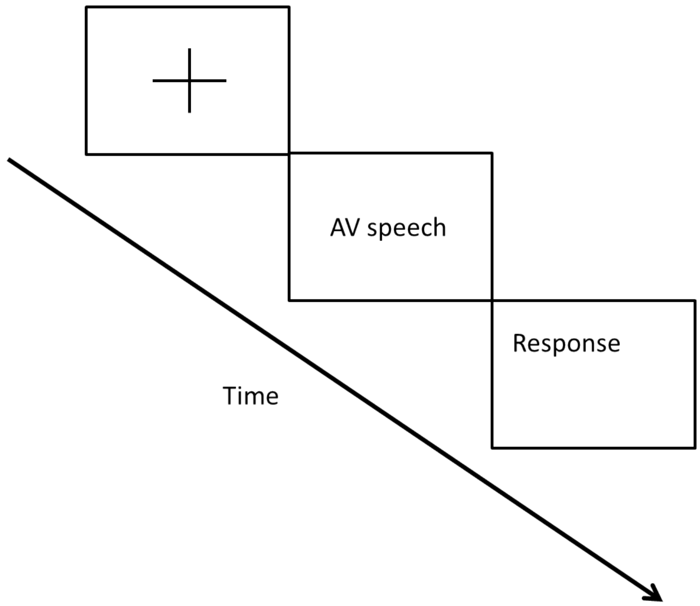

Figure 1: Time-line of one trial

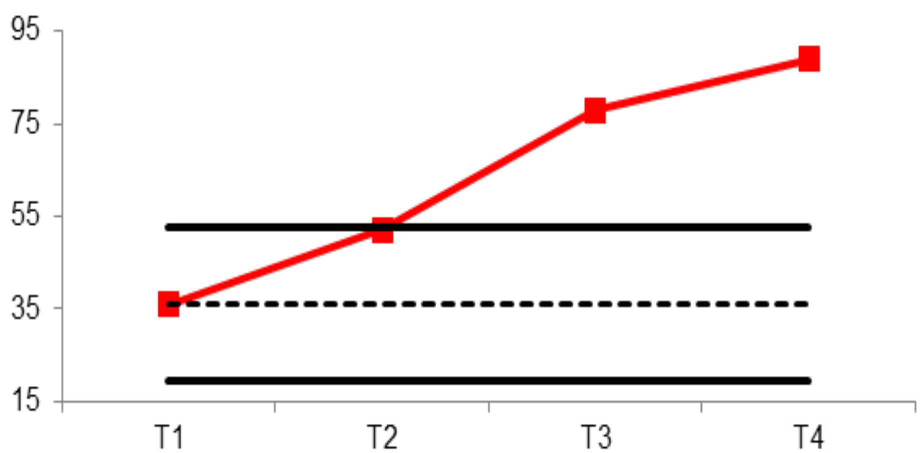

Figure 2: Session-by-session increases in percent correct responses. This chart shows the percentage correct responses for perceptual training in +4dB multitalker babble noise over four training sessions. The mean across three pre-training baselines is shown (dark dotted line) along with the statistical process control lines (solid lines) representing 3SD from the mean.

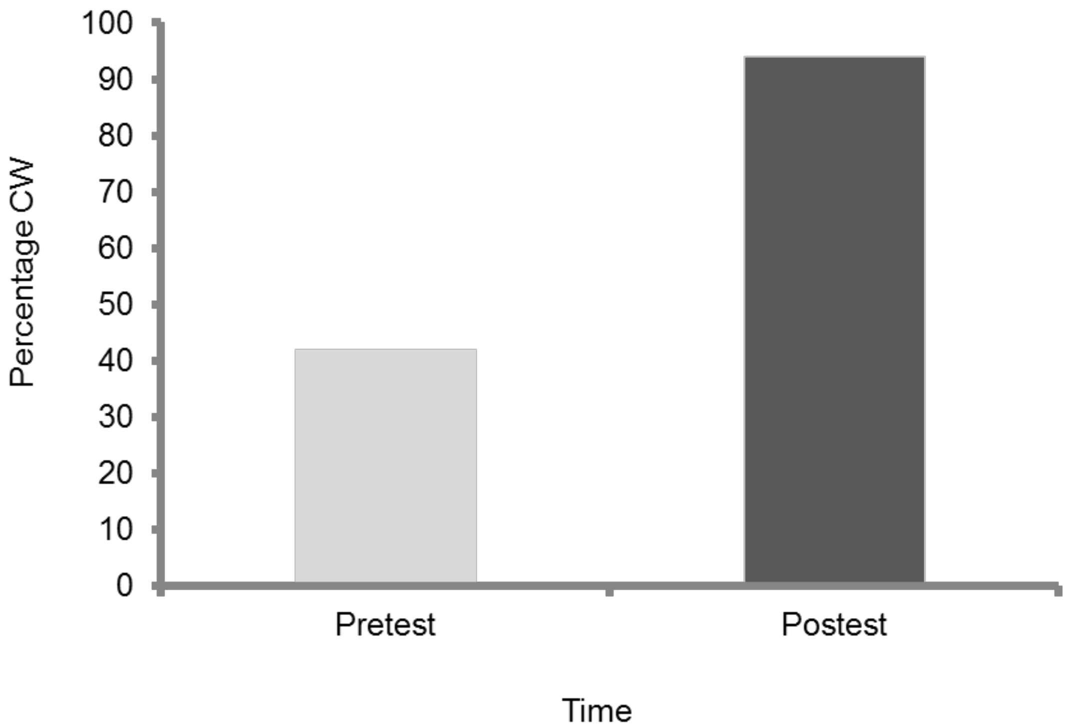

Figure 3: Pre-training and post-training measures. This chart shows pre and post-training percentage of correct responses in the auditory-visual modality.