Audio-visual speech perception training for a child with Autism Spectrum Disorders: A Case Study

April Parkridge

University of Arkansas

Abstract

Children with Autism Spectrum Disorders (ASD) often have difficulty acquiring normal language skills, a deficit that may stem from an inability to integrate audiovisual (AV) information. Audiovisual information is important for language development and enhances language skills. When added to speech, visible facial movements are known to enhance speech perception in noisy environments, removing the equivalent of 20 decibels of auditory background noise. As such, AV integration is critical to the acquisition of an adequate vocabulary in natural, noisy environments. Due to weaker lip reading skills than typically developing children, it has been demonstrated that children with ASD are less able to integrate matched AV speech in noisy environments. Despite these findings, few intervention studies have focused on addressing AV speech perception in children with ASD.

The purpose of the current study is to teach lip reading in a changing criterion, multiple baseline design. During the baseline phase, a child with ASD was asked to identify auditory only, visual only, and bimodal consonant-vowel (CV) syllables (e.g. /bi/ and /di/). McGurk fusion stimuli were also tested in the pre-training and post-training periods. Following the initial baseline phase, the child was taught to read lips in increasingly progressively higher multitalker background noise in five stages, with a the criterion for moving to the next stage set at 100% accuracy. The expected outcome was that the child would demonstrate the ability read lips in background noise with greater accuracy compared to baseline.

The participant showed increases in all modalities from pre to post-treatment. In addition, there was a consistent increase in correct responses over 4 of the 5 training sessions. Although the data shows general increase throughout the study, there was no change for the McGurk effect. The results of the study suggest that it is possible to train attention to the mouth to

enhance perceptual performance using a changing criterion design. Findings are discussed with respect to previous studies of AV perception in ASD and future directions.

Acknowledgements

To my advisor, Mr. Aslin: Thank you for convincing me to apply to the Honors College and stick with it. It's been one of the best decisions I've ever made.

To Dr. Frazier: Thank you for all your expertise, advice, and finding our participant. None of this would be possible without your cooperation.

To Dr. Hagstrom: Thank you for your patience and guidance through the first half of this process and for answering all my questions and emails. I would not have known where to begin this whole thing without you. Also, thank you for always pushing me to step outside of my comfort zone.

To Dr. Bowers: Thank you for all the long hours and hard work. You never failed to answer my ridiculous questions or most random emails, even on paternity leave. Without you, I would have been lost throughout this entire process. Thank you for your support and guidance. I couldn't have asked for a better mentor.

To my wonderful friends and family: Thank you for always being there and supporting me. You keep me going in the most difficult of times.

To my mom: Thanks for all your love and support throughout this journey. You've done your best to keep me sane from 300 miles away, and I love and thank you for that.

Lastly, to my loving fiancé: Thank you for always being there, for supporting me, listening to me rant, taking my stressed out phone calls, and calming me down when I'm freaking out. I look forward to you doing all these things for years to come and I can't wait to see where the future takes us.

Table of Contents

List of Figures

Audio-visual speech perception training for a child with Autism Spectrum Disorders: A Case Study

Introduction

ASD are a group of developmental disorders that cause impairment in multiple developmental and functional areas. Deficits in social interaction, communication, and behavior issues tend to characterize the disorder (DSMV-VI-TR). Although deficits associated with ASD are thought to originate in the brain and have been associated with a number of genes, the central cause remains elusive. One major feature of individuals with ASD is they lack the ability to empathize with others, contributing to their social and pragmatic communication difficulties (Baker, 2013). In addition, ASD is associated with differences in executive function, including higher level thinking skills such as planning, problem solving, goal-directed behavior, emotion regulation, arousal, and verbal working memory (Baker, 2013). However, it is unclear what processes are critical for the development of functional communication skills required for successful communication with peers and caregivers. The identification of these skills is critical, as interventions may target underlying developmental differences in ASD to improve communication outcomes.

One of the earliest theories of autism is based on the notion that this group processes sensory information in a way that is different from their typically developing peers (Dawson, Soulieres, Hubert, & Burack, 2005). The most current theories (e.g. weak central coherence) have continued to suggest sensory atypicalities are at the heart of ASD (Iarocci & McDonald, 2006). Central to these theories is the idea that multisensory binding is different in ASD, perhaps due to lack of orienting or attention to the most important features of the multisensory signal.

Importantly, theories (e.g. dynamic systems theory) suggest identifying the perceptual level at which specific perceptual deficits interact with other critical systems in development (Thelen & Smith, 1994). Recently, it has been suggested that one critical influence on atypical language development in ASD is the lack of ability to integrate auditory and visual information in real-world environments in which auditory background noise is ubiquitous. This is a critical ability for normal language acquisition, as the integration of auditory and visual information can remove the equivalent of 20 decibels of auditory noise from the environment (Sumby & Pollack, 1954).

Several lines of evidence support the proposal that children with ASD integrate auditory and visual information differently than their typically developing peers. A.S. Carter, J.R. Irwin, C. Klaiman, E.A. Mongillo, R.T. Schultz, and D.H. Whalen (2008) compared audiovisual processing in children with and without Autism Spectrum Disorders. Their purpose was to explore how each group uses visual information differently and address whether children with Autism Spectrum Disorders perform typically on similar kinds of tasks with nonhuman stimuli. The findings of their study suggest that children with Autism Spectrum Disorders may use visual information for speech differently than typically developing children (Carter, et al., 2008). T. Charman, K. Davies, M. Elsabbagh, J.A. Guiraud, and M.H. Johnson (2012) studied how the inability to integrate audiovisual information might affect language development. Their research successfully showed that infants at low risk for ASD can integrate audiovisual speech and perceive mismatched auditory and visual cues. However, according to their looking behavior, infants at high-risk for ASD cannot properly integrate audiovisual speech which could have a negative impact on the acquisition of functional vocabulary (Charman et al., 2012).

In addition to evidence directly comparing the audio-visual integration in speech perception to their typically developing peers, it has been demonstrated children with ASD do

not perceive what is known as the McGurk Effect. The McGurk Effect is a compelling example of how perception and vision play a role in communication. Harry McGurk and John MacDonald discovered that when individuals are shown a film of a talking head or face, in which repeated utterances of the syllable [ba] were dubbed onto the lip movements of the syllable [ga], typical adults will hear the syllable [da] (MacDonald & McGurk, 1976). The McGurk Effect is essentially two phonemes put together. It has been demonstrated that children with ASD do not receive the visual fusion effect noted in most typical listeners. Since Autistic individuals lack the ability to fully observe a situation in its entirety, the individuals may only be focusing on either the auditory or visual signals. In support of the notion that children with ASD are not integrating the visual part of the signal, Massaro and Bosseler (2003) demonstrated that children with ASD can be trained to lip read via a computerized face ('Baldi'), improving their audio-visual integration skills. Thus, children with Autism are influenced by speech information in the face and these individuals can be taught to improve sensitivity to visible speech.

Although modest gains in lip reading and attention to a computerized talking head were demonstrated, the Massaro and Bossler (2003) study used unnatural stimuli, a factor that has been shown to influence speech perception generally (Kuhl & Meltzoff, 1984a) and may have an impact on the efficacy of therapeutic approaches more specifically (Kuhl & Meltzoff, 1984b). In addition, that study did not evaluate whether lip reading training improved performance in auditory background noise, a feature that may be critical for perceptual improvements in natural environments. Further, the methods used in that study have not translated into formal treatment approaches for speech-language pathologists, the professionals who are charged with treating speech and language deficits. As such, the aim of the current proposal is to develop and evaluate a novel, audio-visual training approach in which a child diagnosed on the spectrum progressively

learns to read lips by increasing attention to the face and decreasing the amount of target auditory information via background noise. A changing criterion, multiple baseline design was used to evaluate the efficacy of the treatment approach with the following specific aims: 1) to train accurate lip reading of consonant-vowel (CV) syllables with 100% accuracy in 5 successive stages using natural auditory and visual stimuli (i.e. AV recordings of a female speaker); 2) to evaluate treatment outcomes using a multiple baseline. Given previous findings and theoretical predictions, the individual with ASD was expected to show significant gains in language and social skills.

Methods

Participant

One child diagnosed with ASD from the Northwest Arkansas area was used for this study. The child met criteria for ASD based on expert clinical diagnosis. The child was verbal and eleven years old. The child spoke American English and had no other diagnosis of impaired neurological function, psychiatric conditions, attention deficit disorder, or cognitive impairment. Written informed consent approved by the University of Arkansas Institutional Review Board was obtained from a responsible guardian.

Stimuli

Figure 1 depicts visual stimuli. The stimuli used in behavioral training consisted of consonant-vowel (CV) syllables /bi/, /gi/, /ki/, /mi/, and /di/. The duration of the test syllables was between 500 and 600ms. Stimuli were recorded from a female speaker using a Cannon HD Vixia M32 video recorder. The stimuli were low-pass filtered with a cutoff at 5,000 Hz and were root-mean-square (RMS) normalized so that no one recorded stimulus was perceived as louder than any other. Multi-talker babble noise was taken from a standard recording used in

audiological testing (the Hearing-in-noise Test). Multi-talker babble noise was low-pass filtered with the same cut-off as the speech stimuli. Speech and noise recordings were merged using Adobe Sound edit on a Dell desktop computer and RMS normalized to have the same amplitude. In order to train lip reading in varying levels of background noise, speech stimuli was generated with the following signal-to-noise ratios (SNRs): +4dB, +14dB, and 0dB (Binder, Liebenthal, Possing, Medler &Ward, 2004).

Procedure

Speech training was implemented at the Speech and Hearing clinic in Epley Center for Health Professions and supervised by a trained speech-language pathologist. The speech reading training began with a bimodal (i.e. auditory and visual) presentation of each syllable. The intensity of the auditory stimuli used during training was based on the student's performance on the previous training task. If a student attained a passing score in a given training session, the SNR was decreased, increasing reliance on the visual speech signal for successful task performance. The SNR stayed the same in the event that the child did not perform at 100% accuracy. The student was instructed to watch or listen to the recording and indicate the syllable spoken. A cross was presented on the screen prior to each presentation indicating the start of the trial. Following the presentation, the child was instructed to circle the appropriate syllable on a document given at the beginning of the session. The document contained four syllables to choose from. Feedback was given in the form of verbal praise and edible treats. The child was required to give a response prior to the onset of the next trial.

Analysis

Accuracy data was calculated using percentage change and was analyzed using single subject statistical procedures. Changes in accuracy were analyzed using both traditional visual

inspection and statistical process control (SPC) (Phadt & Wheeler, 1995). For the purposes of visual inspection, data was charted over time using line-graphs. SPC analysis theory suggests that predictable variation will fall within a specified range for a given subject. Data falling outside the range of predictable variability indicates a change related to treatment procedures. This procedure has been used to provide a stronger indication to treatment related changes than visual inspection alone in the context of health-care service delivery.

Results

The percentage of correct responses (% CR) was evaluated for both pre and post-training. Percentage correct responses (% CR) for the pre-training and post-training auditory (AUD), visual (VIS) and audio-visual (AV) baselines are displayed in Figure 2. Percentage correct responses for perceptual training on 0dB MBN over five training sessions are displayed in Figure 3. For each modality, percentage change from pre to post-training was: AUD 78%, VIS 19%, and AV 66%. During training, the participant quickly learned to identify syllables in the +14dB and +4dB conditions, scoring 100% for each SNR on the first training session. In the 0dB condition, the participant showed a successive increase from 29% initially to a peak performance of 88% in session 4. There was a slight decrease on the fifth session (71%). Equivalent trends appeared evident for both highly visual contrasts (e.g., /bi/ versus /di/) and syllables with less salient visual information (e.g., /gi/ versus /ki/). Response to fused McGurk stimuli did not change pre to post-training. The participant reported an auditory capture effect when presented with the McGurk Effect for both pre and post-training.

Discussion

The study evaluated the efficacy of the treatment approach with the following specific aims: 1) to train accurate lip reading of consonant-vowel (CV) syllables with 100% accuracy in 5

successive stages using natural auditory and visual stimuli; 2) to evaluate treatment outcomes using a multiple baseline design. In accordance with initial expectations, findings suggest that AV speech perception training enhanced perceptual performance from the initial pre-training baseline. Findings also suggest that training improved significantly above chance over 5 training sessions using progressively higher levels of background noise. However, the participant identified McGurk fusion stimuli by the auditory percept only (i.e., /ba/) both at the initial pre-training baseline and following AV speech perception training. Overall, the results suggest that while training attention to visual features of the mouth can enhance perceptual performance, the underlying mechanisms may be different from typically developing children who have been shown to perceive the McGurk effect without training. In the discussion following, findings will be discussed with respect to the previous findings, limitations, and future directions.

The results of the current study were consistent with Massaro and Bossler (2003) in which AV training resulted in increases in perceptual performance in ideal listening conditions. The current findings suggest that, at least for this single participant, AV speech perception can be trained in high levels of background noise. This is important because perception in natural environments is often accompanied by varying levels of background noise. According to current models of perception in ASD, underlying perceptual deficits in integrating multimodal stimuli may have cascading effects on the later development of vocabulary (Charman et al., 2012; Iarocci & McDonald, 2006). This notion implies that children with ASDs may miss critical perceptual cues in background noise that decrease the comprehension of new vocabulary. Importantly, the training approach used in the current study may also be implemented without much difficulty in therapy sessions suggesting a novel avenue for training perception in noise.

Although findings in this single subject study suggest that AV training can enhance perceptual performance in noise, findings also support the notion that the underlying mechanism may be different than in typically developing children as suggested in theories (e.g., weak central coherence) (Iarocci & McDonald, 2006). Three findings support the notion that underlying mechanisms may be different. First, AV perception in the post-training measure was enhanced by only 6% over the auditory only modality. Perceptual performance in the auditory modality was enhanced by 78%, more than any other modality in the current study despite the fact that it was not explicitly trained. Further, relative to the other two modalities, only modest gains in visual perception were observed from the initial pre-training baseline to the post-training period (19%). Second, the participant identified only the auditory percept for McGurk fusion stimuli, suggesting an underlying mechanism weighted toward the auditory modality. Third, there was no difference in identification performance between highly visual and less visually informative AV stimuli, reinforcing the notion that the participant may have relied more heavily on the auditory portion of the signal.

Limitations and conclusions

Although the results support the previously listed specific aims, limitations to the study do exist. One aspect of the study limiting generalizability is that the study was a single subject design. As such, the findings cannot be extended to the larger population of children with ASD. Several aspects related to the participant might limit generalizability. First, the child is considered high-functioning, suggesting that training may not be as effective for children at different place along the spectrum. Second, this participant was enthusiastic about perceptual training in most of the sessions and motivation (i.e., salience) may be a significant factor in perceptual training. Third, this child was also engaged in other treatment throughout the training

period including a focus on listening skills. Thus, multiple treatment interference effects limit the extent to which findings from this study can be attributed to the particular training approach. Finally, maturation is a limitation common to time-series treatment designs. In the current study, maturation cannot explain the findings as a combination of multiple baseline testing, changing criterion, and statistical process control (SPC) were used. However, the interaction of maturation and treatment effects cannot be completely ruled out.

Limitations notwithstanding, findings from this single subject design suggest that AV training can indeed enhance perceptual performance in high levels of background noise for a child diagnosed on the spectrum. This simple and brief training protocol may be easily implemented as one aspect of therapy in speech and hearing clinics with the potential for generalization to pragmatic aspects of communication. Consistent with Massaro and Bosseler (2003) training attention to the face generally may also result in the acquisition of pragmatic communication skills.

Future directions

Several aspects of the training protocol might be modified in subsequent studies. First, given the increased performance over time, larger group studies are warranted to show generalization to the larger population. Second, the requirement that the subject achieve 100% accuracy for moving on to the next SNR may have been too high. To increase attention to the visual aspects of the signal for task performance, using lower SNRs in which auditory performance has been shown to be at chance levels may significantly increase attention to the visual aspect of the signal (Binder et al., 2004). Finally, given the proposal that lower level perceptual processes influence vocabulary development, it may be critical to test the effects of AV training in more naturalistic words and sentence level contexts.

References

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders-Text revision (DSM-IV-TR; 4th ed.). Washington, DC, American Psychiatric Association, p. 943.

Baker, K. (2013, February). *Autism Spectrum Disorders.* Lecture presented for educational purposes in language disorders, University of Arkansas, Fayetteville, AR.

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

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

Charman, T., Davies, K., Elsabbagh, M., Guiraud, J. A., Johnson, M. H., et al. (2012). Atypical audiovisual speech integration in infants at risk for autism. *Plos One, 7,* 1-6.

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

Kuhl, P.K., & Meltzoff, A.N. (1984a). *Infants' representations of events: Studies in imitation, cross-modal perception, and categorization.* Paper presented at the fourth international conference on infant studies, New York.

Kuhl, P.K., & Meltzoff, A.N. (1984b). The intermodal representation of speech in infants. *Infant Behavior and Development,* 7, 361-381.

Lord, C., Rutter, M., DiLavore, P.C., & Risi, S. (1999). *Autism Diagnostic Observation Schedule-WPS (ADOS-WPS).* Los Angeles, CA: Western Psychological Services.

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

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

Mottron, L., Dawson, M., Soulieres, I. Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: an update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders,* 36, 27-43.

Pfadt, A., & Wheeler, D. J. (1995).Using statistical process control to make data based clinical decisions. *Journal of Applied Behavior Analysis*, *28*(3), 349–370.

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

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

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: MIT Press.

Appendix

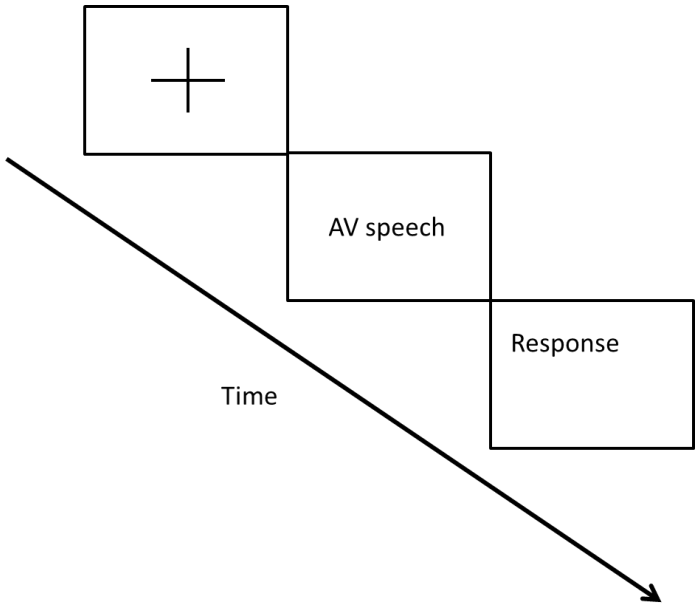

Figure 1: Time-line of one trial

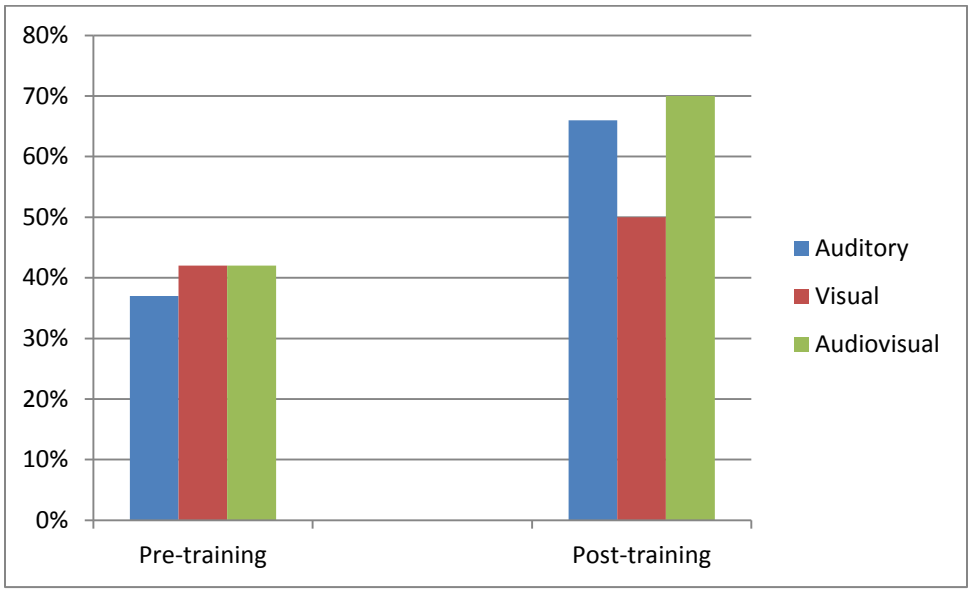

Figure 2: Pre-training and post-training measures. This chart shows pre and post-training percentage of correct responses in the auditory, visual, and audio-visual modalities

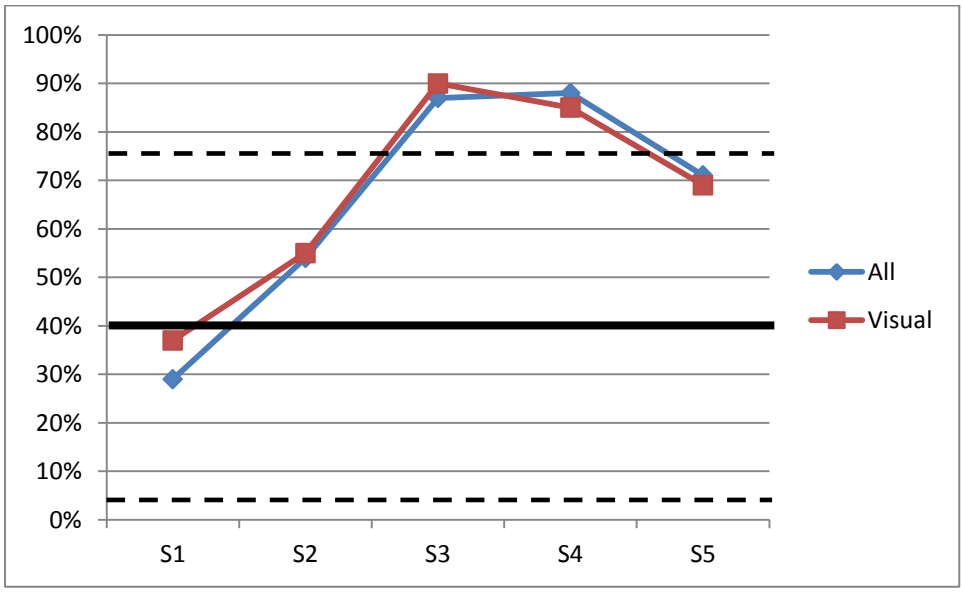

Figure 3: Session-by-session increases in percent correct responses. This chart shows the percentage correct responses for perceptual training in 0dB multitalker babble noise over five training sessions. The mean across three pre-training baselines is shown (dark solid line) along with the statistical process control lines (dotted lines).