A Randomized Controlled Trial for Audiovisual Multisensory Perception in Autistic Youth

Jacob I. Feldman[1,2] · Kacie Dunham[3,4] · Gabriella E. DiCarlo[3,5,6] · Margaret Cassidy[7,8] · Yupeng Liu[7,9] ·
Evan Suzman[10,11] · Zachary J. Williams[2,3,4,6] · Grace Pulliam[7] · Sophia Kaiser[12] · Mark T. Wallace[2,3,4,13,14,15,16] ·
Tiffany G. Woynaroski[1,2,3,13]

Abstract
Differences in audiovisual integration are commonly observed in autism. Temporal binding windows (TBWs) of audiovisual speech can be trained (i.e., narrowed) in non-autistic adults; this study evaluated a computer-based perceptual training in autistic youth and assessed whether treatment outcomes varied according to individual characteristics. Thirty autistic youth aged 8–21 were randomly assigned to a brief perceptual training ($n = 15$) or a control condition ($n = 15$). At post-test, the perceptual training group did not differ, on average, on TBWs for trained and untrained stimuli and perception of the McGurk illusion compared to the control group. The training benefited youth with higher language and nonverbal IQ scores; the training caused widened TBWs in youth with co-occurring cognitive and language impairments.

Keywords multisensory integration · audiovisual integration · perceptual training · autism spectrum disorder · temporal binding window

✉ Jacob I. Feldman
j.i.feldman@vumc.org

[1] Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, MCE 8310 South Tower, 1215 21st Avenue South, Nashville, TN 37232, USA

[2] Frist Center for Autism & Innovation, Vanderbilt University, Nashville, TN, USA

[3] Vanderbilt Brain Institute, Vanderbilt University, Nashville, TN, USA

[4] Department of Hearing and Speech Sciences, Vanderbilt University, Nashville, TN, USA

[5] Present Address: Mass General Brigham Neurology Residency Program, Harvard Medical School, Boston, MA, USA

[6] Medical Scientist Training Program, Vanderbilt University, Nashville, TN, USA

[7] Neuroscience Undergraduate Program, Vanderbilt University, Nashville, TN, USA

[8] Present Address: National Institutes of Health, Bethesda, MD, USA

[9] Present Address: Washington University School of Medicine, Washington University in St. Louis, St. Louis, MO, USA

[10] Master's Program in Biomedical Science, Vanderbilt University, Nashville, TN, USA

[11] Present Address: Southwestern School of Medicine, University of Texas, Dallas, TX, USA

[12] Cognitive Studies Undergraduate Program, Vanderbilt University, Nashville, TN, USA

[13] Vanderbilt Kennedy Center, Vanderbilt University Medical Center, Nashville, TN, USA

[14] Department of Psychology, Vanderbilt University, Nashville, TN, USA

[15] Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN, USA

[16] Department of Pharmacology, Vanderbilt University, Nashville, TN, USA

Disruptions in sensory functioning are commonly observed in autistic individuals[1]. These differences have been observed via a broad range of measurement techniques and across all sensory modalities (Baum et al., 2015; Schaaf & Lane, 2015; Schauder & Bennetto, 2016). Audiovisual integration, or the ability to combine information from auditory and visual sensory inputs, has been particularly well studied in this population (Soto-Faraco et al., 2012; see Feldman et al., 2018 for a review). The ability to integrate the visual and auditory components of social stimuli, such as speech (Bahrick & Todd, 2012), is theorized to be particularly critical to developing accurate, unified representations of the sensory world and foundational to the development of higher-order social, communication, and cognitive skills (Wallace & Stevenson, 2014; Wallace et al., 2020).

In one frequently replicated finding, (e.g., Noel et al., 2017; Stevenson et al., 2014; Woynaroski et al., 2013), autistic individuals tend to present with enlarged temporal binding windows (TBWs; the period of time over which individuals tend to integrate related sensory information from multiple modalities) relative to non-autistic comparisons. Enlarged TBWs in autistic children (i.e., less acute temporal binding of audiovisual stimuli) have been interpreted as maladaptive and are hypothesized to produce cascading effects on development in a number of domains in this clinical population (Cascio et al., 2016). Larger TBWs for speech stimuli are associated with increased features of autism and decreased language abilities (Feldman et al., 2019a; Smith et al., 2017), lending some empirical support to this theory of cascading effects.

Perceptual Training of Temporal Binding of Audiovisual Stimuli

The substantial evidence for altered audiovisual temporal binding in autistic youth, as well as observed relations between TBWs for audiovisual speech and other domains of functioning in autistic youth, has engendered increasing interest in the possibility of training audiovisual integration in autistic youth (e.g., Bahrick & Todd 2012; Cascio et al., 2016; Feldman et al., 2018; Zhou et al., 2020). A number of training studies targeting audiovisual temporal binding have been conducted in non-autistic adults, and have been shown to narrow TBWs in a relatively short period of time (i.e., 3–5 sessions; De Niear et al., 2016, 2018; McGovern et al., 2016; Powers et al., 2009; Setti et al., 2014; Sürig et al., 2018; Zerr et al., 2019). These training paradigms provide automated feedback after each trial of a computerized task wherein participants must make judgements about the synchrony or temporal order of audiovisual stimuli, such as flashes and beeps (e.g., Powers et al., 2009; Setti et al., 2014; Sürig et al., 2018) and audiovisual speech, such as a speaker saying the syllable "ba" (De Niear et al., 2018).

Limitations of this Literature

There are several limitations to the literature on perceptual trainings for audiovisual stimuli in non-autistic adults. First, the majority of these perceptual training studies have found no evidence for generalization to untrained multisensory tasks (De Niear et al., 2018; Powers et al., 2016; Setti et al., 2014) or limited evidence for generalization (i.e., training effects only observed at limited SOAs or conditions in untrained tasks; Zerr et al., 2019). To date, only Sürig et al., (2018) have found strong evidence for generalization. They hypothesized that the adaptive difficulty in their simultaneity judgment (SJ) training (i.e., the perceptual training task was designed to be challenging for each participant rather than utilizing consistent difficulty) resulted in strong learning, enabling gains made on their perceptual training to generalize to an audiovisual localization task. Though other perceptual training studies provide evidence that increasing difficulty does increase learning (De Niear et al., 2016), no other study has evaluated whether adaptive difficulty results in generalization following a perceptual training for temporal binding of audiovisual stimuli.

The intervention literature may provide additional explanations for the lack of generalization in previous studies. First, prior studies may not have found evidence for generalization because they were evaluating effects on outcomes that were very distal to their training paradigms (i.e., those that were too far beyond what was directly taught in their training; Yoder et al., 2013). It may be necessary to assess a variety of outcomes that differ in various degrees from the stimuli and/or the task trained to detect distal or generalized outcomes of perceptual trainings. For example, a perceptual training in the context of an SJ task for audiovisual speech may be more likely to generalize another task utilizing the same instructions with slightly different stimuli (e.g., an SJ task with different stimuli than those utilized in training) than to another task that utilizes different instructions. Alternatively, a task measuring perception of the McGurk effect, wherein incongruent audiovisual stimuli (e.g., visual "ka" and auditory "pa") induce a fused percept (i.e., "ta" or "ha"; McGurk & MacDonald 1976), indexes the influence of vision on auditory speech. The McGurk effect could measure multisensory speech integration in a slightly different context, thus representing a more slightly more distal outcome.

---

[1] There are currently ongoing, complex discussions regarding the use of person-first language (e.g., individuals with autism, individuals with ASD) versus identity-first language (e.g., autistic individuals; see Robison 2019). Clinicians and researchers tend to prefer person-first language, while many autistic individuals and their allies prefer and advocate for identity-first language (e.g., Gernsbacher 2017; Kenny et al., 2016). In keeping with current recommendations (Bottema-Beutel et al., 2021), we have opted to use identity-first language.

Additionally, the intervention literature recommends training with diverse stimuli, which leads to greater generalization (Stokes & Osnes, 1989; Swan et al., 2016). Though using the same stimuli (i.e., the same auditory tones, the same visual flashes, the same speaker) across all trials in an experiment allowed previous researchers to maintain a very high degree of experimental control, it may have been at the expense of generalization.

Perceptual Training of Temporal Binding of Audiovisual Stimuli in Autism

To date there have been very few studies on training audiovisual speech perception in autistic youth. Two studies (Irwin et al., 2015; Williams et al., 2004) have utilized quasi-experimental designs and found that autistic children improved their audiovisual speech perception following brief computerized training. However, due to the small sample sizes and the nature of the quasi-experimental designs, it is difficult to make conclusions about the effectiveness of those training programs.

One additional study (Feldman et al., 2020) adapted the procedures utilized in some of the previously discussed perceptual training studies (e.g., De Niear et al., 2018; Powers et al., 2009) for autistic children. This study utilized a multiple baseline across participants design, a single-case experimental research design (see Ledford et al., 2019); all three of the trained subjects in this experiment demonstrated extreme widening of their TBWs during the extended baseline condition and subsequently exhibited highly variable responses to the perceptual training. Because it was difficult maintain adequate experimental control in the context of a single-case experimental research design, Feldman et al., (2020) were unable to detect a functional relation (i.e., an effect of the training condition). This limitation can only be addressed by follow-up research utilizing group treatment designs.

Additionally, though the differential responses to the perceptual training condition (Feldman et al., 2020) were somewhat expected given the high degree of heterogeneity in both presentations of autism and responses to intervention exhibited by autistic youth (e.g., Marcus et al., 2001; Vismara & Rogers, 2010), single-case experimental research designs are unable to determine whether individual characteristics may have influenced the differential responses. The authors speculated that chronological age may have influenced treatment responses, as the older participants demonstrated more immediate and pronounced responses to the perceptual training. Similarly, it has been frequently noted that psychophysical tasks assessing audiovisual integration require a relatively high degree of cognitive and language skills to understand the task and directions (e.g., Cascio et al., 2016; Feldman et al., 2018; Feldman, Kuang, et al., 2019; Woynaroski et al., 2013). Accordingly, the aforementioned studies

on perceptual training have been conducted on autistic youth with at least average IQ; thus, no study on autistic youth to date has assessed whether the effect of perceptual training may vary according to cognitive abilities. Hypotheses about factors that might influence treatment effects are best evaluated by measuring and testing putative moderators in the context of group designs (Hayes, 2017).

One final limitation of the extant literature is the lack of data collected on participants' (and their parents', in the case of children) thoughts and experiences related to treatment goals, procedures, and outcomes. The collection of this data, referred to as social validity in the intervention literature, is critical for assessing the acceptability and importance of novel interventions (Foster & Mash, 1999; Gast & Ledford, 2014). Autistic self-advocates have pushed researchers to engage in participatory research (e.g., Raymaker & Nicolaidis 2013; Warner et al., 2018), with the goal of creating interventions that improve quality of life and key outcomes rather than cures for autistic traits (Raymaker, 2019). To date, only one study (i.e., Feldman et al., 2020) collected social validity data, and the authors noted that participants did not consistently rate the perceptual training paradigm as helpful or report that they would utilize the training (i.e., "play the game") in their free time. The authors suggested that future studies should try to increase the perceived helpfulness of the training and also make the training more game-like in order to increase positive perceptions about the procedures and goals of the training.

Purpose

The purpose of this study was to conduct a randomized controlled trial testing the short-term effects of computer-based perceptual training utilizing adaptive difficulty in autistic youth. To address limitations in the extant literature, several changes were made to the perceptual training paradigm utilized in previous research including: (a) implementing a game-like scoring system, (b) providing explicit feedback to incorrect answers, (d) utilizing multiple speaker stimuli during the training, and (e) measuring several outcomes intended to index varying degrees of generalization and distality relative to the training stimuli and task.

The following research questions were posed:

(1) Do autistic youth assigned to the perceptual training experience greater narrowing of their TBW for (a) trained audiovisual speech stimuli, (b) untrained speech stimuli, and/or (c) untrained speakers compared to those assigned to the control group? Does the effect of the perceptual training translate to broader multisen-

Fig. 1 Diagram of participant recruitment
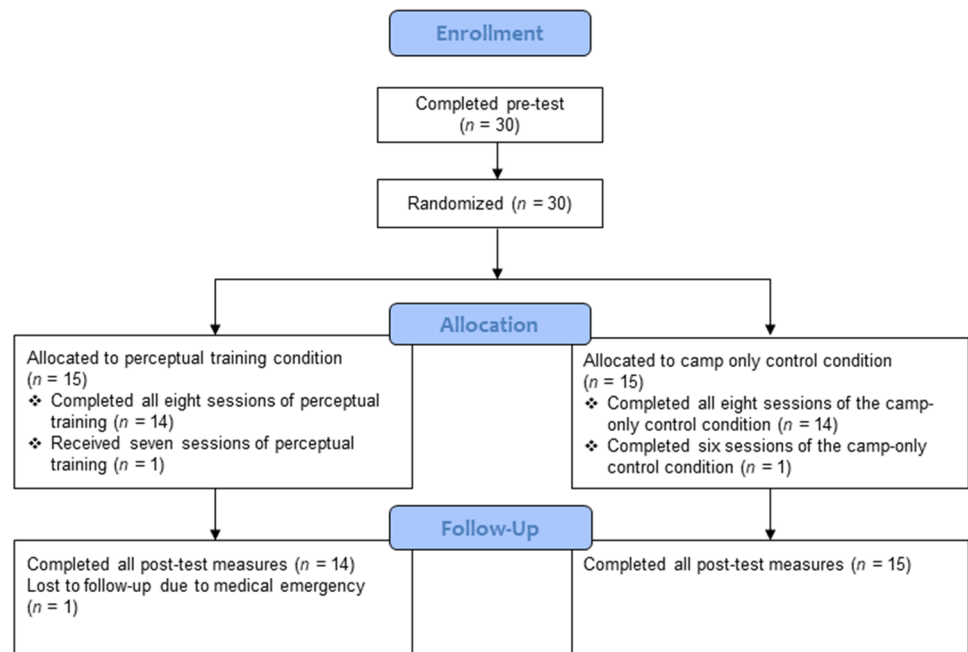
sory integration, specifically perception of the McGurk illusion?

(2) Does the effect of the perceptual training vary according to individual factors, specifically chronological age, nonverbal cognitive ability, and language ability?

Methods

This study was completed at Vanderbilt University Medical Center with procedures approved by the Vanderbilt University Institutional Review Board.

Study Design

To answer these research questions, a randomized controlled trial was conducted with 30 autistic youth (see Participants). After participants consented to participate in the study, they were randomized in pairs (or groups of four, in the case of siblings and individuals who traveled to the study together) matched on chronological age, biological sex, gender, and pre-training TBWs to either the perceptual training condition or the control condition using a random number generator by a naïve member of the study team.

Participants assigned to both groups visited the laboratory for a research camp that ran for four consecutive weekdays over the course of 2 weeks for a total of eight sessions. Each session was a half-day (approximately 2.5–3 h). When participants were not completing research activities

(see Perceptual Training and Camp Only Control Condition), they had access to a variety of preferred activities (e.g., board and video games, toys, music) and completed organized activities in small groups daily. No other therapies or interventions were provided by the study team during the research camp, and parents were asked to report whether their children participated in any outside interventions (e.g., speech-language therapy, occupational therapy, applied behavior analysis consultation or therapy) during the timeframe for the research camp on the last day of the study via REDCap (Harris et al., 2009).

Participants completed the pre-test measures 1 to 3 days prior to the research camp and the post-test measures 1 to 3 days following the research camp. Pre- and post-test measures were collected at the same time of the day for each participant.

The final four participants in this study completed the research camp in June 2020, and thus several modifications to the study protocol were made due to COVID-19 to increase participant safety and reduce the likelihood of virus transmission; the core components of the research camp and both treatment conditions were not impacted by any of the changes. For a list of modifications, see Online Appendix.

Participants

Thirty participants aged 8–21 were recruited from a larger ongoing research project (e.g., Dunham et al., 2020; Feldman 2019; Feldman et al., 2020; see Fig. 1 for a flowchart of participant recruitment and Table 1 for participant

Table 1 Participant characteristics

Characteristic	Perceptual Training Condition ($n = 15$) M (SD) Min–Max	Camp Only Control Condition ($n = 15$) M (SD) Min– Max
Age (Years)	14.2 (4.0) 8.1–21.3	14.0 (3.6) 8.4–19.2
Biological sex	11 Male, 4 female	10 Male, 5 female
Race	12 White 3 Black or African American	13 White 2 Multiple Races
Ethnicity	14 Not Hispanic or Latino 1 Not Reported	14 Not Hispanic or Latino 1 Not Reported
Nonverbal IQ	113.2 (12.14) 93–139	108.7 (24.6) 45–147
Core Language Standard Scores	92.5 (20.3) 48–118	92.0 (25.9) 40–120
TBW$_{trained}$	533.0 (213.0) 173.4–850.0	498.8 (257.1) 191.7–1110.6

TBW$_{trained}$ = Temporal binding window for a speaker from the training saying "ba." Nonverbal IQ measured by the Leiter International Performance Scale, third edition (Roid et al., 2013). Core Language Standard Scores measured by the Clinical Evaluation of Language Fundamentals, fourth edition (Semel et al., 2004) or the Preschool Language Scale, fourth edition (Zimmerman et al., 2011). All standardized language assessments collected 0–30 months prior to the beginning of this study as a part of the larger project. Groups did not differ on any of the above characteristics, $p > 0.5$

demographics). Inclusion criteria were: (a) diagnosis of autism spectrum disorder according to DSM-5 criteria (American Psychiatric Association, 2013) as independently confirmed by a research-reliable administration of the Autism Diagnostic Observation Schedule 2 (Lord et al., 2012) and clinical judgment of a licensed clinician on the research team, (b) normal hearing and normal or corrected-to-normal vision per screening and parent report, (c) no history of seizure disorders, (d) no diagnosed genetic disorders (e.g., Down syndrome, Fragile X), and (e) demonstrated ability to complete an SJ task. Study eligibility was confirmed by members of the research team (i.e., clinical psychologists and speech-language pathologists) during study visits that occurred 0–30 months prior to the beginning of this study as a part of the larger project. Exclusion criteria were medication changes during the perceptual training study. No exclusion criterion based on cognitive ability was imposed. Given that our second research question assessed moderated effects of the perceptual training, we recruited participants who were heterogeneous in regard to putative moderators (i.e., chronological age, nonverbal cognitive ability, and language ability; see Putative Moderators of Training Effects on Outcomes) to permit statistically testing of hypothesized differential effects (Hayes, 2017).

Materials

The perceptual training, as well as the psychophysical data collection (see Pre- and Post-Test Outcomes) occurred in a light- and sound-attenuated booth (WhisperRoom Inc.,

Morristown, TN, USA) with visual stimuli presented on a Samsung Syncmaster 2233RZ 22-inch PC monitor and auditory stimuli presented binaurally via Sennheiser HD559 supra-aural headphones.

Monosyllabic speech stimuli used in the perceptual training and SJ tasks (see Temporal binding window for audiovisual speech) were seven videos obtained from Basu Mallick et al., (2015). For the perceptual training, stimuli included six of those videos (labeled 4.1, 4.2, 4.3, 4.5, 4.6, and 4.7 by Basu Mallick et al., 2015); each video was of a different speaker (three male and three female speakers) saying "ba" in front of a blank (i.e., gray) background with neutral affect. For the pre- and post-test SJ tasks (see Temporal binding window for audiovisual speech), stimuli were videos of one trained speaker (labeled 4.6 by Basu Mallick et al., 2015) saying the (trained) syllable "ba" and the untrained syllable "pa" and a video of a female speaker not included in the perceptual training (the seventh video from Basu Mallick et al., 2015; labeled 4.8) saying the trained syllable "ba."

For the McGurk illusion task, stimuli were videos of a different female speaker saying "pa" and "ka" in front of a neutral background with neutral affect. These stimuli have been utilized in several previous experiments (e.g., Dunham et al., 2020; Feldman et al., 2019a, 2020, 2022; Simon & Wallace, 2018).

All video stimuli were edited in Adobe Premiere to create asynchronous stimuli for the perceptual training and SJ tasks, and incongruent audiovisual stimuli, auditory-only stimuli, and visual-only stimuli for the McGurk illusion task.

Perceptual Training

The perceptual training was a modified SJ task that took approximately an hour to complete. During each trial, participants were asked to indicate whether they perceived the auditory and visual information to have occurred at the same time or at a different time via a serial-response box. Following correct responses, a blue check mark appeared on the screen, accompanied by a non-synchronous sound effect (i.e., a sound effect from the *Mario* series). Following incorrect responses, participants saw a red X on the screen, and received corrective feedback (i.e., "That was same time," "You SAW ba first," and "You HEARD ba first"). Participants were given the choice between visual feedback (i.e., text written below the red X) and auditory feedback (i.e., a recording of a spoken voice) unless participants presented with reduced reading comprehension during the study visits that occurred as a part of the larger project (i.e., standard scores on an age-appropriate reading measure were 1.5 or more standard deviations below the mean; e.g., Reid et al., 2001; Wiederholt & Bryant, 2012; these participants always received auditory feedback).

During each day in the training, participants completed seven rounds of the training. Each round consisted of 48 trials, 50% of which were synchronous. To increase the likelihood of generalization, videos of six different speakers saying "ba" were utilized (see Materials). Each speaker was presented equally across synchronous and asynchronous trials, such that each speaker was utilized four times in synchronous trials and four times in asynchronous trials.

The seven rounds were divided into three levels of difficulty as follows: easy (one round), medium (two rounds), and hard (four rounds). The specific stimulus onset asynchronies (SOAs; i.e., the period of time between the onset of the visual and auditory stimulus; negative SOAs represent auditory-first stimuli and positive SOAs represent visual-first stimuli) at each difficulty level were based on each participants' performance during the previous study day; thus, the task was adaptive. For the first day of the training, participants' performance on the pre-test SJ task utilizing speech stimuli (specifically, $TBW_{trained}$; see Temporal Binding Window for Audiovisual Speech) were utilized to derive initial training SOAs. On all subsequent days (i.e., days 2–8 of the training), the participants' accuracy on the previous day's perceptual training was used to derive new training SOAs. In the easy condition, the training SOAs were the points wherein the psychometric curves fit to the previous day's performance (see Derivation of TBWs) crossed 10%, 20%, and 30% report of synchrony, with a minimum SOA of 133 ms and a maximum SOA of 500 ms. In the medium condition, the SOAs were the points that crossed 40%, 50%, and 60% report of synchrony, with a minimum SOA of 133 ms and a maximum SOA of 400

ms. In the difficult condition, the SOAs were the points that crossed 65%, 75%, and 85% report of synchrony, with a minimum SOA of 133 ms and a maximum SOA of 300 ms. All training SOAs were rounded to the nearest 50 ms or 16.7 ms (i.e., one frame difference between the visual and auditory stimuli). Additionally, all training SOAs were presented equally in both auditory-first (negative) and visual-first (positive) trials so the average of all asynchronous trials equaled 0 ms (i.e., true synchrony).

Participants completed a comprehension check at the start of each day of training. Participants were also able to select images of preferred media or interests (e.g., *Mario*, *Minecraft*, trains, vacuums) that randomly appeared during the training to increase motivation and reinforce on-task behavior.

To make the training feel more game-like, an automated scoring system credited participants' correct answers and their number of correct answers in a row. Participants were shown their scores following each response, and at the end of each round of the training participants were shown their total score for the round and an updated overall total.

Each round of the perceptual training took approximately 6–8 min to complete, depending upon the amount of perceptual feedback delivered. Including the comprehension check and the time required to deliver task instructions and change between difficulty levels, the perceptual training took a total of approximately 45–60 min to complete. If participants finished the perceptual training in less than an hour, they were allowed to rest or choose a quiet activity (see Camp Only Control Condition) until they had been in the WhisperRoom for approximately 1 h.

Camp Only Control Condition

Participants in the camp only condition engaged in quiet activities in the WhisperRoom (i.e., listening to music; simple computer games such as Tetris, snake, solitaire, and minesweeper; card games such as war, Uno, or memory; reading a book to him/herself; puzzles, coloring, napping) for approximately 1 h during each of the eight days of the study. Activities were specifically chosen to be unisensory (i.e., auditory-only or visual-only) and minimally-social. Participants completed these activities in the WhisperRoom in order to keep other members of the research team and the other participants naïve to condition assignment.

Pre- and Post-test Outcomes

All pre-and post-test outcomes were collected by experimenters on the research team naïve to group assignment.

Temporal Binding Window for Audiovisual Speech

The primary outcome was the TBW for audiovisual speech stimuli on which the participants were trained ($TBW_{trained}$; i.e., TBW for a female speaker included in the perceptual training saying "ba"; see Materials). Two types of generalization data were obtained for TBWs utilizing untrained stimuli: one using stimuli featuring a different speaker saying the same syllable ($TBW_{novel\ speaker}$; i.e., a female speaker not included in the perceptual training saying "ba"; labeled 4.8 by Basu Mallick et al., 2015) and one using the trained speaker saying a different syllable ($TBW_{novel\ syllable}$; i.e., the same female speaker mentioned above from the perceptual training saying "pa"). These TBWs were measured via three different SJ tasks in order to evaluate the extent to which training effects were specific to the trained stimuli versus more generalized in nature, in the context of the trained task.

During each SJ task, participants were presented with trials at 15 different SOAs: synchronous (0 ms), ±500 ms, ±400 ms, ±350 ms, ±300 ms, ±250 ms, ±150 ms, and ±100 ms. During each run of the task, each trial was presented two times in random order (total of 30 trials per run). Based on the findings of a stability study and follow-up analyses (Dunham et al., 2020), participants completed ten runs of each SJ task (total of 300 trials, 20 at each SOA) so these variables would be acceptably stable (see Cronbach et al., 1963; Sandbank & Yoder, 2014).

For each trial in each SJ task, participants were instructed to report whether they perceived the auditory and visual stimuli as having occurred at the same time or at different times by pressing "1" and "2," respectively, on the keyboard. To ensure comprehension, each run of each task was preceded by a practice round, consisting of two trials of stimuli presented synchronously and two trials of stimuli presented at an SOA of ±900 ms. Participants were required to correctly respond to all trials of the practice round prior to starting each run.

Derivation of TBWs To derive TBWs (measured in ms), the data from each SJ task were processed in MATLAB. The rate of perceived synchrony across SOAs (i.e., the number of times that the participant indicated that they perceived the stimuli to have occurred at the same time over the total number of trials presented for each SOA) was calculated in MATLAB using an adaptive fit script. The best fit (i.e., the one that resulted in the lowest error term) was chosen between two psychometric functions fit using the *glmfit* function (one for auditory-leading/left trials and one for visual-leading/right trials) and a single Gaussian curve fit using the *fit* function, after normalizing the data (i.e., setting the data to 100%). This approach is consistent with previous perceptual-based training studies targeting temporal binding of audiovisual stimuli (e.g., De Niear et al., 2016;

2018; Feldman et al., 2020; Powers et al., 2009). The TBW for auditory- and visual-leading stimuli were the points at which the curve(s) crossed 75% perceived synchrony, with the overall TBW being the difference between those values.

Data from the perceptual training were processed in the same manner as described above to calculate the next day's training SOAs for the adaptive nature of the training task (see Perceptual Training).

McGurk Illusion

To assess whether gains made in the context of the training translated to untrained tasks that measure broader responses to and integration of audiovisual speech, an additional multisensory task utilizing different task instructions and stimuli than the training was collected, specifically a task measuring perception of the McGurk illusion (McGurk & MacDonald, 1976). Past work suggests that youth who more accurately judge synchronous versus asynchronous audiovisual speech (i.e., those with narrower TBWs) may experience greater perception of the McGurk illusion (Stevenson et al., 2014, 2018); however, it remains to be seen whether training will induce increases in perceptions of the illusion via distal effects on enhanced multisensory integration.

Participants completed a psychophysical task indexing perception of the McGurk illusion with the syllables "pa" and "ka" presented as auditory-only syllables, visual-only syllables, congruent audiovisual syllables, and incongruent audiovisual syllables (i.e., auditory "pa' and visual "ka," which frequently induces an illusory percept of "ta" or "ha"; see Woynaroski et al., 2013 for more information regarding this approach). During each run of the task, participants were presented with 10 trials of each syllable in the auditory-only, visual-only, and matched audiovisual conditions and 10 trials of the incongruent audiovisual (McGurk) stimuli in a randomized order (70 trials per run). Participants completed two runs of the task (i.e., 140 trials total, 20 of each trial type) in order to yield an acceptably stable metric of the perception of the McGurk illusion (Dunham et al., 2020). After each trial, participants reported what syllable they perceived using a 4-button serial-response box. Prior to each run of the task, participants completed a comprehension check wherein they were prompted to press the designated button for each syllable in a random order. Data from this task were processed in MATLAB to obtain the percent of trials for which the participants reported the illusory percept.

Putative Moderators of Training Effects on Outcomes

As a part of the larger project, participants completed cognitive and language testing 0–30 months ($M = 13.6$ months) prior to their participation in this study. Nonverbal cognitive

abilities were assessed using the Leiter International Performance Scale, third edition (Leiter-3; Roid et al., 2013). Language abilities were assessed using the Clinical Evaluation of Language Fundamentals, fourth edition (CELF-4; Semel et al., 2004) for participants who were aged 8–21 years at the time of their assessment ($n = 25$; one participant did not complete the CELF) and the Preschool Language Scale, fourth edition (PLS-5; Zimmerman et al., 2011) for participants who were younger than eight at the time of their assessment ($n = 4$). The core language index score from the CELF-4 and the total language standard score from the PLS-5 were combined to form a single variable of core language ability. Given that standard scores tend to be stable for both language (e.g., Bornstein et al., 2014, 2016a, 2016b; Norbury et al., 2017; Pickles et al., 2014) and cognitive abilities (e.g., Eaves & Ho 1996; Lord & Schopler, 1989; Schneider et al., 2014) over the developmental period of interest to the present study, these scores were considered a suitable proxy for current abilities.

Social Validity

At the end of the final training session, participants completed a questionnaire using REDCap (Harris et al., 2009). This questionnaire was identical to the one used in Feldman et al. (2020). The survey had three questions on a 5-point Likert scale (i.e., "Did you think the game was easy?", "Did you think this game was fun?", and "Did you think this game was helpful?"; pictures of faces were utilized along with the numbers to facilitate comprehension), one yes/no question (i.e., "Would you play this game in your free time?"), and one open-ended question (i.e., "Is there anything else you want to tell us about this game?").

When parent report was available, parents were asked similar questions about their thoughts and experiences. This survey, also administered via REDCap, included four questions that used a 5-point Likert scale (i.e., "Did you notice any change in the way your child interacted with others?", "Did you notice any change in your child's use of language?", "Did you notice any change in your child's communication abilities?", and "Did you notice any change in your child's behavior?"). Each of these Likert questions was accompanied by an open field where parents could describe any changes they saw. One final open-ended question asked parents to describe, "any other changes in your child during sensory camp, either positive or negative, that we have not asked about."

Procedural Fidelity

Procedural fidelity was evaluated for the examiners collecting pre- and post-test data and for the examiners providing the perceptual training and the camp only condition using previously developed checklists of expected behaviors (see Feldman et al., 2020). For the pre- and post-test data collection, expected behaviors included the participant looking at the computer and wearing headphones set to the proper volume, the examiner not providing feedback based on correctness of responses, and the minimization of potential distractors. For the perceptual training, expected behaviors included the participant looking at the computer and wearing headphones set to the correct volume, the examiner setting up the perceptual training correctly, and the examiner not providing additional corrective feedback to the participant (i.e., no feedback beyond what was provided by the computer was given). For the camp only condition, expected behaviors included the participant only engaging in allowed activities, the examiner not providing the training, and the examiner not initiating social interactions with the participant.

Procedural fidelity was evaluated by members of the research team naïve to study hypotheses. For the pre- and post-test data, these data were collected on 20% of all data collection sessions across all examiners and conditions. For the perceptual training and the camp-only condition, these data were collected on 20% of the sessions across all examiners and participants. Sessions checked for procedural fidelity were chosen by random number generators after the training was concluded; thus, the examiners were unaware of which sessions would be selected for procedural fidelity.

Analytic Plan

A series of regression analyses was run to test: (a) the main effects of the perceptual training on post-test outcomes and (b) the effects of the training on outcomes of interest according to the putative moderators. To assess the main effects of the perceptual training, group assignment was assessed as the independent variable for each dependent variable of interest (i.e., $TBW_{trained}$, $TBW_{novel\ speaker}$, $TBW_{novel\ syllable}$, McGurk fusion). To assess moderated effects, group assignment, the putative moderator (i.e., age, nonverbal IQ, language), and the group * moderator interaction term were assessed as the independent variable for each dependent variable of interest. Thus, for each dependent variable, four total regression models were run (i.e., one to test the main effect of group, one to assess each of the three putative moderators).

Prior to conducting these multiple regression analyses, three variables (i.e., nonverbal IQ, pre- and post-test McGurk fusion) were corrected for negative skew with a square transformation in R (R Core Team, 2020). Training and control groups were then compared on all pre-test metrics using independent samples t-tests; groups did not differ on any variables at pre-test. All regression analyses were completed in in SPSS; moderated multiple regression

models were specifically analyzed using the PROCESS macro (Hayes, 2017). Cook's D was calculated for all regression analyses to monitor outliers. Additionally, Hedges' *g* was calculated for each dependent variable to measure the magnitude of the effects of the perceptual training.

Missing Data

Six participants (two perceptual training and four camp only) were missing discrete data points at either pretest or posttest. Three of the six participants were missing some pre-test data, while all six were missing some post-test data. At pre-test, two participants ran out of time during the testing session, and one participant declined to do one task (TBW$_{novel syllable}$); additionally, two of these participants did not produce a TBW during one SJ task due to (apparent) excessive guessing. One participant did not complete any post-testing due to a medical emergency resulting in hospitalization; additionally, one participant ran out of time during the testing session, and four participants did not produce a TBW during at least one SJ task. Participants with missing data did not significantly differ from participants with complete data in age ($t = 0.81$, $p = 0.446$), nonverbal IQ ($t = 0.85$, $p = 0.423$), language ($t = 1.63$, $p = 0.150$), or biological sex ($\chi^2[1] = 0.09$, $p = 0.765$). Given the varied reasons for missing data and lack of systematic differences among participants with and without missing data, these data can be considered missing at random (a core assumption of multiple imputation methods; Enders 2010; Enders et al., 2014).

Missingness ranged from 0 to 17% across all variables. Of note, there were no missing data for the primary dependent variable, TBW$_{trained}$, at pre-test and only one discrete missing data point at post-test (i.e., the participant in training with a medical emergency). In keeping with current recommendations regarding missing data in moderation analyses (Enders et al., 2014; Zhang & Wang, 2017), product terms were calculated prior to imputing the missing data using the *missForest* package (Stekhoven & Bühlmann, 2012).

Results

Adherence to the assigned condition was very high in both conditions. One participant in the perceptual training condition missed 1 day of the training (i.e., Day 7) due to a family emergency. One participant in the camp only condition missed 2 days (i.e., Days 5 and 7) due to parent illness and car troubles, respectively. Attrition was also very low in both conditions; as previously mentioned, only one participant did not complete their post-testing due to a medical emergency.

Table 2 Pre- and post-test outcomes by group

Dependent variable	Perceptual training condition (n=15)		Camp only control condition (n=15)		Hedges' *g*
	Pre-test M (SD)	Post-test M (SD)	Pre-test M (SD)	Post-test M (SD)	
TBW$_{trained}$	533.0 (213.0)	484.2 (295.8)	498.8 (257.1)	632.2 (311.9)	0.47
TBW$_{novel speaker}$	559.0 (348.4)	551.2 (402.1)	522.5 (291.7)	635.6 (347.9)	0.22
TBW$_{novel syllable}$	539.0 (253.3)	557.3 (379.3)	562.4 (251.7)	735.6 (351.6)	0.47
McGurk	0.65 (0.41)	0.66 (0.38)	0.70 (0.28)	0.71 (0.32)	0.13

TBW = Temporal binding window, trained = stimuli were of a speaker included in the training saying the trained syllable (i.e., "ba"), novel speaker = stimuli were of a speaker not included in the training saying the trained syllable (i.e., "ba"), novel syllable = stimuli were of a speaker included in the training saying a novel syllable (i.e., "pa"), McGurk = proportion of trials wherein participants reported perception of the fused percept (i.e., "ta" or "ha"). Imputed data are presented

Table 3 Results from moderated multiple regression models

Dependent variable	Age			NVIQ			Language		
	β_{group}	β_{age}	$\beta_{interaction}$	β_{group}	β_{NVIQ}	$\beta_{interaction}$	β_{group}	$\beta_{language}$	$\beta_{interaction}$
TBW$_{trained}$	244.5	− 25.8	− 7.3	− 888.6*	− 0.07*	0.08*	− 1025.9*	− 9.3*	12.7*
TBW$_{novel\ speaker}$	382.4	− 27.5	− 21.8	− 1187.5*	0.09**	0.10*	− 1177.8*	− 11.4*	13.6*
TBW$_{novel\ syllable}$	679.3	− 23.7	− 35.9	− 869.6	− 0.09*	0.08*	− 743.7	− 9.5	9.9
McGurk	0.22	0.03	− 0.01	− 0.01	0.00	0.00	0.32	0.01*	0.00

NVIQ = Nonverbal IQ measured by the Leiter International Performance Scale, third edition (Roid et al., 2013), Language = Core language standard scores on the Clinical Evaluation of Language Fundamentals, fourth edition (Semel et al., 2004) or the Preschool Language Scale, fourth edition (Zimmerman et al., 2011), β = unstandardized coefficient in the multiple regression model, TBW = Temporal binding window, trained = stimuli were of a speaker included in the training saying the trained syllable (i.e., "ba"), novel speaker = stimuli were of a speaker not included in the training saying the trained syllable (i.e., "ba"), novel syllable = stimuli were of a speaker included in the training saying a novel syllable (i.e., "pa"), McGurk = proportion of trials wherein participants reported perception of the fused percept (i.e., "ta" or "ha")
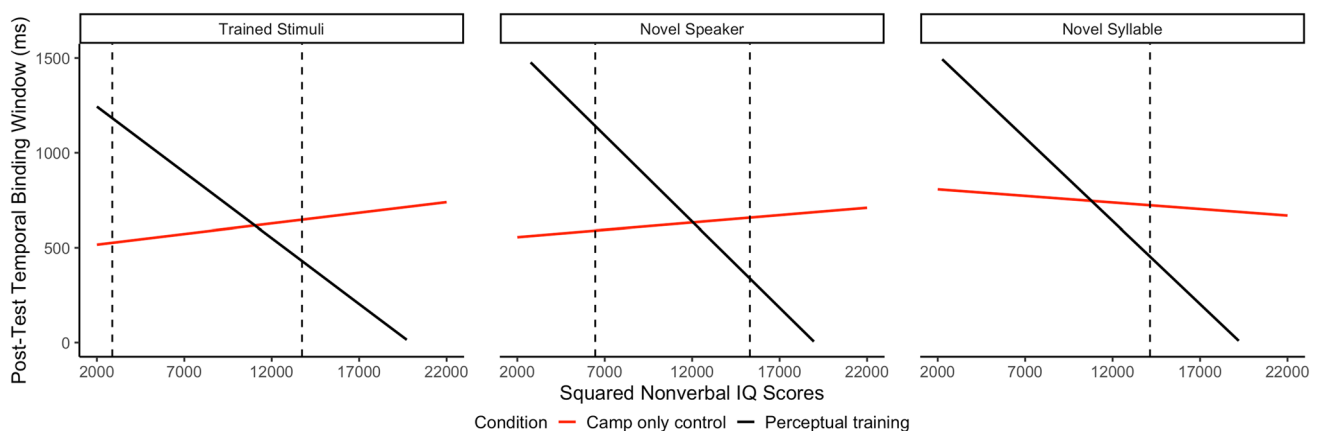
*$p < 0.05$, **$p < 0.01$

Fig. 2 Moderated effect of nonverbal IQ on perceptual training outcomes. *Notes*. Nonverbal IQ scores were derived from the Leiter International Performance Scale, third edition (Roid et al., 2013). Dotted lines represent the cut points identified by the Johnson–Neyman tests. For trained stimuli and the novel speaker stimuli, individuals with nonverbal IQ scores below the dotted line (back-transformed values = 54 and 80, respectively) are likely to experience widening of their temporal binding windows (i.e., a negative or iatrogenic effect of the perceptual training), while individuals with nonverbal IQs above the dotted line (back-transformed values = 117 and 123, respectively) are likely to experience a significant benefit of the perceptual training. For the novel syllable stimuli, individuals with nonverbal IQs above the dotted line (back-transformed value = 118) are likely to experience a significant benefit of the perceptual training

Differences Between Perceptual Training and Camp-Only Control Groups

Pre- and post-test means and standard deviations for all three TBWs (i.e., trained, novel speaker, and novel syllable) and the proportion of reported McGurk illusions according to group are displayed in Table 2. No significant differences between groups were observed, though the unconditional effect of the training (i.e., the group difference without considering covariates or putative moderators) on TBW$_{trained}$ ($\beta = 148.0$, $p = 0.19$, Hedges' $g = 0.47$) and TBW$_{novel\ syllable}$ ($\beta = 178.1$, $p = 0.19$, Hedges' $g = 0.47$) trended in the anticipated direction. These effect sizes were small in magnitude and appeared to be largely driven by widening of the TBW in

the camp only condition rather than narrowing of the TBW in the perceptual training condition. There were additionally no significant unconditional effects of the training on TBW$_{novel\ speaker}$ ($\beta = 84.4$, $p = 0.54$, Hedges' $g = 0.22$) or perception of the McGurk illusion ($\beta = 0.047$, $p = 0.72$, Hedges' $g = 0.13$).

Moderated Effects of Perceptual Training

Effects of the perceptual training, however, varied according to several participant characteristics. Results from all moderated multiple regression models are presented in Table 3.
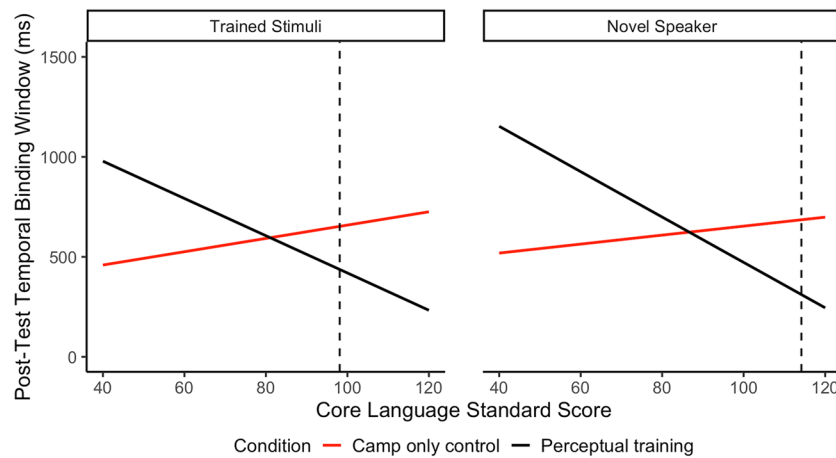
Fig. 3 Moderated effect of language ability on perceptual training outcomes. *Notes.* Core Language Standard Scores were derived from the Clinical Evaluation of Language Fundamentals, fourth edition (Semel et al., 2004) or the Preschool Language Scale, fourth edition (Zimmerman et al., 2011). Dotted lines represent the cut points iden- tified by the Johnson–Neyman tests (standard scores = 98 and 114, respectively); above those points along the continuous moderator, the perceptual training causes a significant reduction in temporal binding window

Age

Age did not moderate the effect of training on any of the outcomes of interest (*p* values for interaction term in the multiple regression models > 0.3).

Nonverbal IQ

Nonverbal IQ significantly moderated the effect of the per- ceptual training on all of the TBW outcomes (*p* values for interaction term in the multiple regression models < 0.05; see Table 3; Fig. 2). For TBW$_{trained}$, Johnson–Neyman tests utilized to derive precise cut points along the continuous moderator of squared nonverbal IQ scores indicated that the training resulted in a significant reduction in TBW$_{trained}$ for individuals with nonverbal IQ scores above 117 and that there was a significant widening of TBWs (i.e., a negative or iatrogenic effect) for individuals with nonverbal IQ scores below 54. Similar results were also found for TBW$_{novel speaker}$, wherein a benefit of training was observed for individuals with nonverbal IQ scores above 123, and a significant widen- ing of TBWs was observed for individuals with nonverbal IQ scores below 80. For TBW$_{novel syllable}$, results of the John- son–Neyman tests indicated a significant benefit of training for individuals with nonverbal IQs above 118. Nonverbal IQ did not moderate the effect of the perceptual training on report of the McGurk illusion.

Language

Language scores also significantly moderated the effect of the perceptual training on TBW$_{trained}$ and TBW$_{novel speaker}$ out- comes (*p* values for interaction terms in the multiple regres- sion models < 0.05; see Table 3). Johnson–Neyman tests indicated that the training resulted in a significant reduction in TBW$_{trained}$ for individuals with language standard scores above 98 and a significant reduction in TBW$_{novel speaker}$ for individuals with language standard scores above 114 (see Fig. 3). Language did not moderate the effect of the per- ceptual training on either TBW$_{novel syllable}$ or report of the McGurk illusion.

Procedural Fidelity

Procedural fidelity was checked for 20% of WhisperRoom sessions for both groups (*n* = 60 sessions) by an observer naïve to the hypotheses. The perceptual training was admin- istered with 98.5% fidelity, and the camp-only session was administered with 100% fidelity. The average fidelity was very high for all four of the examiners who administered these sessions (98.7–100%). Of the 60 sessions, 20% (*n* = 12 sessions) were rated by two naïve observers; the agreement was excellent (ICC = 0.86).

Procedural fidelity was also checked for 23% of the pre- and post-testing sessions (*n* = 29 sessions) by an observer naïve to group assignment and hypotheses. The average fidelity was very high overall at 98.0% and across all six assessors (range = 91.7–100%) and did not differ according to condition (*p* = 0.82; 98.1% for participants assigned to

perceptual training versus 97.9% for participants assigned to camp-only control) or timepoint ($p = 0.83$; 97.9% at pre-test versus 98.1% at post-test).

Social Validity

Participants who completed the perceptual training on average reported that the training was neither easy nor hard ($M = 2.8$) and neither fun nor boring ($M = 2.9$). Most of the participants also rated the training as "kind of helpful" ($M = 2.2$), though three participants responded that they weren't sure how helpful the activity was. None of the participants reported that they would do the training in their free time, though three were unsure.

The parent report survey was collected from 19 parents (8 perceptual training, 11 camp only). Parents in both groups reported on average that they noticed between no change and a slight positive change in their children's social interactions ($M = 2.6$ and 2.3 for perceptual training and camp only, respectively), language ($M = 2.5$ for both groups), and communication ($M = 2.5$ for both groups). In regards to their children's behavior, parents in the camp only condition on average reported a slight positive change ($M = 2.1$) while parents in the perceptual training reported somewhere between no change and a slight positive change ($M = 2.6$). Only two parents (1 perceptual training, 1 control) reported a slight negative change in their children's behavior; in both cases, they reported no change in the other domains.

Discussion

The purpose of the present study was to assess a computer-based perceptual training program designed to narrow TBWs for audiovisual speech in autistic youth. On average, participants assigned to the perceptual training did not differ from the participants assigned to the camp-only control condition at post-test on any of the dependent variables, though unconditional effects did trend towards a narrowing of TBWs for trained stimuli and novel syllable stimuli. Importantly, though, effects of the training varied according to participant characteristics, such that youth who had average to above average language and cognitive ability appeared to benefit from the perceptual training paradigm, but youth who were less cognitively or linguistically able displayed lesser benefit and/or even adverse effects when assigned to the perceptual training condition.

Unconditional Effects of Perceptual Training Were Small and Non-significant

Unconditional effects of the perceptual training versus control condition were non-significant. Notably, even the few

small effect sizes trending in favor of the perceptual training program in the data across all participants appeared to be driven largely by an increase in TBW in the camp only control group: though participants in the perceptual training condition decreased in $TBW_{trained}$ by 49.3 ms on average, participants in the camp only condition on average increased their $TBW_{trained}$ by 133.4 ms. Previous studies have reported a similar increase in TBW following exposure to asynchronous speech and SJ tasks (e.g., Feldman et al., 2020; Powers et al., 2009), and participants in both conditions completed over 1000 SJ trials as part of the testing procedure. These results suggest that the tested perceptual training paradigm does not yield favorable effects on temporal binding of audiovisual speech across all youth on the autism spectrum.

Outcomes are Moderated by Nonverbal IQ and Language Ability

Results from multiple regression models indicated, however, that the perceptual training paradigm narrowed TBWs on trained stimuli in some autistic youth, specifically those with above average nonverbal IQ and average language abilities. Feldman et al., (2020) previously observed highly variable responses to a similar intervention in autistic children and hypothesized that some individual differences might have contributed to these differential responses, but this study is the first to statistically test factors that moderate the effects of perceptual training on TBWs. Notably, all participants in the prior study had nonverbal IQs that were between 93 and 108 (i.e., within the range where individuals would be unlikely to benefit from the perceptual training), possibly explaining why a functional relation (i.e., a clear effect of the intervention) was not observed in the previous investigation of this perceptual training.

Equally importantly are the implications of this study for intervention sciences. This study indicated which individuals were unlikely to derive benefit from the perceptual training (i.e., individuals with nonverbal IQs between 55 and 116 and language standard scores below 98) and which individuals were likely to experience negative or iatrogenic effects as a result of this training paradigm (i.e., individuals with nonverbal IQs below 55). Thus, future studies targeting audiovisual integration in autistic youth with low nonverbal cognitive abilities should either utilize different intervention techniques or implement further adaptations to the perceptual training for this population.

Although outcomes were moderated by both nonverbal IQ and language and Johnson–Neyman tests identified different cut-points for these two moderators, it is notable that these scores were highly correlated in the present sample ($r = 0.70$, $p < 0.001$). The intercorrelation between these scores does limit our ability to determine which factor may

most influence responses to treatment; nevertheless, it is paramount that practitioners understand for whom interventions may be most effective given the heterogeneous nature of autism. This study critically adds to a growing body of literature suggesting that interventions for autistic individuals may be most effective for subsets of the population with certain characteristics (e.g., Carter et al., 2011; Ledford et al., 2016; Marcus et al., 2001; Sandbank et al., 2020; Vismara & Rogers, 2010; Yoder & Compton, 2004). Though previous reviews of technology-based interventions have largely failed to find effects on core and related features of autism, it is notable that prior research on interventions mediated through technology have generally not targeted or measured effects on sensory function (i.e., have focused on social communication targets such as social skills or emotion recognition; Barton et al., 2017; Fletcher-Watson, 2014; Grynszpan et al., 2013) and have not considered individual characteristics that may moderate the effects of such intervention. The present findings underscore the need for future trials of candidate interventions geared towards autistic youth to consider the phenotypic variation that may lead to differential response to treatment, and to employ study designs and analytic approaches that allow for such moderated effects to be evaluated.

Some Evidence for Generalization in Individuals with Higher Nonverbal IQ and Language Ability

Moderation models also indicated that there was some generalization to untrained stimuli in individuals with above average nonverbal IQ and language abilities. This study is the first to indicate that perceptual trainings for audiovisual speech stimuli can generalize to untrained speakers and untrained syllables. Importantly, the language standard score identified as a cut-off by the Johnson–Neyman test for $TBW_{novel\ speaker}$ (114) was one standard deviation higher than the cut-off score for $TBW_{trained}$ (98). A similar pattern was also observed for nonverbal IQ scores, as the cutoff score for likely benefit for $TBW_{novel\ speaker}$ (123) was half a standard deviation higher than the cut-off score for $TBW_{trained}$ (117). Thus, generalization to untrained stimuli, specifically untrained speakers, requires even higher language and cognitive abilities than required to derive any benefit from the training. Additionally, the perceptual training appeared to be more likely to induce widening (i.e., iatrogenic effects) on untrained speakers for individuals with below average nonverbal IQs, further limiting the profile of individuals likely to benefit from the perceptual training. It is not surprising that this perceptual training paradigm requires high nonverbal IQ and average language ability in order for participants to improve their temporal binding of audiovisual speech, given the complexity inherent to the perceptual training

(e.g., Cascio et al., 2016; Feldman et al., 2018; Feldman, Kuang, et al., 2019; Woynaroski et al., 2013).

In the present study, evidence for generalization was limited to SJ tasks, as no effect of the perceptual training was observed for the McGurk effect on any subset of the participants. This finding accords with previous studies of TD adults that found limited to no evidence for generalization to untrained multisensory tasks (e.g., De Niear et al., 2018; Powers et al., 2016; Setti et al., 2014; Zerr et al., 2019). Notably, our generalization task (i.e., measuring the McGurk effect) differed in both the instructions given to participants and stimuli, whereas the generalization task utilized by Sürig et al., (2018; i.e., the previous study on non-autistic adults that found evidence for generalization to untrained tasks) differed only in the instructions given to participants. Given that the McGurk task as employed here did not measure temporal properties of multisensory integration, it is difficult to conclude from the present study whether the candidate perceptual training improved temporal aspects of audiovisual integration that could be detected beyond the specific task (i.e., simultaneity judgements) utilized in the context of training and outcome measurement. Future studies may wish to assess effects of perceptual training on temporal processing of multisensory information utilizing other stimuli (e.g., an SJ task with flash and beep stimuli) and other tasks (e.g., temporal order judgment tasks) or to evaluate effects of the training on broader multisensory integration (e.g., inverse effectiveness via listening in noise; Foxe et al., 2015; Ross et al., 2006; spatial localization; Sürig et al., 2018). Such work would advance our understanding of the degree to which this training has the potential to yield more distal and generalized effects for the subgroup of youth who appear, based on the present results, to derive some benefit.

Social Validity Data Suggest Largely Neutral Impressions Regarding Perceptual Training

On average, participants in the perceptual training reported that the training was neither easy nor hard, neither fun nor boring, and kind of helpful. Though none of the participants reported that they wanted to "play the game" in their free time, several participants did note that they liked how they could "set a goal" for themselves using the scoring system. Additionally, none of the participants reported being confused or frustrated by the training, indicating that the explicit feedback provided in this new instantiation may have improved the training at least to some degree over the previous iteration, though anecdotally several participants did still appear frustrated or confused during the latter days of the training.

Parents reported roughly equal changes in their children's behavior in both groups. Though parents' positive perceptions of both conditions likely had more to do with

the activities done outside the context of the study as a part of the larger research camp, it is important that parents reported, on average. no change or slightly positive changes in their children, given that there were some iatrogenic effects of the training and widening of TBWs in the camp only participants.

Although perceptions of the training largely represent an improvement from prior work (i.e., Feldman et al., 2020), participants and their parents still did not report that the goals and outcomes of the training are meaningful. Future studies should evaluate the attitudes and perceptions of autistic self-advocates, particularly those with higher cognitive and language abilities, towards the perceptual training and evaluate how the perceptual training might be further modified to better meet the needs of this community. Adapting the perceptual training into activities that are meaningful to participants may increase the likelihood for sensory-based neuroplasticity (see Lane & Schaaf 2010), and thus narrowing of the TBW.

Limitations and Future Directions

There are several limitations of the present study. First, the sample size of this study was small, which may have limited our ability to detect effects of interest. The study was further limited by the use of non-concurrent language and cognitive testing, and the concatenation of language scores from multiple measures (i.e., four participants were administered the PLS-5; 25 participants were administered the CELF-4). However, this limitation is mitigated by the previously demonstrated stability of language and cognitive abilities in this age-span (e.g., Bornstein et al., 2014, 2016a, 2016b; Eaves & Ho, 1996; Lord & Schopler, 1989; Norbury et al., 2017; Pickles et al., 2014; Schneider et al., 2014). Additionally, the pre- and post-test data collection required rather long testing sessions to obtain stable estimates for TBW variables, which perhaps caused testing, fatigue, and/or exposure effects in at least some participants.

Though the present study does demonstrate some moderated effects of the training, it is unclear whether there are any factors that mediate the effect of the training. The cascading effects hypothesis posits that sensory interventions may improve behavior via altered neural processing (Cascio et al., 2016), and while perceptual training has been shown to alter neural function in non-autistic adults (La Rocca et al., 2020; Powers et al., 2012), no study to date has assessed whether altered neural function is the mechanism by which the training alters perceptual abilities. Theory would also suggest that audiovisual speech trainings may also be mediated by altered patterns of looking during audiovisual speech processing. One might expect training paradigms to narrow TBWs for audiovisual speech via increased attention to the mouth, the source of multisensory

redundancy; increased looking to the mouth has been linked to increased language and prelinguistic communication in children diagnosed with or at high-familial risk for autism (Santapuram et al., 2022; Woynaroski et al., 2019). Alternatively, these training paradigms may facilitate increased looking to the eyes, which is a sign of mature audiovisual processing (Lewkowicz & Hansen-Tift, 2012; Soto-Faraco et al., 2012) and is associated with narrower TBWs in autistic children and non-autistic peers (Liu et al., 2020). No study to date has assessed whether looking patterns during audiovisual speech are modified by audiovisual training for speech stimuli. Future work should evaluate whether neural processing of audiovisual speech (e.g., the P3b waveform, believed to represent evidence accumulation in the context of decision making; Twomey et al., 2015) or attention to the regions of the face during audiovisual speech mediates intervention outcomes.

The evidence here for moderated effects suggest two divergent paths for further research into perceptual trainings for temporal binding of audiovisual speech. First, future work must further evaluate whether such perceptual trainings yield more distal effects for autistic youth with high nonverbal IQ and average language ability. Though results of this study indicate that this perceptual training results in improvements in TBWs for trained stimuli that may generalize to at least some untrained stimuli in this population, these perceptual trainings must result in at least some gains in distal outcomes deemed critical by the autistic community (e.g., improvements on language, social communication, or behavioral responses to sensory stimuli) to maximize their utility. Evidence of generalization to distal effects such as language or social communication would provide increased support for "sensory-first" hypotheses of autism, which posit that sensory differences emerge early and contribute to or cause the core differences observed in autism (Cascio et al., 2016; Robertson & Baron-Cohen, 2017; Wallace et al., 2020). Although it is unlikely that the brief perceptual training would result in significant improvements in broader features of autism, it is possible that the training may result in slight improvements in some aspects of language and communication. Thus, future studies should endeavor to assess effects of perceptual training on language and communication changes via standardized behavioral samples at post-test.

Future work must also evaluate treatment approaches for audiovisual speech perception for autistic youth with below average language and cognitive ability. For example, future research could work to reduce the language and cognitive requirements of extant perceptual training paradigms to best reach these individuals, who are arguably most likely to benefit from or need these types of interventions. It is possible that setting the maximum SOA for training stimuli at ±500ms may have been too challenging for some of the participants; using a higher maximum training SOA (e.g.,

$\pm$ 900ms, which was used as the comprehension check) may better scaffold narrowing for the subset of the sample who experienced iatrogenic effects. Alternatively, future research could develop and assess novel approaches to treatment that may improve audiovisual speech perception (e.g., Tenenbaum et al., 2017).

Summary

The brief computer-based perceptual training for temporal binding of asynchronous audiovisual speech resulted in small but non-significant changes in the TBW in autistic youth, on average, compared to a group of participants assigned to a camp-only control condition. Effects of the training program varied according to participant profiles, however, with significant effects in favor of the training apparent for participants with nonverbal IQs above 117 and language standard scores above 98. There was also evidence for generalization to untrained stimuli in the subgroup of participants with above average language and nonverbal IQ scores. However, participants with nonverbal IQs below 54 were likely to experience widening of their TBW on trained stimuli, and participants with nonverbal IQs below 80 were likely to experience widening of their TBW on untrained stimuli. Thus, the candidate training paradigm is contraindicated for autistic youth diagnosed with co-morbid intellectual impairments. Future studies should evaluate (a) whether factors such as neural processing of audiovisual speech and/or attention to regions of the face during audiovisual speech mediate outcomes of the perceptual training, (b) whether perceptual trainings can improve more distal outcomes in autistic youth with higher cognitive and language ability, and (c) novel approaches to improving audiovisual speech perception in autistic youth with lower cognitive and language ability.

Declarations

References

American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders-5

Bahrick, L. E., & Todd, J. T. (2012). Multisensory processing in autism spectrum disorders: Intersensory processing disturbance as atypical development. In B. E. Stein (Ed.), *The new handbook of multisensory processes* (pp. 657–674). MIT Press.

Barton, E. E., Pustejovsky, J. E., Maggin, D. M., & Reichow, B. (2017). Technology-aided instruction and intervention for students with ASD: A meta-analysis using novel methods of estimating effect sizes for single-case research. *Remedial and Special Education, 38*(6), 371–386. https://doi.org/10.1177/0741932517729508

Basu Mallick, D., Magnotti, J. F., & Beauchamp, M. S. (2015). Variability and stability in the McGurk effect: Contributions of participants, stimuli, time, and response type. *Psychonomic Bulletin & Review, 22*, 1299–1307. https://doi.org/10.3758/s13423-015-0817-4

Baum, S. H., Stevenson, R. A., & Wallace, M. T. (2015). Behavioral, perceptual, and neural alterations in sensory and multisensory function in autism spectrum disorder. *Progress in Neurobiology, 134*, 140–160. https://doi.org/10.1016/j.pneurobio.2015.09.007

Bornstein, M. H., Hahn, C. S., & Putnick, D. L. (2016a). Long-term stability of core language skill in children with contrasting language skills. *Developmental Psychology, 52*(5), 704–716. https://doi.org/10.1037/dev0000111

Bornstein, M. H., Hahn, C. S., & Putnick, D. L. (2016b). Stability of core language skill across the first decade of life in children at biological and social risk. *Journal of Child Psychology and Psychiatry, 57*(12), 1434–1443. https://doi.org/10.1111/jcpp.12632

Bornstein, M. H., Hahn, C. S., Putnick, D. L., & Suwalsky, J. T. D. (2014). Stability of core language skill from early childhood to adolescence: A latent variable approach. *Child Development, 85*(4), 1346–1356. https://doi.org/10.1111/cdev.12192

Bottema-Beutel, K., Kapp, S. K., Lester, J. N., Sasson, N. J., & Hand, B. N. (2021). Avoiding ableist language: Suggestions for autism researchers. *Autism in Adulthood, 3*(1), 18–29. https://doi.org/10.1089/aut.2020.0014

Carter, A. S., Messinger, D. S., Stone, W. L., Celimli, S., Nahmias, A. S., & Yoder, P. (2011). A randomized controlled trial of Hanen's 'More Than Words' in toddlers with early autism symptoms. *Journal of Child Psychology and Psychiatry, 52*(7), 741–752. https://doi.org/10.1111/j.1469-7610.2011.02395.x

Cascio, C. J., Woynaroski, T., Baranek, G. T., & Wallace, M. T. (2016). Toward an interdisciplinary approach to understanding

sensory function in autism spectrum disorder. *Autism Research, 9*, 920–925. https://doi.org/10.1002/aur.1612

Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology, 16*(2), 137–163. https://doi.org/10.1111/j.2044-8317.1963.tb00206.x

De Niear, M. A., Gupta, P. B., Baum, S. H., & Wallace, M. T. (2018). Perceptual training enhances temporal acuity for multisensory speech. *Neurobiology of Learning and Memory, 147*, 9–17. https://doi.org/10.1016/j.nlm.2017.10.016

De Niear, M. A., Koo, B., & Wallace, M. T. (2016). Multisensory perceptual learning is dependent upon task difficulty. *Experimental Brain Research, 234*, 3269–3277. https://doi.org/10.1007/s00221-016-4724-3

Dunham, K., Feldman, J. I., Liu, Y., Cassidy, M., Conrad, J. G., Santapuram, P., Suzman, E., Tu, A., Butera, I. M., Simon, D. M., Broderick, N., Wallace, M. T., Lewkowicz, D. J., & Woynaroski, T. G. (2020). Stability of variables derived from measures of multisensory function in children with autism spectrum disorder. *American Journal of Intellectual and Developmental Disabilities, 125*(4), 287–303.

Eaves, L. C., & Ho, H. H. (1996). Brief report: Stability and change in cognitive and behavioral characteristics of autism through childhood. *Journal of Autism and Developmental Disabilities, 26*(5), 557–569. https://doi.org/10.1007/bf02172276

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.

Enders, C. K., Baraldi, A. N., & Cham, H. (2014). Estimating interaction effects with incomplete predictor variables. *Psychological Methods, 19*(1), 39–55. https://doi.org/10.1037/a0035314

Feldman, J. I., Cassidy, M., Conrad, J. G., Dunham, K., Kuang, W., Liu, Y., Simon, D. M., Tu, A., Wallace, M. T., & Woynaroski, T. G. (2019a). *Audiovisual multisensory integration in children with autism spectrum disorder: Associations with broader symptomatology* [Poster presentation]. Gatlinburg Conference on Research and Theory in Intellectual Disabilities, San Antonio, TX, United States

Feldman, J. I., Conrad, J. G., Kuang, W., Tu, A., Liu, Y., Simon, D. M., Wallace, M. T., & Woynaroski, T. G. (2022). Relations between the McGurk effect, social and communication skill, and autistic features in children with and without autism. *Journal of Autism and Developmental Disorders*, *52*, 1920–1928. https://doi.org/10.1007/s10803-021-05074-w

Feldman, J. I., Dunham, K., Cassidy, M., Wallace, M. T., Liu, Y., & Woynaroski, T. G. (2018). Audiovisual multisensory integration in individuals with autism spectrum disorder: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews, 95*, 220–234. https://doi.org/10.1016/j.neubiorev.2018.09.020

Feldman, J. I., Dunham, K., Conrad, J. G., Simon, D. M., Cassidy, M., Liu, Y., Tu, A., Broderick, N., Wallace, M. T., & Woynaroski, T. G. (2020). Plasticity of temporal binding in children with autism spectrum disorder: A single case experimental design perceptual training study. *Research in Autism Spectrum Disorders, 74*, 101555. https://doi.org/10.1016/j.rasd.2020.101555

Feldman, J. I., Kuang, W., Conrad, J. G., Tu, A., Santapuram, P., Simon, D. M., Foss-Feig, J. H., Kwakye, L. D., Stevenson, R. A., Wallace, M. T., & Woynaroski, T. G. (2019b). Brief report: Differences in multisensory integration covary with differences in sensory responsiveness in children with and without autism spectrum disorder. *Journal of Autism and Developmental Disabilities, 49*, 397–403. https://doi.org/10.1007/s10803-018-3667-x

Fletcher-Watson, S. (2014). A targeted review of computer-assisted learning for people with autism spectrum disorder: Towards a consistent methodology. *Review Journal of Autism and Developmental Disorders, 1*(2), 87–100. https://doi.org/10.1007/s40489-013-0003-4

Foster, S. L., & Mash, E. J. (1999). Assessing social validity in clinical treatment research: Issues and procedures. *Journal of Consulting and Clinical Psychology, 67*(3), 308–319. https://doi.org/10.1037/0022-006X.67.3.308

Foxe, J. J., Molholm, S., Bene, D., Frey, V. A., Russo, H. P., Blanco, N. N., Saint-Amour, D., & Ross, L. A. (2015). Severe multisensory speech integration deficits in high-functioning school-aged children with autism spectrum disorder (ASD) and their resolution during early adolescence. *Cerebral Cortex, 25*(2), 298–312. https://doi.org/10.1093/cercor/bht213

Gast, D. L., & Ledford, J. R. (2014). *Single case research methodology: Applications in special education and behavioral sciences* (2nd ed.). Routledge.

Gernsbacher, M. A. (2017). Editorial perspective: The use of person-first language in scholarly writing may accentuate stigma. *Journal of Child Psychology and Psychiatry, 58*(7), 859–861. https://doi.org/10.1111/jcpp.12706

Grynszpan, O., Weiss, P. L., Perez-Diaz, F., & Gal, E. (2013). Innovative technology-based interventions for autism spectrum disorders: A meta-analysis. *Autism, 18*(4), 346–361. https://doi.org/10.1177/1362361313476767

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, *42*, 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). Guilford Press.

Irwin, J., Preston, J., Brancazio, L., D'angelo, M., & Turcios, J. (2015). Development of an audiovisual speech perception app for children with autism spectrum disorders. *Clinical Linguistics & Phonetics, 29*, 76–83. https://doi.org/10.3109/02699206.2014.966395

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism, 20*(4), 442–462. https://doi.org/10.1177/1362361315588200

La Rocca, D., Ciuciu, P., Engemann, D. A., & van Wassenhove, V. (2020). Emergence of β and γ networks following multisensory training. *Neuroimage*. https://doi.org/10.1016/j.neuroimage.2019.116313

Lane, S. J., & Schaaf, R. C. (2010). Examining the neuroscience evidence for sensory-driven neuroplasticity: Implications for sensory-based occupational therapy for children and adolescents. *The American Journal of Occupational Therapy, 64*(3), 375–390. https://doi.org/10.5014/ajot.2010.09069

Ledford, J. R., Barton, E. E., Severini, K. E., & Zimmerman, K. N. (2019). A primer on single-case research designs: Contemporary use and analysis. *American Journal on Intellectual and Developmental Disabilities, 124*, 35–56. https://doi.org/10.1352/1944-7558-124.1.35

Ledford, J. R., King, S., Harbin, E. R., & Zimmerman, K. N. (2016). Antecedent social skills interventions for individuals with ASD: What works, for whom, and under what conditions? *Focus on Autism and Other Developmental Disabilities, 33*(1), 3–13. https://doi.org/10.1177/1088357616634024

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences, 109*(5), 1431–1436. https://doi.org/10.1073/pnas.1114783109

Liu, Y., Feldman, J. I., Wallace, M. T., Lewkowicz, D. J., & Woynaroski, T. G. (2020). *Atypical multisensory integration is linked with immature gaze patterns across groups of children with and without autism spectrum disorder* [Poster presentation].

Gatlinburg Conference on Research and Theory in Intellectual Disabilities, San Diego, CA, United States (Conference canceled)

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. L. (2012). *Autism diagnostic observation schedule, second edition (ADOS-2) manual (Part I): Modules 1–4*. Western Psychological Services.

Lord, C., & Schopler, E. (1989). The role of age at assessment, developmental level, and test in the stability of intelligence scores in young autistic children. *Journal of Autism and Developmental Disabilities, 19*(4), 483–499. https://doi.org/10.1007/bf02212853

Marcus, L. M., Garfinkle, A., & Wolery, M. (2001). Issues in early diagnosis and intervention with young children with autism. In E. Schopler, N. Yirmiya, C. Shulman, & L. M. Marcus (Eds.), *The research basis for autism intervention* (pp. 171–185). Springer. https://doi.org/10.1007/0-306-47946-X_12

McGovern, D. P., Roudaia, E., Newell, F. N., & Roach, N. W. (2016). Perceptual learning shapes multisensory causal inference via two distinct mechanisms. *Scientific Reports, 6*, 24673. https://doi.org/10.1038/srep24673

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746–748. https://doi.org/10.1038/264746a0

Noel, J. P., De Niear, M. A., Stevenson, R., Alais, D., & Wallace, M. T. (2017). Atypical rapid audio-visual temporal recalibration in autism spectrum disorders. *Autism Research, 10*, 121–129. https://doi.org/10.1002/aur.1633

Norbury, C. F., Vamvakas, G., Gooch, D., Baird, G., Charman, T., Simonoff, E., & Pickles, A. (2017). Language growth in children with heterogeneous language disorders: a population study. *Journal of Child Psychology and Psychiatry, 58*(10), 1092–1105. https://doi.org/10.1111/jcpp.12793

Pickles, A., Anderson, D. K., & Lord, C. (2014). Heterogeneity and plasticity in the development of language: A 17-year follow-up of children referred early for possible autism. *Journal of Child Psychology and Psychiatry, 55*(12), 1354–1362. https://doi.org/10.1111/jcpp.12269

Powers, A. R., Hevey, M. A., & Wallace, M. T. (2012). Neural correlates of multisensory perceptual learning. *The Journal of Neuroscience, 32*, 6263–6274. https://doi.org/10.1523/JNEUROSCI.6138-11.2012

Powers, A. R., Hillock-Dunn, A., & Wallace, M. T. (2016). Generalization of multisensory perceptual learning. *Scientific Reports, 6*, 23374. https://doi.org/10.1038/srep23374

Powers, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *Journal of Neuroscience, 29*, 12265–12274. https://doi.org/10.1523/JNEUROSCI.3501-09.2009

Raymaker, D., & Nicolaidis, C. (2013). Participatory research with autistic communities: Shifting the system. In J. Davidson & M. Orsini (Eds.), *Worlds of autism: Across the spectrum of neurological difference* (pp. 169–188). University of Minnesota Press. https://doi.org/10.5749/j.ctt4cggsn.11

Raymaker, D. M. (2019). Reclaiming research for the autistic adult community. *Autism in Adulthood, 1*(3), 160–161. https://doi.org/10.1089/aut.2019.29005.dra

R Core Team (2020). *R: A language and environment for statistical computing*. (Version 4.0.2) Vienna, Austria. https://www.R-project.org/

Reid, D. K., Hresko, W. P., & Hammill, D. D. (2001). *Test of early reading ability* (3rd ed.). Pro-Ed.

Robertson, C. E., & Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews Neuroscience, 18*, 671–684. https://doi.org/10.1038/nrn.2017.112

Robison, J. E. (2019). Talking about autism—Thoughts for researchers. *Autism Research, 12*(7), 1004–1006. https://doi.org/10.1002/aur.2119

Roid, G. H., Miller, L. J., Pomplun, M., & Koch, C. (2013). *Leiter international performance scale* (3rd ed.). Western Psychological Services.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2006). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex, 17*, 1147–1153. https://doi.org/10.1093/cercor/bhl024

Sandbank, M., Bottema-Beutel, K., Crowley, S., Cassidy, M., Feldman Jacob, I., Canihuante, M., & Woynaroski, T. (2020). Intervention effects on language in children with autism: A project AIM meta-analysis. *Journal of Speech Language and Hearing Research, 63*(5), 1537–1560. https://doi.org/10.1044/2020_JSLHR-19-00167

Sandbank, M., & Yoder, P. (2014). Measuring representative communication in young children with developmental delay. *Topics in Early Childhood Special Education, 34*, 133–141. https://doi.org/10.1177/0271121414528052

Santapuram, P., Feldman, J. I., Bowman, S. M., Raj, S., Suzman, E., Crowley, S., Kim, S. Y., Keceli-Kaysili, B., Bottema-Beutel, K., Lewkowicz, D. J., Wallace, M. T., & Woynaroski, T. G. (2022). Mechanisms by which early eye gaze to the mouth during multisensory speech influences expressive communication development in infant siblings of children with and without autism. *Mind Brain and Education, 16*(1), 62–74. https://doi.org/10.1111/mbe.12310

Schaaf, R. C., & Lane, A. E. (2015). Toward a best-practice protocol for assessment of sensory features in ASD. *Journal of Autism and Developmental Disorders, 45*, 1380–1395. https://doi.org/10.1007/s10803-014-2299-z

Schauder, K. B., & Bennetto, L. (2016). Toward an interdisciplinary understanding of sensory dysfunction in autism spectrum disorder: An integration of the neural and symptom literatures. *Frontiers in Neuroscience, 10*, 268. https://doi.org/10.3389/fnins.2016.00268

Schneider, W., Niklas, F., & Schmiedeler, S. (2014). Intellectual development from early childhood to early adulthood: The impact of early IQ differences on stability and change over time. *Learning and Individual Differences, 32*, 156–162. https://doi.org/10.1016/j.lindif.2014.02.001

Semel, E. M., Wiig, E. H., & Secord, W. A. (2004). *Clinical evaluation of language fundamentals* (4th ed.). The Psychological Corporation.

Setti, A., Stapleton, J., Leahy, D., Walsh, C., Kenny, R. A., & Newell, F. N. (2014). Improving the efficiency of multisensory integration in older adults: Audio-visual temporal discrimination training reduces susceptibility to the sound-induced flash illusion. *Neuropsychologia, 61*, 259–268. https://doi.org/10.1016/j.neuropsychologia.2014.06.027

Simon, D. M., & Wallace, M. T. (2018). Integration and temporal processing of asynchronous audiovisual speech. *Journal of Cognitive Neuroscience*, *30*, 319–337. https://doi.org/10.1162/jocn_a_01205

Smith, E. G., Zhang, S., & Bennetto, L. (2017). Temporal synchrony and audiovisual integration of speech and object stimuli in autism. *Research in Autism Spectrum Disorders, 39*, 11–19. https://doi.org/10.1016/j.rasd.2017.04.001

Soto-Faraco, S., Calabresi, M., Navarra, J., Werker, J. F., & Lewkowicz, D. J. (2012). The development of audiovisual speech perception. In A. J. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 207–228). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199586059.003.0009

Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics, 28*, 112–118. https://doi.org/10.1093/bioinformatics/btr597

Stevenson, R. A., Segers, M., Ncube, B. L., Black, K. R., Bebko, J. M., Ferber, S., & Barense, M. D. (2018). The cascading influence of multisensory processing on speech perception in autism. *Autism, 22*, 609–624. https://doi.org/10.1177/1362361317704413

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *Journal of Neuroscience, 34*, 691–697. https://doi.org/10.1523/JNEUROSCI.3615-13.2014

Stokes, T. F., & Osnes, P. G. (1989). An operant pursuit of generalization. *Behavior Therapy, 20*(3), 337–355. https://doi.org/10.1016/S0005-7894(89)80054-1

Sürig, R., Bottari, D., & Röder, B. (2018). Transfer of audio-visual temporal training to temporal and spatial audio-visual tasks. *Multisensory Research, 31*, 556–578. https://doi.org/10.1163/22134808-0000261

Swan, A. J., Carper, M. M., & Kendall, P. C. (2016). In pursuit of generalization: An updated review. *Behavior Therapy, 47*(5), 733–746. https://doi.org/10.1016/j.beth.2015.11.006

Tenenbaum, E. J., Amso, D., Righi, G., & Sheinkopf, S. J. (2017). Attempting to "increase intake from the input": Attention and word learning in children with autism. *Journal of Autism and Developmental Disorders, 47*(6), 1791–1805. https://doi.org/10.1007/s10803-017-3098-0

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience, 42*, 1636–1643. https://doi.org/10.1111/ejn.12936

Vismara, L. A., & Rogers, S. J. (2010). Behavioral treatments in autism spectrum disorder: What do we know? *Annual Review of Clinical Psychology, 6*(1), 447–468. https://doi.org/10.1146/annurev.clinpsy.121208.131151

Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia, 64*, 105–123. https://doi.org/10.1016/j.neuropsychologia.2014.08.005

Wallace, M. T., Woynaroski, T. G., & Stevenson, R. A. (2020). Multisensory integration as a window into orderly and disrupted cognition and communication. *Annual Review of Psychology, 71*, 193–219. https://doi.org/10.1146/annurev-psych-010419-051112

Warner, G., Parr, J. R., & Cusack, J. (2018). Workshop report: Establishing priority research areas to improve the physical health and well-being of autistic adults and older people. *Autism in Adulthood, 1*(1), 20–26. https://doi.org/10.1089/aut.2018.0003

Wiederholt, J. L., & Bryant, B. R. (2012). *Gray oral reading tests* (5th ed.). Pro-Ed.

Williams, J. H. G., Massaro, D. W., Peel, N. J., Bosseler, A., & Suddendorf, T. (2004). Visual–auditory integration during speech imitation in autism. *Research in Developmental Disabilities, 25*, 559–575. https://doi.org/10.1016/j.ridd.2004.01.008

Woynaroski, T. G., Feldman, J. I., Edmunds, S. R., Simon, D. M., Tu, A., Kuang, W., Conrad, J. G., Santapuram, P., Broderick, N., & Wallace, M. T. (2019). *Audiovisual speech processing and attentionare linked with language in children with and without autism* [Paperpresentation]. Society for Research in Child Development, Baltimore, MD, UnitedStates

Woynaroski, T. G., Kwakye, L. D., Foss-Feig, J. H., Stevenson, R. A., Stone, W. L., & Wallace, M. T. (2013). Multisensory speech perception in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 43*, 2891–2902. https://doi.org/10.1007/s10803-013-1836-5

Yoder, P., & Compton, D. (2004). Identifying predictors of treatment response. *Mental Retardation and Developmental Disabilities Research Reviews, 10*(3), 162–168. https://doi.org/10.1002/mrdd.20013

Yoder, P. J., Bottema-Beutel, K., Woynaroski, T., Chandrasekhar, R., & Sandbank, M. (2013). Social communication intervention effects vary by dependent variable type in preschoolers with autism spectrum disorders. *Evidence-Based Communication Assessment and Intervention, 7*, 150–174. https://doi.org/10.1080/17489539.2014.917780

Zerr, M., Freihorst, C., Schütz, H., Sinke, C., Müller, A., Bleich, S., Münte, T. F., & Szycik, G. R. (2019). Brief sensory training narrows the temporal binding window and enhances long-term multimodal speech perception. *Frontiers in Psychology, 10*, 2489. https://doi.org/10.3389/fpsyg.2019.02489

Zhang, Q., & Wang, L. (2017). Moderation analysis with missing data in the predictors. *Psychological Methods, 22*(4), 649–666. https://doi.org/10.1037/met0000104

Zhou, H. Y., Cheung, E. F. C., & Chan, R. C. K. (2020). Audiovisual temporal integration: Cognitive processing, neural mechanisms, developmental trajectory and potential interventions. *Neuropsychologia, 140*, 107396. https://doi.org/10.1016/j.neuropsychologia.2020.107396

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). *Preschool language scale* (5th ed.). Harcourt.