Multisensory Perceptual Training in Youth with Autism Spectrum Disorder: A Randomized Controlled Trial

By

Jacob Ian Feldman

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Hearing and Speech Sciences

August 31, 2020

Nashville, Tennessee

Approved:

Tiffany Woynaroski, Ph.D.

Mark Wallace, Ph.D.

Jim Bodfish, Ph.D.

Tyler Reimschisel, MD, MHPE

Acknowledgments

This work was funded by an award from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) to the Vanderbilt Kennedy Center (U54 HD083211; PI: Neul), National Institute on Deafness and Other Communication Disorders (NIDCD; 1R21DC016144; PI: Woynaroski), and National Science Foundation Research Traineeship (NSF NRT) grant DGE 19-22697 (PI: Wallace). The content is solely the responsibility of the author/s and does not necessarily represent the official views of the National Institutes of Health or the NSF.

When I was completing a systematic review as a part of my second pre-dissertation project (a concept that no longer exists in my department; thanks Melissa Duff!), I had to read a lot of dissertations. My favorite part to read was always the acknowledgements, because behind every successful dissertation is a large group of people who made it possible. If you're anything like me, buckle in and enjoy my effusive praise (also, add nine to all of the page numbers in my table of contents if you downloaded this paper off of the Vanderbilt ETD).

First and foremost, this project would not have been possible without the time and effort of my participants and their families. As a part of the larger project, I was fortunate enough to have worked with over 200 participants and over 450 participants as a part of my PhD program. Some of these families have become friends and even treat me like family; they are the reason why I loved collecting data.

This project also required a large team of researchers. The pilot work for my dissertation would not have been possible without David Simon, Jill Omer, Sarah Bowman, Bahar Keceli-Kaysili, Julie Conrad, Kacie Dunham, and Alex Tu. The work you are about to read about would not have been possible without Gabby DiCarlo, Yupeng Liu, Maggie Cassidy, Kacie, Laney Woynaroski, Evan Suzman, Zack Williams, Peter Abdelmessih, Grace Pulliam, and Sophie Kaiser. It was a pleasure to work with each and every one of you!!

Thank you also to Debshila Basu Mallick and John Magnotti for making their stimuli publicly available answering my questions about their stimuli.

I have been very fortunate to have been a part of Vanderbilt University and the Department of Hearing and Speech for the past seven years! I am forever grateful for the support of the MS-SLP Class of 2015, my officemates, Brian Weiler, Dillon Pruett and Emily Kimball; my honorary officemate (and former roommate), Cara Singer; my fellow PhD students in the department, most especially Carlos Benítez-Barrera, Natalie Covington, Sarah Diehl, Alison Hessling, Sarah Schneck, and Allie Whitten; the clinical faculty and staff (most especially Barb Jacobson, Mrs. Hale, Penny Welch, and Mike de Riesthal) who always supported me and let me sit with them at clinical case conference; and the communications team, Kate and Keli, who made sure I always had pens and went home with leftovers. To my colleagues in the Wallace Lab, David Simon, Aaron Nidiffer, Jean-Paul Noel, Gabby, iliza butera, and Collins Opoku-Baah, it was a pleasure to just be in the same room with some of the smartest people I've ever met. Thanks for letting me bring you down to my level.

Thank you to my dissertation committee for helping to push me and my science to be best it could be. Tyler, you were among the first professors at Vanderbilt to learn my name. Thank you for including me in the LEND program and teaching me how to be a leader — I promise that I'll sit next to you if we ever end up in the same lecture hall! Jim, thank you for all of your encouragement, your service on nearly all of my pre-doctoral committees, your invaluable feedback, and the many, many times that you helped me get back into some room I'd accidentally locked. Mark, you opened the door to so many opportunities that would have never been available to me otherwise. Thank you for believing in me, trusting me, paying for all of my research, and anything else I'm forgetting. It was a pleasure each and every time you fought with Tiffany about who was my mentor!

I'm so grateful to all of my friends, near and far, who offered me moral support. To my trivia team, which at various times consisted of Emily, John, Jena, Tate, Natasha, Christian, Thomas, Caitlin, Katy, and Chris, I'm 20 points positive that you're the best. Thank you to my game night crew of Sarah, Susannah, Morgan, Sutter, M, Scott, and Rodney for so many fun nights. Thank you to my former mentors and forever friends from Special Equestrians, most especially Anne, Sam, Diane, Angelica, Deb, and Tracy, who introduced me to the disability community and knew that I would be an expert in developmental disabilities before I did. Thank you to all of my former students and riders, who pushed me to be the best clinician, therapist, and teacher I could be. I had the best group of cheerleaders from afar in my high school and college friends. They say that once you're friends with someone for 10 years you're stuck with them for life. So, to Erin, Rachel, Michelle, Ben and Rosie, Agnew, Marc, Steph, Nate, Nick and Bailey, Adam, and Liz, I'm so glad that you're stuck with me and I love you all! And to Bruce, Laney, Teelyn, and Éowyn, hopefully I'll be a better Uncle Jacob now that this paper is finished.

I definitely would not be alive today if not for my physical and mental health care team. Dr. Gorden and Dr. Fiske, you literally saved my life. Dr. Nunn and Dr. Raj, without your counseling and care, I would not have made it through my PhD program.

I could never sufficiently describe how grateful I am to my mentor Tiffany. If you had not come into my life when you did, I would never have finished my PhD program. You believed in me whenever I didn't believe in myself and you treated me like a member of your family. You were generous with your time, your feedback, and your praise. I could work in your lab for the rest of my career and still not repay your kindness, but one day I might get close.

I have been blessed by my supportive family. Thank you to all of my relatives who always asked me about my research whenever they saw me. To the other Dr. Feldman, my sister Rose, I'm sorry that our "final" graduations won't be what we were hoping for, but I look forward to celebrating our terminal degrees when we can. Thank you for your support and for my loving dog-niece. Thank you to my parents for supporting me in every way possible throughout my entire life. Dad, I am proud that I will never be the biggest Feldman in the disability community in Philadelphia. Thanks for pushing me towards this path; it definitely opened more doors than it closed. Mom, thank you for all of our phone calls, for listening to me vent, for encouraging me, for telling me all the news, for flying down for my procedures, and for everything else you've done for me. Last but not least, to my Poppop: I finished my paper and I luv ya!

Finally, if you're still reading these acknowledgements and you haven't seen your name on the list, know that you are important to me and I appreciate you. There's no other reason why you'd still be reading!

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Evidence for Altered Temporal Binding of Audiovisual Speech in Autism

Disruptions in sensory functioning are commonly observed in individuals with autism[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). For example, researchers have investigated temporal binding of audiovisual stimuli in autistic individuals. 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 typically developing (TD) individuals.

The ability to integrate the visual and auditory components of social stimuli, such as speech, is theorized to be particularly critical to developing accurate, unified representations of

[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 this dissertation, I will primarily refer to youth using person-first language and adults using identity-first language. Though this decision may not satisfy all readers and stakeholders, and is unlikely to stand the test of time, I look forward to watching this dialogue between researchers and the autistic community continue to unfold.

the sensory world and foundational to the development of higher-order social, communication, and cognitive skills (Wallace & Stevenson, 2014). Enlarged TBWs in children with autism have thus been interpreted as maladaptive and hypothesized to produce cascading effects on development in a number of domains in this clinical population (Bahrick & Todd, 2012; Cascio et al., 2016). Larger TBWs for speech stimuli have been associated with increased autism symptoms and decreased language abilities (Feldman et al., 2019a; Smith et al., 2017), lending some empirical support to this theory.

Perceptual Training of Temporal Binding of Audiovisual Stimuli

The substantial evidence for altered audiovisual temporal binding in children with autism, as well as observed relations between TBWs for audiovisual speech and other domains of functioning in children with autism, has engendered increasing interest in the possibility of training audiovisual integration in children with autism (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 TD 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 (De Niear et al., 2018).

Limitations of this Literature

There are several limitations to the literature on perceptual trainings for audiovisual stimuli in TD adults. First, the vast majority of perceptual training studies on TD adults have found no evidence for generalization to untrained multisensory tasks (De Niear et al., 2018; Powers et al., 2016; Setti et al., 2014) or very limited evidence for generalization (Zerr et al., 2019). Sürig et al. (2018) found strong evidence for generalization; they hypothesized that the adaptive difficulty in their simultaneity judgment (SJ) training, wherein the task was designed to be challenging for each participant rather than utilizing consistent difficulty, resulted in strong learning, enabling gains made on their SJ task 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). In order to best detect distal or generalized outcomes of perceptual trainings, it may be necessary to assess a variety of outcomes that differ in various degrees from the stimuli and/or the task trained. 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 (e.g., a task measuring perception of the McGurk effect, when incongruent audiovisual stimuli induces a fused percept; McGurk &

Macdonald, 1976)

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.

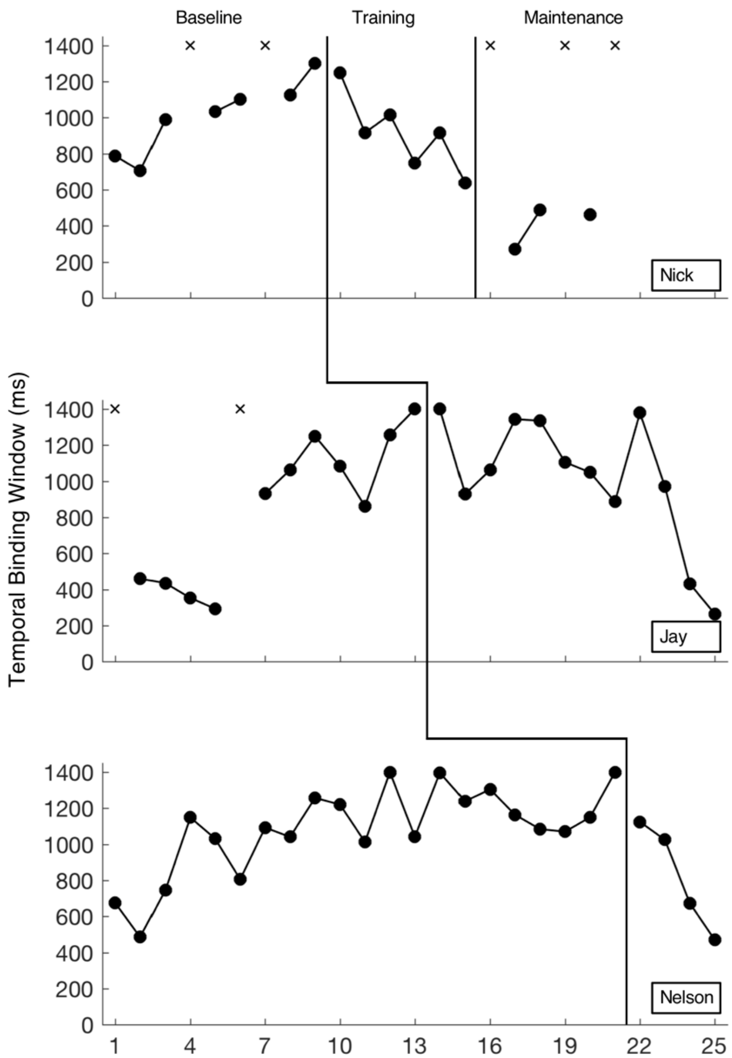
Perceptual Training of Temporal Binding of Audiovisual Stimuli in Autism

To date there have been very few studies on training audiovisual speech perception in children with autism. Two studies (Irwin et al., 2015; Williams et al., 2004) have utilized quasi-experimental designs and found that children with autism 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 children with autism. This study utilized a multiple baseline across participants design, a single-case experimental research design (see Ledford et al., 2019) wherein participants complete an extended baseline condition in order to (a) serve as their own control and (b) serve as a control for other participants via staggered introduction of the experimental condition (Gast & Ledford, 2014). 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 (see Figure 1).

Figure 1

Results from Feldman et al. (2020)



Note. Figure adapted from Feldman et al. (2020) depicts temporal binding windows (TBWs) for trained participants during baseline, perceptual training, and maintenance phases. X represents TBWs that could not be derived. "Nick" showed a clear and immediate response to the perceptual training. "Jay" showed a delayed response to the perceptual training. Though "Nelson" showed the most immediate response to the perceptual training, this effect is confounded due to the lack of independence between his response and "Jay's" response. Data collection was terminated earlier for "Nick" than for others based on school calendars/start dates.

There are a few limitations to this previous study based on the study design. First, it was difficult maintain adequate experimental control in the context of a single-case experimental research design. As a result, 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 (see Figure 1; Feldman et al., 2020) were somewhat expected given the high degree of heterogeneity in both presentations of autism and responses to intervention exhibited by children with autism (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 in order to understand the task and directions (e.g., Cascio et al., 2016; Feldman et al., 2018, 2019b; Woynaroski et al., 2013). 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. Additionally, two participants commented that they were confused about the automated feedback from the computer, perhaps indicating that participants required greater scaffolding to be successful at the task. 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 children with autism. 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 diverse 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:

a) Do children with autism assigned to the perceptual training experience perceptual narrowing within the context of the training?

b) Do children with autism assigned to the perceptual training experience greater narrowing of their TBW for audiovisual speech compared to children with autism assigned to a control group?

c) Does the effect of the perceptual training on TBWs generalize to untrained speech stimuli and/or untrained speakers?

d) Does the effect of the perceptual training translate to broader multisensory integration, specifically perception of the McGurk illusion?

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

Chapter 2

Method

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 youth with autism (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 Vanderbilt together) matched on chronological age and biological sex 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 two weeks for a total of eight sessions. 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 all of the pre-test measures one to three days prior to the research camp and all of the post-test measures one to three 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 Appendix.

Participants

Thirty participants aged 8-21 were recruited from a larger ongoing research project (see Figure 2 for a flowchart of participant recruitment and Table 1 for participant demographics). Inclusion criteria were: (a) diagnosis of autism spectrum disorder according to DSM-5 criteria (APA, 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.

Figure 2

Diagram of Participant Recruitment

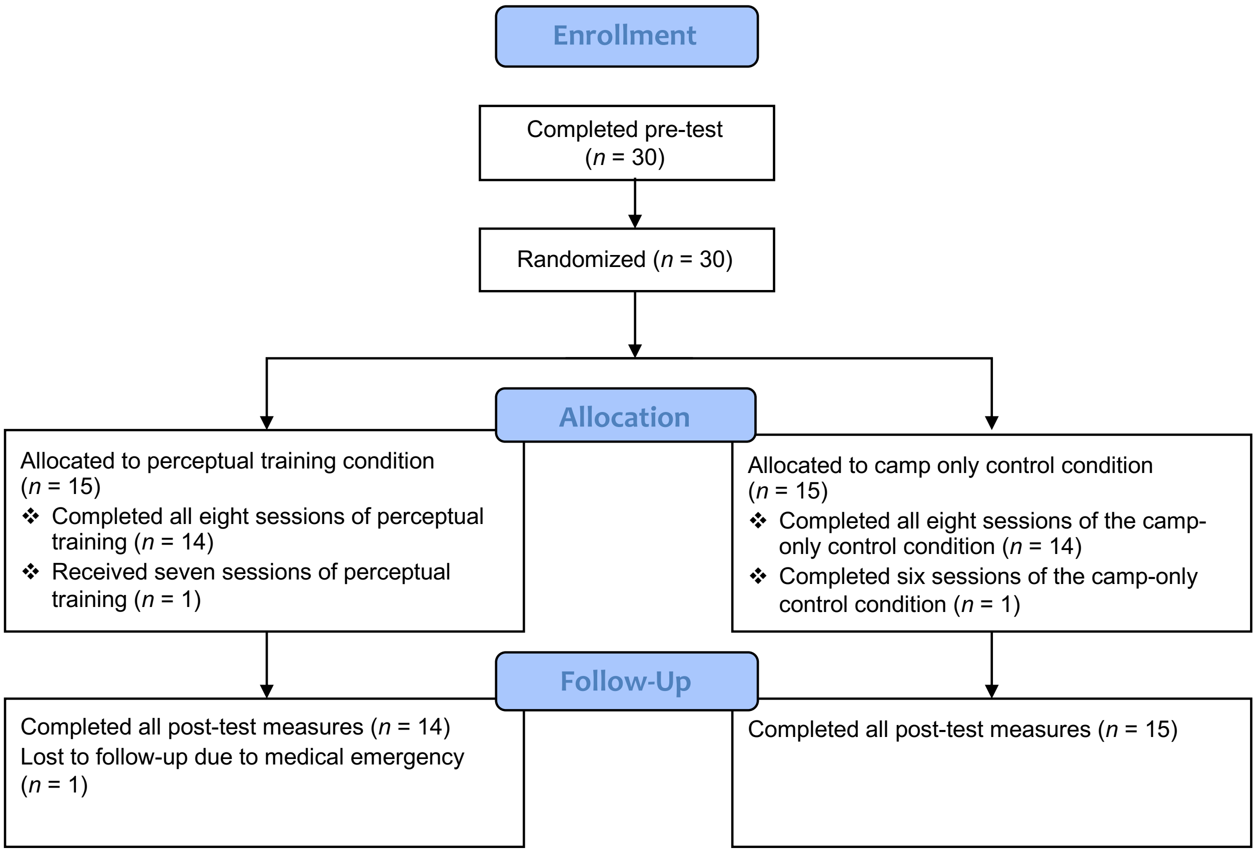


Table 1

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

Note. $TBW_{trained}$ = 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 1-33 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, $p > 0.5$.

Materials

The perceptual training, as well as all of 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 HD550 series supra-aural headphones.

Monosyllabic speech stimuli used in the perceptual training and SJ tasks (see Temporal binding window for audiovisual speech) were obtained from Basu Mallick et al. (2015). For the perceptual training task, stimuli were videos of six individuals (labeled 4.1, 4.2, 4.3, 4.5, 4.6, and 4.7 by Basu Mallick et al., 2015; 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 (i.e., 4.6) saying the trained syllable "ba" and the untrained syllable "pa", and a video of a female speaker not included in the perceptual training (i.e., 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; Simon & Wallace, 2018).

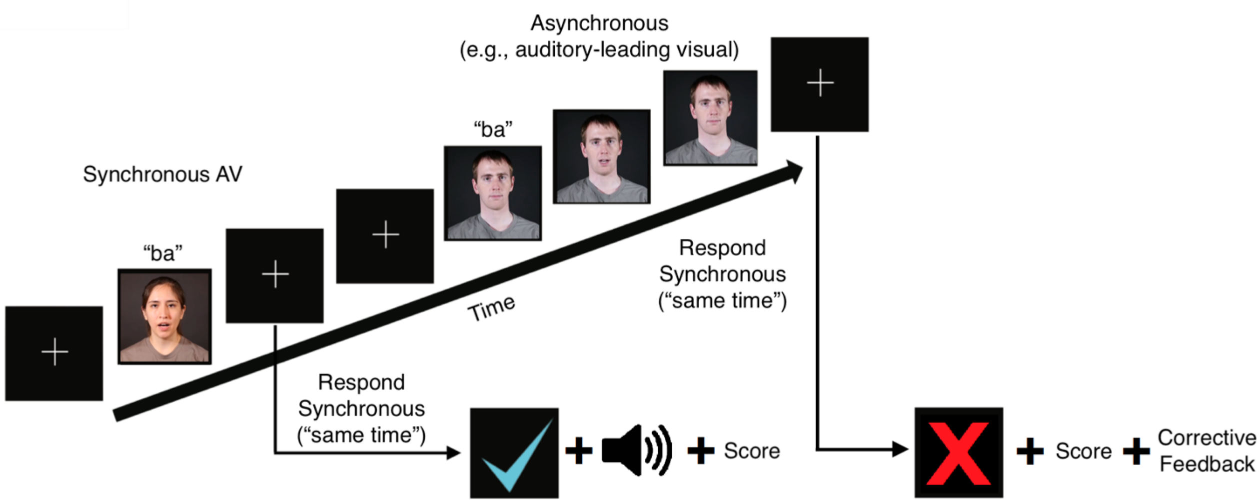
All video stimuli were edited in Adobe Premiere to create asynchronous stimuli for the perceptual training and SJ tasks, and incongruent audiovisual stimuli, auditory-only stimuli, and visual-only stimuli for the McGurk illusion task.

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 the trial, participants received computer-delivered feedback (see Figure 3). 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).

Figure 3

Depiction of the Perceptual Training



Note. AV = audiovisual.

During each day in the training, participants completed seven rounds of the training. Each round consisted of 48 trials, 50% of which were synchronous. In order 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 difficulties: one round of "easy" difficulty, two rounds of "medium" difficulty, and four rounds of "hard" difficulty (see Table 2). 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, adaptive in nature. For the first day of the training, participants' performance on the pre-test SJ task utilizing speech stimuli on which the participants were trained ($TBW_{trained}$; see Temporal Binding Window for Audiovisual Speech); on subsequent days, the participants' accuracy on the previous day's training task was used. 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 (see Figure 4). 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).

Table 2

Difficulty Levels Utilized in the Perceptual Training Paradigm

Difficulty	Number of Levels	% Reported Synchronous	Min SOA	Max SOA
Easy	1	10%, 20%, 30%	133 ms	500 ms
Medium	2	40%, 50%, 60%	133 ms	400 ms
Hard	4	65%, 75%, 85%	133 ms	300 ms

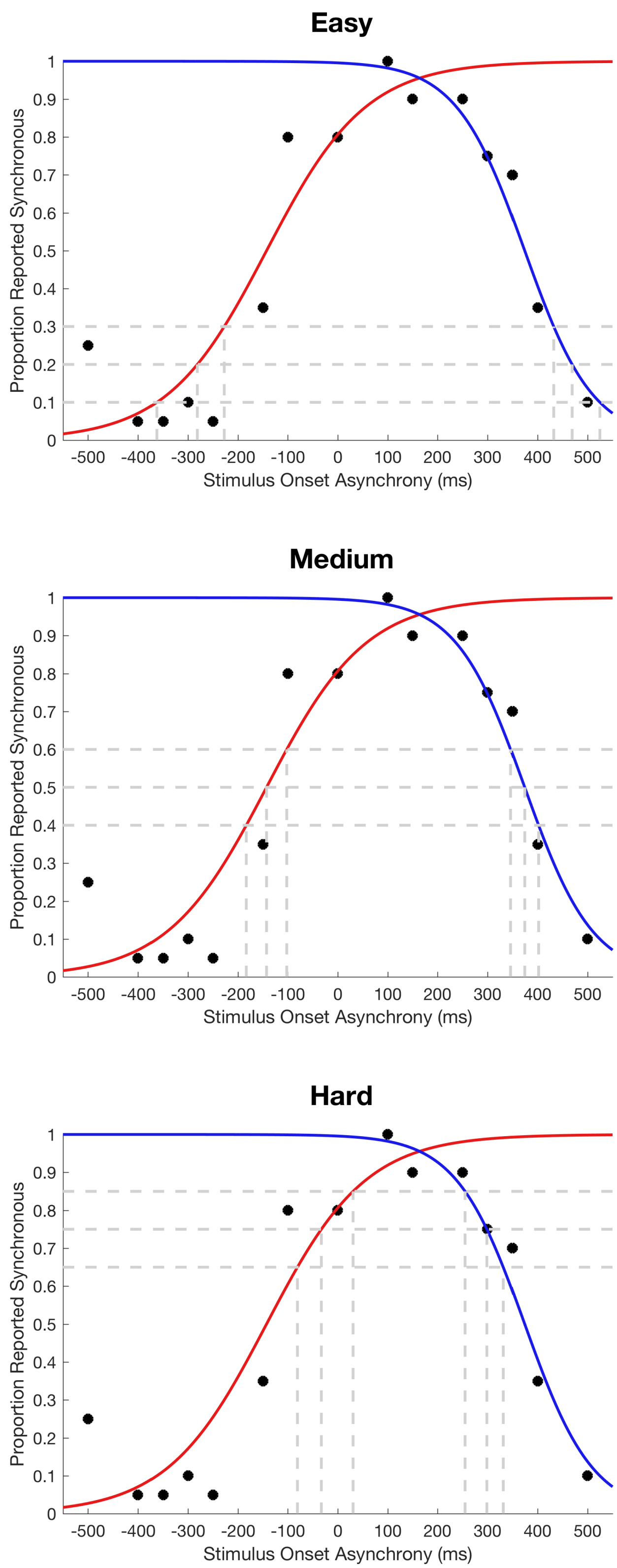
Note. Number of Levels = the number of times that this condition was presented in each training session; SOA = Stimulus onset asynchrony; % Reported Synchronous = the level of reported synchrony on psychometric curves fit to participant data and used to derive training SOAs (Note that a small % reported synchronous at non-zero SOAs represents accurate perception of asynchrony). Training SOAs were derived as the percent perceived synchronous, rounded to the nearest multiple of 50 ms or 16.7 ms (i.e., frame) or to the minimum or maximum values set for that difficulty level. Table originally published in Feldman et al. (2020).

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.

Figure 4

Example Training Stimulus Onset Asynchronies (SOAs)



Note. In each graph, the proportion of perceived synchrony from a non-training simultaneity judgment task is plotted against each SOA. The blue line represents the psychometric function fit to the right data points (visual first trials). The red line represents the psychometric function fit to the left data points (auditory first trials). The temporal binding window (TBW) is derived from the points at which these psychometric functions cross 0.75.

In the easy condition, the training SOAs would be ±433 ms, ±466 ms, and ±500 ms based on the right curve (blue; based on visual-first trials; note the original value of 516 ms was rounded down to the maximum value of 500 ms) and ±183 ms, ±283 ms and ±366 ms based on the left curve (red; based on auditory first-trials). In the medium condition, the training SOAs would be ±350 ms, ±366 ms, and ±400 ms based on the right curve and ±133 ms, ±150 ms and ±183 ms based on the left curve (note the original value of 100 ms was rounded up to the minimum value of 133 ms). In the hard condition, the training SOAs would be ±250 ms, ±299 ms, and ±300 ms based on the right curve (note the original value of 350 ms was rounded down to the maximum value of 300 ms) and ±133 ms based on the left curve (note the original values of 33 ms and 83 ms were rounded up to the minimum value of 133 ms).

Figure originally published in Feldman et al. (2020).

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 one hour 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

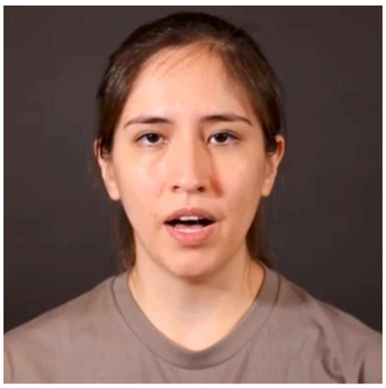
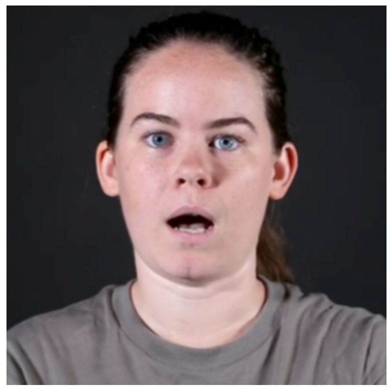
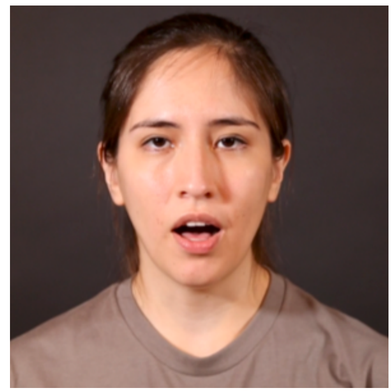
All pre-and post-test outcomes were collected by experimenters on the research team naïve to group assignment.

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"; see Figure 5). 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.

Figure 5

Stimuli Used in the Simultaneity Judgement (SJ) Tasks

$\text{TBW}_{\text{trained}}$	$\text{TBW}_{\text{novel speaker}}$	$\text{TBW}_{\text{novel syllable}}$
 Speaker included in the training saying “ba”	 Speaker not included in the training saying “ba”	 Speaker included in the training saying “pa”

Note. TBW = Temporal binding window.

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 of the task, 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, 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 child indicated that he/she 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%; see Figure 4 for an example). 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 task was processed in the same manner as described above in order to (a) estimate the effect of the perceptual training on changing the TBW within the context of the training task and (b) calculate the next day's training SOAs for the adaptive nature of the training task. Performance in the training task represents an estimated or pseudo TBW.

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. Perception of the McGurk illusion is purported to index the influence of vision on auditory speech, and can thus be considered a measure of multisensory speech perception (McGurk & Macdonald, 1976). Past work suggests that children 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 child looking at the computer and wearing headphones set to the proper volume, the examiner not providing feedback based on correctness of child response, and the minimization of potential distractors. For the perceptual training, expected behaviors included the child looking at the computer and wearing headphones set to the correct volume, the examiner setting up the task correctly, and the examiner not providing additional corrective feedback to the child (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 child.

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

To answer the first research question (i.e., regarding the effect of the perceptual training within the context of the training task), each day of the training was compared with the previous day using paired samples t-tests. $TBW_{trained}$ from pre-test and post-test was used to compare the training to the first and final days of training, respectively. Pairwise deletion was used to handle missing data for these analyses.

To answer the remaining research questions, 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. Prior to conducting these multiple regression analyses, variables that were not normally distributed were transformed in R (R Core Team, 2017). 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). Training and control groups were then compared on all pre-test metrics using independent samples t-tests; any analyses involving variables that were significantly different at pre-test included the pre-test performance as a covariate. 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, Hedge's *g* was calculated for each dependent variable to measure the magnitude of the effects of the perceptual training.

Chapter 3

Results

Preliminary Analyses

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

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. As previously mentioned, one participant did not complete any post-testing due to a medical emergency; 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 ($\chi^2 = 0.09$, $p = 0.765$). 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% - 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).

Transformation of Variables

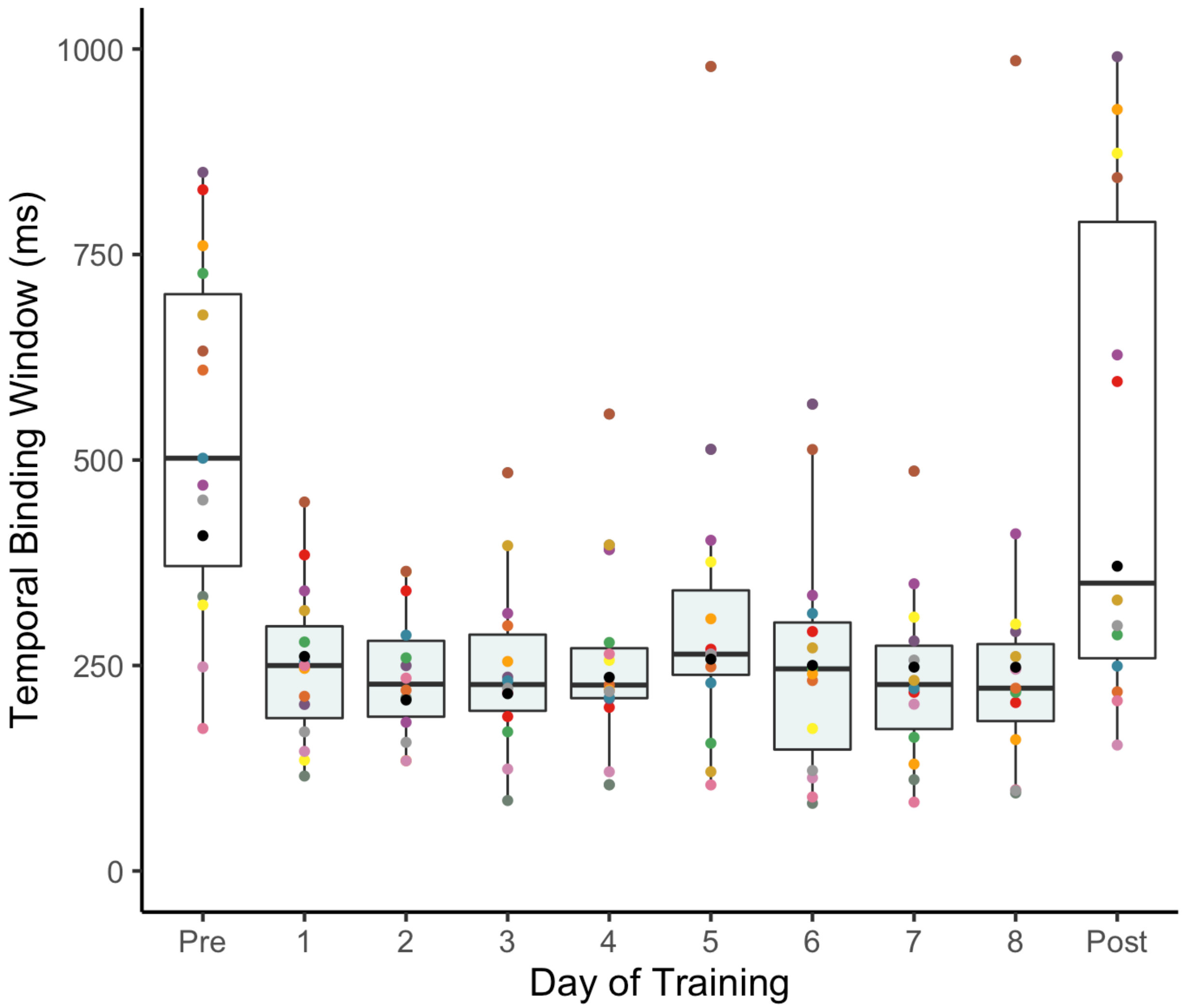
Prior to imputing missing data, all variables of interest were evaluated for normality, specifically for skewness > |1.0| and kurtosis > |3.0|. Three variables (i.e., nonverbal IQ, pre- and post-test McGurk fusion) were corrected for negative skew with a square transformation.

Change within Perceptual Training

Participants' performance in the perceptual training are plotted in Figure 6. On the first day of the training, participants significantly lowered their TBW compared to the $TBW_{trained}$ collected at pre-test, $t\,(14) = -5.86$, $p < 0.001$. Participants' performance did not change between any of the remaining days in the training, *p* values for the contrast of current versus immediately preceding day > 0.1. However, participants' TBW as measured during the last day of the training was significantly lower than the $TBW_{trained}$ collected at post-test, $t\,(14) = 2.81$, $p = 0.015$.

Figure 6

Participants' Performance in the Perceptual Training



Note. At pre- and post-test (white boxes), participants' temporal binding windows (TBW) derived from the simultaneity judgement task of the speaker included in the training saying "ba" ($TBW_{trained}$) are plotted. The days in training (light blue boxes) represent estimated TBWs from performance in the training task. Each participant's individual performance is plotted in a different color. A longer break (i.e., a weekend) did occur between Days 4 and 5 of the training.

Differences Between 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 3. Unconditional effects of the training on $TBW_{trained}$ (β = 148.0, p = 0.19, Hedge's g = 0.47) and $TBW_{novel\ syllable}$ (β = 178.1, p = 0.19, Hedge's g = 0.47) trended in the anticipated direction, but did not reach statistical significance. 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 $TBW_{novel\ speaker}$ (β = 84.4, p = 0.54, Hedge's g = 0.22) or perception of the McGurk illusion (β = 0.047, p = 0.72, Hedge's g = 0.13).

Table 3

Pre- and Post-Test Outcomes by Group

Dependent Variable	Perceptual Training Condition (n = 15) Pre-Test M (SD)	Perceptual Training Condition (n = 15) Post-Test M (SD)	Camp Only Control Condition (n = 15) Pre-Test M (SD)	Camp Only Control Condition (n = 15) Post-Test M (SD)	Hedge's g
$TBW_{trained}$	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

Note. 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.

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 4.

Table 4

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

Note. 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., 2002), 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”).

*$p < 0.05$, **$p < 0.01$

Age

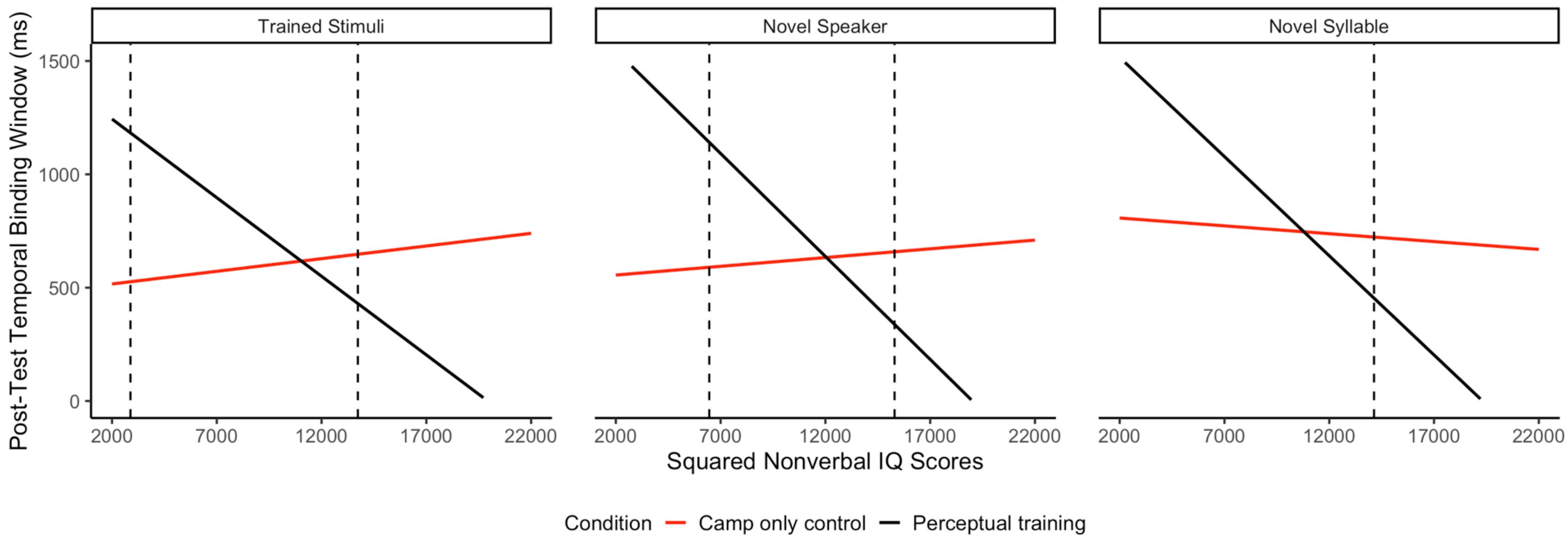
Age did not moderate the effect of training on any of the outcomes of interest (p values for interaction term in the multiple regression models > 0.3).

Nonverbal IQ

Nonverbal IQ significantly moderated the effect of the perceptual training on all of the TBW outcomes (p values for interaction term in the multiple regression models < 0.05; see Table 4). 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 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 iatrogenic effect was observed for individuals with nonverbal IQ scores below 80. For $TBW_{novel\ syllable}$, results of the Johnson-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.

Figure 7

Moderated Effect of Nonverbal IQ on Perceptual *Training Outcomes*



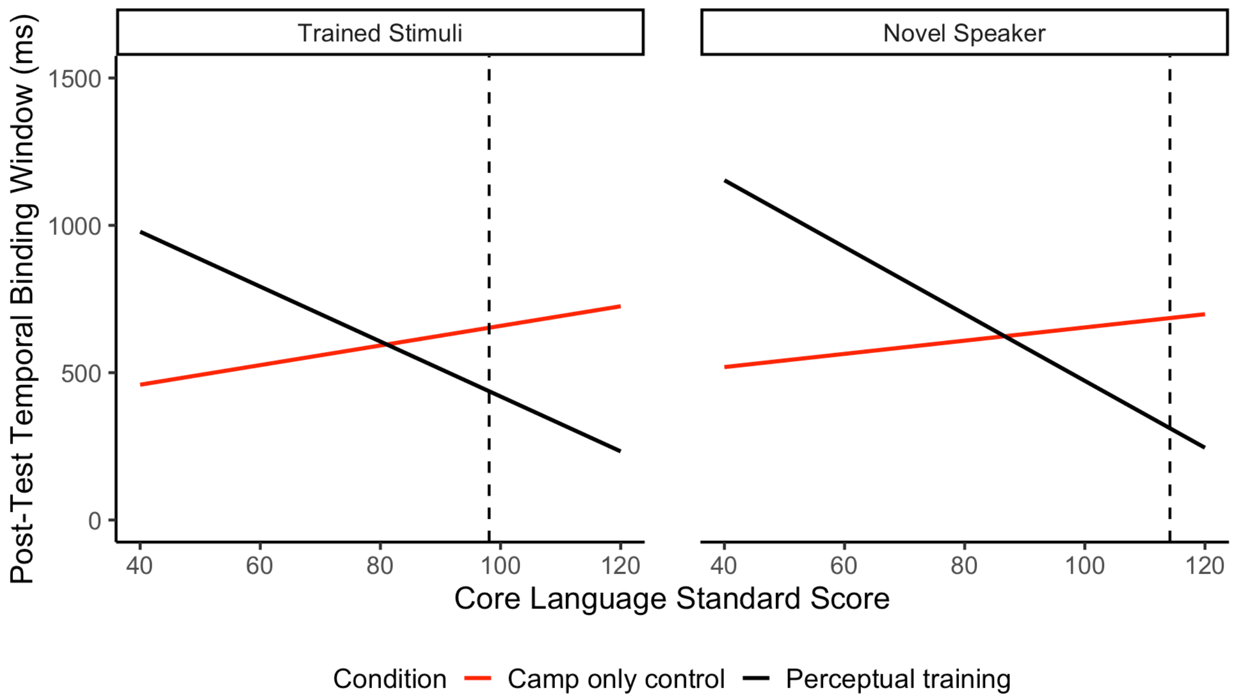
Notes. Nonverbal IQ scores were derived from the Leiter International Performance Scale, third edition (Roid et al., 2013). Dotted lines represent the cut points identified by the Johnson-Neyman tests. For trained stimuli and the novel speaker stimuli, individuals with nonverbal IQ scores below the dotted line (back-transformed values = 54 and 80, respectively) are likely to experience an 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.

Language

Language scores also significantly moderated the effect of the perceptual training on $TBW_{trained}$ and $TBW_{novel\ speaker}$ outcomes (p values for interaction terms in the multiple regression models < 0.05; see Table 4). 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 Figure 8). Language did not moderate the effect of the perceptual training on either $TBW_{novel\ syllable}$ or report of the McGurk illusion.

Figure 8

Moderated Effect of Language Ability on Perceptual Training Outcomes



Note. 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 identified 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.

Procedural Fidelity

Procedural fidelity was checked for 20% of WhisperRoom sessions for both groups who completed the study in 2019 ($n =$ 26). The perceptual training was administered with 98.6%

fidelity, and the camp-only session was administered with 100% fidelity. The average fidelity was very high for all three of the examiners who administered these sessions (98.7% – 99.6%).

Procedural fidelity was also checked for 20% of the testing sessions that occurred in 2019. The average fidelity was very high overall at 98.2%. Fidelity was very high across all five assessors (range = 97.7% – 100%) and did not differ according to condition ($p = 0.70$; 98.0% for perceptual training versus 98.4% for camp-only control) or timepoint ($p = 0.37$; 97.7% at pre-test versus 98.6% 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 ($M = 2.6$).

Chapter 4

Discussion

The purpose of the present study was to assess a computer-based perceptual training program designed to narrow TBWs for audiovisual speech in children with autism. On average, participants assigned to the perceptual training appeared to display a significant decrease in TBW size over pre-treatment measures that was detectable immediately following the onset of training and maintained over the course of the training period. At post-test, however, there were no significant effects of the training on TBWs in the perceptual training group, on average, compared to the control group. Importantly, though, effects of the training varied according to participant characteristics, such that children who had average to above average language and cognitive ability appeared to benefit from the perceptual training paradigm, but children 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 1,000 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 children 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 individuals with autism, specifically those with above average nonverbal IQ and average language abilities. Equally importantly for intervention sciences, this study also indicated individuals who are 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 a subset of individuals who are likely to experience iatrogenic effects as a result of this training paradigm (i.e., individuals with nonverbal IQs below 55).

Feldman et al. (2020) previously observed highly variable responses to a similar intervention in children with autism and hypothesized that some individual differences might have contributed to these differential responses, but this study is the first to statistically test child factors that moderate the effects of perceptual training on TBWs. Notably, all of the participants in the prior study had nonverbal IQs that were between 93 – 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.

Given the heterogeneous nature of autism, it is necessary for practitioners to understand for whom interventions may be most effective. This study critically adds to a growing body of literature suggesting that interventions for individuals with autism 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 children with autism 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 task (e.g., Cascio et al., 2016; Feldman et al., 2018, 2019b; Woynaroski et al., 2013).

In the present study, 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, this task differed in both the instructions given to participants and stimuli, whereas the generalization task utilized in the previous training study reporting generalized effects utilized a task that differed only in the instructions given to participants (i.e., Sürig et al., 2018). 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 children 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, is important that parents reported on average no 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 the pilot study (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.

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 in order 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 TD 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 with or at high-familial risk for autism (Santapuram et al., 2019; 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 children with autism and TD 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. Participants in the present study did complete an event-related potential task to index neural processing of audiovisual speech in the context of an SJ task (Simon, 2018; Simon et al., 2017, 2018; Simon & Wallace, 2018) and an eye tracking task to index attention to audiovisual speech (Dunham et al., 2020; Lewkowicz & Hansen-Tift, 2012) at the midpoint of the intervention. Thus, future work may 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 youth with autism 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) in order 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 autism symptomatology, 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 children with autism 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 children, who are arguably most likely to benefit from or need these types of interventions. 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 children with autism 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 children with autism and 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 children with autism with higher cognitive and language ability, and (c) novel approaches to improving audiovisual speech perception in children with autism with lower cognitive and language ability.

Chapter 5

References

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

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. https://doi.org/d2xz

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., 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). Perceptual training in children with autism spectrum disorder: A single-case treatment study. *Research in Autism Spectrum Disorders, 74*, 1-13. 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., Del Bene, V. A., Frey, H. P., Russo, N. N., Blanco, D., 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*, 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*. 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, 206*, Article 116313. https://doi.org/10.1016/j.neuroimage.2019.116313

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*, 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 US. 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*, Article 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, 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

Powers, A. R., Hillock-Dunn, A., & Wallace, M. T. (2016). Generalization of multisensory perceptual learning. *Scientific reports, 6*, 1-9. https://doi.org/10.1038/srep23374

R Core Team. (2017). *R: A language and environment for statistical computing* (Version 3.4.1) Vienna, Austria. https://www.R-project.org/

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

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., Dunham, K., Suzman, E., Augustine, A. E., Bowman, S. M., Golden, A. J., Keceli-Kaysili, B., Broderick, N., Cascio, C. J., Lewkowicz, D. J., & Woynaroski, T. G. (2019). *Links between eye gaze patterns and vocal development in infants at high and low risk for autism spectrum disorder* [Poster presentation]. Gatlinburg Conference on Research and Theory in Intellectual Disabilities, San Antonio, TX, United States.

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*, 1-18. 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. (2018). *Electrophysiological signatures of multisensory temporal processing in the human brain* [Doctoral dissertation, Vanderbilt University].

Simon, D. M., Nidiffer, A. R., & Wallace, M. T. (2018). Single trial plasticity in evidence accumulation underlies rapid recalibration to asynchronous audiovisual speech. *Scientific Reports, 8*, 1-14. https://doi.org/10.1038/s41598-018-30414-9

Simon, D. M., Noel, J. P., & Wallace, M. T. (2017). Event related potentials index rapid recalibration to audiovisual temporal asynchrony. *Frontiers in Integrative Neuroscience, 11*, 1-12. https://doi.org/10.3389/fnint.2017.00008

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*. Oxford.

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 attention are linked with language in children with and without autism* [Paper presentation]. Society for Research in Child Development, Baltimore, MD, United States.

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*, Article 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*, Article 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.

Appendix

Changes to the Research Camp Protocol in 2020 due to COVID-19

2019	2020
There were three sessions of the research camp. During each session, 4-5 participants came into the lab in the morning and afternoon each day of the training.	There was one session of the research camp. Only two participants came into the lab in the morning and afternoon each day of the training.
Participants were in the lab for 3 hours during each training day. During that time, participants did about 1 hour of WhisperRoom activities, 1 hour of arts and crafts, and 30 minutes of organized group activities.	Participants were in the lab for 2.5 hours during each training day. During that time, participants did about 1 hour of WhisperRoom activities and 1.25 hours of group activities, including games and arts and crafts.
Participants did their WhisperRoom activities with the door shut and the examiner in the room with them. Both WhisperRooms were located in one larger room.	Participants did their WhisperRoom activities with the door open and the examiner seated outside the room at least six feet away. Each WhisperRoom was located in a separate room.
Each day, participants would interact with 3-5 trained research assistants.	Each day, participants would interact with up to 2 trained research assistants.
	Participants wore masks during all activities.

2019	2020
	Every day, participants had their temperatures checked and their parents were asked about their symptoms of, travel related to, and exposure to COVID-19.