Title

Effect of Autistic-Like Traits on Novel Second Language Phonetic Trait Imitation Ability in Children with Varying Autism Quotient Scores

Author

Dunlap, Madison Raelene

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Effect of Autistic-Like Traits on Novel Second Language Phonetic Trait Imitation Ability in Children with Varying Autism Quotient Scores

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy in Hispanic Languages and Literatures

by

Madison Raelene Dunlap

2024

ABSTRACT OF THE DISSERTATION

Effect of Autistic-Like Traits on Novel Second Language Phonetic Trait Imitation Ability in Children with Varying Autism Quotient Scores

by

Madison Raelene Dunlap
Doctor of Philosophy in Hispanic Languages and Literatures
University of California, Los Angeles, 2024
Professor Ji Young Kim, Chair

With the increasing prevalence of autism diagnoses globally, the overall aim of this dissertation is to improve our understanding of the linguistic aptitude of autistic individuals. Given that the autism diagnosis is a spectrum, we rely on the Autism Quotient (AQ) test, which measures the quantity of autistic-like traits an individual possesses, to provide a quantitative value for the diagnosis. The goal of this study is to better understand how autistic-like traits may affect a child's phonological abilities at the segment-level, measured by their ability to imitate voice onset time (VOT) when producing stop consonants of a novel second language (L2).

Our participants are both typically-developing (TD) and neurodiverse children, all 18 of whom are English-dominant children ages seven to 11 years old based in New South Wales, Australia. The methodology includes a parent survey completed on behalf of the child and experimental tasks, completed in person by the child, across two sessions that were separated by four to seven days. The experimental tasks consist of both English and Spanish read-aloud tasks, as well as Spanish immediate imitation tasks, and perception tasks featuring manipulated syllables spanning the full range of Spanish and English stop consonant VOT values. These tasks allow us to evaluate how well the participants imitate and learn to produce Spanish words with the appropriate, shortened VOT values.

Since individuals with higher AQ scores have more autistic-like traits, such as greater attention to detail and focus, we predict these individuals will be better L2 phonetic learners. The results show that autistic-like traits are not predictive of a child's L2 phonetic imitation, nor short-term or long-term L2 phonetic learning abilities. That being said, regardless of diagnosis, our participants show a significant VOT reduction on the L2 phonetic imitation task and they learn to maintain a reduced VOT when producing Spanish word-initial stops, in both the short-term and the long-term. All of the children in the present study are capable L2 phonetic imitators and learners; thus, we encourage L2 learning in classrooms to offer an additional communicative outlet and improve children's social skills, neurodivergent or not.

The dissertation of Madison Raelene Dunlap is approved.

Benjamin Aaronson

Jesse A. Harris

Antonio C. Quicoli

Victoria Mateu Martin

Ji Young Kim, Committee Chair

University of California, Los Angeles

2024

To my little writing buddy, Joey.

Table of Contents

List of Tables

List of Figures

Acknowledgments

First, I would like to thank my committee for their support and feedback throughout my graduate education, and specifically in writing this dissertation. I have learned so much from each of the members and greatly appreciate the wisdom they have shared with me. Thank you to my chair, Dr. Ji Young Kim, for the many meetings, hours of coding, and unwavering support.

I would not be in this PhD program if it were not for the research experience and encouragement I received at UC Berkeley from Dr. Justin Davidson. Thank you for inspiring me to go to graduate school and further explore my passion for Spanish linguistics.

I am incredibly grateful for the mentorship I have received from Dr. Ben Aaronson in both academic and clinical settings. He has shown me what it truly means to be a leader and work as a team on multiple occasions, lessons which I will carry with me through life.

This dissertation would not have been possible without the support of the Bright Eyes Early Intervention clinics in Sydney and Blue Mountains, Australia. I am immensely grateful to Dr. Alex Brown who welcomed me into her clinic without hesitation and connected me with all of the incredible kiddos that participated in my research. I hope to reunite on the slopes soon.

I am grateful for the communities that have allowed me to connect with fellow students including UCLA Club Track and Field, APEX, ADCC, UC-LEND, Chi Alpha, Cal Lights, and, of course, my fellow cohort-mates in the Spanish department. I have learned so much from my peers and I will cherish the lifelong friendships I've gained through these communities.

I could not be more thankful for the love and support my family has provided throughout my PhD. My dad can succinctly say what my research is better than I can and has networked on my behalf for years. Thank you to my mom for attending conferences and being my biggest fan. Thank you to my best friend, parents, and brother for inspiring me to work as hard as they all do.

Biographical Sketch

2015-2019	B.A. Hispanic Languages and Bilingual Issues, University of California, Berkeley
2015-2019	B.A. Molecular and Cellular Biology, University of California, Berkeley
2019-2021	Dean’s Scholar Fellowship
2019-2021	M.A. Hispanic Languages and Literature, University of California, Los Angeles
2021-2022	Ben and Rue Pine Travel Award to attend New Sounds International Conference in Barcelona, Spain
2020-2022	Teaching Assistant for the UCLA Department of Spanish and Portuguese
2022-2023	Graduate Research Mentorship Fellowship
2023-2024	Ben and Rue Pine Travel Award to collect dissertation data in Sydney, Australia
2023-2024	UC Leadership Education in Neurodevelopmental Disabilities Trainee

Chapter 1: Introduction

1.1. Introduction

As autism awareness and prevalence continues to grow, both nationally and globally, it is crucial our understanding of the autism diagnosis parallels that growth such that we are better able to provide lifelong support to autistic individuals (Maenner, 2021). The autism diagnosis varies individually and spans a wide range of symptoms, severity, and needs for intervention, yet our understanding of the linguistic profile of autistic individuals is particularly lacking across the spectrum. Through this doctoral dissertation, we hope to contribute to the limited research that currently exists related to the linguistic profile of autism, specifically with respect to the bilingualism potential of these individuals. We will utilize the Autism Quotient (AQ) test as a lens through which we can understand how accurately individuals with more autistic-like traits are able to imitate phonetic traits of a foreign language, Spanish in this case. The AQ test evaluates the cognitive areas that are more commonly affected in autistic individuals; we will focus on two of these measurable cognitive areas in particular: attention to detail and attention switching. (Baron-Cohen et al., 2001; Ruzich et al., 2015b). We aim to determine whether individuals with a higher number of autistic-like traits and, thus, higher AQ scores, are able to learn some features of a second language (L2) as well, or even better, than individuals without a diagnosis, given their heightened attention to detail and greater ability to focus, corresponding to a higher attention switching score.

1.1.1. Motivation

The results from this study will hopefully offer a new perspective on L2 acquisition and teaching in clinical populations. Despite previous concerns regarding raising autistic children, who tend to struggle with communication in general, in bilingual environments, more recent studies have shown that bilingually-exposed autistic children are not at a disadvantage in terms of their early language development compared to their monolingually-exposed autistic peers (Jordaan, 2008; Ohashi et al., 2012). Researchers found that Spanish-English bilingual autistic children used more gestures and produced more vocalizations than their monolingual English-speaking counterparts; given that there were no other significant language skill differences between the two groups, the conclusion was that bilingualism did not negatively affect the autistic children (Valicenti-McDermott et al., 2013). With this in mind, parents and families who speak a language other than English at home can feel comfortable speaking to their child in their native language without worrying about the effect this may have on their ability to acquire English in school. That being said, we would like to propose a shift in perspective; rather than perceiving bilingualism as a neutral option when raising autistic children, perhaps we should encourage and promote bilingual language acquisition for these children. We know that autistic children are not hindered by dual language acquisition, but is it possible these children may experience social and communicative benefits by having not one, but two, outlets for communication? To understand whether or not we should actually promote bilingualism in autistic populations, we first hope to gain a better understanding of the linguistic profile of autism, specifically from the perspective of L2 phonological acquisition.

1.1.2. Terminology

In order to make meaningful contributions to the field of autism research, it is important to keep autistic individuals at the center of said research. In the past, autism researchers have referred to autism as a “disorder” that needs to be “cured;” using these ableist terms to describe autism negatively impacts the view society has towards autism and presents autism from a deficit-based perspective, instead of a strength-based perspective (Bottema-Beutel et al., 2021; Dwyer et al., 2022). The Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5) is used to diagnose autism, listed as “Autism Spectrum Disorder (ASD)” in this manual; according to the DSM-5, autism presents as persistent impairments in social interactions and restricted, repetitive behaviors, interests, or activities (American Psychiatric Association, 2013). The use of terms such as “impairments” and “restricted, repetitive” in the DSM-5 gives autism a negative connotation, but these terms can be replaced with “areas of challenge” and “focused, intense, passionate,” respectively (Bottema-Beutel et al., 2021; Dwyer et al., 2022).

More recently, researchers have turned to consulting autistic individuals and advocates to determine the best language to use when discussing research conducted in the field (Bottema-Beutel et al., 2021; Dwyer et al., 2022). In accordance with the findings from these studies, in the present paper we will use identity-first language (i.e. “autistic child” instead of “child with autism”) and we will refer to autism using the preferred language and terminology of autistic individuals (Bottema-Beutel et al., 2021; Dwyer et al., 2022). Through this study, we hope to reduce stigma towards autism by focusing on a strength-based approach to autism research; we will consider the advantages of autism in making our predictions, while maintaining a neutral perspective while interpreting the results. The present study aims to contribute to the growing number of studies conducted with autistic individuals at the center of autism research.

1.2. Literature Review

1.2.1. Language Acquisition in Autism

1.2.1.1. L1 Acquisition in Autism

Although the DSM-5 does not recognize these classifications, Tager-Flusberg (2006) identified two language classifications of autism: Autism with Language Impairments (ALI) and Autism with Language Normal (ALN). For the purpose of the present study, we are interested in children with ALN, or autistic children without Specific Language Impairment (SLI) or any additional speech disability diagnoses; we would like to understand how autistic-like traits, rather than SLI or other speech disabilities, affect phonological acquisition abilities.

It is not uncommon for autistic children, or children with ALN to be more specific, to experience language challenges as a result of their disability, even without a diagnosis specifically related to language challenges; as a matter of fact, Eigsti et al., (2011) state that language challenges are consistently present in nearly all autistic children, with one of the most common areas of challenge being delayed language development. Autistic children say their first words around the average age of 38 months old, while typically-developing (TD) children say their first words in the eight to 14 month old range (Howlin, 2003).

Given the challenges autistic individuals often face with respect to social interactions, there are several studies that look into first language (L1) acquisition and speech in autism. According to Eigsti et al. (2011), autism is characterized by challenges or delays in two main areas: social skills and language and communication skills. In fact, many children are evaluated for autism after missing language milestones early on in childhood (De Giacomo & Fambonne, 1998). Although delayed language acquisition is often associated with language development in autism, several studies have found that autistic children acquire language in a way that is both aligned with typical development and entirely unique compared to typical development (Wolk & Edwards, 1993; Wolk & Giesen, 2000). Since early language skills and development are crucial for lifelong communication abilities, there needs to be more research related to language acquisition in autism.

Across the spectrum, many autistic children tend to struggle with pragmatics (Eigsti et al., 2011); this is likely a result of their purported challenges with "Theory of Mind," i.e. the ability to understand what others are thinking (Baron-Cohen, 1988). According to this theory, an autistic child's inability to comprehend what others intend to communicate interferes with their ability to understand the pragmatics of language (Baron-Cohen, 1988; Eigsti et al., 2011). An alternative theory, the "Executive Functions" (EF) theory, postulates that autistic children have core cognitive deficits, which then affect their language and communication skills, among other skills required for typical functioning (Eigsti et al., 2011; Pennington & Ozonoff, 1996).

With respect to prosody, or the stress and intonation patterns of the spoken words, several studies have found that autistic children struggle, and perform worse on experimental tasks, compared to their TD counterparts (Diehl et al., 2008; Eigsti et al., 2011; McCann et al., 2007). In the study by McCann et al. (2007), the autistic group performed significantly worse than the TD group on 11 of 12 prosody assessments, while the study by Diehl et al. (2008) showed that the autistic group performed significantly worse than the TD group when tasked with discerning syntactic ambiguities. More recently, researchers relied on prosody abnormalities typically found in autistic children (echolalia, monotonous intonation, atypical pitch, and irregular linguistic stress patterns) to detect autism in natural speech recordings using a machine learning approach (Chi et al., 2022). These researchers were successful in using machine learning methods to detect autism automatically from natural, home-based audio recordings, which may be useful for detecting autism without the need for specialized equipment.

For the most part, and in the more recent studies, autistic children have shown similar syntax acquisition to that of other children with disabilities such as Down syndrome[1] and Dysphasia[2], a language disability (Cantwell et al., 1978; Eigsti et al., 2011; Tager-Flusberg et al., 1990). Other studies have shown that autistic children have less syntactic complexity and variety (Eigsti et al., 2011; Shapiro & Kapit, 1978). Interestingly, Shapiro and Kapit (1978) showed that autistic children were superior imitators compared to their neurotypical peers, based on their syntactic and semantic analyses. Eigsti et al. (2011) concludes that the past studies are not in total agreement regarding the effect of autism on syntax acquisition, but it appears autistic children experience at least some type of delay or deficit within this area of language acquisition. According to Koizumi et al. (2019), children with a variety of intellectual disabilities, as well as autism, show delayed syntactic development with notable improvement once they reach a mental age of seven to nine years old; in this study, the participants with intellectual disabilities were reported to reach a mental age of seven to nine when they were between 11;3 and 16;10 in chronological age.

While there is limited research related to morphology acquisition in autistic populations, in the study by Cantwell et al. (1978), they found that autistic children performed similarly to their peers with Dysphasia, in that they used several of the same morphemes in their spontaneous speech (Eigsti et al., 2011; Cantwell et al., 1978). Bartolucci et al. (1980) found that autistic children omit morphemes such that they produce ungrammatical sentences more than their TD

[1] Down syndrome is a genetic disability that typically involves mild to moderate intellectual disability and distinct physical characteristics (Antonarakis et al., 2004).

[2] Dysphasia is a language disability induced by brain damage, which results in difficulty understanding and using language (Albert et al., 2013).

peers and their developmentally delayed peers, with IQ scores of between 50 and 70, matched on nonverbal mental age. It is important to note that prepositions, conjunctions, and pronouns are particularly challenging morphemes for autistic children (Eigsti et al., 2011; Ricks & Wing, 1975).

Autistic children proved that they were able to associate novel words with novel objects as well as TD children (Eigsti et al., 2011; Tek et al., 2008). Autistic children performed similarly to both children with other developmental delays and TD children when asked to sort pictures according to their category membership (Eigsti et al., 2011; Tager-Flusberg, 1985). Still, Kamio et al. (2007) found that autistic children are less successfully primed by semantically-related words, even though they were successfully primed with pictures (Eigsti et al., 2011). Eigsti et al. (2011) claims that autistic children's understanding of mental state verbs, as well as their semantic organization, is evidently different than their TD peers; despite this, autistic children perform similarly to their mental age-matched TD peers on vocabulary standardized tests, basic categorization tasks, and have age-appropriate lexicon size.

Some researchers propose oral motor dysfunction could be the source of speech difficulties in autism, as children with better motor imitation skills responded better to speech production treatment (Rogers et al., 2006). Despite this theory, several studies have found that autistic children have mostly intact, though potentially delayed, phonological systems (Bartak et al., 1975; Eigsti et al., 2011; Kjelgaard & Tager-Flusberg, 2001; MacMillan, 2015). One study, which nicely summarized the general conclusions related to phonological acquisition in autism, found that most autistic individuals do not have specific phonological areas of challenge, but autistic individuals with higher support needs (previously coined "low-functioning autistic individuals"), may experience both phonological and articulatory deficits (Eigsti et al., 2011;

Lord & Paul, 1997). Wu et al. (2020) focused on monolingual Putonghua-speaking autistic children and their phonological acquisition compared to both age-matched and receptive-language-age-matched TD peers. They used a picture-naming task to elicit the children's speech and measured phonological acquisition by the initial sounds, final sounds, and tones that each child produced. Wu et al. (2020) found that autistic children have similar phonological development to their receptive-language-age-matched peers, but have significantly lower phonological development than their age-matched peers. Schoen et al. (2011) corroborates these findings; they found that autistic children performed similarly to their language-age-matched TD peers who were younger in chronological age, but they performed significantly worse than their chronological age-matched TD peers, in terms of their phonological and vocal productions. In this case, the autistic children ranged in age from 18 to 36 months, while the language-age-matched children were 11 to 13 months old; thus, there was up to a 25 month, just over two years, age difference between the experimental group and this control group (Schoen et al., 2011). These studies validate the claim that language acquisition in autism is notably delayed compared to TD children, and this delay extends to phonological language abilities.

Despite potential language delays, Kjelgaard and Tager-Flusberg (2001) claim that the acquisition of phonetic traits tends to be the most spared linguistic ability in autism. Kjelgaard and Tager-Flusberg (2001) found that expressive phonology at the one word level is intact in the speech of autistic children, meaning that these children had native-sounding articulation. This finding confirmed an earlier study that found autistic children have superior articulation compared to their TD peers (Boucher, 1976). Boucher (1976) was the first study to date to test articulation in autism using a standardized test, the Edinburgh Articulation Test (EAT) (Anthony et al., 1971). In this study, the experimental group consisted of autistic children, while the two control groups were TD children matched to the experimental group according to both chronological age and vocabulary skills, as measured by the Peabody Picture Vocabulary Test (PPVT), and children with dysphasia matched to the experimental group according to both chronological age and nonverbal IQs. In comparing the experimental group's performance on the EAT to that of the control groups, the researchers concluded that autistic individuals showed a particular strength in their articulation language skills (Boucher, 1976).

We may attempt to summarize the phonology-specific findings as follows: autistic individuals with low support needs and strong motor skills exhibit strong articulation abilities with similar, but delayed, phonological development to their TD peers. With the aforementioned experiments and theories in mind, we will focus on phonology for the purpose of the present study since there is substantial evidence to support the belief that autistic children will ultimately develop an intact, complete phonological system.

1.2.1.2. L2 Acquisition in Autism

While many autistic children may be able to speak their native language fluently and with the proper pronunciation eventually, this delayed development, or acquisition, causes complications with respect to the child's ability to communicate and relate to his or her peers and family. This concern is amplified when it comes to raising autistic children in a bilingual environment.

Kremer-Sadlik (2005) acknowledges that parents of autistic children in the U.S. have historically been instructed to speak English to their autistic children, such that the child is exposed to the same language both inside and outside of the home. Thus, instead of learning their home language and being able to communicate fluently with their family members, these children are only receiving input in English, the language that they hear at school. The results of this study show that there is not enough evidence to suggest that multilingualism worsens the language acquisition of children with delayed or impaired language abilities (Kremer-Sadlik, 2005; Paradis et al., 2003; Peña et al., 2020). It is important for these children to feel equipped to communicate with their family members in the language that their loved ones speak at home. Given that it is already challenging for autistic children to understand the world from other perspectives and to communicate effectively, limiting their language-speaking abilities to the language that they use outside of the home will only distance them further from their family members.

Hambly and Fombonne (2012) found similar results; this study compared the language levels and social abilities of autistic children raised in monolingual versus bilingual environments, which was further separated into simultaneous bilinguals who learned both languages prior to the age of 12 months and sequential bilinguals who learned the L2 after the age of 12 months. All of the children in this study spoke French, English, Chinese, Farsi, Hebrew, Italian, Romanian, Spanish and/or Tamil. According to the results, children diagnosed with autism do not have additional delays in language development, compared to their monolingual counterparts, when exposed to bilingual environments as either simultaneous or sequential L2 learners. This emphasizes the fact that autistic children are not negatively impacted by learning an L2 nor by being exposed to two different language environments. This study also concludes that there is limited research in the area of Second Language Acquisition (SLA) in autistic populations. Although families do not need to be worried about introducing an L2 or hosting a bilingual environment, we still do not know much about the pace of L2 learning by autistic children (Hambly & Fombonne, 2012).

It has also been shown that autistic children benefit from individualized tutoring and support in L2 learning (Barletta, 2018; Bradley, 2019). Barletta (2018) performed a case study with a child with "high-functioning autism." The researcher observed a six-year-old child with L1-Spanish, L2-English; the child recently arrived in the U.S. from South America and, thus, was learning the L2 in an immersive setting. The child received 16 tutoring sessions, which were recorded, transcribed, and evaluated after each individual session, such that future sessions could be altered to match the child's needs. Barletta (2018) concluded that the child with high-functioning autism followed a similar SLA path as TD children, but his success in learning the L2 seemed to be dependent on the one-on-one nature of the tutoring sessions. They also concluded that autistic characteristics could actually be seen as assets in SLA, if they are understood and enabled appropriately by L2 teachers.

Asperger Syndrome (AS) now falls within the same autism diagnosis listed in the DSM-5, though it was once recognized as its own, unique diagnosis; like autism research, the field of AS research is also extremely lacking. Bradley (2019) performed a case study of an L1-Spanish, L2-English undergraduate student in Costa Rica to evaluate his experience learning an L2 in a classroom setting, given his AS diagnosis. The student was observed in three different English courses, over the span of two years, and the student's performance was evaluated via an Oral Proficiency Exam, a survey following completion of the English courses, and class observations and interactions between the researcher, who was the student's professor, and the student. In consultation with online forums about people with AS and their experiences learning an L2, this study concludes that individuals with AS, though varied in their individual language skills and interests, are certainly capable L2 learners, especially with an inclusive, supportive classroom environment.

Nordgren (2015) performed a case study of an L1-Swedish boy who was five-years-old and diagnosed with autism. In addition to Swedish, the child was exposed to both English and his parents' L1, an unspecified non-Germanic language. He experienced language delays at 1;6, so the family prioritized speaking only their L1 at home up until their child was age 2;5, when they switched to only speaking their L2, their child's L1, of Swedish at home. For the experiment, the child received a training package consisting of minimal pairs, or pairs of words that differ by only one sound; the goal of this training package was to introduce the boy to L2 phonological contrasts and the letters that represent sounds. They recorded the boy's speech when interacting with both speech pathologists and teachers, and found that he was able to develop and produce new speech sounds, but he experienced the phonological delays that are typically observed in autism with respect to syllables and both word segments and boundaries.

Although previous concerns regarding teaching autistic children an L2 have resulted in limited research regarding bilingualism in autism, more recent studies have shown that bilingualism does not negatively impact language in autistic children (Hambly & Fambonne, 2012; Kremer-Sadlik, 2005; Ohashi et al., 2012; Valicenti-McDermott et al., 2013). These studies leave us with two questions: Do all autistic children experience phonological delays? Is there individual variability dependent on where each autistic individual falls along the spectrum? Before we can address these questions, it is important to point out that the DSM-5 autism diagnosis does not offer a quantitative value; thus, clinicians typically diagnose cases categorically, not using a scale, making it difficult to look at the degree of autistic-like traits.

1.2.2. Autism Quotient Score

The AQ test was first published in 2001 and has since been used in both clinical and research settings to assess the presence of autistic traits in clinical and nonclinical populations (Baron-Cohen et al., 2001; Ruzich et al., 2015b). Although the AQ score is not diagnostic, it is a reliable measure of autistic-like traits (Hoekstra et al., 2008). The AQ score is determined using a forced-choice questionnaire, which is typically self-reported or parent-reported, depending on the age and literacy of the participant. The questionnaire consists of 50 statements to which the respondent selects one of the following options: “Definitely Agree,” “Slightly Agree,” “Slightly Disagree,” or “Definitely Disagree.” The AQ test evaluates five relevant subcategories of cognitive abilities that are often affected by autism including social skills, attention switching, attention to detail, communication, and imagination. The interpretation per subcategory is as follows: a higher social skill score indicates weaker social skills, a higher attention switching score indicates weaker attention switching skills and heightened focus, a higher attention to detail score indicates stronger attention to detail skills, a higher communication score indicates weaker communication skills, and a higher imagination score indicates weaker imagination skills. There are 10 statements related to each subcategory, though these subcategories are not explicitly indicated on the questionnaire (Appendix A). For each answer that aligns with the neurodivergent response, regardless of whether they responded strongly or not, the participant’s overall score increases by one point (Appendix B). Thus, the participant receives a sum score ranging from zero to 50, with a higher score corresponding to a higher number of autistic-like traits. For individuals who score 26 or higher, it is likely they have been evaluated for or diagnosed with autism, while lower scores indicate that one has probably neither been considered for a diagnosis nor diagnosed (Engelbrecht, 2020).

Since it was first published, there have been several new versions of the AQ test created. For example, there is an AQ test intended specifically for children and a test specifically for adolescents (Engelbrecht, 2020). The test has also been translated into other languages, such that it can be used in populations other than English-speaking populations. Thus, it is evident the AQ test and AQ scores can be useful in a wide range of research studies. For example, a study from 2015 found that higher AQ scores are correlated with males and STEM professionals, and lower AQ scores are correlated with females and non-STEM professionals (Ruzich et al., 2015a). In other recent studies, researchers found that AQ scores are also related to computer skills, in that those with higher AQ scores demonstrate more advanced technological knowledge (Harvey et al., 2016; Seigfried-Spellar et al., 2015). That being said, individuals that reported more advanced computer skills also reported having poorer social skills, communication skills, and imagination skills (Seigfriend-Spellar et al., 2015). Thus, while the individuals with higher AQ scores in these studies have enhanced cognitive strengths, such as attention to detail and attention switching, these strengths are present at the expense of their social and communication skills, which in turn can impact certain linguistic abilities. Bishop (2012) utilized a cross-modal lexical decision task to understand how AQ scores play a role in English-speaking adults' ability to disambiguate sentences using prosody, namely prenuclear accents. The results showed that participants with high AQ scores, on the communication subscale specifically, relied less on prosody than participants with lower scores; in other words, participants with lower communication skills were more susceptible to priming, regardless of prosodic features, while participants with stronger communication skills (i.e. lower communication subscale score) only demonstrated priming when prenuclear accent was absent. The results were interpreted from a variety of angles, including in the context of potential pragmatic challenges or limited processing resources of individuals with more autistic-like traits.

Through the present study, we hope to propose the opportunity to harness the heightened cognitive abilities often present in individuals with high AQ scores to actually improve their social and communication skills, through the introduction of an L2. While AQ scores can be useful in a wide range of research studies, we will focus on their use in linguistic studies, specifically phonetic imitation studies, which do not rely heavily on a strong understanding of pragmatics.

1.2.3. Phonetic Imitation

Phonetic imitation is the process by which the speech sounds and patterns a speaker produces converge with those produced by a narrator (Yu et al., 2013). Phonetic imitation tasks are useful in understanding the inventory of speech sounds that speakers are able to produce. Furthermore, phonetic imitation tasks may be useful in introducing, or even teaching, new speech sounds to speakers in a way that relies on both perception and production ability. The AQ test and scores can be useful in predicting this ability, as it assesses attention to detail and attention switching, or focus, which may be useful skills for perceiving the fine-grained details and differences between the speech sounds of two unique languages.

Mielke et al. (2013) used a phonetic imitation task to evaluate the role of memory in autism compared to TD individuals. Based on the results of this study, the researchers were able to confirm that both autistic individuals and TD individuals with high AQ scores relied on declarative (explicit) memory to complete the phonetic imitation task, while TD individuals with low AQ scores relied on procedural (implicit) memory to complete the phonetic imitation task. The procedure included a read-aloud task, a recording with extended voice onset time (VOT) of /p/, followed by another read-aloud task. VOT is the duration of time between the initiation of vocal cord vibration and the release burst of sound when producing stop consonants. The study showed that autistic individuals, or high AQ-scoring individuals, only extended the VOT of the stop consonant /p/ and did not generalize to the unaltered stop consonant /k/. On the contrary, the TD individuals generalized the VOT extension to /k/ (Mielke et al., 2013). Thus, the results from this study may suggest that individuals with more autistic-like traits are worse at applying what they learned to novel words. On the other hand, the results may also suggest that individuals with more autistic-like traits are able to notice and produce subphonemic details with great specificity. Because explicit memory relies on actively recalling the learning event, while implicit memory relies on repeated past exposure, the results from Mielke et al. (2013) may highlight the heightened attention to detail and focus, which corresponds to a higher attention switching score, that autistic individuals possess during the learning event compared to their peers with low AQ scores (Cohen et al., 1997).

Snyder et al. (2019) used AQ scores to determine cognitive processing style to then predict phonetic imitation of both human and digital device voices, such as Apple's Siri, by TD individuals. Based on the results, individuals with a lower imagination AQ score (meaning they were better at imagination) were better at imitating vowel duration, while individuals with a higher attention to detail AQ score (meaning they were more detail-oriented) were better at imitating vowel fundamental frequency (f0). The authors understand these results to suggest that individuals have unique cognitive profiles which affect their interactions with digital device voices, as evidenced by their imitation performance in the study. These findings corroborate the findings by Yu et al. (2013) in that individuals with higher AQ scores may be more successful at certain aspects of phonetic imitation than individuals with lower AQ scores.

Yu et al. (2013) conducted a study which assessed the phonetic imitation abilities of several adults. In this study, monolingual English-speaking participants completed a series of three tasks: a production task, followed by a perception task, followed by another production task. For the production tasks, the participants were asked to read a word list containing 72 target words, which were /p, t, k/-initial, real, English words. The researchers recorded the participants' speech and measured the VOT for each of the target initial consonants. For the perception task, the participants listened to a first-person narrative in which the same 72 /p, t, k/-initial words were embedded, with VOT values that were extended by 100%; there were four different storylines that varied in perceived sexual orientation of the male narrator (homosexual or heterosexual, manipulated by changing the pronouns of the narrator's partner) and story outcome (positive or negative, manipulated by the results of the date in the story). The results from the study showed that individuals with higher AQ scores, specifically on the subscale of attention switching, had higher phonetic imitation abilities. In other words, the individuals with a greater number of autistic traits were able to mimic and appropriately extend their VOT values when producing the initial consonants /p, t, k/ better than individuals with fewer autistic traits. In evaluating the effect of subjective attitude towards the speaker on phonetic imitation ability, the results showed a significant correlation between story outcome and degree of phonetic imitation; thus, the participants demonstrated greater imitation for positive outcomes. In addition, the participant's subjective attitude towards the speaker and how open they are as an individual were both significantly correlated with phonetic imitation ability; thus, there is a connection between "liking" someone and the perception-production link (Yu et al., 2013).

1.2.4. Phonetics of Plosives

When discussing consonants and the different ways in which this airflow is obstructed, we should pay attention to three main parameters: place of articulation, manner or mode of articulation, and voicing or vocal cord action. Place of articulation refers to the articulators used to produce a consonant sound, mode or manner of articulation refers to the type of obstruction to the airflow, and voicing refers to the way in which the vocal cords vibrate when producing each consonant. We will focus on voiceless plosives, or stop consonants, in Spanish and English, both of which have voiceless stops with three different places of articulation. All stop consonants, regardless of language, are produced when there is total obstruction of airflow. According to both Hualde (2014) and Cho and Ladefoged (1999), stop consonants fall into three voicing categories: voiced, in which the vocal cords begin vibrating prior to the release burst of the stop; voiceless *unaspirated*, in which the vocal cords begin vibrating at the time of the release burst or shortly after; and voiceless aspirated, in which there is a delay after the release burst before the vocal cords begin vibrating. Although we are focusing on voiceless stop consonants in Spanish and English in general, it is important to point out that the voiceless stop consonants /p, t, k/ are unaspirated in Spanish, but aspirated in English. The three different places of articulation relevant to voiceless stop consonants in Spanish include bilabial, dental, and velar, while the three places of articulation relevant to voiceless stop consonants in English include bilabial, alveolar, and velar (Hualde, 2014). The voiceless bilabial stop /p/ is produced when the lower lip makes contact with the upper lip, causing an obstruction and a resulting explosion of air following the obstruction. The voiceless dental stop /t/ is produced when the tip of the tongue makes contact with the back of the upper front teeth to stop the airflow. Lastly, the voiceless velar stop /k/ is produced when the back part of the tongue, the dorsum, makes contact with the velum, which is the back part of the roof of the mouth, to create the necessary obstruction.

The phonetic properties that may be useful in distinguishing the place of articulation, in both Spanish and English, include the VOT and the formant transitions, as well as the f0, the frequency at which the vocal cords vibrate when producing voiced sounds, at the onset of the following vowel after the stop consonant (Hualde, 2014). For the purpose of the present study, we are interested in VOT specifically. It is important to acknowledge that the average VOT values for Spanish stop consonants /p, t, k/ tend to be much shorter than the average VOT values for English stop consonants /p, t, k/, due to the aspiration involved in producing these sounds in English. In word-initial position, the Spanish voiceless unaspirated stop consonants /p, t, k/ are produced with short-lag VOT values (the red highlight in Figure 1), which means that the vocal cords vibrate briefly after the explosion of air that occurs when the articulators separate. On the other hand, the voiceless aspirated stops /p, t, k/ have a long-lag VOT in English (the red highlight in Figure 2); thus, there is a longer time between the occlusion and the initiation of vocal cord vibration.

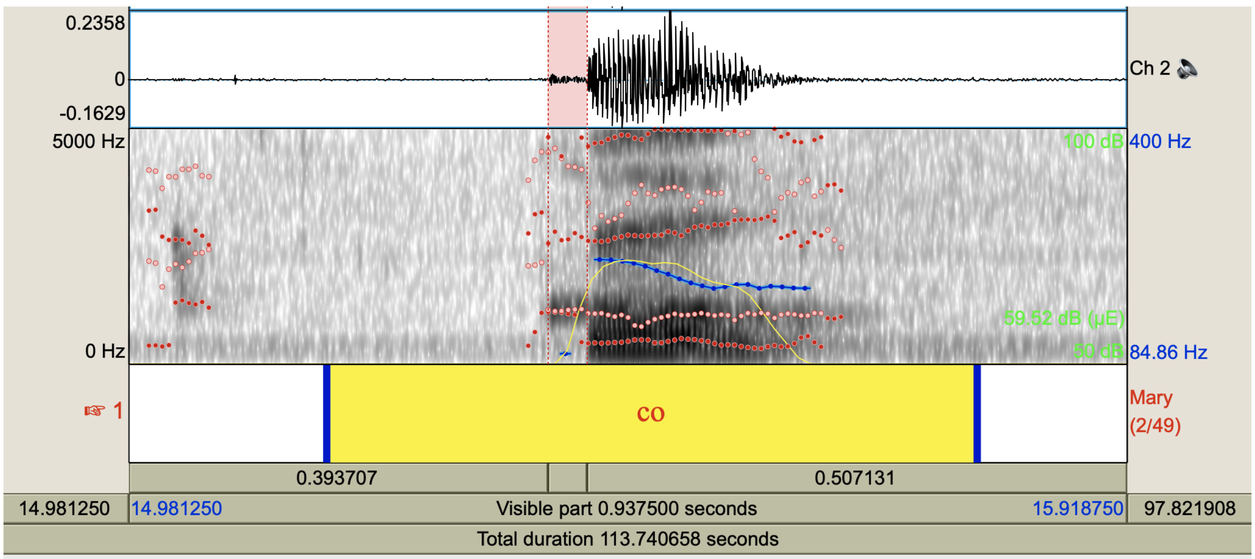


Figure 1. Example of Spanish /ko/, with the red highlight indicating /k/ VOT = 36.662 ms.

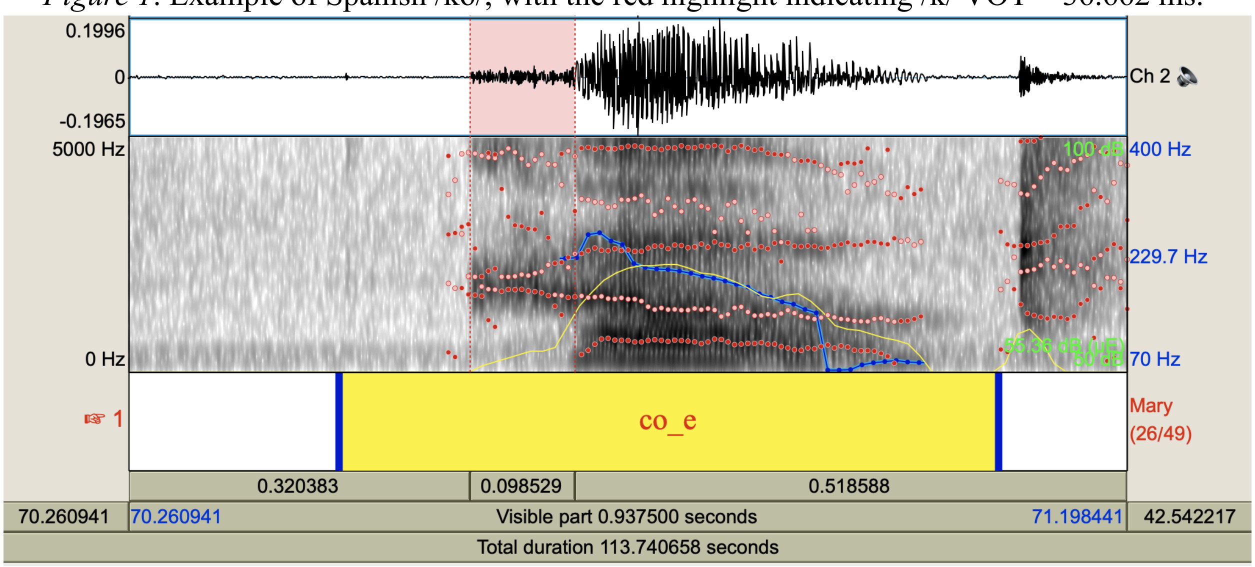


Figure 2. Example of English /ko/, with the red highlight indicating /k/ VOT = 98.529 ms.

Although these VOT values can vary, Castañeda (1986) found the following average VOT values for Castilian peninsular Spanish: 6.5 milliseconds (ms) for /p/, 10.4 ms for /t/, and 25.7 ms for /k/. These average values are similar to those reported by Lisker and Abramson (1964), which were four ms for /p/, nine ms for /t/, and 29 ms for /k/. Michnowicz and Carpenter (2013) reported the following ranges and averages for Spanish in the Yucatan peninsula: 9.17 to 22.4 ms as the range and 17.21 ms as the average for /p/, 14.27 to 27.53 ms as the range and 22.02 ms as the average for /t/, and 23.09 to 41.12 ms as the range and 34.04 ms as the average for /k/. Finally, Zampini (2013) offers a similar short-lag VOT value range of zero to 25 ms. With respect to English, Lisker and Abramson (1964) reports averages of 58 ms for /p/, 70 ms for /t/ and 80 ms for /k/; these align well with the long-lag VOT value range of 30 to 100 ms in English, reported by Zampini (2013). All of these VOT values are summarized in Table 1 below.

Source	**Average VOT of Spanish voiceless stops**	**Average VOT of English voiceless stops**
Castañeda (1986)	(Castilian Spanish) /p/ 6.5 ms /t/ 10.4 ms /k/ 25.7 ms	N/A
Lisker & Abramson (1964)	(Puerto Rican Spanish) /p/ 4 ms /t/ 9 ms /k/ 29 ms	/p/ 58 ms /t/ 70 ms /k/ 80 ms
Michnowicz & Carpenter (2013)	(Yucatan peninsula Spanish) /p/ 17.21 ms (range: 9.17-22.24 ms) /t/ 22.02 ms (range: 14.27-27.53 ms) /k/ 34.04 ms (range: 23.09-41.12 ms)	N/A
Zampini (2013)	/p, t, k/ 0-25 ms	/p, t, k/ 30-100 ms

Table 1. Summary of VOT values for Spanish and English voiceless stops.

We can measure VOT (in ms) by analyzing the spectrograms of the productions using software such as Praat (Boersma & Weenink, 2023). On the spectrogram, the stop consonant will be recognizable as a darker vertical line indicating the release burst (Figure 3). The VOT is measured as the distance between this vertical line and the beginning of the vowel, which will be identifiable by additional shading and clear formants, or horizontal lines showing the energy bands at consistent frequencies, typical of vowels. Formants are labeled first, second, and third from the bottom to the top of the spectrogram, as the left axis represents frequency in Hertz.

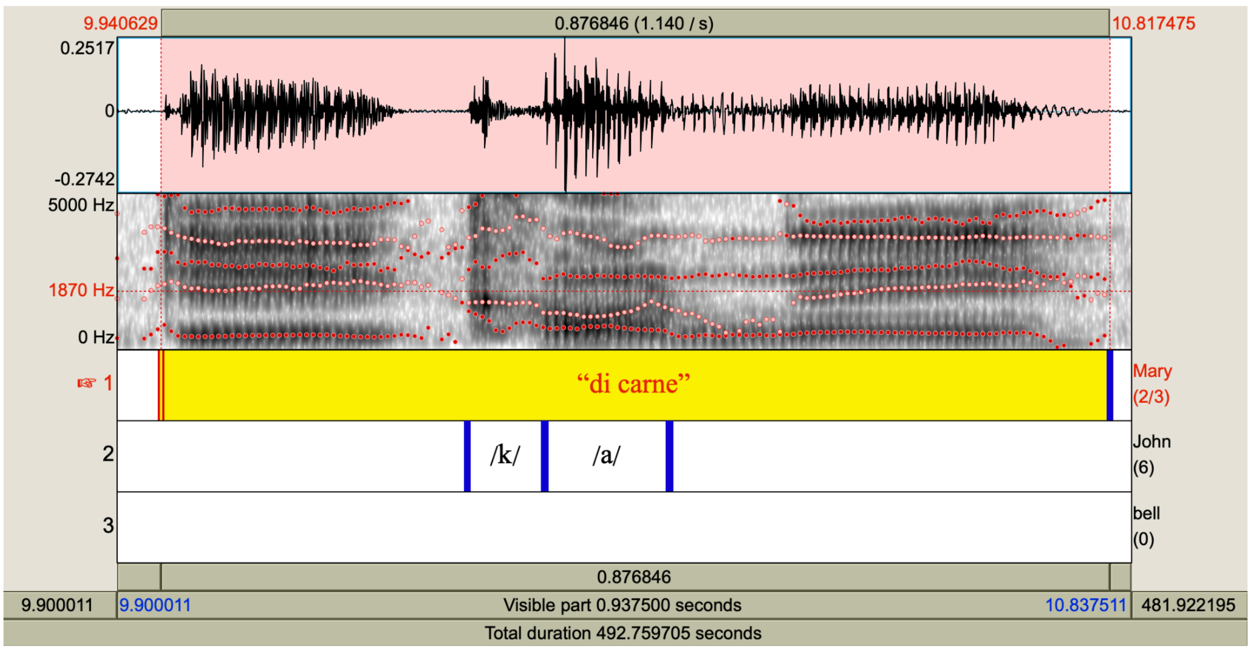


Figure 3. Spectrogram and oscillogram of the phrase "di carne" (*"say meat"*) indicating the VOT of the stop consonant /k/, as well as the vowel /a/.

Since VOT differs according to both place of articulation and language, it is a useful phonetic property to study in order to analyze the place distinction. While VOT appears to be the most commonly studied phonetic property with respect to distinguishing place of articulation, formant transitions and formant values immediately following stop consonants are also informative (Hualde, 2014). For example, the studies by Dmitrieva et al. (2015) and Schertz et al. (2015) both utilized the f0 at the onset of the vowel following the stop consonant to better understand voiceless stop consonants. In Dmitrieva et al. (2015), the aim of the study was to determine whether VOT and onset f0 covary automatically (the phonetic approach) or intentionally (the phonological approach). Lead VOT (voiced stop consonants) versus short-lag VOT (voiceless unaspirated stop consonants) is contrastive in Spanish, but not in English (which has voiceless unaspirated and voiceless aspirated stop consonants). Additionally, the short-lag VOT voiceless unaspirated stop consonants in Spanish are /p, t, k/, while the short-lag VOT voiceless unaspirated stop consonants in English are /b, d, g/ (Figure 4). Thus, the researchers were able to distinguish the phonetic effects of VOT on onset f0 from the phonological effects by comparing these phonetic properties in these two languages. The results showed that lead and short-lag VOT were only distinguished by f0 in Spanish, not English; since the associated stop consonants differ phonologically in Spanish, the results of this study provide evidence in favor of the phonological approach (Dmitrieva et al., 2015). Following voiceless consonants in either language, the f0 was higher than it was following voiced consonants (Dmitrieva et al., 2015). In Schertz et al. (2015), onset f0 was evaluated as one of the phonetic cues bilinguals, with Korean as their L1 and English as their L2, used to perceive and produce English stop consonants. While the Korean three-way stop contrast can be perceived and produced using both VOT and onset f0, VOT tends to be a more useful cue than onset f0 in distinguishing English stops, specifically with respect to the /b/ and /p/ stop contrast (Schertz et al., 2015). Schertz et al. (2015) concluded that participants varied, unpredictably so, in which phonetic cues they used to perceive L2-English stop contrasts.

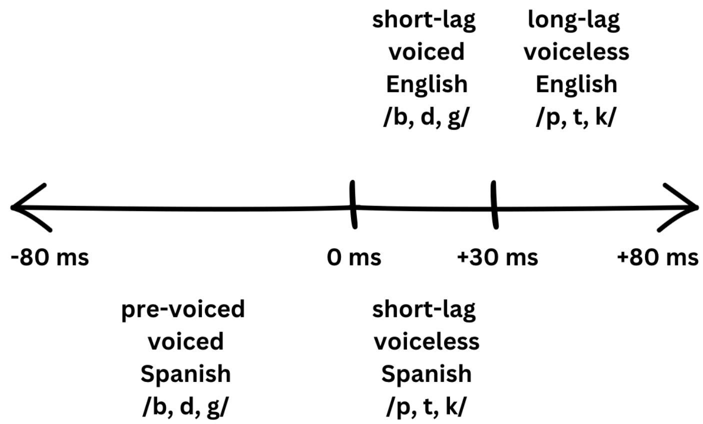


Figure 4. VOT range for Spanish and English voiced and voiceless stop consonants - adapted from Stölten et al. (2014).

While the aforementioned studies focused on onset f0 as a cue associated with voicing and distinguishing stop consonants with varying types of voicing, formant transitions may be more useful in distinguishing the place of articulation of both Spanish and English voiceless stop consonants. In Kissling (2015), the goal was to evaluate whether or not explicit phonetic instruction improved L1-English, L2-Spanish bilinguals' ability to perceive the sounds of Spanish. In their analysis of the results, they found that the participants were able to discriminate /t/ in Spanish versus English better than /k/ (Kissling, 2015). They attribute this to place of articulation, claiming that the dental /t/ of Spanish has different formant transitions, specifically in the second formant (F2) of the following high vowels /i, u/. As stated by Moffitt (1971), one of the phonetic properties of stop consonants is the presence of energy transitions at the start of the first formant (F1) and the F2. If the F1 transition is present, a voiced stop consonant is produced, while the absence or reduction of this transition indicates the production of a voiceless stop consonant. Moffitt (1971) also claims that variation in F2 transitions contributes to differences in place of articulation within either voiced or voiceless stop consonants, which corroborates the findings presented in Kissling (2015).

Although Spanish and English both have voiceless stop consonants with three places of articulation, the phonetic realization of these consonants may vary depending on the neighboring vowels. In Spanish, only the voiced stop consonants have multiple allophones depending on their location both within words and phrases (Hualde, 2014). That being said, the Spanish voiceless velar stop consonant /k/ is produced further forward on the velum/roof of the mouth before middle vowels /i, e/ and further back on the velum/roof of the mouth before high vowels /o, u/. In English, the voiceless aspirated stop consonants appear in both word-initial position and in stressed syllables that are not word-initial, but these voiceless stop consonants are no longer aspirated when they follow a stressed vowel or /s/; in fact, following /s/, the distinction between English /p, t, k/ and /b, d, g/ neutralize such that word pairs like <*sdeak> and <steak> would be realized in the same way. In addition to losing their aspiration post-stressed vowel, the duration of the vowel before English /p, t, k/ affects how the distinction between these consonants is perceived since vowels are produced with longer duration before voiced stop consonants than voiceless stop consonants. This is not observed in the Spanish voiceless stop consonants /p, t, k/ to the same degree. In English, if either the voiced stop consonant /d/ or the voiceless stop consonant /t/ is produced after a stressed vowel and before an unstressed vowel, it will likely be produced as a "flap," or a stop with shortened occlusion; in this case, the distinction between /t/ and /d/ is maintained by the preceding vowel (Hualde, 2014). Again, the Spanish voiceless stop consonants do not convert to a flap in any vowel contexts.

With this basic understanding of VOT and voiceless stop consonants, summarized in Table 1 above, we can focus on how these speech sounds are acquired in both TD children and autistic children. Given our interest in L2 phonetic imitation and acquisition of voiceless stop consonants, we will provide a brief review of the relevant and recent studies on this topic.

1.2.4.1. Acquisition of Voiceless Stops

1.2.4.1.1. TD Populations

Hoonhorst et al. (2009) conducted an experiment with babies raised in a French-speaking environment with the goal of developing a better understanding of the shift from language-general (-30 ms to +30 ms) to language-specific perceptual boundaries (Lisker & Abramson, 1970). Figure 4 above shows the language-specific boundaries for Spanish (0 ms) and English (+30 ms), as an example. Other studies found (1) this shift occurs around the age of 10 to 12 months and (2) babies raised in a Spanish-learning environment seem to experience this shift earlier, around six to eight months, than English-learning babies (Eilers et al., 1979; Kuhl et al., 2006). Furthermore, Kuhl and Miller (1978) showed that chinchillas have comparable auditory perception skills to English-speaking adults and share similar VOT boundaries, which suggests the transition from language-general to language-specific boundaries corresponds to a transition from psychoacoustic to linguistic perceptual processing. Hoonhorst et al. (2009) tested four-month-old and eight-month-old babies on their ability to perceive voicing contrasts in French, which has a 0 ms VOT boundary like Spanish. They presented the babies with a /də-tə/ VOT continuum and monitored their heart rates throughout the experiment, with a change in heart rate associated with the ability to discriminate sounds. The results showed a decreased sensitivity to the language-general perceptual boundaries and an increased sensitivity to the language-specific perceptual boundary, when comparing the four-month-olds to the eight-month-olds. Thus, they concluded that French-learning babies demonstrate the same shift from language-general to language-specific perceptual boundaries as Spanish-learning babies: before eight months old. The researchers suggest that this shift occurs earlier in French and Spanish because there is greater distance between the voiced and voiceless categories in these languages compared to English, in which the voiced and voiceless categories are adjacent and, thus, require greater perceptual acuity (Hay, 2005). The findings from the existing literature suggest that initial perceptual boundaries form in infancy and shift to accommodate the boundaries that exist in the language to which each baby is exposed.

Relevant to the present study, Escudero et al. (2011) and Chládková et al. (2022) found that exposure to the distinct categorical boundaries of a novel L2 aid in boundary formation in adults. Escudero et al. (2011) introduced L1-Spanish L2-Dutch adults to stimuli with either average vowel productions or extreme vowel productions of a complex Dutch vowel contrast; the results showed that the adults benefitted from the extreme vowel productions and showed improved categorization of this L2 vowel contrast. Chládková et al. (2022) introduced functional Spanish monolinguals to altered Spanish /e-i/ phonemic boundaries and found that, while they were able to shift their boundaries to match those to which they were exposed, it might be easier for these adults to acquire L2 contrasts that are similar to the contrasts they already established in their L1, as opposed to novel contrasts that do not exist in their L1.

In terms of production, based on a longitudinal study by Macken and Barton (1980a) involving 1.5-year-olds, the acquisition of L1 English word-initial stop consonants by monolingually-raised children tends to occur in the following pattern: initially the child does not have any voicing contrast; the child is able to produce contrasting VOT values, but the values fall within the same perceptual boundary for a single phoneme for adults, so this contrast is not yet perceptible by adults; and then the child is finally able to produce stop consonants in a way that is similar to adult production of these consonants. Macken and Barton (1980a) found that some children are able to make decent progress towards producing adult-like stop consonants as early as 2;0, though other children may need several more months, or even years, before they are able to produce this voicing contrast regularly. Macken and Barton (1980b) repeated a similar voicing contrast analysis of Mexican-Spanish word-initial stops. They recorded the speech of children ages 1;7 and 3;10, approximately, who spoke Mexican-Spanish and found that, even by the age of 3;10, three of the four children did not consistently produce the voicing contrast in an adult-like manner; the one child that demonstrated adult-like VOT only demonstrated this accurate voicing contrast for the bilabial stops (Macken & Barton, 1980b). This may suggest that the age of acquisition of voicing contrast is delayed in Spanish compared to English. One of the proposed reasons for this observation is that lead voicing may be more difficult to acquire than aspiration; regardless of the reason, Macken and Barton (1980b) conclude that the spirants [β,ð,ɣ] may be more relevant in the early acquisition of Spanish phonology, while VOT may be a less influential feature affecting the Spanish voicing contrast.

Although the results from the studies by Macken and Barton (1980a, 1980b) offer a nice foundation for L1 acquisition of voiceless plosives, we would like to focus on SLA of voiceless plosives. Fellbaum (1996) studied L1 English children learning L2 Spanish, and L1 Spanish children learning L2 English. They predicted that L1 English children would shorten the VOT values of the voiceless plosives when producing L2 Spanish, in order to match the Spanish duration. In accordance with the Markedness Differential Hypothesis (MDH), they also predicted the L1 Spanish children would struggle with lengthening their VOT values to match the appropriate VOT duration of L2 English voiceless plosives (Eckman, 1977; Fellbaum, 1996). According to the MDH, aspirated stops are more challenging to acquire than unaspirated stops for L2 learners because languages that contain aspirated stops also utilize unaspirated stops, while languages that contain unaspirated stops do not always make use of aspirated stops as well (Eckman, 1977). The results from this study failed to reject the null hypothesis; in other words, the L1 English L2 Spanish speakers and the L1 Spanish L2 English speakers did not produce the voiceless plosives significantly differently from each other (Fellbaum, 1996). This study also found that it is important to use an acceptable range of VOT values, as opposed to a mean VOT value, when determining accurate productions; Fellbaum (1996) used both a range of VOT values and an average VOT value for comparison, but only found meaningful results with the VOT range.

Fabiano-Smith and Bunta (2012) focused on the voiceless velar and bilabial stop consonants, /k/ and /p/, respectively. There is a VOT value contrast between these two consonants in both English and Spanish, as mentioned previously, yet there are also different VOT values for each consonant when comparing Spanish and English. For this study, the researchers recruited three-year-olds who were bilingual English-Spanish speakers, as well as their age-matched monolingual English and monolingual Spanish peers. The primary finding of this study was that the monolingual Spanish-speaking children and bilingual children produced Spanish words with similar VOT values, but the monolingual English-speaking children and bilingual children did not produce English words, specifically those with /k/, with similar VOT values. Furthermore, the bilingual children did not produce the Spanish and English voiceless stop consonants significantly differently, while the monolingual English and monolingual Spanish controls did produce these consonants significantly differently; on average, the VOT values produced by the bilingual speakers were shorter, more Spanish-like than the monolingual English speakers (Fabiano-Smith & Bunta, 2012). Fabiano-Smith and Bunta (2012) interpreted their findings within the context of Flege's original Speech Learning Model (SLM) and assumed the results support his theory of equivalence classification, or phonetic category assimilation (Flege, 1987b). Based on the results, it is possible bilingual children assimilate their production of the English voiceless plosives to Spanish, even though they learned both languages at a young age. These results may also offer support for the MDH, though it is still unclear whether equivalence classification or MDH, or another hypothesis, is most likely in play with respect to L2 phonetic trait acquisition.

Another study focused on acquisition of VOT by both monolingual and bilingual French-speaking children (Kehoe & Kannathasan, 2021). In this study, the bilinguals were divided into two groups: bilinguals whose L2 was a long-lag VOT language, including English, German, Norwegian and Swedish, and bilinguals whose L2, like French, was a lead VOT language, including Catalan, Italian, Portuguese, Spanish, and Romanian. The researchers utilized a memory game picture-naming task to elicit word-initial stop consonants. The results showed that the bilinguals in the long-lag subgroup had longer VOT values in French, if French was not their dominant language, and, compared to the monolinguals, they produced fewer of the target voiced stops with lead voicing. Thus, the researchers concluded that their results suggest cross-linguistic interaction in the phonetic domain is not always a byproduct of bilingual acquisition but that it can be observed if language dominance is considered a factor (Kehoe & Kannathasan, 2021).

1.2.4.1.2. Autistic Populations

Given that phonetic traits are often well-preserved in autistic populations, despite potential language delays and other hindrances, there is very limited research on the acquisition of voiceless stop consonants in autistic children (Eigsti et al., 2011). Furthermore, given the historical lack of bilingualism in studies of clinical populations, there does not seem to be any research related to L2 acquisition of voiceless stop consonants in autistic children. Thus, we will instead provide an overview of L1 acquisition of voiceless stops in autistic populations.

Beginning with perception, Bourdeau (2009) evaluated autistic children's ability to perceive English voiced and voiceless stop consonants compared to their age-matched TD peers, as well as younger TD peers. They recruited autistic children between eight to 14 years old, as well as TD peers that were either seven, or fell into the same eight to 14 age group. The stimuli consisted of the voiced and voiceless stop consonants [da] and [ta], respectively, with VOT values ranging from -15 ms to 40 ms. The results showed that the autistic children struggled to distinguish the voiced and voiceless consonants, and performed more similarly to the seven-year-old TD children than to their age-matched TD peers; thus, autistic children may be delayed in their categorical perception of voiced and voiceless stops (Bourdeau, 2009). Expanding on the potential delays observed in autistic populations, You et al. (2017) set out to answer the question of whether speech perception challenges in autism are related to difficulties with categorical precision, meaning they struggle to perceive the differences between phoneme categories, or an allophonic mode of speech perception, in which they have enhanced perception of acoustic differences of allophones, or phonemes in the same category, but lack coherence in distinguishing and identifying speech sounds; these potential sources for the difficulties observed in autistic populations were based on results from previous studies (DePape et al., 2012; Happé & Frith, 2006). The results of You et al. (2017) indicate that categorical precision difficulties are the source of perceptual challenges in autistic children, as they were less precise in their ability to identify the consonants and vowels, presented along three different continua, compared to their TD peers. Lastly, Chen and Peng (2021) found that Mandarin-speaking autistic adolescents, who are not severely impacted by their diagnosis, tend to have intact categorical perception abilities; that being said, they showed reduced perceptual abilities with respect to VOT, but not lexical tones, when compared to their TD peers. In conclusion, it seems as though autistic individuals may experience challenges and delays with respect to their ability to perceive voicing contrasts in their L1, but are expected to develop intact categorical perception abilities with time.

Shifting our focus towards production studies, both Wolk and Edwards (1993) and Wolk and Giesen (2000) investigated the phonological systems of monolingual English-speaking autistic children. Wolk and Edwards (1993) thoroughly studied the phonological system of a single eight-year-old boy. They elicited the child's speech via delayed imitation, object naming, and connected speech sample tasks, and then performed both a phonetic inventory analysis and a phonological process analysis. The results from this study showed that the child had a phonological system that was like his TD peers in some ways and different from his TD peers in other ways. While the boy exhibited some language delays (i.e., errors that persisted beyond the typical age) and some uncommon sound changes (e.g., cluster reduction, glottal replacement, segment coalescence, consonant epenthesis, reduplication, and metathesis), his production of nasals, glides, and, crucially, stop consonants was generally intact (Wolk & Edwards, 1993). Wolk and Giesen (2000) conducted a similar case study in which they analyzed the phonological systems of four siblings with autism diagnoses, ranging in ages from 2.3 to 9.0 years old. After analyzing the children's spontaneous speech and speech elicited via an object naming task, again, they performed a phonetic inventory analysis and a phonological process analysis (Wolk & Giesen, 2000). Their analyses revealed that autistic children do not only experience delayed language acquisition, rather they have acquisition patterns that are both aligned with those of their TD peers and atypical compared to their TD peers; these findings corroborated the previous findings from Wolk and Edwards (1993), with a single subject. The four siblings showed different degrees of restricted phonetic inventories and were categorized from oldest to youngest as having an almost full phonetic inventory with a few exceptions (age 9.0), a severely restricted phonetic inventory (age 5.9), a moderately restricted phonetic inventory (age 3.9), and a moderately to severely restricted phonetic inventory (age 2.3; Wolk & Giesen, 2000). Of the four children, the two with moderately restricted to full phonetic inventories produced both voiced and voiceless stops correctly and exhibited typical acquisition of these consonants. The other two children, with severely restricted phonetic inventories, did not show proper acquisition of the voiceless stop consonants; the youngest child, who was verbal, but had a severely restricted phonetic inventory, only produced word-initial, voiceless velar stops /k/ (Wolk & Giesen, 2000). Most importantly for our study, both the study by Wolk and Edwards (1993) and the study by Wolk and Giesen (2000) found that autistic children with moderately restricted to full phonetic inventories typically acquire voiceless stop consonants similar to their TD peers; thus, we would expect to observe a similar developmental trajectory in children with ALN, with mostly intact language skills and without an SLI diagnosis.

Chenausky and Tager-Flusberg (2017) also studied L1 acquisition of the contrasting VOTs between English /p/ and /b/ in children with a low risk of autism, children with a high risk of autism but no diagnosis by 36 months, and children with a high risk of autism and a diagnosis at 36 months. They evaluated the spontaneous speech of toddlers with these varying risk levels at the ages of 18, 24, and 36 months. The results showed that the high risk toddlers that received an autism diagnosis at 36 months old were the only participants that did not produce the acoustic distinction between /p/ and /b/; this may imply that the autistic toddlers either have not fully developed the ability to produce this contrast or they use alternate strategies, other than VOT, to produce this distinction. While Chenausky and Tager-Flusberg (2017) do not elaborate on these potential alternate strategies, they propose several reasons for the observed lack of acoustic distinction in autistic children including differences in brain structure and organization in autism, speech-related motor challenges, and difficulties perceiving speech sounds which may affect their ability to then imitate and produce said speech sounds. Despite the lack of bilabial voicing contrast produced by the high-risk and diagnosed toddlers in this study, the aforementioned studies by Wolk and Edwards (1993) and Wolk and Giesen (2000) found that the older children, ages eight and nine, and those without severely impacted phonetic inventories acquired voiceless stops similar to their TD peers. Therefore, it seems as though autistic children's ability to produce the bilabial voicing contrast eventually aligns with that of their peers.

While these studies are interesting and improve our understanding of L1 acquisition of voiceless stops by autistic individuals, we are primarily interested in L2 acquisition of voiceless stops in autism. We did not find any studies that specifically focused on the acquisition of L2 voiceless stop consonants by autistic children; we hope to contribute to filling this gap in the literature.

1.2.5. Gaps in the Literature

While all of the aforementioned studies are insightful in their own ways, we have identified five additional gaps in the literature, all of which we hope to address in the present study. We are primarily focused on the phonetic imitation studies that also relied on AQ scores, mentioned in Section 1.2.3., as our study will follow a similar design. The first gap in the literature is that all of the phonetic imitation studies focus on L1 phonetic imitation tasks of manipulated L1 speech sounds, but not L2 phonetic imitation. The second gap in the literature is that the studies using AQ scores only used these scores to assess the number of autistic-like traits TD individuals possess, but they were not used within autistic populations. The third gap in the existing studies that used AQ scores to predict phonetic imitation abilities is the use of adults only, but not children. The fourth gap we have identified in the imitation studies, and hope to address, is the evaluation of short-term imitation exclusively, but not long-term or ‘learning’. Finally, the imitation studies reviewed above focus exclusively on production tasks, not perception tasks. In order to address these gaps, we developed a similar design to the studies outlined in Section 1.2.3., with some additions and alterations. More specifically, in order to address the first two gaps, we conducted an L2 (Spanish) phonetic imitation task in both autistic and TD populations. We specifically recruited individuals with ALN, or autistic individuals without any existing speech disability diagnoses; thus, we were able to cautiously assume that these individuals would have mostly intact phonetic abilities. We recruited English-dominant individuals without any Spanish exposure, which limited potential pragmatic concerns. We further limited the effect of pragmatics by focusing on VOT as the dependent variable, like Yu et al. (2013), as opposed to prosodic prominence, like Bishop (2012). In order to address the third gap, we recruited child participants instead of adults. We specifically examined children at ages prior to the potential critical period for L2 acquisition. There is considerable evidence children tend to achieve better ultimate attainment in the L2 acquisition process compared to adults (Granena & Long, 2013; Hartshorne et al., 2018; Johnson & Newport, 1989; Lenneberg, 1967, a.o.). In any case, recruiting children ages seven to 12, without prior Spanish exposure, will reduce the concern of confounds. Additionally, we hope that the results from this study may influence how bilingualism is viewed and potentially implemented as a social intervention in schools and clinics that cater to autistic children, so we prioritized recruiting individuals belonging to the population whom we hope to serve through the present study. To address the fourth gap, we added a delayed imitation task to evaluate long-term learning, in addition to short-term learning, since language acquisition relies on long-term memory of the learned language skills, as opposed to only short-term imitation abilities. Lastly, to address the fifth gap, we included perception tasks throughout the experiment to gauge participants' perceptual abilities prior to exposure to Spanish, immediately following exposure, and delayed post-exposure.

1.3. Research Questions and Hypotheses

The research questions we hope to answer through the present study, and a brief preview of the comparisons that we will complete, across participant AQ scores, are as follows:

1. Do autistic-like traits predict L2 phonetic imitation ability?

   Comparison: participant's VOT during an immediate imitation task compared to their baseline VOT, as well as the VOT of the model speaker they are imitating
2. Do autistic-like traits predict short-term L2 phonetic learning?

   Comparison: participant's VOT shortly after exposure to the model speaker's speech compared to their baseline VOT
3. Do autistic-like traits predict long-term L2 phonetic learning?

   Comparison: participant's VOT delayed a few days after exposure to the model speaker's speech compared to their baseline VOT

   a. Is long-term L2 phonetic learning generalizable to new words?

   Comparison: participant's VOT when producing words they've seen before compared to novel words, delayed after exposure to the model speaker's speech
4. If short-term and long-term learning are predicted by autistic-like traits, is it due to enhanced perceptual abilities?

   Comparison: participants' perceptual boundaries compared to their VOT productions across the tasks specified in the previous two questions

Given that diagnosed autistic individuals and individuals with higher AQ scores have more autistic-like traits, traits including higher attention to detail and higher focus, we have generated the following predictions related to our research questions outlined above. We predict that individuals with more autistic-like traits will have stronger imitation abilities, will be better at short-term L2 phonetic learning, and will be better at long-term L2 phonetic learning. Although You et al. (2017) determined that autistic individuals have difficulties perceiving different phoneme categories, they also stated that there needs to be further research to understand the potential allophonic perception present in autism as well. Given that Happé and Frith (2006) and DePape et al. (2012) found that autistic individuals showed strong perceptual abilities of non-native speech sounds, it is possible they are better at noticing the details of a novel language, which could be a result of their heightened attention to detail. While there are a variety of second language learning models and theories, there is an argument for the use of articulatory gestures in perceiving novel sounds, as in the Perception Assimilation Model (PAM; Best, 1995), as well as an argument for the necessity of perception to precede production, as in the SLM (Flege, 1987a). Regardless of which models apply, any heightened awareness of the details of a novel language that autistic individuals may possess could allow for improved phonetic perception and production of the sounds of a novel L2. While accurate production of the sounds of a novel language is not indicative of an individual's ability to acquire the language fluently, it is certainly a crucial first step in being able to speak and produce that language in a way that can be understood by other speakers of that language. Given the findings presented by Mielke et al. (2013), we expect individuals with more autistic-like traits will not generalize any long-term L2 phonetic learning to novel words. Autistic individuals demonstrated a stronger dependency on explicit memory, which may reflect their heightened attention to detail and focus, and, thus, their ability to recall the fine-grained phonetic details they learned upon initial exposure (Cohen et al., 1997). Based on the findings of previous studies that relied on AQ scores, we may also expect to see a positive linear correlation between AQ score, or number of autistic-like traits, and both imitation ability and L2 phonetic learning (Snyder et al., 2019; Yu et al., 2013).

In order to address the research questions outlined above, we conducted two experiments: Experiment 1 and Experiment 2. Experiment 1 is designed to address the research questions from the perspective of short-term learning, while Experiment 2 is designed to address the research questions from the perspective of long-term learning. In order to answer the fourth question, we included perception tasks across both experiments.

1.4. Pilot Study

We conducted a pilot study with a very similar procedure to both Experiment 1 of the present study and that which was used by Yu et al. (2013). The purpose of the pilot study was to assess the feasibility of the study, refine the methodology, assess the time and cost associated with data collection, and allow the research team to gain experience running the experiment in order to improve reliability[3].

The pilot study had six American adult participants ages 18 to 25 who were all based in Los Angeles, CA and identified as monolingual English speakers. First, the participants completed a survey consisting of demographic questions, questions about their linguistic profile, and the AQ test. The participants were reminded of the purpose of the study, the tasks involved in the study, and that their participation was optional, before consenting to participate in the experiment. Following the survey, participants completed three production tasks in the following

---

[3] The results from this pilot study were presented at the New Sounds 2022 - 10th International Symposium on the Acquisition of Second Language Speech, the Meeting on Language in Autism 2023, and the Western Conference on Linguistics 2023.

order: a Spanish read-aloud pre-test, a Spanish immediate imitation task, and a Spanish read-aloud post-test. Each task had a brief practice activity prior to beginning the actual task. For the first task (pre-test), the participants read a list of Spanish words, each presented individually on a slide in the carrier phrase "di ___ de nuevo," ("*say ___ again*"). The stimuli consisted of 30 Spanish target words that began with /p/, /t/, or /k/ and 30 filler words with other word-initial consonants, repeated three times each, for a total of 180 words. All the token words were disyllabic nouns, with penultimate stress, e.g. padre (*father*), toro (*bull*), cama (*bed*). For the second task, the participants heard a recording of a native Spanish speaker saying the same words presented in the carrier phrase, in a different order, and repeated each phrase immediately after hearing the native speaker's speech. The stimuli for the second task were produced by a female in her twenties; she is a native Spanish speaker from Mexico who moved to the U.S. when she was 17. The third task (post-test) was the same as the first task, but the stimuli were presented in a randomized order, once again.

All of these tasks were recorded via the experimenter's personal computer, a MacBook Pro with macOS Monterey, using a microphone headset provided by the research team and sanitized between participants. The microphone headset used throughout the study was the Logitech H340 which has a frequency response of 20 to 20,000 Hz for the headset and 100 to 10,000 Hz for the microphone, and a sensitivity of -40 dBV/Pa +/- 3 dB. Each participant's speech was recorded in Audacity at a sampling rate of 44.1kHz and a sample size of 32 bits (Audacity Team, 2023). Qualtrics was used to upload and store the .wav audio recordings for each of the three tasks, per participant (Qualtrics XM, 2024). The participants were allowed to choose the location of the recording, such as at their own home or in a common space on UCLA's campus, but were instructed to choose a quiet, furnished space. Because data collection took place during the COVID-19 pandemic, only remote data collection was available. On average, the study took approximately 45 minutes for each adult participant to complete, with the survey taking approximately 10 minutes and the experimental tasks taking closer to 35 minutes. Each participant received $15 in the form of a Target e-gift card as compensation.

For the analysis, we used Praat to annotate each of the target features (e.g., VOT) before using a Praat script (adapted from a custom script by Christopher Carignan of University of Illinois at Urbana-Champaign) to get the duration of each of the labeled segments (Boersma & Weenink, 2023). We then calculated the average VOT for each initial consonant of interest, for each participant, for both the pre-test and post-test production tasks. We used RStudio to perform paired t-tests to compare the significance of the average VOT difference between the pre-test and post-tests for each participant (RStudio Team, 2020). With respect to the AQ test, each participant scored below the 26 point cutoff for being considered for an autism diagnosis, with the scores ranging from 11 to 24. We analyzed the effect of AQ scores on participants' VOT values by running a correlation test between the AQ scores and the VOT difference between pre-test and post-test. We also ran correlation tests between each of the AQ subscale scores of interest, attention to detail and attention switching, and the pre-test VOT.

Table 2 reports each participant's AQ score (lowest to highest), mean VOT per place of articulation for the pre-test and post-test, the standard deviation associated with each mean value reported, as well as the p-value showing the significance of the difference between the pre-test and post-test VOT values. Based on the results reported in Table 2, we can see that five out of the six participants reduced their VOT values when producing Spanish /p, t, k/-initial words between the pre-test and post-test. Only one participant produced longer VOT values in the third task compared to the first task. The three participants with the highest AQ scores showed a non-significant difference in their VOT values for some places of articulation: /p/, /t/, or /k/. The difference between pre-test and post-test performance was significant (indicated by an asterisk in Table 2) in 10/18 of the task comparisons carried out per consonant per participant. We adjusted the p-value to be 0.003 through Bonferroni correction (0.05/18 = 0.003).

AQ	Participant #	Task	PoA	Mean VOT	SD	p-value
11	2	pre	p	68.58	20.68	.000*
11	2	post	p	27.18	13.84	
11	2	pre	t	79.47	27.61	.000*
11	2	post	t	23.66	8.12	
11	2	pre	k	76.89	19.34	.000*
11	2	post	k	47.45	25.36	
12	3	pre	p	37.48	20.79	.002*
12	3	post	p	22.18	13.66	
12	3	pre	t	56.53	18.62	.000*
12	3	post	t	35.75	16.75	
12	3	pre	k	59.99	15.23	.000*
12	3	post	k	40.85	10.78	
15	4	pre	p	26.18	17.93	.002*
15	4	post	p	14.52	6.59	
15	4	pre	t	41.76	23.31	.000*
15	4	post	t	14.91	23.31	
15	4	pre	k	51.17	27.05	.000*
15	4	post	k	26.14	7.22	
15	6	pre	p	13.83	5.78	.009
15	6	post	p	18.43	7.17	
15	6	pre	t	16.95	3.77	.074
15	6	post	t	20.79	6.68	
15	6	pre	k	38.17	8.05	.831
15	6	post	k	37.66	8.61	
17	5	pre	p	36.41	21.32	.003
17	5	post	p	21.98	10.20	
17	5	pre	t	47.02	21.58	.216
17	5	post	t	40.07	15.77	
17	5	pre	k	63.56	22.31	.077
17	5	post	k	54.47	16.28	
24	1	pre	p	84.37	18.21	.276
24	1	post	p	76.37	20.69	
24	1	pre	t	109.59	35.22	.002*
24	1	post	t	81.46	19.34	
24	1	pre	k	106.25	18.31	.004
24	1	post	k	89.96	26.04	

Table 2. Pilot data: comparing pre-test to post-pest mean VOTs.

Each of three graphs below represents the relationship between the VOT difference (in ms) between the pre-test and post-test, and the corresponding AQ score, for /p/-initial words (Figure 5), /t/-initial words (Figure 6), and /k/-initial words (Figure 7). All three trend lines were negative; in other words, as the AQ score increased, the difference between the VOT production on the post-test to the pre-test decreased, for all three consonants tested. That being said, only the /t/-initial words showed a significant negative correlation between VOT difference and AQ score. This finding contradicted our initial hypothesis that monolingual English speakers with higher AQ scores would learn to produce Spanish /p, t, k/-initial words with a shorter, more accurate, native-like VOT than English speakers with lower AQ scores.

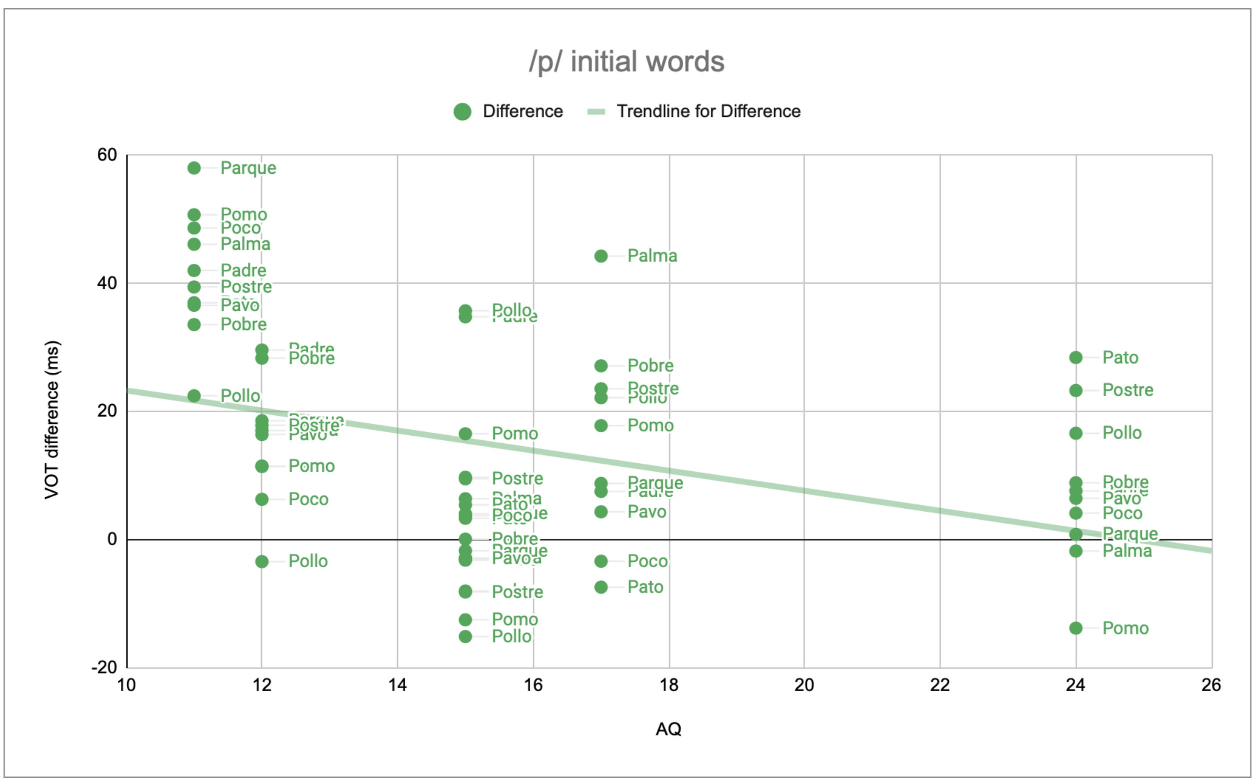


Figure 5. Pilot data: VOT difference for /p/-initial words by AQ score.

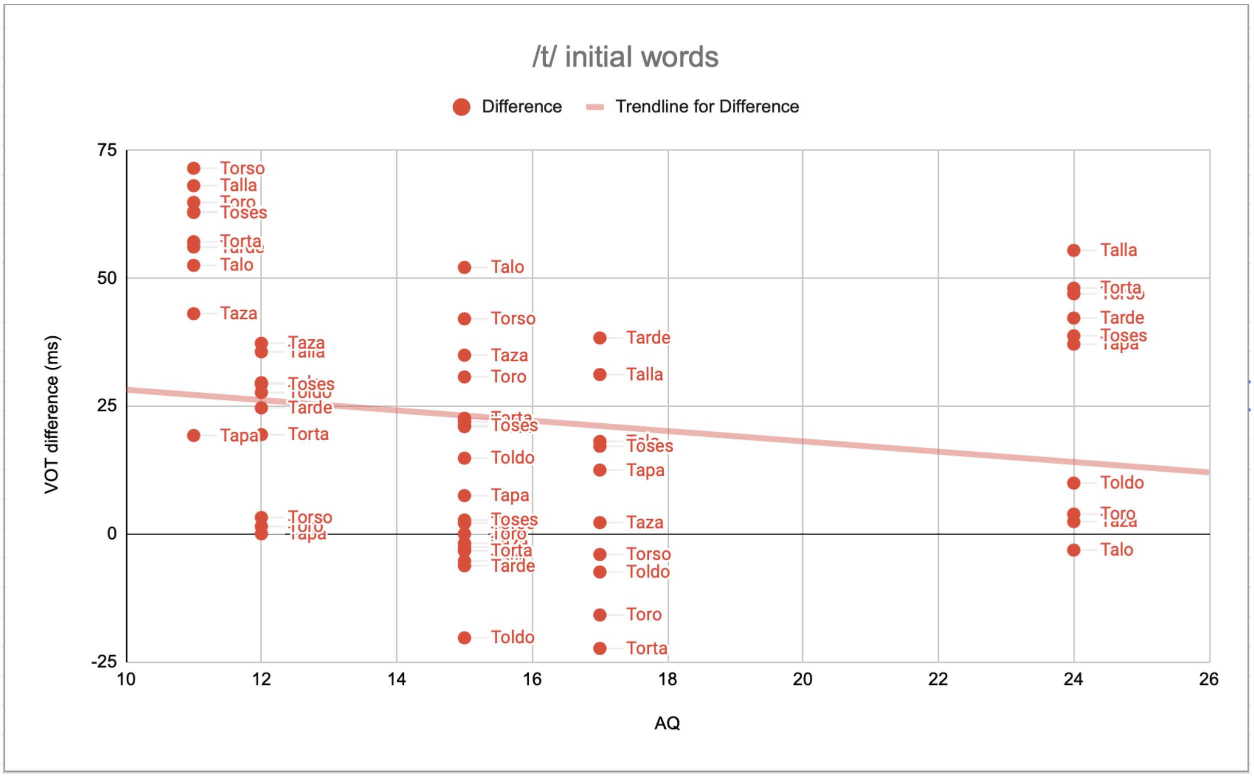


Figure 6. Pilot data: VOT difference for /t/-initial words by AQ score.

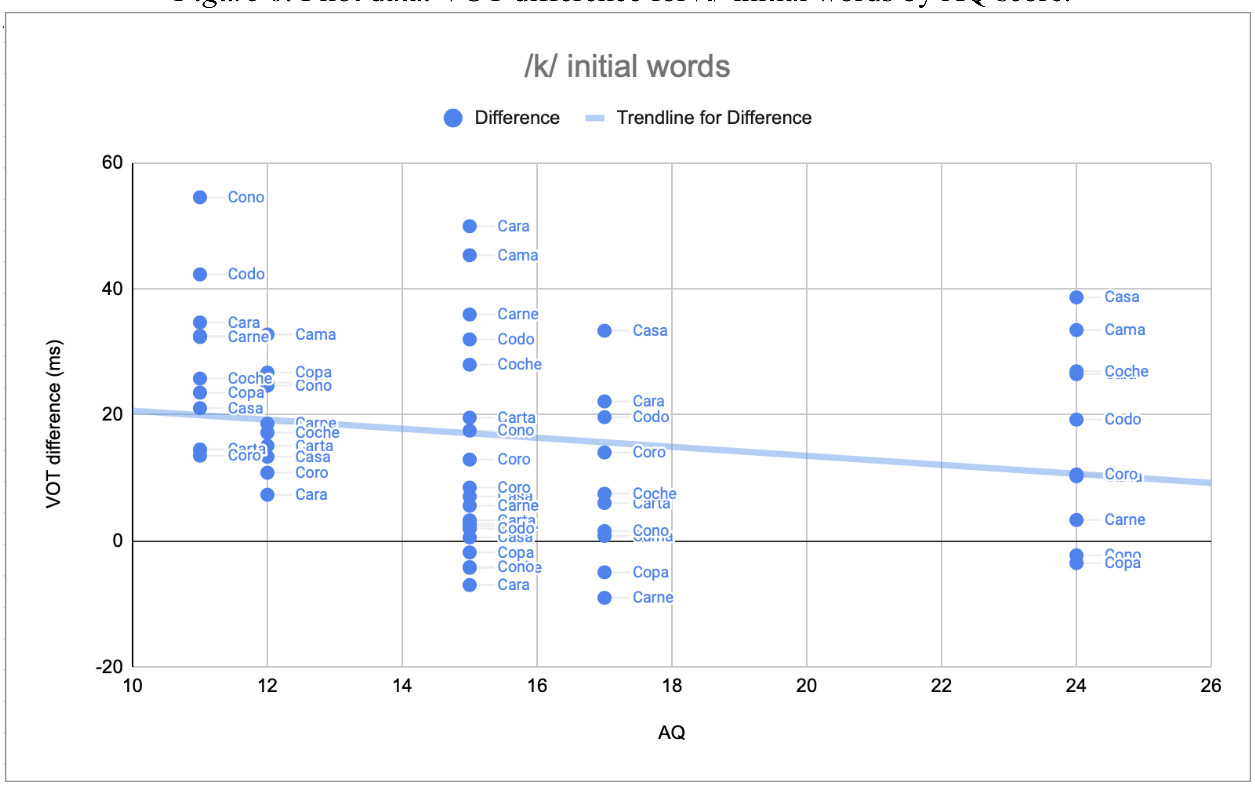


Figure 7. Pilot data: VOT difference for /k/-initial words by AQ score.

Although the majority of the participants did reduce their VOT values by the post-test, the need for this reduction varied across participants. Several of the participants produced relatively low and, in some cases, accurate VOTs when producing the Spanish words in the pre-test. It is possible these participants already picked up on and acquired the proper pronunciation of Spanish /p, t, k/ through their indirect Spanish exposure. Despite self-identifying as monolingual English speakers, the majority of our participants were raised in California or Texas, where they most likely received at least some Spanish exposure throughout their upbringing. AQ score may aid in this potential heightened ability to pick up on the sounds of a language indirectly, as we saw a decreased need to accommodate VOT in the three participants that scored 15 to 17 on the AQ test, as compared to their lower-scoring peers. That being said, the participant with the highest AQ score did not seem to experience this same perk. Perhaps this highlights the individual variation in either exposure to Spanish or ability to learn the fine-grained phonetic details of a novel language via this exposure. It is also possible we observed this trend by chance; it would be beneficial to expand the sample size of future studies.

Because the difference between pre-test and post-test performance was significant in the majority of the comparisons, participants generally seemed to perceive and learn to shorten the VOT when producing Spanish /p, t, k/-initial words. In our initial hypothesis, we predicted we would see the following trend: as AQ score increases, VOT difference increases. This prediction was based on the assumption that our participants were naive Spanish learners who would produce the pre-test stimuli with an English-like VOT, but would correct to a more Spanish-like VOT proportional to their AQ score. The results did not support our initial hypothesis. In fact, the following trend was observed for /t/: as AQ score increased, VOT difference decreased.

Through this study, we found a potential confound in our results: exposure to Spanish without explicit, formal education of the language.

With the results of the pilot study taken into consideration, we designed the present study with the intention of minimizing the confounds discovered in the pilot study.

Chapter 2: Experiment 1

2.1. Introduction

The overall goal of Experiment 1 is to gather baseline speech perception and production data from each of the participants, as well as introduce the novel L2, Spanish, to evaluate their ability to imitate and learn Spanish in the short-term. More specifically, we will address the following three of our four overall research questions, outlined in Section 1.3:

1. Do autistic-like traits predict L2 phonetic imitation ability?
2. Do autistic-like traits predict short-term L2 phonetic learning?

4. If short-term learning is predicted by autistic-like traits, is it due to enhanced perceptual abilities?

In this chapter, we will discuss both the methodology and results of Experiment 1 before summarizing the findings as a precursor to Experiment 2.

2.2. Methodology

2.2.1. Recruitment

For the purpose of the present study, we recruited English-dominant children ages seven to 12 that were either TD or neurodiverse with an autism or ADHD diagnosis. Since most autistic children are not diagnosed with the ALN or ALI distinction, we specified in our recruitment information that we were seeking autistic children, without a language disability diagnosis. The recruitment materials also specified that the children should not have any known hearing challenges, nor other neurological challenges; that being said, we recognize that autism and ADHD are often co-occurring, with similar, and sometimes overlapping, symptoms, so we did not exclude participants with either, or both, diagnoses (Antshel et al., 2013; Antshel &

Russo, 2019; Craig et al., 2015). According to Antshel and Russo (2019), autism and ADHD result in similar challenges related to social and executive functioning, but should still be considered distinct diagnoses. Craig et al. (2015) compared children with autism and ADHD to children with either diagnosis, as well as a control group. The results from this study showed that the individuals with both autism and ADHD had lower IQ scores and more severe autism symptoms than their peers with just one of these diagnoses, but behaved similarly to the ADHD group with respect to inattention and difficulties with hyperactivity, as well as challenges with emotions and behaviors. The co-occurring diagnosis group demonstrated challenges with adaptive behavior, similar to the autistic group. Thus, individuals with co-occurring autism and ADHD may show overlapping symptoms of both diagnoses, as well as additional symptoms due to their dual diagnoses (Craig et al., 2015). We also acknowledge that there is not a "pure" view of autism, as every autistic individual is unique and may have co-occurring diagnoses, whether that is ADHD, gastrointestinal (GI) issues, bipolar disorder, etc. As a common saying in the field of autism research goes, "if you've met one person with autism, you've met one person with autism" (Shore, 2016).

Given the nature of the study, we needed participants who were able to read short phrases independently. Nation et al. (2006) found that autistic children, ages six years old and up, performed similarly to their TD peers on reading accuracy, word recognition, and nonword decoding tasks. According to normative data, English phonology is mostly acquired by the age of six (Priester et al., 2011). Recognizing the language delays that are common in autism, we recruited participants ages seven to 12 to allow a one-year buffer for potential language delays or individual variation in L1 phonological system establishment (Eigsti et al., 2011). In addition to recruiting children without any speech, hearing, nor other neurological impairments, we specified the tasks that the child would complete in the recruitment materials so that parents could decide whether or not their child would be successful in completing the experiment.

Our initial recruitment took place in Los Angeles. By recruiting children with, theoretically, less societal-based language exposure due to their younger age and education level as compared to adults, we hoped to avoid the same potential confound discovered in the pilot study (Section 1.4): indirect exposure to Spanish. Unfortunately, even with the first participant, we noticed that the monolingual English-speaking child raised in Los Angeles was already unintentionally exposed to Spanish enough to know some of the features of the language and the differences between Spanish and English sounds. In an effort to minimize, and eliminate, this confound, we decided to take our study outside of, not only Los Angeles, but the U.S.; the present study was conducted entirely in Australia in order to reach a population that is more isolated from predominantly Spanish-speaking populations.

In 2019, 61.6% of U.S. households that reported speaking a non-English language at home reported Spanish as their primary home language, making it the second most popular language in the U.S. (Dietrich & Hernandez, 2022). In Australia, on the other hand, only 22.3% of households speak a language other than English at home, with Spanish being the 10th most popular non-English language spoken at home (Australian Bureau of Statistics, 2021; Zhang et al., 2023). We primarily recruited Australian participants from Sydney through Bright Eyes Early Intervention, an autism clinic based in two cities across New South Wales, Australia: Sydney and Katoomba. Although there may be minor dialectal and accent variation across these recruitment sites and across different speakers, we still expect their VOT values to be within the acceptable and expected ranges for English (Allen et al., 2003).

We did not require that the children were monolingual English speakers, as long as they were not exposed to Spanish, nor any other voicing languages, or languages with voiceless plosives with short-lag VOT. The other languages that our participants were exposed to include Mandarin, Cantonese, and Farsi, all of which are aspirating languages like English (e.g., Mandarin: Jensen, 1993; Cantonese: Lisker & Abramson, 1964; Farsi: Saljughian, 2012). We also ensured that the multilingual participants only spoke languages in language families that differ from Spanish; in other words, we would have excluded participants that reported prior exposure to any Romance language in the Indo-European family since we wanted to limit exposure to languages with similar roots to those of Spanish.

We recruited participants by sending recruitment emails to clinics and schools in the specified locations. We had virtual meetings and maintained ongoing communication with the lead psychiatrist at Bright Eyes Early Intervention prior to traveling to Australia for data collection purposes. This clinic led all participant recruitment efforts on the ground prior to our arrival; the participants that were interested in the study signed up in advance via Calendly in order to facilitate smooth data collection upon arrival (Calendly, 2024). We advertised the compensation rate of $25 USD, which is roughly $37 AUD, per parent-child pair following completion of Experiment 1, in order to drive recruitment (Appendix C).

Initially, our target recruitment size was roughly 26 to 34 children, ideally with a wide range of AQ scores. In order to determine the required sample size for the study, we completed a power analysis using the “pwr” package in R (Champely, 2020; v4.2.3; R Core Team, 2023). Given the nature of the study, we predicted we would utilize a paired, two-tailed t-test for the statistical analysis of the data. Thus, for the power analysis, we utilized the standard power and significance level values of 0.80 and 0.05, respectively. Based on our pilot data, which was conducted on neurotypical adults, we set the effect size at 0.576, which falls in the range of a medium effect size. The required sample size at this effect size resulted in 26 participants. We also set the effect size to 0.5, the standard Cohen's d value for an anticipated medium effect size, and found a required sample size of 34 participants (Cohen, 2013).

Despite this target recruitment size, we shifted our expectations to a slightly lower number of participants once we were in the field collecting data and have redirected our analysis to focus more heavily on case studies. Given the heterogeneity of autism, this analytical approach seems the most appropriate.

2.2.2. Participants

We recruited 22 participants ages seven to 12, all of whom identified as either monolingual English speakers or English-dominant multilingual speakers, with no exposure to Spanish. All of our participants fell within the seven to 11 age range, as no 12-year-olds participated in the study. We excluded the single participant from Los Angeles, who we determined had too much Spanish exposure, despite his identification as a monolingual English speaker. Three of the children that attended the sessions were not able to read the short phrases independently, rather required assistance in sounding out the words and frequently paused while reading each phrase; thus, we had to remove their data as well. After excluding the participants that did not meet all of the inclusion criteria, we were left with 18 participants.

All 18 of our participants were based in New South Wales, Australia, either in Sydney or Katoomba. As evidenced by the following table (Table 3), which shows the demographics and linguistic profiles of our participants, four speakers were exposed to a language other than English at home: Cantonese (n = 1), Farsi (n = 1), and Mandarin (n = 2). Only the participants that were exposed to Mandarin at home also spoke the language, while the other participants had receptive knowledge of the home language. We had nine neurodiverse participants with an autism (*n* = 6), ADHD (*n* = 2), or co-occurring autism and ADHD (*n* = 1) diagnosis, and nine TD participants. Given the similarities and overlap between autism and ADHD diagnoses, we welcomed participants with either, or both, diagnoses (see Antshel et al., 2013; Antshel & Russo, 2019; Craig et al., 2015). We later categorized participants by their AQ scores, as opposed to their diagnosis, as discussed in our analysis (Section 2.2.6). Although all of our recruitment efforts occurred in collaboration with Bright Eyes Early Intervention clinic, the psychiatrists at each location invited friends and family members who were eligible to participate in the study regardless of whether or not they were receiving services from the clinic, allowing us to capture a more developmentally diverse population. We would have excluded autistic children diagnosed with an ongoing speech disability or language challenge, as well as children with neurological or hearing challenges; however, none of these concerns were applicable to our participants. Of the 18 participants whose data we kept, Participant 219 (P219) had disordered speech and echolalia at age three and P211 had a previous speech disability diagnosis, also in the early stages of language acquisition. It was unclear whether either child still experienced any language-related challenges as a result of their prior diagnoses, but neither child struggled to complete the experimental tasks; thus, we included their data in our analysis.

Participant Code	Biological Sex	Age	Diagnosis	Languages
P202	Female	9	autism	English, Cantonese *exposure*
P203	Male	10	autism	English
P204	Male	10	autism	English, Mandarin
P205	Female	7	TD	English, Mandarin
P206	Female	9	TD	English
P208	Prefer not to say	8	autism, ADHD	English
P209	Male	7	autism	English, Farsi *exposure*
P210	Female	10	ADHD	English
P211	Male	9	autism	English
P212	Male	9	TD	English
P213	Male	7	TD	English
P214	Male	11	ADHD	English
P215	Female	7	TD	English
P217	Male	11	TD	English
P218	Female	8	TD	English
P219	Male	9	autism	English
P220	Female	9	TD	English
P221	Female	11	TD	English

Table 3. Participant demographics and linguistic profiles.

2.2.3. Research Materials

Experiment 1 included a parent survey completed on behalf of the child, and one in-person experimental session consisting of both perception and production tasks. The parent survey was created via Qualtrics (Qualtrics XM, 2024). It was composed of both an online version of the AQ test and a questionnaire with demographic questions, diagnosis questions, and questions related to the child's linguistic profile (Appendix D). The survey also included the research information sheet, both for viewing and downloading, which reminded the parent of the purpose of the study, the tasks that their child would complete, and that both their child's and their participation was completely optional (Appendix E).

2.2.3.1. Perception Task Materials

While we were primarily interested in how well the participants were able to produce, learn to produce, and imitate Spanish stop consonants, it was crucial to evaluate whether the participants were able to perceive these fine phonetic details. We evaluated each participant's perceptual abilities prior to exposure to the model speaker's speech, as well as immediately following this exposure. The perception tasks allowed us to understand each participant's baseline perceptual boundaries and, if any, their short-term learned perceptual boundaries, though this was not the main focus of the study. To examine each participant's VOT category boundary between voiceless stops /p, t, k/ and voiced stops /b, d, g/, we conducted a perception task with 12 stop-initial syllables, using the six aforementioned stop consonants, followed by either /o/ or /a/, with varying VOT values. We manipulated the VOT values of the initial stop consonants in the following minimal pairs: /pa/ and /ba/, /ta/ and /da/, /ka/ and /ga/, /po/ and /bo/, /to/ and /do/, and /ko/ and /go/.

To elaborate, we recorded a female in her twenties who is a native bilingual speaker of both Spanish and English. Our model speaker grew up in the U.S., though both of her parents were raised in Mexico. We used a bilingual native speaker to record the experimental stimuli, as she was able to produce both Spanish and English sounds native-like (see Table 6 in Section 2.2.3.2 below). We had our model speaker produce the syllables /pa/, /ba/, /po/, /bo/, /ta/, /da/,

/to/, /do/, /ka/, /ga/, /ko/, and /go/ with both Spanish-like and English-like pronunciation in order to gather a full range of VOT values for each minimal pair, which we could then modify to create the necessary continua, discussed below. The syllables were presented on Google Slides, one syllable per slide, in the same size and font throughout. The native speaker was asked to read all of the syllables in Spanish three times for the first task before repeating the same syllables in English three times for the second task (Figures 8 and 9). This task design allowed us to ensure we captured the full range of VOT values for both Spanish and English.

Figure 8. Read-aloud directions for the Spanish syllables.

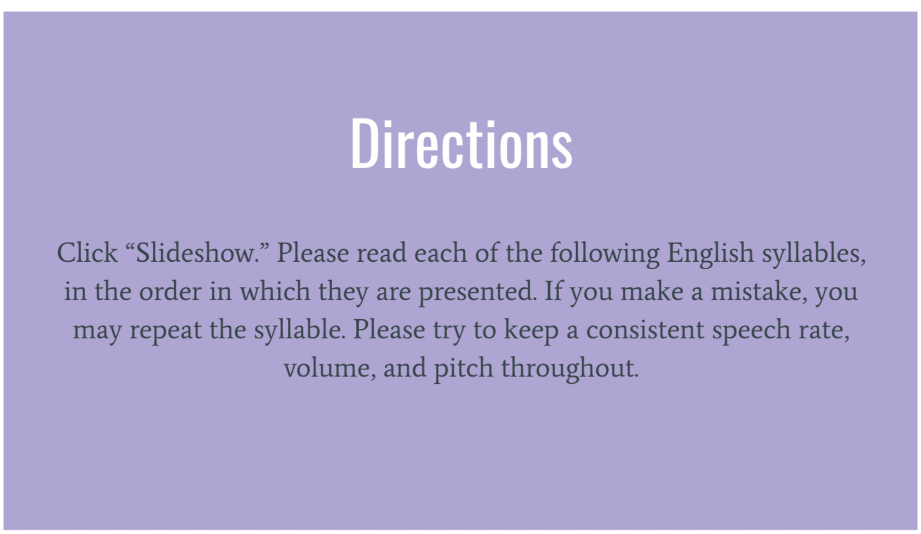


Figure 9. Read-aloud directions for the English syllables.

For each place of articulation we created a set of stimuli with a unique range of VOT values, modeled after Kellogg and Chang (2023), which captured the full range of VOT values by using the highest to lowest values per boundary (see Figure 4 in Section 1.2.4 above). In other words, for each place of articulation, the VOT range included the lowest VOT value typical of Spanish voiced stop consonants /b, d, g/ and the highest VOT value typical of English voiceless stop consonants /p, t, k/. For the bilabial consonants /b/ and /p/, we manipulated the model speaker's productions to include VOT values ranging from -70 ms to 60 ms (Figure 10). For the coronal consonants /d/ and /t/, the VOT values included ranged from -80 ms to 70 ms (Figure 11). For the velar consonants /g/ and /k/, the VOT values ranged from -60 ms to 80 ms (Figure 12). We included stimuli for each range in 10 ms increments (Lozano-Argüelles et al., 2021); thus, the bilabial VOT range included 13 steps, the coronal VOT range included 15 steps, and the velar VOT range included 14 steps, for a total of 42 steps, or stimuli, across all three places of articulation.

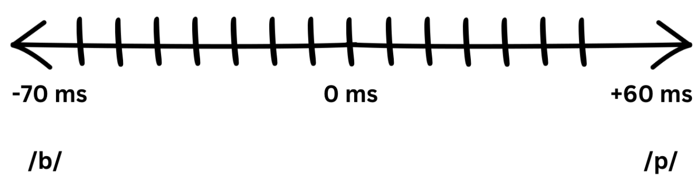


Figure 10. Bilabial VOT range.

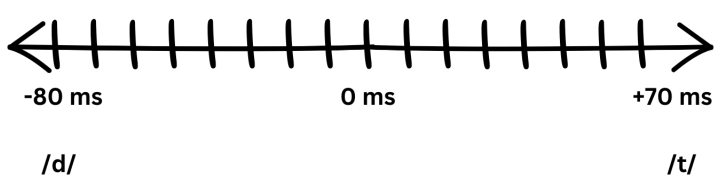


Figure 11. Coronal VOT range.

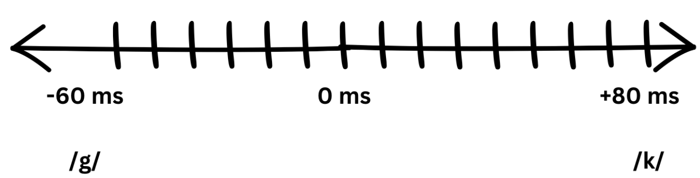


Figure 12. Velar VOT range.

We had 84 stimuli total since we had 42 steps across the three consonants and syllables with both /o/ and /a/ as the vowel following the consonant. Using Praat, we performed cross-splicing in order to ensure the vowel sounds were the same across the minimal pairs; all of the vowels used were from the /b, d, g/-initial syllables when produced in Spanish (Boersma & Weenink, 2023). For all 84 stimuli, we manipulated the pitch, vowel duration, and intensity of the stimuli in an effort to eliminate secondary perceptual cues, as recommended by Winn (2020a). We used Praat to set the pitch range from 70 to 200 Hz, with the average pitch per stimuli set at 180 Hz. The vowel duration was 150 ms for all of the stimuli. The intensity was set at 62 dB for all of the stimuli, using a Praat script (Winn, 2020b). In order to test the full range of VOT values and the participant's perceptual boundary across all three places of articulation, the participants were presented with 42 stimuli in a randomized order for each of the perception tasks. We randomly selected whether the participant would hear syllables with /o/ versus /a/ for each place of articulation for each perception task, but the vowel following the consonant was consistent per place of articulation per perception task, and all three places of articulation were tested per perception task. We did not specify the language of the perception task, rather presented the stimuli in a kid-friendly manner, stating in the directions that they would select the sound they heard our "robot friend" saying per audio file (Figure 13). These directions remained consistent across all of the perception tasks.

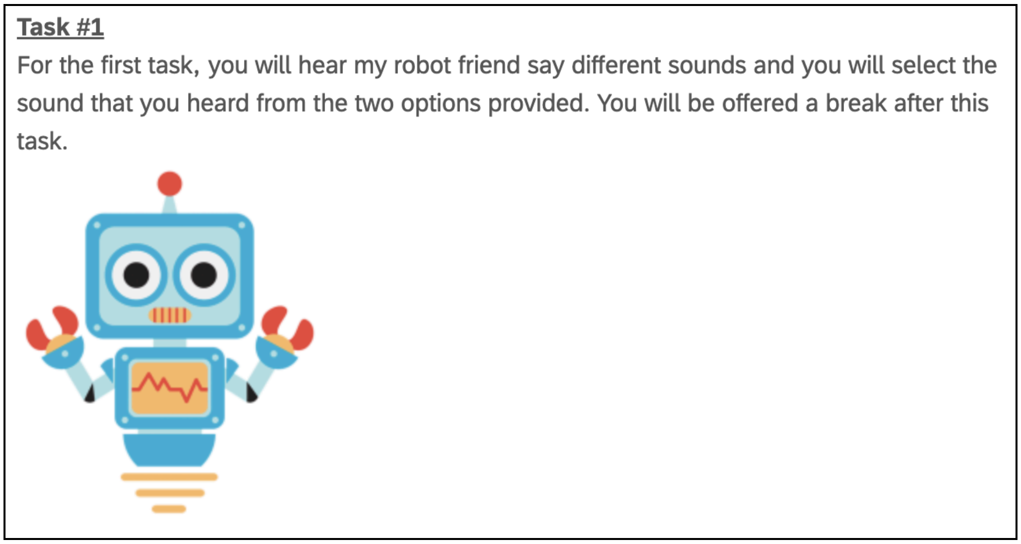


Figure 13. Directions for perception tasks.

In order to confirm that the perception stimuli were manipulated well and free of secondary cues, we created an online survey through Amazon Web Services Mechanical Turk (AWS MTurk) featuring all 84 of the perception stimuli. We asked participants for their biological sex, age, native language, as well as other languages that they have either been exposed to or spoke. While we asked the participants for their language experience to confirm that they only spoke English, we also restricted participation in the online survey to areas in the U.S. with minimal Latino, Hispanic, and Spanish-speaking populations, per the census data (Berry, 2018; Berry, 2022). For example, we invited participants from North Dakota, Ohio, and Oklahoma; North Dakota and Ohio have less than a 5% Latino population, while Oklahoma has less than a 13% Latino population. Furthermore, just over 2% of the population of Ohio speaks Spanish, around 4% of North Dakota residents speak Spanish, and less than 8% of Oklahoma's population speaks Spanish. We paid participants $1 each, using personal funds unrelated to any funds designated for the present study, to complete the 10-minute survey. Through this survey, we were able to confirm that there were not any secondary cues negatively impacting our perception stimuli; in other words, the perception stimuli seem to effectively and accurately represent the minimal pairs presented. We determined this by evaluating which sounds our monolingual English-speaking participants, with presumably minimal Spanish exposure, heard for each stimuli across the VOT range. It seemed as though our participants were randomly guessing which sound they heard at times, as their responses were not consistent with the VOT values presented. Although there did not appear to be a definitive trend in the data, it seemed as though the syllables that were the easiest for most participants to distinguish were syllables with more clearly voiced or voiceless VOT values. We interpreted this to mean that the participants were not utilizing secondary cues to accurately predict the sound and, thus, concluded that we had eliminated any potential secondary cues from the perception stimuli. We did not perform any further analysis with the MTurk survey data.

2.2.3.2. Production Task Materials

We have included a variety of production tasks to understand both the participants' baseline VOT values as well as their learned VOT values. The production tasks include both English and Spanish read-aloud tasks and a Spanish immediate imitation task, all of which are self-paced. We have included baseline production tasks in both English and Spanish in an effort to understand each participant's baseline VOT values, from the perspective of their native language and prior to any exposure to a model Spanish speaker, in order to confirm that their baseline VOT in Spanish is the same as their baseline VOT in English, assuming the participant is not attempting to "sound foreign," altering their VOT.

We presented 12 English words in the carrier phrase "say ___". Six of these words began with /p, t, k/ and six were filler words, which began with consonants other than /p, t, k/ (Table 4). We used disyllabic nouns with CVCV structure for all of the English words, and each consonant was followed by the vowel "o" or "a". All of the words had initial stress.

Target Words	Filler Words
patty	baby
pogo	saga
taxi	lady
tofu	hobo
camo	logo
cola	soda

Table 4. English production task stimuli.

We presented both a directions slide and an example slide prior to the experimental slides (Figures 14 and 15); the example slide followed the directions slide.

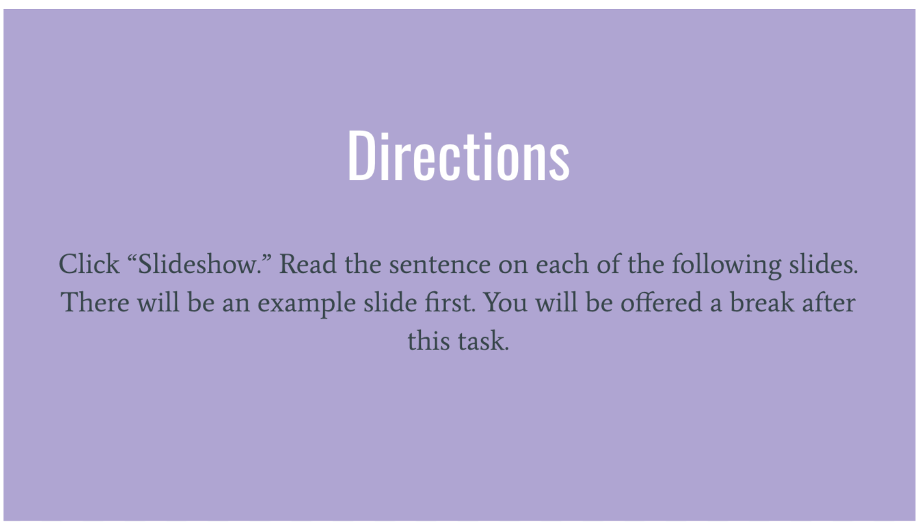


Figure 14. Directions for production tasks.

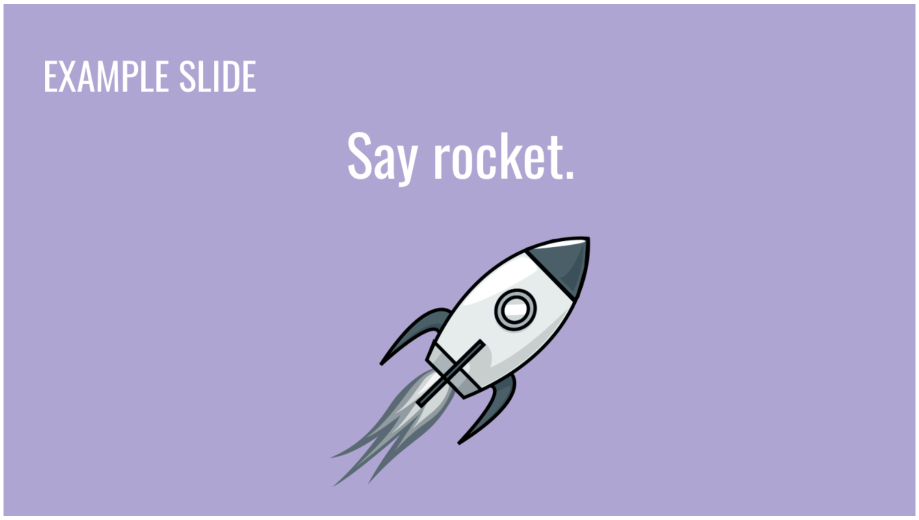


Figure 15. English example slide.

The Spanish production tasks included a randomized list of 18 /p, t, k/-initial words and 18 non-/p, t, k/-initial filler words in Spanish, in the carrier sentence “di _____” (“*say*___”) (Table 5). The token and filler Spanish words were all disyllabic nouns, with CVCV structure, controlled for penultimate stress.

Target Words	**Filler Words**
pato	mano
palo	gato
pavo	lago
pomo	vaca
poco	gafa
poro	mapa
talo	bala
taza	baba
tapa	rana
tomo	mono
toro	lobo

topo	ropa
cama	moda
cara	goma
casa	moto
codo	boca
copa	loro
coro	foso

Table 5. Spanish production task stimuli.

While Spanish has a high grapheme to morpheme correspondence, we also provided an instructional slide (Figure 16), after the directions slide (Figure 14), with an example sentence prior to beginning the Spanish production tasks since it was a novel language for all of our participants. However, the training item did not include any word-initial stops. The participants were not given any feedback regarding their pronunciation; rather, the participants were reminded to read the full phrase, as it was presented on the slide, without skipping the carrier phrase nor translating the word based on the accompanying image. We included six words that began with each of the three voiceless stop consonants, for a total of 18 Spanish target words in each of these tasks. Furthermore, we utilized three words with the vowel “o” following the stop consonant and three words with the vowel “a” following the stop consonant, for each of our three stop consonants of interest (Table 5). Half of the filler words consisted of the vowel “o” following the non-/p, t, k/-initial word, while the other half had the vowel “a” following the consonant.

Figure 16. Spanish example slide.

For the immediate imitation production task, the participants heard an audio recording of the model Spanish speaker reading the same 18 target words and 18 distractor words, in the same carrier phrase, in a randomized order (Table 5). We've included an example of the directions slide that preceded the imitation task below (Figure 17).

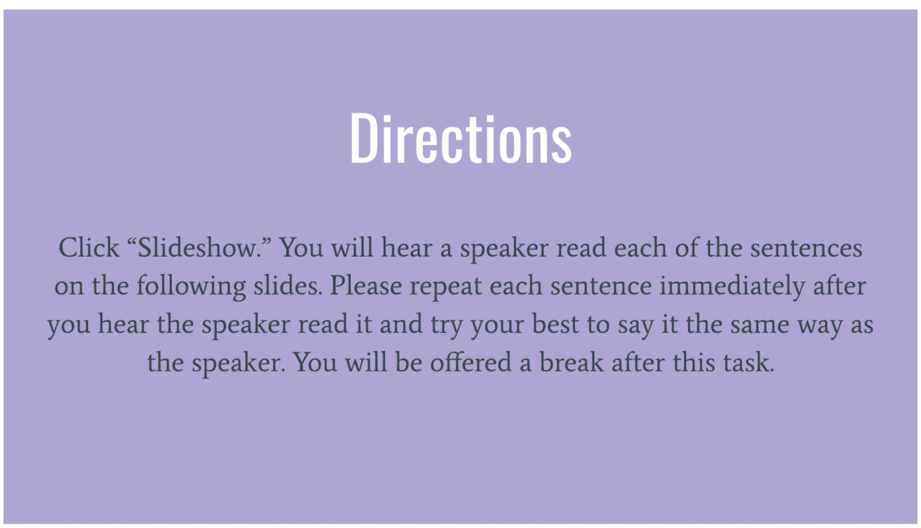


Figure 17. Directions for immediate imitation task.

The model speaker recorded the stimuli using the same microphone that the participants used and she recorded the stimuli in a quiet, furnished room to reduce background and echo noise. We segmented and labeled the model speaker's speech using Praat; we also used this software to confirm that her VOT values were close to, or within, the acceptable range for Spanish. (Boersma & Weenink, 2023; Castañeda, 1986; Lisker & Abramson, 1964; Michnowicz & Carpenter, 2013; Zampini, 2013). Our model speaker's average VOT values per place of articulation are included in the table below (Table 6). We included the acceptable VOT values provided by Michnowicz and Carpenter (2013) since our model speaker's family is from Michoacán, México and the Yucatan peninsular Spanish is likely the most similar to that of the Michoacán variant[4]. We also included the typical English VOT values provided by Lisker and Abramson (1964), for comparison sake. Zampini (2013) did not specify VOT values per place of articulation, but they reported an acceptable range for English VOT values for /p, t, k/ of 30 to 100 ms.

Place of Articulation	**NS Average VOT (ms)**	**Acceptable Spanish VOT (ms)**	**Typical English VOT (ms)**
/p/	16.8	9.17 - 22.24	58 ms
/t/	22.1	14.27 - 27.53	70 ms
/k/	42.6	23.09 - 41.12	80 ms

Table 6. Model speaker's average VOT values.

Similar to the parent survey, we utilized a Qualtrics survey to consolidate all of the tasks for the child participants, as well as present the information about the experiment to the child prior to assenting to participate in the experiment (Appendix F; Qualtrics XM, 2024). All of the production task stimuli were presented on Google Slides, with the same font style and size throughout, linked to the Qualtrics survey (Figures 18, 19, 20, and 21).

[4] Our model speaker's average VOT value when producing /k/ was 1.48 ms above the range expected for Yucatan peninsular Spanish, meaning some of her productions had longer VOT values than would typically be considered native-like. This may be the result of dialectal variation, not encompassed by the values for the Spanish of the Yucatan peninsula.

Figure 18. Example of English target stimuli slide.

Figure 19. Example of English filler stimuli slide.

Figure 20. Example of Spanish target stimuli slide.

Figure 21. Example of Spanish filler stimuli slide.

Each slide also included a single, black and white clipart image related to the presented token so as to encourage the learning of the word, i.e., a concept-label association. The tasks were recorded as .wav files via the research team's password-protected MacBook Pro with macOS Monterey and Audacity recording software. Each participant's speech was recorded in Audacity at a sampling rate of 44.1kHz and a sample size of 32 bits (Audacity Team, 2023). All of these tasks were recorded using a microphone headset provided by the research team and sanitized between participants. The microphone headset used throughout the study was the same as that of the pilot study: the Logitech H340, which has a frequency response of 20 to 20,000 Hz for the headset and 100 to 10,000 Hz for the microphone, and a sensitivity of -40 dBV/Pa +/- 3 dB. Qualtrics was used to upload and store the .wav audio recordings, as well as an .mp3 backup recording, for each of the tasks, per participant (Qualtrics XM, 2024). We presented the perception stimuli audio files within Qualtrics, with the corresponding responses recorded as part of the survey. There was not any identifiable information linked to the parent survey nor the participant's recordings, but we did link the two using a designated participant code such as "P201," "P202," etc. All of the stimuli were presented in a random order. Although we acknowledge the low number of target stimuli utilized in this study, we were working with a group with limited patience and therefore made this decision strategically in an effort to encourage participants to complete all of the tasks.

2.2.4. Procedure

The experiment procedure had two parts: the parent survey, which the parent completed at their own convenience and on behalf of the child, and the in-person experimental tasks completed by the child. The parents received the Qualtrics survey link via email upon scheduling and receiving confirmation of their child's experimental session through Calendly (Calendly, 2024). We required that parents completed the survey prior to the child's experimental session, as the parent provided their consent to allow their child to participate in the experiment within the survey. Parents were asked to spend 10 to 15 minutes completing their survey prior to their child's experimental session; according to the survey responses, it took parents anywhere from four to 25 minutes to complete this survey. Throughout our communication with the parents, we made sure to welcome questions at any point before, during, or after their involvement in the experiment.

At each clinic location, we collected the data in a quiet, furnished room in order to minimize background and echo noise. Parents were invited to be present in the room while their child completed the experimental tasks, or they were allowed to wait outside the room in the lobby, depending on their child's preference and their own preference. We estimated the experimental session would take between 20 and 30 minutes, depending on how many breaks each child decided to take. Each participant was given as much time as they wanted and needed to read through the information about the experiment, ask any questions they had before beginning the experiment, and assent (Appendix F). Similar to the parent survey, each child agreed to participate in the experiment by proceeding with their Qualtrics survey (Qualtrics XM, 2024). Prior to beginning the experiment, the research team set up the headset microphone and made sure the mouthpiece was in the appropriate placement, within a couple of inches of the child's mouth. The research team also explained the function of Audacity in recording the child's speech and began the recording at the start of the experimental session (Audacity Team, 2023).

The child completed the perception and production tasks in the following order:

1. English baseline: read-aloud task (*production task #1*)
2. Perception task #1
3. Spanish pre-test: read-aloud task (*production task #2*)
4. Spanish immediate imitation task (*production task #3*)
5. Perception task #2
6. Spanish immediate post-test: read-aloud task (*production task #4*)

We offered the child breaks between each of the six experimental tasks outlined above, with a recommended break time of two minutes; we provided them with the option to color during these breaks and we supplied the coloring materials. We've included an example of the break slide that was shown after each of the task slideshows, except for the last task at which point the child was excused to leave (Figure 22). For the children that decided to take a break, we set a timer for two minutes on the research team's personal cellular device in order to keep the child on time with respect to completing the experiment in 30 minutes maximum.

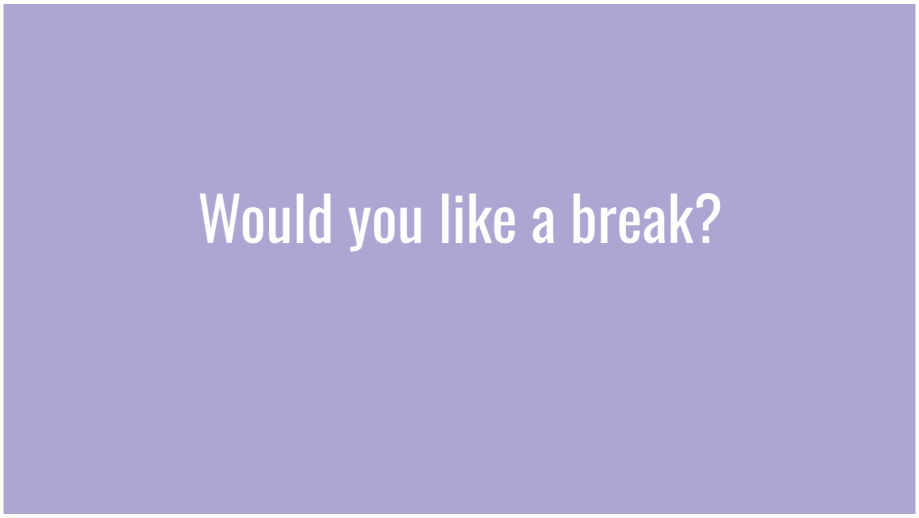


Figure 22. Break slide shown between experimental tasks.

For all of the experimental tasks, regardless of task type, we allowed the child to choose whether they wanted to read the directions aloud or have the research team member read the directions to them. The research team member conducting the experiment ensured each participant understood the task at hand and was ready to participate.

For the perception tasks, the participant heard the audio at least once before selecting which sound they heard in the format of a two-alternative forced-choice task. If the child wanted to replay the audio, they were allowed to do so. We allowed the child to choose whether they wanted to scroll through the audio tasks and use the trackpad on their own to select the option they heard, or if they preferred to have the research team member scrolling and selecting the sound for them. In the latter case, the research team member asked the child if they heard the first or second option listed, as opposed to repeating the syllables shown.

For the English and Spanish production tasks, barring the immediate imitation task, the participant simply read the stimuli presented in the carrier phrase, in either English or Spanish, as each sentence appeared per slide. The immediate imitation task had a similar design, but the participant first heard an audio recording of the native Spanish speaker reading each sentence before immediately repeating the sentence after them. Again, the participant was allowed to choose whether they wanted to be responsible for navigating through the slides presenting the production stimuli independently, or with the support of the research team member present. The research team read the example phrase on the English example slide to the child in order to demonstrate that the child would read the word "say," in addition to the target word, for each of the stimuli that followed (Figure 15). For the Spanish example slide, we specifically highlighted the pronunciation of the word "di" given its crucial role as the carrier phrase throughout the study (Figure 16). For the imitation task, the participant was asked to repeat the phrase to sound like the model speaker, per the directions slide (Figure 17).

The research team monitored the flow of the experimental tasks and breaks, and ensured all of the tasks were recorded properly. At the end of the experiment, we provided the parent and child with our contact information and reminded them that they were more than welcome to contact us to ask any questions they had following the study. Upon completion of all parts of the experiment, the parent-child pair received $25 USD (approximately $37 AUD) in the form of an Amazon gift card as compensation.

2.2.5. Variables

Of the research questions outlined above (Section 1.3), Experiment 1 addressed questions (1), (2), and part of question (4), reiterated below:

1. Do autistic-like traits predict L2 phonetic imitation ability?

2. Do autistic-like traits predict short-term L2 phonetic learning?
4. If short-term learning is predicted by autistic-like traits, is it due to enhanced perceptual abilities?

For the purpose of Experiment 1, we were interested in measuring one dependent variable: VOT. In measuring VOT across each of the production tasks, we could then compare VOT values from these different tasks to answer the research questions outlined above in a stepwise, strategic approach. To answer question (1) listed above, we compared the participants' VOT values on the Spanish immediate imitation task (production task #3) to the Spanish pre-test (production task #2) to understand their ability to imitate the model speaker by shifting their VOT values when speaking the novel language, upon exposure to the model speaker. Next, to answer question (2), we compared the participants' VOT values on the Spanish immediate post-test (production task #4) to their reference level VOT values, Spanish pre-test (production task #2), in order to understand their imitation ability. Furthermore, we compared our participants' VOT values on the Spanish immediate imitation task (production task #3) and Spanish immediate post-test (production task #4) to understand whether changes in VOT, if any, are maintained shortly after the imitation task (i.e., short-term learning). By relying on VOT as our dependent variable, we were able to analyze whether or not the participants were sensitive to fine-grained details; in other words, we were able to evaluate how well they accommodated their baseline VOT values and imitated the shortened VOT values of the tokens produced by the model speaker.

The independent variables in this study included general AQ score or specific AQ subscale score (e.g., attention to detail), task (English baseline, Spanish pre-test, Spanish immediate imitation, Spanish immediate post-test), and place of articulation (/p/, /t/, or /k/). AQ score and AQ subscale score were treated as continuous variables, as the sum test scores range from zero to 50 and the subscale scores range from zero to 10. For the AQ subscale score, we looked at two of the five AQ subscales: attention switching and attention to detail. We utilized the participants' diagnoses, or lack thereof, to validate their AQ scores, though we did not treat diagnosis as an independent variable. Task was a categorical variable and the number of levels was two, but the two selected levels varied. For example, we compared the English baseline and Spanish pre-test to understand the participants' English and Spanish baseline VOT categories, the Spanish immediate post-test and the Spanish pre-test to understand short-term learning ability, and the Spanish immediate imitation task and Spanish pre-test to understand participants' ability to imitate the model Spanish speaker. Place of articulation was also a categorical variable with three levels: /p/, /t/, and /k/, and we compared all three levels to each other.

The random effects for this study included participant and the experimental item, while the covariate included the vowel, either /o/ or /a/, immediately following the target stop consonant.

In order to answer question (4) listed above, we compared and graphed the perception task responses per place of articulation, per task, and inspected each visualization in order to estimate the boundary, or 50% crossover point, and the boundary slope. The 50% crossover point was the step at which the mean of the group's responses to the stimuli for a given place of articulation, for a given task, were half voiced and half voiceless responses. The boundary slope was calculated as the slope at the crossover point, per place of articulation, per task. We utilized these graphs and data points to understand the location of the participants' categorical boundary before (perception task #1) and after (perception task #2) exposure to the model speaker's speech. This is discussed further in the analysis section (Section 2.2.6) that follows.

2.2.6. Analysis

We began our analysis by confirming that the AQ scores were validated by each participant's diagnosis. In referring to Table 7 below, we found that all of the participants with either an autism or ADHD diagnosis scored higher than the TD participants. The neurodiverse participants had AQ scores ranging from 24 to 43, while the TD participants' AQ scores ranged from two to 14. Thus, we were able to conclude that the participants' diagnoses, or lack of diagnosis, validated their AQ scores and we relied on AQ scores alone in the rest of our analysis. We also reported each participant's score on the attention switching and attention to detail subscales of the AQ test, as we will refer to these scores throughout our analysis.

Participant Code	**Diagnosis**	**AQ Score**	**Attention Switch Score**	**Attention to Detail Score**
P202	autism	43	10	9
P209	autism	39	9	10
P211	autism	36	10	7
P219	autism	32	7	9
P208	autism, ADHD	31	8	6
P210	ADHD	28	8	9
P204	autism	27	6	5
P214	ADHD	27	7	4
P203	autism	24	6	9
P218	TD	17	6	7
P206	TD	14	4	6
P220	TD	14	5	2
P205	TD	13	4	6
P217	TD	10	2	6

P215	TD	8	2	4
P212	TD	6	1	1
P221	TD	4	1	2
P213	TD	2	1	0

Table 7. Participant AQ scores.

2.2.6.1. Perception Analysis

The first step of our analysis consisted of graphing the participants' responses to the perception tasks to evaluate the point at which their perceptual boundary exists between the two options of the two-alternative forced-choice task. In order to create each of these graphs, we coded each participant's responses such that a voiced response was labeled as "1" and a voiceless response was labeled as "0." The participants heard one stimulus per step, or one stimulus per VOT value on the VOT spectrum presented per place of articulation, for each of the perception tasks. For participants that showed categorical perception, we expected to see their responses create an inverse S-shape pattern for each place of articulation. We graphed each individual's responses per place of articulation and task in order to visually inspect their response pattern and, thus, their categorical perception. However, in our group analyses, we only included participants that demonstrated the expected inverse S-curve with either a perfect inverse S-curve or a nearly perfect S-curve with at most one response straying away from the S-curve (i.e. they reported one "voiceless" response in the middle of several "voiced" responses, or vice versa), across both tasks. We assumed that participants with a nearly perfect S-curve and just one deviation simply made a mistake in their response to that token, while participants with multiple (more than one) deviations from an S-curve were likely guessing. With this more conservative inclusion criteria, our group analyses included the data of 13 participants for the bilabial analysis, nine participants for the dental analysis, and three participants for the velar analysis.

We graphed the group's responses per place of articulation and checked for both a task effect and an AQ score effect. In other words, we compared the group's VOT boundary values pre-exposure (perception task #1) and post-exposure (perception task #2), as well as across AQ scores. In order to create each of these graphs, we grouped responses by step and either task or AQ score before taking the mean; this allowed us to gather the mean percent of voiced responses per step, as well as per task or AQ score (Appendix G). For the group analysis, we calculated the estimated boundary[5], or 50% crossover point, as well as the slope[6] at this boundary to better understand the effect of task. For the individual task effect analysis, we also reported the participants' boundaries and boundary shift across tasks. These comparisons allowed us to evaluate whether or not the participants' ability to decipher the boundary between different voiced and voiceless stop consonants improved with exposure to the model speaker's speech. If the participants demonstrated improved perception, we expected to see the boundary shift to the left and the slope grow steeper. The AQ score analysis allowed us to evaluate whether or not AQ score affected perceptual abilities and boundaries on each task individually. Again, we reported the participants' boundaries and the distance of each of these boundaries from the expected VOT boundary at 0 ms. Nearly every participant had their own unique AQ score, but there were two participants with an AQ score of 14 and two with an AQ score of 27. We provide an analysis of each of the visualizations in Section 2.3. As shown in Appendix G, we used the "ggplot2" package in R to generate the visualizations, after summarizing the data using the "dplyr" package (v4.2.3; R Core Team, 2023; Wickham, 2016; Wickham et al., 2023).

[5] Example of boundary formula, for a boundary between steps 7 and 8: boundary = 7 + (% Voiced at Step 7 - 0.50)/(% Voiced at Step 7 - % Voiced at Step 8)

[6] Example of slope formula, for slope at boundary between steps 7 and 8: slope = (% Voiced at Step 8 - % Voiced at Step 7)/(Step 8 - Step 7)

The next step of our analysis involved running logistic regressions both across[7] the three places of articulation and per place of articulation: bilabial[8], dental[9], and velar[10] (Appendix G). We used the “lme4” package within R to run the logistic regression analyses (Bates et al., 2015; v4.2.3; R Core Team, 2023). The logistic regression analyses allowed us to evaluate the effect of AQ score, place of articulation, and exposure (i.e., task) on the participants’ voiced responses on both the first (pre-exposure) and second (post-exposure) perception tasks.

2.2.6.2. Production Analysis

In order to analyze the production data, we began by uploading the .wav files into Praat software (Boersma & Weenink, 2023). Using Praat, we labeled the target sounds, namely the stop consonants /p, t, k/ in both English and Spanish, directly on the textgrid (Figure 3). We relied on both the spectrogram and the oscillogram to determine the VOT onset and offset. Once we labeled each participant’s productions, we extracted the duration in seconds, or the VOT value, of each target stop consonant production using a Praat script adapted from a custom script developed by Christopher Carignan of the University of Illinois at Urbana-Champaign Department of French in 2009.

We utilized the test-retest approach to confirm reliability of our data analysis. The initial test and the retest were separated by exactly six weeks. We randomly selected 20 /p/-initial, 20

[7] Across POA model syntax: glm(formula = Voiced ~ AQ + POA + factor(Task) * Step, family = binomial(link = "logit"), data = new.total)

[8] Bilabial model syntax: glm(formula = Voiced ~ AQ + factor(Task) * Step, family = binomial(link = "logit"), data = data.bilabial)

[9] Dental model syntax: glm(formula = Voiced ~ AQ + factor(Task) * Step, family = binomial(link = "logit"), data = data.dental)

[10] Velar model syntax: glm(formula = Voiced ~ AQ + factor(Task) * Step, family = binomial(link = "logit"), data = data.velar))

/t/-initial/, and 20 /k/-initial words from our data, across all participants and both experiments, and repeated the VOT measure. We then used the cor.test() function in R to check the correlation between the original measures of VOT for the 60 randomly selected tokens to the repeated measures of VOT and found a strong correlation between the two measures ($r = 0.9896$, $p < 0.001$; v4.2.3; R Core Team, 2023). Thus, we concluded that our VOT values and data collection methods were both reliable and accurate.

We began by comparing our participants' performance on the English baseline task (production task #1) to their performance on the Spanish pre-test (production task #2) to confirm whether these values were similar or if their English and Spanish /p, t, k/ categories were distinct, despite being naive Spanish learners. By including the English task, we gained a better understanding of how each participant produced /p, t, k/-initial words in both languages. While the Spanish pre-test baseline was the reference level throughout the experiment, we included the English baseline task in order to confirm that participants used their L1 English VOT values to produce the novel L2 Spanish /p, t, k/ and did not alter their baseline VOT when producing the novel language, Spanish, in an effort to sound different or foreign. If our participants already had distinct English and Spanish VOT categories, we would have used their English baseline VOT values as their true baseline in understanding their ability to imitate a model speaker and their short-term learning skills. We also would have reevaluated our recruitment efforts in an attempt to find truly naive learners of Spanish.

We used linear mixed-effects models with the "lme4" and "lmerTest" packages within R to analyze the performance on the first two production tasks, using the English baseline task (production task #1) as the reference level for the Spanish pre-test (production task #2), to evaluate whether there was any difference between these two tasks (Bates et al., 2015;

Kuznetsova et al., 2017; v4.2.3; R Core Team, 2023). We have included the full code and output in Appendix H. We ran linear mixed-effects models to check for an effect of AQ score[11], the attention switching subscale[12], and the attention to detail subscale[13] on VOT in the English baseline task and Spanish pre-test, and we evaluated any effect of place of articulation across all three of these models (Appendix H).

Using the Spanish pre-test (production task #2) as our reference level, we repeated our linear mixed-effects models analysis using the "lme4" and "lmerTest" packages, with the addition of the "emmeans" package that allowed us to do pairwise comparisons[14] (Bates et al., 2015; Kuznetsova et al., 2017; Lenth, 2023; v4.2.3; R Core Team, 2023). In this second round of analysis, we compared our participant's Spanish immediate imitation VOT values (production task #3) and their Spanish immediate post-test VOT values (production task #4) to the Spanish pre-test VOT values (production task #2), as well as each other, across AQ score[15], their attention switching subscale score[16], their attention to detail subscale score[17], and place of articulation[18,19].

---

11 AQ model syntax: lmer(VOT ~ AQ + POA * Time + (1|Speaker) + (1|Word), data = data.pre)

12 Attention switching model syntax: lmer(VOT + POA ~ Sub_switch * Time + (1|Speaker) + (1|Word), data = data.pre

13 Attention to detail model syntax: lmer(VOT + POA ~ Sub_attention * Time + (1|Speaker) + (1|Word), data = data.pre

14 Pairwise comparison model syntax: emmeans(fit.sp, pairwise ~ Time)

15 AQ model syntax: lmer(VOT ~ AQ * Time + (1|Speaker) + (1|Word), data = data.sp)

16 Attention switching model syntax: lmer(VOT ~ Sub_switch * Time + (1|Speaker) + (1|Word), data = data.sp)

17 Attention to detail model syntax: lmer(VOT ~ Sub_attention * Time + (1|Speaker) + (1|Word), data = data.sp)

18 Addition of place of articulation to AQ model syntax: lmer(VOT ~ AQ + POA * Time + (1|Speaker) + (1|Word), data = data.sp)

19 Place of articulation pairwise comparison model syntax: emmeans(fit.sp, pairwise ~ POA)

As mentioned in Section 2.2.5 above, we utilized a stepwise approach in comparing VOT values, our dependent variable, across all of the tasks in order to answer the intended research questions. As part of the production analysis, we also compared each participant's VOT on the Spanish immediate imitation task (production task #3) to the native speaker's Spanish VOT values, as produced in the recording. We used Welch two sample t-tests for this analysis[20]. All of the code and its output is reported in Appendix H.

In addition to the group analyses outlined above, we will include a summary of the individual results in Section 2.3.2. We will elaborate on a few of the participant's perception and production task performance individually in Chapter 3, which utilizes more of a case study approach.

2.3. Results

As mentioned above, we have broken down the following analysis such that the group trends are addressed before taking a more personalized approach in analyzing the participants' performance individually. Within each of these sections, the perception results are presented first, with the production results following.

2.3.1. Group Results

The group results focus on a summarized review of the participants' performance across the perception and production tasks. We have included a report of the major trends observed as a group below.

[20] Welch two sample t-test model syntax: t.test(new.data.sp$VOT, new.data.sp$Native.VOT)

2.3.1.1. Group Perception Results

As discussed in Section 2.2.6.1., our perception analysis began with us coding and graphing each participant's responses to the perception stimuli (Appendix G, Appendix I). After excluding participants that did not show categorical perception, or categorical perception with just one mistake or guess, we checked for both a task effect and an AQ effect. For the participants whose data was excluded from the group analysis, most of their miscategorizations occurred in the lower VOT values, before the anticipated transition from voiced to voiceless Spanish consonants at 0 ms. In a few cases, the participants made more miscategorizations in the area where their English boundary likely exists, especially in response to the velar stimuli; it is possible these participants were not making "mistakes," rather combined their novel perceptual boundaries with their existing English-like boundary, resulting in VOT boundaries near the steps corresponding to both 0 ms and 30 ms. However, it appears as though the majority of the deviations that occurred in the later steps, at higher, more English-like VOT values, occurred during the second task (post-exposure), but, crucially, not during the first task (pre-exposure); if these miscategorizations were representative of their existing, English VOT boundary, we would expect to see these deviations before exposure to Spanish as well. Thus, there is not strong evidence to support these deviations represent a combination of Spanish and English VOT boundaries and are instead interpreted as mistakes.

As noted in the graphs below, 72.22% of participants (*n* = 13) demonstrated bilabial categorical perception, 50% of participants (*n* = 9) demonstrated dental categorical perception, and 16.67% of participants (*n* = 3) demonstrated velar categorical perception. Of the participants that demonstrated bilabial categorical perception, seven of the 13 participants were neurodiverse and the remaining six were TD. With respect to dental categorical perception, five of the nine participants were neurodiverse, while the remaining four participants were TD. Lastly, two of the three participants that demonstrated velar categorical perception were neurodiverse, while one was TD. Thus, there is a mix of both neurodiverse and TD participants that showed strong perceptual abilities across all three places of articulation. The neurodiverse and TD participant data included in the group analyses is nearly balanced; however, there is consistently one more neurodiverse participant's data included in each analysis, per place of articulation.

Beginning with the task effect analysis, we have included three graphs below, one per place of articulation (Figures 20, 21, 22). These graphs show the group mean percent of "voiced" (/b, d, g/) responses per step for each of the respective places of articulation, across the two perception tasks completed during Experiment 1: perception task #1 (pre-exposure) and perception task #2 (post-exposure). The data from both perception tasks are graphed together so that we can assess the effect of exposure on our participants' perceptual boundaries; as a reminder, the first and second perception tasks were separated by two Spanish production tasks, the Spanish pre-test (production task #2) and the Spanish immediate imitation task (production task #3), the latter of which included exposure to the model speaker's speech. Figure 23 is a graph showing the mean percent voiced responses per step for the bilabial range; to reiterate, each step corresponds to a VOT manipulation in 10 ms increments, with step one being -70 ms and step 13 being 60 ms. The color and style of each line corresponds to a task, either Task 1 (pre-exposure; purple dashed line) or Task 2 (post-exposure; green solid line). Similarly, Figure 24 is a graph showing the mean percent voiced responses per step for the dental range, with step one being -80 ms and step 15 being 70 ms. Finally, Figure 25 is a graph showing the mean percent voiced responses per step for the velar range, with step one being -60 ms and step 14 being 80 ms.

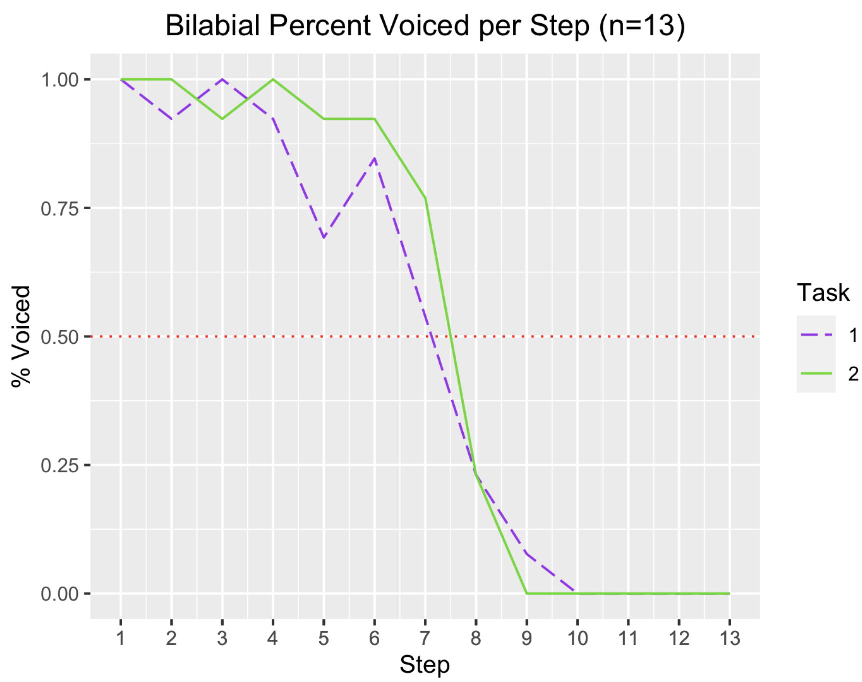


Figure 23. Percent bilabial voiced responses per step per task.

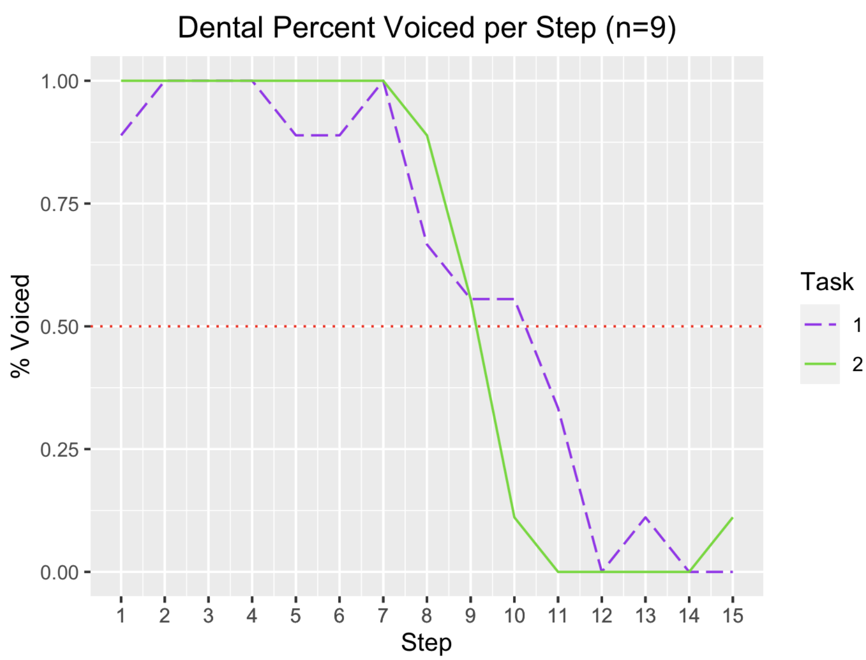


Figure 24. Percent dental voiced responses per step per task.

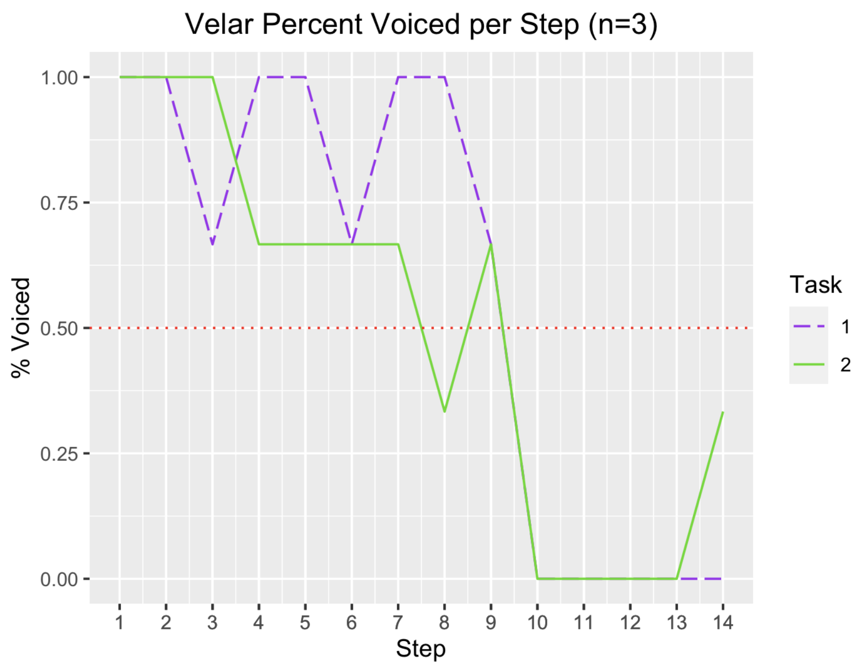


Figure 25. Percent velar voiced responses per step per task.

Based on the details of the perception task materials (Section 2.2.3.1.) and the number of steps per place of articulation, the point at which we expect to see the transition between voiced and voiceless responses is at 0 ms; in other words, we expect the boundary, or crossover point, between voiced and voiceless responses to occur between steps seven and eight for the bilabial stimuli, between steps eight and nine for the dental stimuli, and between steps six and seven for the velar stimuli. We also expect to see the steepest slope at this crossover point, in the transition between the negative and positive VOT values. We have included a table below with the estimated boundary, or crossover point, as well as the estimated slope at the boundary point, per task and place of articulation (Table 8).

PoA	Task	Boundary/50% crossover (step)	Slope (% voiced/step)
Bilabial	1	7.129	-0.310
	2	7.491	-0.550
Dental	1	10.217	-0.230
	2	9.114	-0.440
Velar	1	9.231	-0.650
	2	9.231	-0.650

Table 8. Boundary and slope estimates per task and place of articulation.

Based on the graphs and table above, it seems there was a clear pattern in how well our participants were able to categorize the perception stimuli based on the place of articulation presented. The categorization pattern followed such that the bilabial stimuli were the easiest to decipher, followed by the dental stimuli, and, lastly, the velar stimuli were the most difficult for our participants to distinguish and correctly categorize. This is evidenced by the number of participants included in each place of articulation analysis, as a result of their perfect or nearly perfect perceptual categorization abilities, the group's boundary/crossover point per task, and the steepness of the slope at this boundary.

The boundary on the bilabial graph is between steps seven and eight, as predicted. The lines for the first and second perception tasks are very similar for the bilabial stimuli, though the slope appears steeper by the second task. This may suggest that the participants were able to adjust their perceptual boundaries and categories through exposure to our model speaker's speech, resulting in more clear boundaries.

The boundary on the dental graph appears to be shifted to the right such that it is between steps 10 and 11 in the first task and between steps nine and 10 in the second task, instead of eight and nine as predicted. That being said, the boundary is shifted back towards the left by the second perception task. Again, this may suggest that the participants, as a group, learned to shift their perceptual boundary accordingly for the Spanish sounds they were exposed to during the immediate imitation task. Because English voiced and voiceless stop consonants have positive, greater VOT values than Spanish stop consonants, it is not surprising our participants started with a perceptual boundary that is further to the right, and shifted their boundary towards the left, in the direction of Spanish, upon hearing the language for the first time.

Lastly, the velar graph shows a rather unpredictable pattern, even after excluding the 15 participants that did not show categorical perception of the velar stimuli. The line for the first perception task appears to follow the inverse S-shaped pattern, though it is shifted several steps to the right, with the boundary between steps nine and 10, instead of between steps six and seven as predicted. Even more surprisingly, the line for the second perception task crosses the 50% crossover point three times and even turns upwards for the last two steps, indicating that participants reported hearing the stimuli as the voiced stop consonant for both the earlier and later steps, spanning negative and positive VOT values. It is possible the contrast between the velar stop consonants was more difficult for our participants to perceive. Although we checked for the naturalness of the stimuli beforehand, some participants still commented that some of the velar tokens sounded unnatural. At times, participants reported that the audio they heard did not seem to match either option presented in the two-alternative forced-choice task; however, we were not able to identify tokens that were consistently difficult for our participants.

As mentioned in Section 2.2.6.1, we ran logistic regression analyses to evaluate the effect of AQ score, place of articulation, and task on participants' categorization of the perception stimuli, across tasks, pre-exposure and post-exposure (Appendix G). The broad logistic regression analysis, across all three places of articulation, revealed a task effect, or an effect of exposure (β = 2.290, SE = 1.053, z = 2.174, $p < 0.05$). In other words, the number of voiced responses was significantly different across all AQ scores and places of articulation, when comparing the first and second perception tasks. This general logistic regression analysis also revealed an effect of place of articulation (dental ~ bilabial: β = 2.024, SE = 0.328, z = 6.171, $p < 0.0001$; velar ~ bilabial: β = 0.891, SE = 0.433, z = 2.058, $p < 0.05$), meaning that the number of voiced responses across participants was significantly different across these places of articulation. It is important to note that each place of articulation utilizes a unique number of steps to cover the full range of possible VOT values, such that we expect seven voiced and six voiceless responses for Spanish-like perception of the bilabial stimuli, eight voiced and seven voiceless responses for Spanish-like perception of the dental stimuli, and six voiced and eight voiceless responses for Spanish-like perception of the velar stimuli. Lastly, this broad analysis revealed no effect of AQ score (β = 0.002, SE = 0.011, z = 0.187, p = 0.852).

We then ran logistic regressions per place of articulation. The bilabial logistic regression analysis showed no effect of AQ score (β = 0.001, SE = 0.017, z = 0.085, p = 0.933) nor task (β = 3.969, SE = 2.165, z = 1.834, p = 0.067), meaning the number of voiced responses to the bilabial stimuli presented did not significantly change across either of these factors. The dental logistic regression analysis revealed no effect of AQ score (β = -0.007, SE = 0.016, z = -0.447, p =0.655), but a task effect for the dental stimuli (β = 5.938, SE = 2.759, z = 2.152, $p < 0.05$); in other words, the participants' voiced responses to the dental stimuli were significantly different before and after exposure to a native speaker's speech. This finding is evidenced by the leftward shift in boundary/crossover point shown in Figure 24. Lastly, the velar logistic regression analysis showed neither an effect of AQ score (β = 0.031, SE = 0.029, z = 1.098, p = 0.272) nor a task effect (β = -2.667, SE = 2.197, z = -1.214, p = 0.225). In conclusion, the bilabial and velar analyses did not show a significant difference in the number of voiced responses across tasks nor AQ score, while the dental analysis showed a task effect, but not an AQ effect. Thus, AQ score did not affect perception according to any of our analyses. As noted in Appendix G, there was a significant effect of step across all of our analyses; this is expected and, therefore, uninformative since each step corresponds to a different VOT value, in increasing order, across all of the tasks and places of articulation.

Continuing our analysis, we then tested for an AQ effect. In other words, we graphed the data such that all of the AQ scores are represented individually so that we could evaluate the effect of AQ on our participants' perceptual boundaries. We have included six graphs below, two per place of articulation for the two different perception tasks, pre-exposure and post-exposure. Figures 26 and 27 show the percent voiced responses for the bilabial tokens for tasks one and two, respectively. Figures 28 and 29 show the percent voiced responses for the dental tokens for tasks one and two, respectively. Lastly, Figures 30 and 31 show the percent voiced responses for the velar tokens for tasks one and two, respectively. As indicated within each graph, each colored line represents a different AQ score, with lower AQ scores on the red side of the rainbow and higher AQ scores on the violet side of the rainbow. Again, these graphs only include the responses from the participants that exemplified categorical perception with a single mistake allowance.

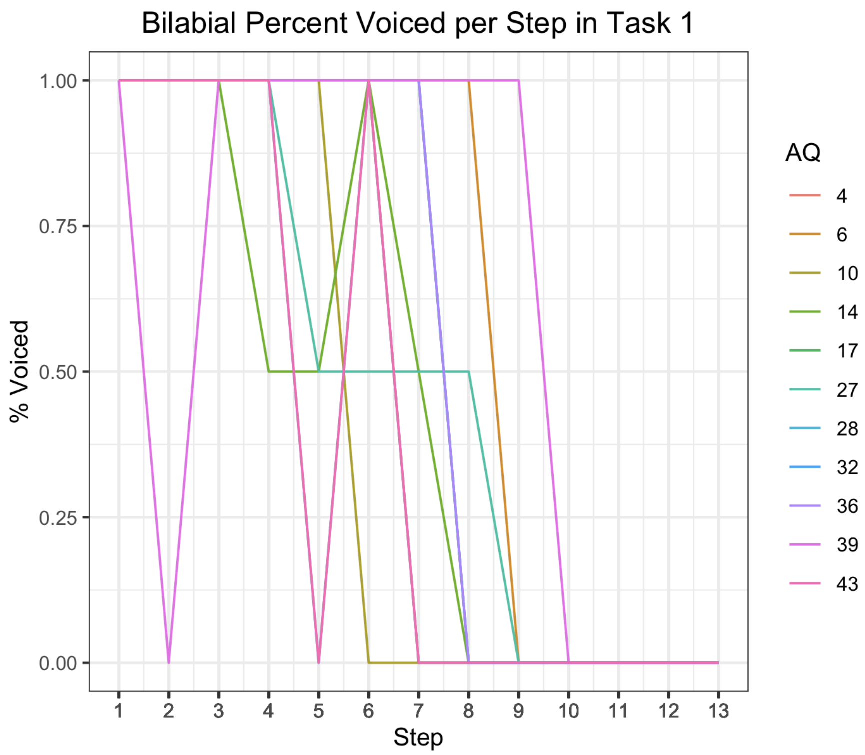


Figure 26. Percent bilabial voiced responses per step in perception task #1.

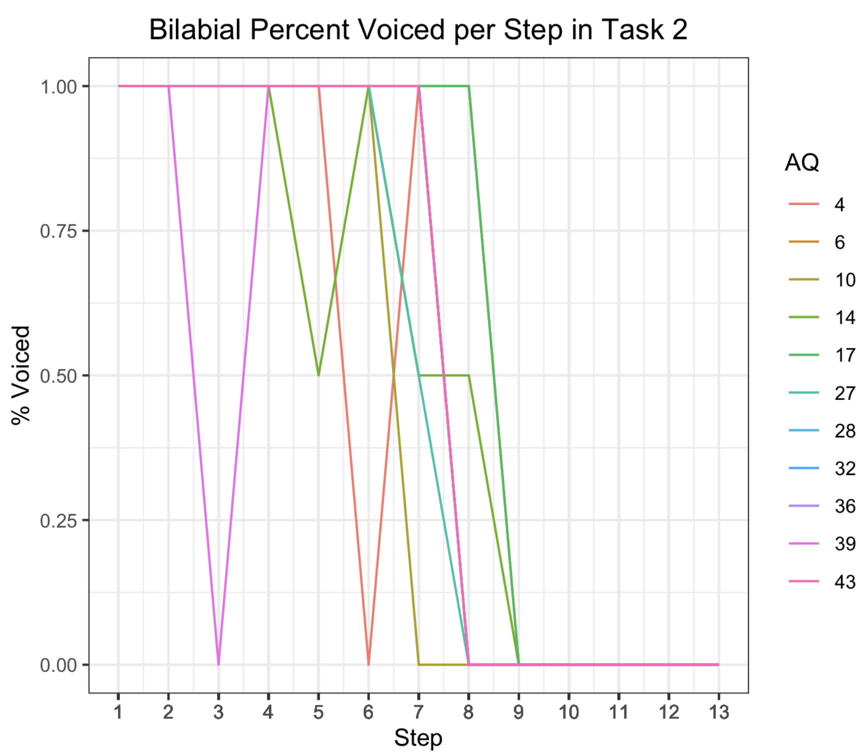


Figure 27. Percent bilabial voiced responses per step in perception task #2.

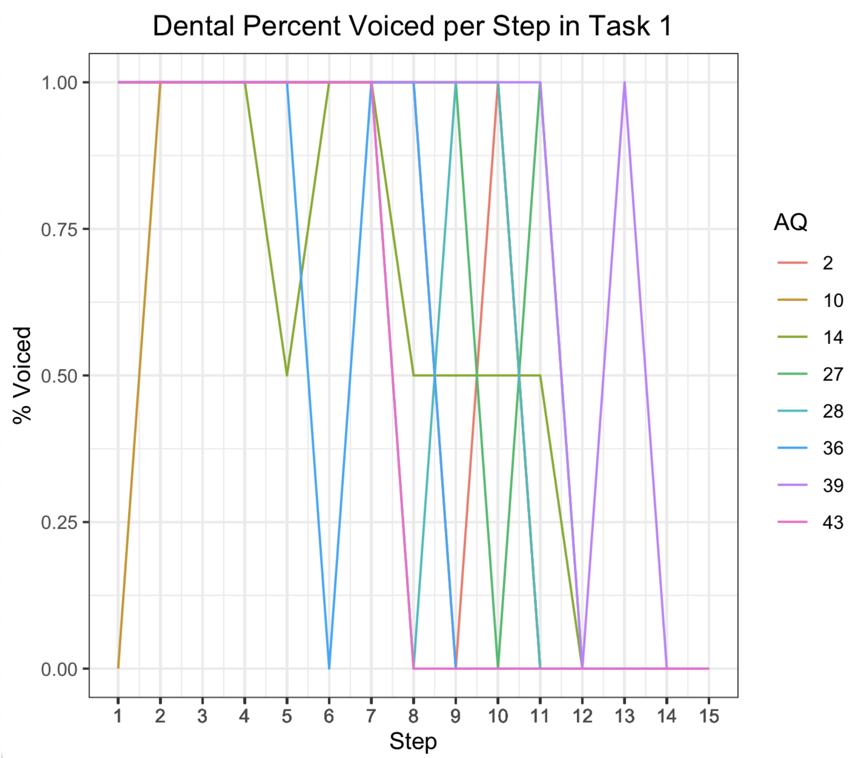


Figure 28. Percent dental voiced responses per step in perception task #1.

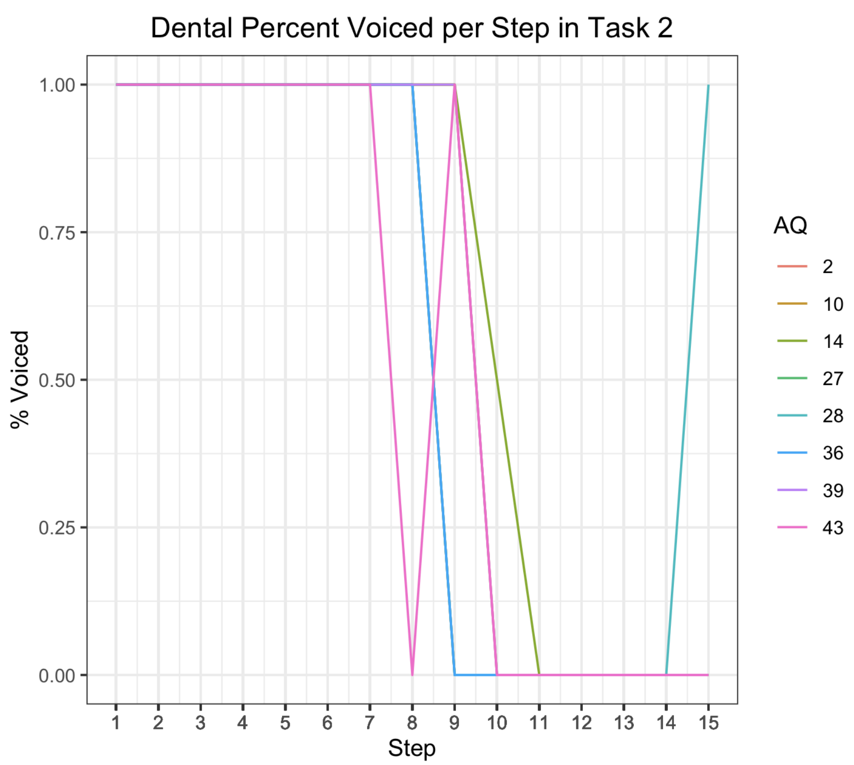


Figure 29. Percent dental voiced responses per step in perception task #2.

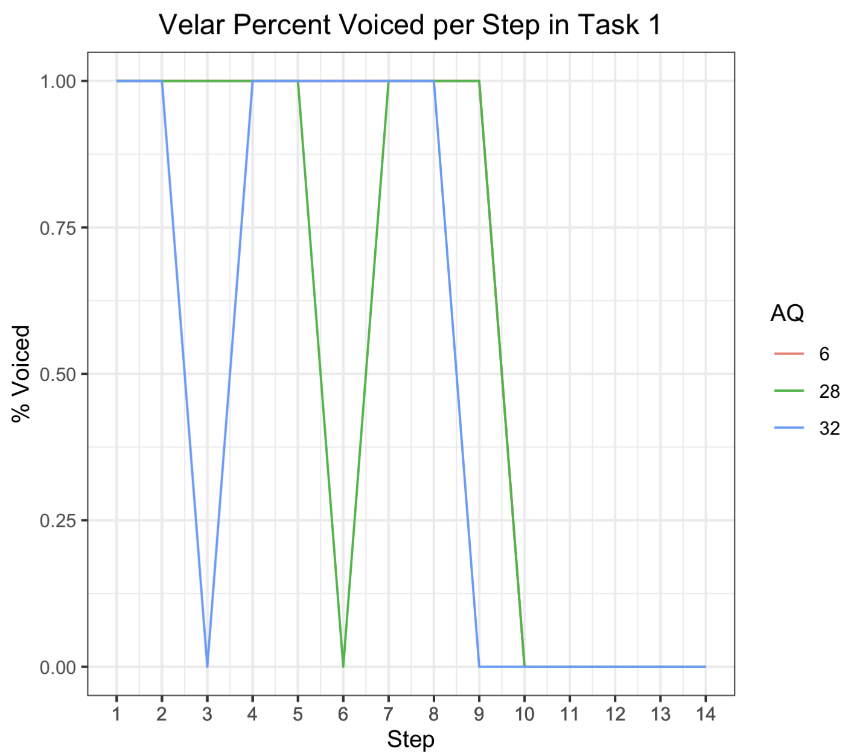


Figure 30. Percent velar voiced responses per step in perception task #1.

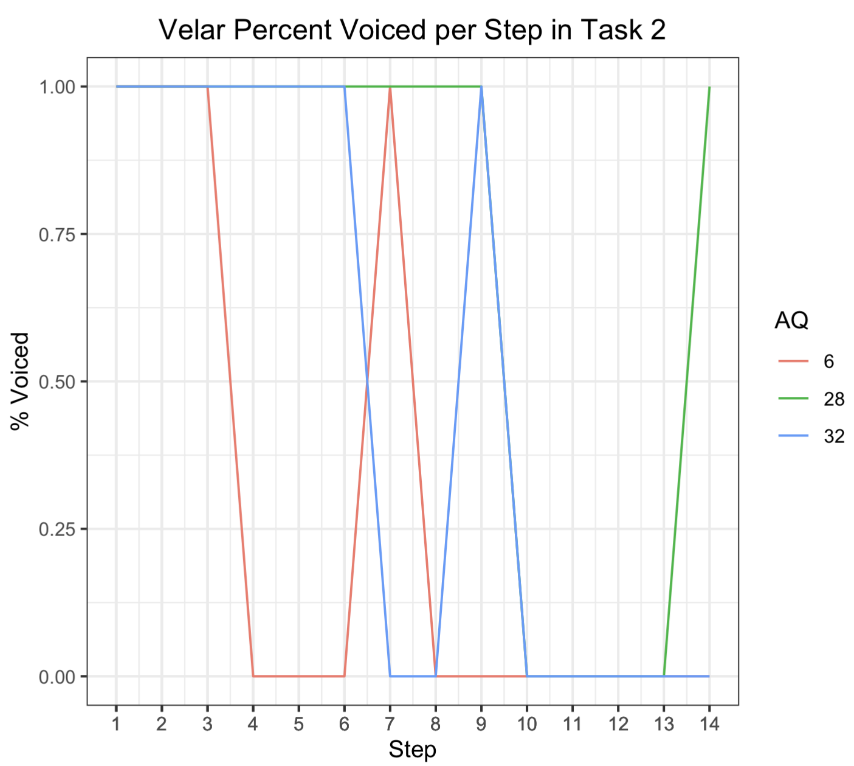


Figure 31. Percent velar voiced responses per step in perception task #2.

We have summarized the findings from the graphs in Table 9 below, which includes the VOT boundaries across the two tasks, per place of articulation and per AQ score, as well as the distance from the expected VOT boundary at 0 ms. The negative values in the distance column indicate a VOT boundary that is below the Spanish-like boundary at 0 ms, while positive values indicate a VOT boundary that is more English-like, in the direction of the 30 ms boundary. Because there were two participants with AQ scores of 14 and 27, the boundary presented for each of these scores in the table below is an average of the participants' individual boundaries.

AQ	PoA	Task	Boundary/50% crossover (step)	Distance from 0 ms (steps)
43	**Bilabial**	1	6.5	-1.0
		2	7.5	0
39	**Bilabial**	1	9.5	+2.0
		2	7.5	0
36	**Bilabial**	1	7.5	0
		2	7.5	0
32	**Bilabial**	1	7.5	0
		2	7.5	0
28	**Bilabial**	1	6.5	-1.0
		2	7.5	0
27	**Bilabial**	1	6.5	-1.0
		2	7.0	-0.5
17	**Bilabial**	1	7.5	0
		2	8.5	+1.0
14	**Bilabial**	1	5.5	-2.0
		2	6.5	-1.0
10	**Bilabial**	1	5.5	-2.0
		2	6.5	-1.0
6	**Bilabial**	1	8.5	+1.0
		2	8.5	+1.0
4	**Bilabial**	1	6.5	-1.0
		2	5.5	-2.0
43	**Dental**	1	7.5	-1.0
		2	7.5	-1.0
39	**Dental**	1	11.5	+3.0

		2	9.5	+1.0
36	**Dental**	1	8.5	0
		2	8.5	0
28	**Dental**	1	10.5	+2.0
		2	8.5	0
27	**Dental**	1	9.5	+1.0
		2	8.5	0
14	**Dental**	1	9.5	+1.0
		2	10.0	+1.5
10	**Dental**	1	10.5	+2.0
		2	9.5	+1.0
2	**Dental**	1	8.5	0
		2	8.5	0
32	**Velar**	1	8.5	+2.0
		2	6.5	0
28	**Velar**	1	9.5	+3.0
		2	9.5	+3.0
6	**Velar**	1	9.5	+3.0
		2	3.5	-3.0

Table 9. VOT boundaries by AQ score, per place of articulation and task.

We will present a more individualized approach to interpreting the perception data per AQ score in Section 2.3.2, but we have summarized the group trends here. In comparing the bilabial perception task #1 (Figure 26) and perception task #2 (Figure 27) graphs, we see a general consolidation to the left as a group (n = 13). Table 9 validates this finding, as evidenced by the reduced distance from 0 ms and the increased accuracy of the VOT boundaries falling at exactly 0 ms. In the first task, the participant with the second highest AQ score reported voiced responses until the transition between steps nine and 10, when they switched to reporting voiceless responses (Figure 26). That being said, we do not see a consistent trend of higher AQ-scoring individuals showing a more English-like perceptual boundary than lower AQ-scoring individuals, since the next right-most slope belongs to the second lowest scoring participant. By the second task, all of the participants reported hearing voiceless stops by step nine, making the transition from voiced to voiceless between steps eight and nine (Figure 27). This is only slightly shifted compared to the expected steepest slope between steps seven and eight for the bilabial stop consonants. Thus, as a group, our participants seemed to learn to shift their perceptual boundary in the direction of the lower VOT values that Spanish favors. In both the first and second perception task graphs for the bilabial stop consonants, we see several deviations from the expected inverse S-shape pattern, despite excluding participants that were clearly guessing. These deviations do not seem to be predictable per end of the AQ score spectrum; in other words, participants with a wide range of high and low AQ scores appear to make mistakes in their responses in an uncorrelated manner.

The dental graphs include the results from 50% of the participants (n = 9). Similar to the bilabial graphs, we see the same leftward consolidation from perception task #1 to perception task #2 for the dental graphs (Figures 28 and 29). Table 9 shows the distance from 0 ms decreased from the first to second tasks for four of the AQ scores. While there are both voiced and voiceless dental responses reported through step 13 in the first task, all but one of the participants reported only voiceless responses as of step 11 and beyond by the second task. It is also evident there are fewer deviations from the expected inverse S-curve in the second task, meaning fewer participants guessed the opposite stop consonant as they heard the continuum of VOT values by task two. Additionally, the graph for perception task #2 shows fewer lines, meaning there was overlap between several of the participants. Again, we did not notice any particularly clear color patterns in either graph such that we cannot conclude if there is any specific effect of AQ on the observed results. We hope to be able to provide more details in the individual analysis that follows.

Lastly, the velar graphs show more complicated patterns compared to the bilabial and dental graphs, and include only 16.67% of the participants (n = 3). In the first task, we see some inconsistent categorizations up until step nine, but there are two clear slopes that all three of the participants shared in transitioning from reporting voiced to voiceless responses (Figure 30). These two steepest slopes fall between steps eight and nine, and nine and 10. None of the participants reported hearing a voiced stop consonant for the last four steps. However, the perception task #2 graph shows the confusion that the stimuli for the velar stop consonants of task two presented (Figure 31). In this graph, we see several deviations from the inverse S-curve, or spikes in the opposite direction, across a wide range of steps. As a group, there is a lack of consistency in reporting voiced or voiceless responses, even after excluding participants that seemed to be guessing. It appears AQ score does not add any level of clarity in predicting our participants' responses, as there is not a visible color trend for task one nor task two. Table 9 shows that the highest AQ-scoring participant achieved accurate velar categorization by the second task, while the other two participants were consistently three steps away from the 0 ms boundary.

To summarize the perception task results presented here, we found that, as a group, the participants showed that they had a more clear categorization of the VOT boundary, as indicated by a steeper slope and leftward boundary shift, in the second perception task for the bilabial and dental stimuli. The group showed a general trend towards shifting their perceptual boundary to be more Spanish-like, with lower voiceless VOT values; the logistic regression analysis revealed that this task effect was significant, specifically for the dental stimuli. While we did see these changes across tasks, we did not notice any particular trend of AQ affecting the predictability of the steeper slope or shifted boundary. Both neurodiverse and TD participants showed strong categorical perception skills across all three places of articulation, with almost equal numbers of neurodiverse and TD participants' data included in each of the group analyses. These results suggest that, regardless of AQ score, or diagnosis, our participants' L2 perceptual abilities improved after brief exposure to the model speaker's speech. These results also suggest that individuals with higher AQ scores do not have enhanced perceptual abilities compared to their lower-scoring peers; rather, perceptual abilities are comparable across AQ scores.

2.3.1.2. Group Production Results

As mentioned in Section 2.2.6.2, our analysis began by comparing the English baseline task (production task #1) and Spanish pre-test (production task #2) VOT values, across a variety of variables, via linear mixed-effects models. There was not a significant difference between our participants' VOT values on these two tasks when evaluated as a function of AQ score (β = -0.065, SE = 0.613, t = -0.105, p = 0.917) nor either subscale score (attention switching: β = 0.022, SE = 2.531, t = 0.009, p = 0.993; attention to detail: β = 1.111, SE = 3.324, t = 0.334, p = 0.743). Our analysis also revealed that there was no effect of place of articulation on the resulting English baseline task (production task #1) and Spanish pre-test (production task #2) VOT values (/p/ ~ /k/: β = -8.980, SE = 13.102, t = -0.685, p = 0.502; /t/ ~ /k/: β = -14.586, SE = 13.102, t = -1.113, p = 0.280). Through this analysis, we were able to confirm that there was no significant difference between our participants' VOT values when producing English /p, t, k/-initial words and producing Spanish /p, t, k/-initial words in the baseline tasks. Thus, we concluded that children use their native English VOT when producing novel Spanish words. We utilized the Spanish pre-test (production task #2) as our baseline for the remainder of the analysis.

We performed further linear mixed-effects models to compare VOT values on the Spanish immediate imitation task (production task #3) and the Spanish immediate post-test (production task #4) to the Spanish pre-test (production task #2), as well as each other. We found that AQ score (β = -0.118, SE = 0.622, t = -0.189, p = 0.852) and both subscale scores (attention switching: β = -0.124, SE = 2.566, t = -0.048, p = 0.962; attention to detail: β = 0.700, SE = 2.617, t = 0.267, p = 0.792) did not have a significant effect on VOT. Interestingly, there was a main effect of task (production task #3 ~ production task #2: β = 24.99, SE = 2.87, t = 8.714, p < 0.0001; production task #4 ~ production task #2: β = 18.25, SE = 2.87, t = 6.363, p < 0.0001; production task #3 ~ production task #4: β = -6.75, SE = 2.87, t = -2.352, p < 0.05). These results indicate that Spanish imitation and Spanish immediate post-test VOTs were significantly shorter than those of the Spanish pre-test, following exposure to a native speaker (Appendix H). Additionally, there was a significant difference between the Spanish immediate imitation and the Spanish immediate post-test VOT values. These results suggest a trend towards reducing VOT appropriately, upon exposure to the model speaker's speech, and the effect of imitation that lasted until the immediate post-test; however, this VOT reduction was not maintained to the same degree in the immediate post-test compared to the immediate imitation task, as the VOT values across these two tasks were significantly different, with an increase in VOT in the immediate post-test. Because there was no effect of AQ score nor subscale scores on VOT, the autistic children were just as good as their TD peers at imitating the novel sounds of a second language and maintaining this knowledge, at least in the short-term.

We have included a box plot below that shows a visualization of the main effect of task, or time, as it is labeled on the x-axis (Figure 32). As indicated below, the English baseline task ("english," production task #1) and Spanish pre-test ("spanish1," production task #2) VOT values are not significantly different. Using the Spanish pre-test for comparison sake, it is evident the Spanish immediate imitation ("spanish2," production task #3) and Spanish immediate post-test ("spanish3," production task #4) VOT values are both significantly lower than the average VOT value for the Spanish pre-test. Furthermore, the Spanish immediate imitation and immediate post-test VOT values are significantly different from one another, with higher VOT values in the immediate post-test.

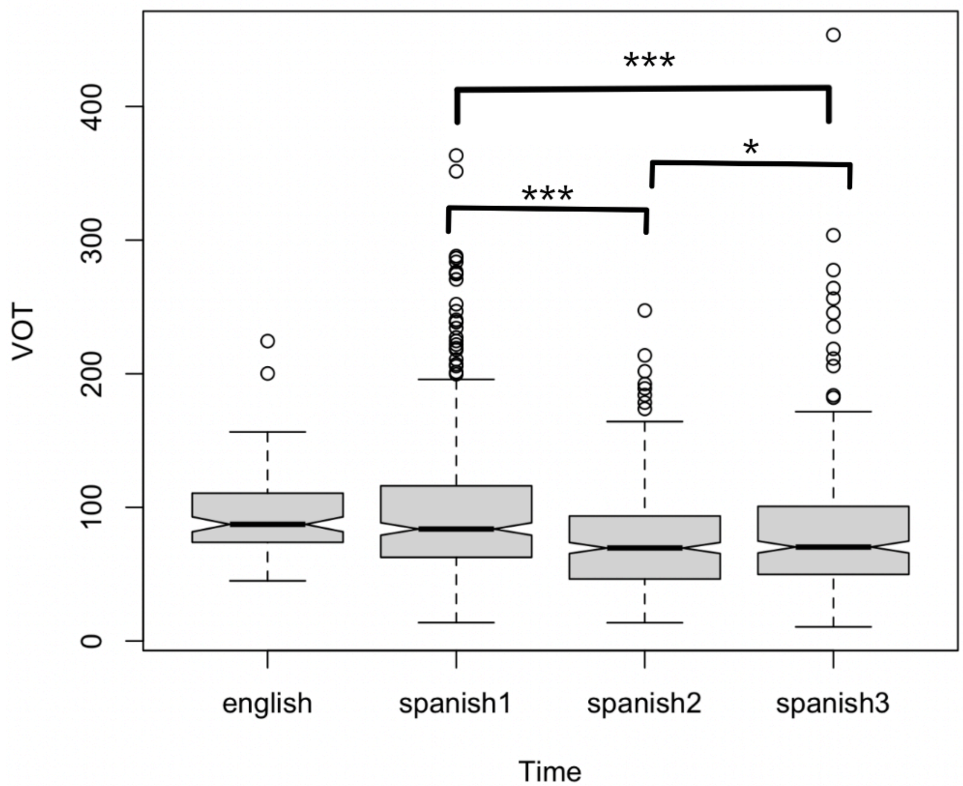


Figure 32. A box plot showing the distribution of VOT values across each of the four production tasks.

To support the box plot visualization above, we have also included the group mean VOT values presented numerically in the table below (Table 10). The table further shows the sharp drop in VOT value from the English baseline and Spanish pre-tests, which show similarly high values, to the Spanish immediate imitation and immediate post-tests, which show the significantly decreased VOT values. Corroborating the results presented in the box plot, the table below shows the most dramatic VOT drop between the Spanish pre-test and the Spanish immediate imitation task, and this lowered VOT value persists through the Spanish immediate post-test, though increases slightly compared to the Spanish immediate imitation task.

Task	Place of Articulation	PoA Mean VOT (ms)	Task Mean VOT (ms)
English baseline	/p/	91.109	92.234
English baseline	/t/	85.503	
English baseline	/k/	100.089	
Spanish pre-test	/p/	93.128	99.026
Spanish pre-test	/t/	94.697	
Spanish pre-test	/k/	109.254	
Spanish immediate imitation	/p/	63.178	74.033
Spanish immediate imitation	/t/	76.028	
Spanish immediate imitation	/k/	82.892	
Spanish immediate post-test	/p/	70.802	80.778
Spanish immediate post-test	/t/	79.562	
Spanish immediate post-test	/k/	91.969	

Table 10. Group mean VOT values across tasks.

Continuing our analysis, we added place of articulation as an independent variable to our linear mixed-effects model and conducted a pairwise analysis between the places of articulation. These analyses revealed a significant difference in VOT value between /p/ and /k/ (β = 19.05, *SE*

= 6.88, t = 2.768, p = 0.0361), but not between /p/ and /t/ (β = -7.85, SE = 6.88, t = -1.140, p = 0.0505) nor /t/ and /k/ (β = 11.20, SE = 6.88, t = 1.628, p = 0.265). This finding is not surprising given that /p/ tends to have the shortest VOT value, followed by /t/, and, lastly, /k/ tends to have the longest VOT value; this pattern follows for both Spanish and English voiceless stop consonants (Lisker & Abramson, 1964; Zampini, 2013).

Lastly, we compared our participant's VOT values on the Spanish immediate imitation task (production task #3) to the native speaker's VOT values when producing these stop consonants in the recording that the participants were imitating. The result of the t-tests showed a significant difference between our participants' VOT and the native speaker's VOT across all three places of articulation (/p/: t = 13.431, p < 0, /t/: t = 14.042, p < 0, /k/: t = 11.011, p < 0). Thus, although our participants significantly reduced their VOT values across tasks, from the Spanish pre-test (production task #2) to the Spanish immediate imitation task (production task #3), and maintained a reduced VOT value through the Spanish immediate post-test (production task #4), their VOT values were still significantly different from those of the native speaker they were imitating. This may suggest that, while our participants were capable of perceiving and reducing their VOT accordingly, they were not able to do this to a native level of accuracy.

2.3.2. Individual Results

Given the heterogeneity of the autism diagnosis and the frequency of co-occurring diagnoses, an individualized approach is very fitting for studies involving autistic children, or even children with a wide range of AQ scores.

2.3.2.1. Individual Perception Results

For each of the 18 participants, we generated three graphs visualizing their perceptual boundary across the two perception tasks, with each graph representing a different place of articulation (Appendix I). Similar to the group perception graphs, each line is color-coded to match one of the two tasks (pre-exposure perception task #1: purple line; post-exposure perception task #2; green line). All of the graphs show the percent of voiced (/b, d, g/) responses reported per participant, per step, per task. Again, we expect to see an inverse S-shape for participants that heard voiced stop consonants for the lower, negative VOT values and heard voiceless stop consonants for the higher, positive VOT values. For clarity sake, we will reiterate the meaning of each of the steps, per place of articulation: for the bilabial stop consonants, there were 13 steps ranging from -70 ms to 60 ms, with each step separated by 10 ms; for the dental stop consonants, there were 15 steps, ranging from -80 ms to 70 ms; and for the velar stop consonants, there were 14 steps, ranging from -60 ms to 80 ms. Thus, we expect to see the transition between voiced and voiceless responses to occur at step 7.5 for the bilabial stimuli, at step 8.5 for the dental stimuli, and at step 6.5 for the velar stimuli.

As with the group analysis, we have generated the following table to provide a summary of each participant's boundary per task and boundary shift from perception task #1 to perception task #2, per place of articulation (Table 11); Table 11 is similar to Table 9 in Section 2.3.1.1, with the addition of the participant code, as well as the boundary shift per participant as opposed to the distance of each boundary from the expected boundary at 0 ms. In this more individualized analysis, the boundary crossover point is more exact, since the boundary always occurs exactly between a 100% voiced response and 0% voiced response; in other words, the crossover point occurs between exactly two tokens. Thus, the slope for all of these crossover points is exactly

-1.0, or -100% voiced/step. Instead of reporting the same slope across participants, we have included the boundary shift between perception task #1 and perception task #2. We expect to see a negative boundary shift if participants' categorical boundary shifted to the left, to be more Spanish-like, following exposure to the native speaker's speech. On the contrary, a positive boundary shift across tasks would indicate a shift to the right, in the direction of a more English-like categorical boundary. Again, we excluded participants that showed more than one mistake, deviating from the anticipated inverse S-curve we would expect if categorical perception occurred. In determining which boundary to use, in the case of an inverse S-curve with one deviation, we used the boundary that was the closest to that participant's boundary on the other perception task.

Participant	AQ	PoA	Task	Boundary/50% crossover (step)	Boundary Shift (steps)
P202	**43**	**Bilabial**	1	6.5	+1.0
			2	7.5	
P209	**39**	**Bilabial**	1	9.5	-2.0
			2	7.5	
P211	**36**	**Bilabial**	1	7.5	0
			2	7.5	
P219	**32**	**Bilabial**	1	7.5	0
			2	7.5	
P210	**28**	**Bilabial**	1	6.5	+1.0
			2	7.5	
P204	**27**	**Bilabial**	1	4.5	+2.0
			2	6.5	
P214	**27**	**Bilabial**	1	8.5	-1.0
			2	7.5	
P218	**17**	**Bilabial**	1	7.5	+1.0
			2	8.5	
P206	**14**	**Bilabial**	1	3.5	+1.0
			2	4.5	
P220	**14**	**Bilabial**	1	7.5	+1.0
			2	8.5	

P217	**10**	**Bilabial**	1	5.5	+1.0
			2	6.5	
P212	**6**	**Bilabial**	1	8.5	0
			2	8.5	
P221	**4**	**Bilabial**	1	6.5	-1.0
			2	5.5	
P202	**43**	**Dental**	1	7.5	0
			2	7.5	
P209	**39**	**Dental**	1	11.5	-2.0
			2	9.5	
P211	**36**	**Dental**	1	8.5	0
			2	8.5	
P210	**28**	**Dental**	1	10.5	-2.0
			2	8.5	
P214	**27**	**Dental**	1	9.5	-1.0
			2	8.5	
P206	**14**	**Dental**	1	7.5	+2.0
			2	9.5	
P220	**14**	**Dental**	1	11.5	-1.0
			2	10.5	
P217	**10**	**Dental**	1	10.5	-1.0
			2	9.5	
P213	**2**	**Dental**	1	8.5	0
			2	8.5	
P219	**32**	**Velar**	1	8.5	-2.0
			2	6.5	
P210	**28**	**Velar**	1	9.5	0
			2	9.5	
P212	**6**	**Velar**	1	9.5	-6.0
			2	3.5	

Table 11. Summary of estimated boundaries and slopes.

Of the 13 participants whose bilabial data was included in the analysis, three participants showed a leftward more Spanish-like boundary shift post-exposure, seven showed a boundary shift to the right in the English-like direction, and three participants showed no change across tasks. Of the three participants whose boundary shifted to the left, two were neurodiverse and one was TD. Furthermore, four of the 13 participants (neurodiverse: n = 2, TD: n = 2) had a bilabial categorical boundary at the anticipated step 7.5 in the first task and two of those four maintained this boundary through the second task; interestingly, the two participants that maintained this boundary were the neurodiverse participants. By perception task #2, six of the 13 participants' bilabial categorical boundaries occurred at step 7.5; all six of these participants were neurodiverse. Although seven participants showed a shift to the right, in the English-like direction, by perception task #2, five of these participants shifted appropriately closer to step 7.5.

Of the nine participants whose dental data was included in the analysis, five participants' boundaries shifted to the left by the second task (neurodiverse: n = 3, TD: n = 2), one participant's boundary shifted to the right by the second task, and three participants' boundaries did not change across tasks. Compared to the bilabial task (23.08%), a higher percentage (55.56%) of the participants whose data was included in the dental categorical perception analysis showed the appropriate leftward shift. This may suggest that the dental contrast was easier for participants to (1) perceive in the native speaker's stimuli and (2) identify in the perception stimuli presented. We expected the boundary to occur at step 8.5 for the dental stimuli; two of the participants demonstrated a dental categorical boundary at 8.5 in perception task #1 and they both maintained this boundary in perception task #2. Overall, four of the nine participants' dental categorical boundaries occurred at the anticipated step 8.5 by the second perception task (neurodivergent: n = 3, TD: n = 1).

Lastly, there were only three participants' data included in our analysis of the velar categorical boundary; two of these participants showed a leftward, more Spanish-like boundary shift from perception task #1 to perception task #2, and one participant showed no change in their boundary across tasks. We expected the transition from voiced to voiceless velar stops to occur at step 6.5. The participant with the highest AQ score in the velar analysis, P219 (AQ score of 32), was the only participant that appropriately shifted their boundary to the left to step 6.5 by perception task #2. Interestingly, the other participant that showed a negative, leftward shift showed the greatest shift out of any of the analyses across places of articulation; P212 showed a boundary shift of -6.0 steps from perception task #1 to perception task #2.

In this more individualized analysis of the perception data, there seems to be a potential neurodivergent advantage in perceptual abilities. While participants across the full range of AQ scores showed both positive and negative changes in boundary, as well as varying accuracy in boundary/crossover point, six out of the seven (85.71%) participants that showed the anticipated categorical boundary by perception task #2 were neurodiverse. If we break this down by place of articulation, all six participants that ultimately achieved accurate Spanish-like bilabial categorization were neurodiverse, three of the four participants that achieved accurate Spanish-like dental categorization were neurodiverse, and the one participant that achieved accurate Spanish-like velar categorization was neurodiverse. It seems as though the same order of ease of comprehension follows for the ease of achieving accurate Spanish-like categorization by perception task #2. Across all three places of articulation, the participants that were included in each analysis varied greatly in their AQ scores; in other words, the clean data did not result in the exclusion of participants with consistently high nor consistently low AQ scores. Thus, the results from this individual analysis both strengthen and challenge our previous conclusion that autistic-like traits are not predictive of perceptual abilities.

2.3.2.2. Individual Production Results

In this section, we will present a table summarizing each participant's performance across all four of the production tasks presented in Experiment 1. We will share each participant's mean VOT values, both per place of articulation and conflated across places of articulation, per task.

We have included the group means across tasks at the bottom of Table 12 below, to reiterate the values previously presented in Table 10 (Section 2.3.1.2). We have also included the group averages for the neurodiverse and TD participants below the overall group averages.

Participant	AQ	PoA	English baseline VOT (ms)	Spanish pre-test VOT (ms)	Spanish immediate imitation VOT (ms)	Spanish post-test VOT (ms)
P202	**43**	**/p/**	78.839	51.242	37.518	47.795
		/t/	112.243	76.160	25.892	28.850
		/k/	82.556	107.707	77.232	58.810
		Mean	91.213	78.370	46.880	45.152
P209	**39**	**/p/**	84.222	163.276	47.163	102.907
		/t/	91.069	128.590	69.387	91.344
		/k/	128.815	142.669	110.223	130.228
		Mean	101.369	144.845	75.591	108.160
P211	**36**	**/p/**	95.203	54.362	59.697	64.754
		/t/	114.452	88.145	60.759	70.892
		/k/	98.614	105.416	96.587	117.375
		Mean	102.756	82.641	72.348	84.341
P219	**32**	**/p/**	83.348	50.364	59.086	32.693
		/t/	80.418	74.963	70.612	52.068
		/k/	73.916	79.098	65.348	45.139
		Mean	79.227	68.141	65.015	43.300
P208	**31**	**/p/**	86.738	206.162	83.727	113.745
		/t/	52.455	146.269	116.886	185.517
		/k/	131.422	183.821	131.385	149.378
		Mean	90.205	178.751	110.666	149.547
P210	**28**	**/p/**	68.833	66.117	73.739	67.057
		/t/	78.331	82.001	87.569	91.514
		/k/	96.473	88.416	90.613	94.120
		Mean	81.212	78.845	83.974	84.231
P204	**27**	**/p/**	76.656	52.143	63.798	48.124
		/t/	55.960	58.648	71.693	56.329
		/k/	82.895	89.224	82.478	86.338
		Mean	71.837	66.672	72.656	63.597

P214	**27**	**/p/**	82.058	54.372	53.563	63.179
		/t/	68.079	77.659	57.758	75.572
		/k/	64.153	95.635	58.381	93.569
		Mean	71.430	75.889	56.567	77.440
P203	**24**	**/p/**	53.321	67.889	67.076	49.492
		/t/	66.971	76.564	79.705	50.701
		/k/	96.671	83.185	98.675	61.942
		Mean	72.321	75.879	81.818	54.045
P218	**17**	**/p/**	124.899	62.930	67.432	73.526
		/t/	88.920	94.896	67.450	93.299
		/k/	94.450	78.254	61.034	71.758
		Mean	102.756	78.693	65.305	79.528
P206	**14**	**/p/**	122.638	85.659	54.490	58.667
		/t/	105.703	79.694	70.151	70.863
		/k/	121.721	107.428	93.783	88.747
		Mean	116.687	90.927	72.808	72.759
P220	**14**	**/p/**	122.784	81.646	81.141	69.673
		/t/	89.019	80.605	70.278	66.971
		/k/	107.549	80.442	52.733	65.451
		Mean	106.450	80.898	68.051	67.365
P205	**13**	**/p/**	77.954	202.085	115.341	176.973
		/t/	71.114	191.078	154.647	146.278
		/k/	95.189	199.719	132.631	219.314
		Mean	81.419	197.628	134.206	180.855
P217	**10**	**/p/**	93.708	75.577	45.110	57.681
		/t/	99.279	77.227	46.975	54.234
		/k/	102.218	72.224	49.408	57.181
		Mean	98.401	75.009	47.164	56.365
P215	**8**	**/p/**	152.090	151.645	52.736	70.648
		/t/	113.528	136.962	64.086	102.393
		/k/	95.607	183.716	52.060	74.979
		Mean	120.408	157.441	56.294	82.673
P212	**6**	**/p/**	96.355	53.501	58.282	31.456
		/t/	80.919	55.023	58.524	51.536
		/k/	75.831	48.352	72.463	57.028
		Mean	84.368	52.292	63.090	46.673

P221	**4**	**/p/**	70.659	79.741	57.157	80.841
		/t/	103.874	100.076	89.037	84.075
		/k/	133.510	97.420	84.060	82.265
		Mean	102.681	92.412	76.751	82.393
P213	**2**	**/p/**	69.657	117.584	60.145	65.220
		/t/	66.718	79.986	107.091	59.676
		/k/	120.008	123.840	82.970	101.827
		Mean	85.461	107.137	83.402	75.574
Group Averages		**/p/**	91.109	93.128	63.178	70.802
		/t/	85.503	94.697	76.028	79.562
		/k/	100.089	109.254	82.892	91.969
		Mean	92.234	99.026	74.033	80.778
Neurodiverse Group Average		**/p/**	78.802	85.103	60.596	65.527
		/t/	79.998	89.889	71.140	78.088
		/k/	95.057	108.353	90.102	92.989
		Mean	84.619	94.448	73.946	78.868
TD Group Average		**/p/**	103.416	101.152	65.759	76.076
		/t/	91.008	99.505	80.915	81.036
		/k/	105.120	110.155	75.682	90.950
		Mean	99.848	103.604	74.119	82.687

Table 12. Summary of individual production task results.

With this summary of the individual production results readily accessible in Table 12, we will briefly discuss the production results per participant, in order of descending AQ score, as this was the primary focus of Experiment 1. We have also made note of any production errors per participant in the detailed results that follows. Unless otherwise specified, "group average" and "group mean VOT" refers to the combined group averages, as opposed to the average for either the neurodiverse or TD group.

Compared to the group mean VOT values per task, it is evident P202 produced the Spanish stimuli with much lower, more accurate, VOT values (Table 12). While her baseline English VOT values were very comparable to the group mean VOT value for the English task, at just over one ms shorter, her Spanish VOT values across tasks ranged from being 20.656 ms to

35.626 ms shorter than the group averages (Table 12). This may suggest that P202 had a superior ability to pick up on the fine-grained phonetic details of Spanish and was able to both imitate and learn the sounds to a more accurate degree. Given that P202 had the highest AQ score of the group, with very high subscale scores as well, her results alone may support our initial hypothesis: individuals with greater autistic-like traits may be superior imitators and learners of L2 phonetic details. While we did not exclude any of P202's productions, she nearly elided the /p/ in the /po/ token in the pre-test.

P209's mean VOT values were greater than the group mean VOT values across all four tasks (Table 12). With respect to the English baseline task, P209's mean VOT value was only 9.135 ms greater than the group mean VOT value on this task. For the Spanish pre-test, P209's mean VOT value was 45.819 ms greater than the group's average VOT value on this task. This may suggest that P209 was more hesitant, or slower, in reading the Spanish phrases initially, which we noted in our observation notes for the session. Interestingly, for the Spanish immediate imitation task, P209's mean VOT value for the task was just barely above the group's mean VOT value at just 1.558 ms greater than the group average. This further supports the theory that P209 was simply hesitating when reading the Spanish phrases initially, but was able to repeat the phrases successfully after the model speaker and accommodate the shortened VOT value necessary. This trend persists in the Spanish post-test. While P209's mean VOT value jumped back up to being 27.382 ms greater than the group mean VOT value on this task, he did drastically shorten his VOT when producing Spanish sounds, independent of the support of the model speaker's audio; in other words, we see a 36.685 ms decrease in P209's VOT values between the Spanish pre-test and Spanish post-test (Table 12). This may suggest that P209 learned, at least in the short-term, to shorten his VOT when producing Spanish sounds. As mentioned, P209 was hesitating and sounding out the Spanish words throughout the session and he swapped vowels/consonants in seven of his total productions.

As indicated in the table, P211's mean VOT value was greater than the group average on both the English baseline task and the Spanish post-test (Table 12). In the case of the English baseline task, P211's mean VOT value was just over 10 ms greater than the group mean VOT value on this task. P211's mean VOT value on the Spanish post-test was just 3.563 ms greater than the group's mean VOT value on this task. For the Spanish pre-test and immediate imitation tasks, P211's mean VOT value was lower than the group averages. Compared to his English baseline VOT, P211 decreased his VOT when producing Spanish by more than 20 ms. We may interpret this to mean that, even before hearing Spanish via the model speaker, P211 changed the way he was speaking to sound different or "foreign." While P211's mean VOT value on the Spanish immediate imitation task was 16.385 ms lower than the group mean VOT value on this task, his mean VOT value on the immediate imitation task was just 1.685 ms lower than the group mean VOT value. Because P211's mean VOT value increased again in the Spanish post-test, after the Spanish immediate imitation task, to a value that was greater than his Spanish pre-test mean VOT value, it seems as though P211 did not learn to shorten his VOT when producing Spanish sounds, beyond the initial shortening that he did between the English baseline and Spanish pre-tests.

P219's mean VOT values were consistently lower than the group mean VOT values across all four production tasks. Although P219's mean VOT values on the Spanish immediate imitation and post-tests were lower than the group averages, it is important to point out that P219's mean VOT value on the Spanish pre-test was already 11.086 ms lower than his mean VOT value on the English baseline task. P219 only shortened his mean VOT value on the

Spanish immediate imitation task by 3.126 ms upon hearing the model speaker stimuli, but he did still shorten his VOT in the direction of sounding more native-like. By the Spanish post-test, P219's mean VOT value had decreased by 24.841 ms from his mean VOT value on the Spanish pre-test. This may indicate P219's strong ability to recognize the fine-grained details of a novel L2 and learn to adapt his own speech, to sound more native-like, at least in the short-term. His mean VOT value on the Spanish post-test was an impressive 37.478 ms shorter than the group average on this task (Table 12).

P208's mean VOT value was slightly lower than the group's mean VOT value on the English baseline task, but their mean VOT values were much higher than the group's averages on all three of the Spanish production tasks (Table 12). There is a large jump in P208's mean VOT value when comparing the first English task to the first Spanish task, which may imply some hesitation in producing the Spanish words and phrases. P208 then decreased their mean VOT value by 68.085 ms when imitating the model Spanish speaker during the immediate imitation task, compared to their mean VOT value when reading Spanish for the first time independently. It does appear as though P208 learned to shorten VOT when producing Spanish words, at least in the short-term, as their mean VOT in the Spanish post-test was 29.204 ms shorter than their Spanish pre-test mean VOT value. Although P208 had a higher mean VOT value on all of the Spanish tasks compared to the group mean VOT values, ranging from 36.633 ms to 79.25 ms higher, they did show that they learned to decrease their VOT value when producing Spanish compared to the first Spanish read-aloud task (Table 12). In the Spanish pre-test, P208 said "daza" instead of "taza" for one of the stimuli, though we still included this production in their mean VOT value for /t/.

As evidenced by Table 12, P210's mean VOT values were slightly lower than the group mean VOT values on the first two production tasks, the English baseline and Spanish pre-tests, but her mean VOT values were slightly higher than the group averages on the second two production tasks, the Spanish immediate imitation and post-tests. In fact, it seems as though P210's mean VOT values did not change much over the course of the experiment. Her mean VOT values, across all four tasks, range from 78.845 ms to 84.231 ms. Thus, the greatest difference between her mean VOT values is just 5.386 ms and that difference occurred between the Spanish pre-test and post-tests. Because her mean VOT value actually increased gradually across the three Spanish tasks, it does not seem as though P210 noticed a need to shorten her VOT values, or accommodate her VOT when producing Spanish sounds, beyond her already low English baseline and Spanish pre-test mean VOT values (Table 12).

P204's mean VOT values are lower than the group mean VOT values across all four tasks (Table 12). That being said, there is a greater difference between his VOT values and the group's VOT values on all three of the tasks other than the immediate imitation task, which is also the task with the lowest group mean VOT value. It seems P204 already started with a very low VOT value when producing the English stimuli, and this pattern persisted when he produced the Spanish stimuli. The only Spanish task with a slightly higher VOT value than his English baseline VOT value was the Spanish immediate imitation task; however, P204's mean VOT value on this task was less than one ms greater than his mean English baseline VOT value. By the Spanish immediate post-test, P204 produced the stimuli with the lowest mean VOT values and, thus, shifted his production of the Spanish words accordingly (Table 12).

As evidenced by the table above, P214's mean VOT values across all four production tasks are much lower than the group averages per task (Table 12). His mean VOT value on the

English baseline and Spanish pre-tests is 20.804 ms and 23.137 ms lower than the group mean VOT values on these tasks, respectively. It seems like P214 did in fact use his native English VOT in producing the Spanish phrases, given that his mean VOT values on the English baseline and Spanish pre-tests only differ by 4.459 ms. The most drastic drop in P214's mean VOT value occurs between the Spanish pre-test and immediate imitation tasks, where we see his mean VOT value drop by 19.322 ms. Compared to the group mean VOT value, P214's mean VOT value on the immediate imitation task is 17.466 ms lower. This may suggest that P214 had a strong ability to perceive and imitate the reduced VOT value he heard produced by the model Spanish speaker in the immediate imitation task. P214's mean VOT value increased in the Spanish post-test, compared to both the Spanish pre-test and immediate imitation tasks; however, his mean VOT value on this task was still lower than the group average by 3.338 ms (Table 12).

P203 shows much shorter VOT values across three of the four production tasks, including the English baseline task (Table 12). Interestingly, P203's VOT values are shorter than the group mean VOT values on all of the tasks except for the Spanish immediate imitation task, when he repeated the phrases immediately after hearing the model speaker's audio. Compared to the baseline English and Spanish pre-tests, P203's VOT values increased during the Spanish immediate imitation task, but decreased during the Spanish immediate post-test. It is possible P203 was speaking at a high speech rate, but was forced to slow down a bit during the imitation task since he could not speak at his own rate, rather he had to wait for each of the stimuli to play before repeating the phrase. This theory is further supported by the drastic decrease in VOT value by the last production task (Table 12).

While P218's mean VOT value is greater than the group average on the English baseline task, her mean VOT values are lower than the group averages across all three of the Spanish production tasks. Although her mean VOT value on the English baseline task is more than 10 ms above the group mean VOT value on this task, her mean VOT value on the first Spanish task, the pre-test, is more than 20 ms below the group average. This may suggest that P218 was attempting to sound different or "foreign" when reading the Spanish phrases. Between the Spanish pre-test and the immediate imitation task, P218's mean VOT value dropped by 13.388 ms. It seems as though, upon hearing the model speaker's stimuli, P218 dropped her VOT even further in an attempt to sound more native-like. By the Spanish post-test, P218's mean VOT value increased again such that it was just 1.25 ms below the group mean VOT value for this task and 0.835 ms greater than her initial Spanish mean VOT value on the Spanish pre-test.

P206's English baseline mean VOT value is the only value that is greater than the group mean VOT value (Table 12). For all three of the Spanish production tasks, P206's mean VOT value sits just below the group's mean VOT value. In addition, there is a gradual decrease in P206's Spanish VOT values as the tasks proceed. The largest transition between P206's VOT values occurs between the English baseline task and the Spanish pre-test. This may suggest that, even without hearing the model speaker's speech, P206 shortened her VOT values when producing Spanish words in an attempt to make the words sound foreign, or different than English. Interestingly, P206's mean VOT values for the Spanish immediate imitation and immediate post-tests are nearly identical, which may show P206's strength in remembering to shorten her VOT values when producing Spanish even more than she did in the first Spanish task (Table 12). P206 added an additional consonant to one of her /p/ productions in the Spanish post-test, and she also swapped the vowels of two /p/ tokens, which were labeled with the corresponding vowel (i.e. /poto/ was labeled as a /po/ token and /pako/ was labeled as a /pa/ token, even though the target words provided were actually "pato" and "poco").

As indicated in Table 12, P220's mean VOT value was greater than the group mean VOT value on the English baseline task, but lower than the group mean VOT values on all three of the Spanish production tasks. Compared to the group, P220's mean VOT value on the English baseline task was 14.216 ms higher. When she switched to the first Spanish production task, P220's mean VOT value already dropped by 25.552 ms, resulting in a value that was 18.128 ms lower than the group mean VOT value on this task. Since none of our participants had any prior exposure to Spanish, including P220, this drop in VOT from the English baseline to Spanish pre-tests may be indicative of an attempt to sound "foreign." Upon hearing the model speaker's speech in the immediate imitation task, P220's mean VOT value dropped further, specifically by 12.847 ms from the Spanish pre-test to Spanish immediate imitation task. This effect lasted through the Spanish post-test, as P220's mean VOT value on the Spanish post-test remained low and was almost the same as her mean VOT value on the Spanish immediate imitation task.

Per Table 12, it is evident P205 has drastically longer VOT values across all three of the Spanish production tasks compared to her peers. With respect to the English baseline production task, P205's mean VOT value was more than 10 ms shorter than the group's mean VOT value on the same task. Thus, it does not seem like P205's speech rate affected her VOT values across the board, but it is possible her speech rate did slow down when sounding out the novel sounds of Spanish, which we noted observing in our notes for the session. This theory is further supported by P205's Spanish pre-test and post-test mean VOT values, which are both much higher than her Spanish immediate imitation mean VOT value. Across the three Spanish production tasks, P205's mean VOT value ranged from being 60.173 ms to 100.077 ms greater than the group's mean VOT values on each of the respective tasks (Table 12). Because of the drastic shift in P205's mean VOT value from the English task to the Spanish tasks, specifically the first Spanish task, we may interpret this to mean that at least part of P205's extended VOT values were the result of Spanish being such a novel foreign language to her. As mentioned, we noted that P205 was sounding out most of the words, which resulted in a slow, hesitant speech rate and separation of the carrier phrase from the token. We also noticed that P205 substituted consonants or added consonants erroneously four times across the tasks presented in the first experiment.

P217 had much shorter mean VOT values compared to the group averages for all three of the Spanish production tasks, but a longer mean VOT value compared to the group on the English baseline task. It appears as though P217 already started with a low mean VOT value when producing Spanish, as his VOT dropped by 23.392 ms when he switched from English to Spanish stimuli. Even so, P217 shortened his mean VOT value further in the Spanish immediate imitation task, indicating he may have noticed the model speaker's shortened VOT values. Compared to the group mean VOT value on the immediate imitation task, P217's mean VOT value was 26.869 ms shorter, reduced in the direction of sounding more native-like. While his mean VOT value increased between the Spanish immediate imitation and post-tests, it still remained 18.644 ms shorter than his mean VOT value on the Spanish pre-test. Thus, it seems as though P217 recognized the need to shorten his VOT in order to produce more native-like Spanish sounds, which he did in both the imitation task, as well as the post-test. This shows P217's ability to learn to shorten his VOT value, at least in the short-term.

P215's mean VOT values were much higher than the group mean VOT values on the English baseline and Spanish pre-tests. Compared to the group averages, P215's mean VOT value was 28.174 ms greater on the English baseline task and a shocking 58.415 ms greater on the Spanish pre-test. Given that both values were much higher than the group averages, we do not believe P215 was simply hesitating when reading Spanish, due to its novelty. Her mean VOT value was still higher than the group average on the Spanish post-test, but less drastically so; in this case, her mean VOT value was only 1.895 ms greater than the group average. Interestingly, her mean VOT value on the Spanish immediate imitation task was much lower than the group mean VOT value on this task, perhaps showcasing her superior imitation skills. Compared to the group mean VOT value on the Spanish immediate imitation task, P215's mean VOT value was 17.739 ms lower, a stark contrast from her performance compared to the group on the three other tasks. It is important to note the huge drop in P215's mean VOT value from the Spanish pre-test to both the Spanish immediate imitation and post-tests. P215's VOT dropped by 101.147 ms in the immediate imitation task, upon hearing the model speaker's speech. P215's mean VOT value increased again once she was reading the Spanish phrases independent of the model speaker stimuli, in the post-test, but her mean VOT value still dropped by 74.768 ms in this task compared to her mean VOT value on the Spanish pre-test. Not only did P215 do well imitating the model speaker and adjusting her VOT to sound more native-like, she also showed that she was able to learn from this exposure to the model speaker's speech, as the effect lasted through the post-test. P215 produced the incorrect vowel sound with two of the /k/ tokens across the first experiment.

P212's mean VOT values were drastically lower than the group mean VOT values across the board (Table 12). While P212's mean VOT value on the English baseline task was just 7.866 ms below the group mean VOT value on this task, his mean VOT value dropped 32.076 ms between the first English and Spanish production tasks, such that his mean VOT value on the Spanish pre-test was 46.734 ms lower than the group average on this task. This may suggest that P212 was attempting to "sound foreign" even before hearing the model Spanish speaker's speech. Regardless, we know there was not a significant difference between the group's English baseline and Spanish pre-test VOT values. During the Spanish immediate imitation task, P212's mean VOT value increased by just over 10 ms compared to the Spanish pre-test, but his mean VOT value on this task still remained just over 10 ms shorter than the group's average. By the Spanish post-test, P212 learned to shorten his VOT even further to be more native-like, and his mean VOT value was 34.105 ms shorter than the group mean VOT value on this task (Table 12).

P221's mean VOT values were greater than the group averages across three out of the four production tasks, with the outlier being the Spanish pre-test. Although P221's mean VOT value was greater than the group mean VOT value on the English baseline task, by just over 10 ms, her mean VOT value on the Spanish pre-test dropped by just over 10 ms compared to her English baseline mean VOT value. Because the difference between her English baseline and Spanish pre-test is not extreme, it is unclear whether this may imply P221 was attempting to sound "foreign" or if this was a regular fluctuation in her baseline, native VOT. Compared to the Spanish pre-test, P221's mean VOT value on the Spanish immediate imitation task dropped by 15.661 ms, putting it just 2.717 ms above the group mean VOT. This may suggest that she was sensitive to the need to shorten her VOT values further in order to sound more like the model speaker she was imitating. Although there was a slight increase in her mean VOT value between the Spanish immediate imitation to post-tests, her mean VOT value on the post-test was still lower than her initial Spanish pre-test mean VOT value. More specifically, her Spanish post-test mean VOT value was 10.019 ms shorter than her Spanish pre-test mean VOT value. This may suggest that P221 was successful in learning to shorten her VOT when producing Spanish, at least in the short-term.

According to Table 12, P213's mean VOT values were higher than the group mean VOT values on the Spanish pre-test and Spanish immediate imitation tasks, but they were lower than the group on the English baseline and Spanish post-tests. Given the 21.676 ms increase in P213's VOT value between the English baseline and Spanish pre-test, it is possible he was slightly more hesitant when reading the novel Spanish phrases initially. By the Spanish immediate imitation task, P213 already decreased his VOT value by 23.735 ms compared to his mean VOT value in the Spanish pre-test. Although his mean VOT value for the Spanish immediate imitation task was still greater than group average by just under 10 ms, P213 then learned to shift his VOT even lower by the Spanish post-test such that his mean VOT value on this task was more than 5 ms lower than the group average on this task. Across the three Spanish production tasks, P213's mean VOT value dropped by 31.563 ms from the first to third task. On the other hand, the group mean VOT value only dropped by 18.248 ms from the first to third Spanish production tasks. Thus, despite P213's high mean VOT values on the Spanish pre-test and immediate imitation tasks, he gradually shortened his VOT across these tasks and showed that he learned to cut back on VOT by the last Spanish task, when he produced more native-like VOT than the group, on average.

In summary, this more in-depth analysis of the participant's individual production task results revealed the same trend we noticed in the group analysis (Section 2.3.1.2): regardless of AQ score, most participants shorten their VOT values after exposure to a native speaker's speech. As Table 12 shows, both participants with high AQ scores and participants with low AQ scores had VOT values below and above the group averages, across tasks. The neurodiverse group averages included in Table 12 are consistently lower than the TD group averages, as evidenced by the mean VOT values per task. Of the neurodiverse participants, all of whom had AQ scores above 24, 44.44% (*n* = 4) produced average VOT values per task that were consistently shorter than the group average VOT values for each respective task. None of the neurodiverse participants produced VOT values below the group average VOT values for only the Spanish tasks, unlike the TD participants. 44.44% (n = 4) of the TD participants produced VOT values below the group average VOT values for the Spanish tasks only. Six participants (neurodiverse: n = 1, TD: n = 5) produced VOT values that were above the group average on the English baseline task, but produced VOT values that were lower than the group average on the Spanish pre-test; perhaps these participants were all attempting to alter their speech in some way to accommodate the foreign language they were reading. It seems the neurodiverse participants were less likely to make this same effort to alter their speech, rather they showed more consistently lower VOT values across tasks, including the English baseline task. Only one of the TD participants showed average VOT values that were lower than the group's average VOT values across all four tasks. Interestingly, P215 was the only participant that produced VOT values that were shorter than the group's average VOT values on just the Spanish tasks post-exposure to a native speaker's speech. That being said, per the group analysis in Section 2.3.1.2, we know that, as a group, our participants reduced their VOT values significantly following exposure to the model speaker in the Spanish immediate imitation task, and this reduction was maintained into the Spanish immediate post-test. Another trend that we noticed was that some participants reduced their speech rate and seemed to hesitate when reading the Spanish sounds, which resulted in higher average VOT values. Again, this hesitation was not predictable by AQ score.

2.4. Discussion

The overarching goal of Experiment 1 was to answer the following three of our four main research questions (Section 1.3):

1. Do autistic-like traits predict L2 phonetic imitation ability?

2. Do autistic-like traits predict short-term L2 phonetic learning?
4. If short-term learning is predicted by autistic-like traits, is it due to enhanced perceptual abilities?

We will summarize Chapter 2 by answering these three questions. We will address each question from the perspective of both the group results and the individualized results to provide a more robust view.

In order to assess whether or not autistic-like traits predict imitation ability, we looked at how our participants' VOT values shifted upon exposure to the model speaker's speech. We analyzed this VOT shift with respect to our participants' overall AQ scores, as well as their scores on the attention switching and attention to detail subscales of the AQ test. Based on the group production results presented in Section 2.3.1.2, we found that there was no significant difference between our participants' VOT values on the English baseline and Spanish pre-test. Thus, our participants did not have separate VOT boundaries for the two languages, which confirmed that they were truly naive Spanish learners. Utilizing the Spanish pre-test as our baseline, we then found that there was a significant difference between the Spanish pre-test and Spanish immediate imitation task VOT values. Although this task effect was significant, there was no effect of AQ score, attention switching subscale score, nor attention to detail subscale score, as a group. We also confirmed that our participants' VOT values on the Spanish immediate imitation task were significantly different from the native speaker's VOT values when producing the same tokens; thus, although our participants were capable of significantly reducing their VOT value post-exposure to the native speaker, this reduction was not to a native level accuracy.

Taking a deeper dive into the individual data, we found a similar pattern in which the majority of participants decreased their VOT when producing Spanish during the immediate imitation task (Section 2.3.2.2). That being said, P204, P210, and P212 all showed a slight increase in VOT between the Spanish pre-test and Spanish immediate imitation tasks. These three participants have AQ scores of 27, 28, and six, respectively. Their attention switching subscale scores are six, eight, and one, respectively, while their attention to detail subscale scores are five, nine, and one, respectively. This subgroup of the participants did not perceive the immediate need to shorten their VOT when imitating the model speaker's speech, but P204 and P212 did eventually decrease their VOT when producing Spanish by the post-test. Because these three participants have varied AQ scores and subscale scores, both in the neurodivergent and TD ranges, there does not appear to be a pattern regarding the number of autistic-like traits one possesses and their ability to imitate novel L2 sounds.

Next, we will look at the participants that showed the greatest shift in VOT from the Spanish pre-test to the immediate imitation task, as these are the participants that potentially noticed the model speaker's reduced VOT more immediately and were able to implement the appropriate reduction in their own speech. Two participants we would like to highlight here are P202 and P213, as these participants showed the largest decrease in VOT from the Spanish pre-test to the immediate imitation tasks, without starting with a drastically high Spanish pre-test VOT value. P202's mean VOT value dropped by 31.49 ms between the baseline and imitation tasks, while P213's mean VOT value dropped by 23.735 ms between these two tasks. P202 is autistic with an AQ score of 43, while P213 is TD with an AQ score of two. P202 has an attention switching subscale score of 10 and an attention to detail subscale score of nine, while P213 has an attention switching subscale score of one and an attention to detail subscale score of zero. Thus, it does not appear AQ score, nor AQ subscale scores, predict one's imitation ability in this case since the participants that made the greatest shift in VOT upon hearing the model speaker's speech were both our highest and lowest AQ-scoring participants.

It is important to point out that several participants showed a more drastic decrease in VOT between the Spanish pre-test and immediate imitation tasks, but they started with exceedingly high mean VOT values in the Spanish pre-test. Thus, these participants had to shift their VOT values quite a bit to approach imitating the model speaker, but their VOT values remained exceptionally high. To be more specific, from the Spanish pre-test and immediate imitation tasks, P205 had a VOT shift of 63.422 ms, P208 had a VOT shift of 68.085 ms, P209 had a VOT shift of 69.254 ms, and P215 had a VOT shift of 101.147 ms. All four of these participants' mean VOT values increased again by the Spanish post-test. While P215 and P209's shifted mean VOT values were closer to the acceptable range of native Spanish VOT values, P205 and P208 did not approach this acceptable range. It is also interesting to note that all four of these participants were on the younger side of our recruitment age range; three of the four participants were seven years old and one was eight years old at the time of the study. It is possible the reason these four participants had such high VOT values in the Spanish pre-test is because of their young age and, thus, less experience reading in general, much less in Spanish. This could also explain their drastic decrease in VOT when imitating the model speaker; they were much stronger imitators than they were readers.

In short, the answer to the first research question above is no - autistic-like traits do not predict imitation ability. As a group, our participants showed a significant decrease in VOT value upon exposure to the model speaker's speech in the imitation task. However, there was no effect of AQ score, nor AQ subscale score, on this decrease in VOT value on either the group or individual level. In other words, there was neither a statistically significant finding that autistic-like traits can predict imitation ability, nor a qualitative pattern in the in-depth review that would suggest the ability of autistic-like traits to predict imitation ability.

In assessing whether or not autistic-like traits predict short-term learning, we will first refer back to the group production results presented in Section 2.3.1.2. As stated previously, there was a significant difference between the Spanish pre-test, used as the reference level, and the Spanish immediate imitation task VOT values, as a group. There was also a significant difference between the Spanish immediate imitation task and Spanish immediate post-test VOT values, with higher VOT values in the immediate post-test compared to the immediate imitation task. It is not surprising that there was a slight increase in VOT value between the immediate imitation task and the immediate post-test, since the participants did not have the native speaker stimuli to imitate in the immediate post-test. Because there was a significant difference between the Spanish pre-test and Spanish post-test VOT values, we conclude that there was a significant task effect that persisted through the immediate post-test, which indicates the group was able to learn to reduce their VOT values to be more Spanish-like, in the short-term. Similar to the results discussed above, there was no effect of AQ score nor AQ subscale score on short-term L2 phonetic learning, as measured by the difference between the Spanish pre-test and post-test VOT values.

Again, we will look into the individual results to expand on the group results (Section 2.3.2.2). We would like to highlight participants that showed the largest decrease in VOT value from the Spanish pre-test to the Spanish post-test. From the pre-test to the post-test, P215's mean VOT decreased by 74.768 ms, P209's mean VOT decreased by 36.685 ms, P202's mean VOT decreased by 33.218 ms, P213's mean VOT decreased 31.563, P208's mean VOT decreased by 29.204 ms, and P219's VOT decreased by 24.841 ms. The overall AQ scores of these participants, in the same order as the participants were just listed, are as follows: eight, 39, 43, two, 31, and 32. Similarly, their attention switching subscale scores are two, nine, 10, one, eight, and seven, respectively. Lastly, their attention to detail subscale scores are four, 10, nine, zero, six, and nine, respectively. While P215, P209, and P208 also showed large VOT reductions, they started with extremely high VOT values in the Spanish pre-test and did not necessarily have to approach a Spanish-like VOT range to show the greatest VOT shift between the pre-test and immediate imitation tasks. However, we have included them in the present short-term learning analysis because regardless of how high their VOT value is in the post-test, we care about the absolute difference between the pre-test and post-test values in this analysis. We would like to highlight P202 and P213 again, as these participants had the most drastic VOT shift from the Spanish pre-test to the immediate imitation tasks, without starting with incredibly high VOT values, and their VOT values decreased even further by the Spanish post-test. As we mentioned previously, these participants have the highest and lowest AQ scores, respectively, of the group. Thus, their results do not suggest a trend of autistic-like traits being able to predict short-term L2 phonetic learning. Of the six participants that showed the greatest VOT shift from the pre-test to post-tests, two of the participants were TD, three of the participants were autistic, and one participant was both autistic and diagnosed with ADHD. It is interesting that more of these top-performing short-term phonetic learners have an autism diagnosis, or co-occurring neurodiverse diagnoses, than do not. These results may suggest that a higher number of autistic-like traits and greater neurodivergence could be useful in short-term learning of L2 phonetics; however, the results here are not robust enough to draw this conclusion with confidence. If we look at the next two participants with the greatest VOT shift from the pre-test to post-test Spanish tasks, we see that P203 had a VOT shift of 21.834 ms and P219 had a VOT shift of 18.644 ms. P203's AQ score is 24, attention switching subscale score is six, and attention to detail subscale score is nine. P217's AQ score is 10, attention switching subscale score is two, and attention to detail subscale score is six. In expanding our sample size, we found an even number of TD and neurodiverse participants were strong short-term L2 phonetic learners.

Interestingly, four of our participants showed a slight increase in VOT from the Spanish pre-test to post-tests. These participants were P210, P211, P214, and P218. Three out of these four participants are neurodivergent, two of which have ADHD. It would be interesting to do a follow-up study in which we recruit more participants with ADHD to see if their short-term L2 phonetic learning ability is any different than autistic children. Based on this small sample size alone, we cannot draw any major conclusions. However, it is possible the children with ADHD, without autism, were less attentive to the model speaker's reduced VOT values and, thus, did not adjust their own VOT values very much between the pre-test and post-test, and even between the pre-test and immediate imitation tasks in the case of P210.

A summarized response to the second question outlined above would, again, be no - the results of the present study do not support the idea that autistic-like traits predict short-term L2 phonetic learning. While our participants did shorten their VOT values significantly between the Spanish pre-test and immediate post-test, as a group, their AQ scores and AQ subscale scores did not affect this. Even upon looking at the participants on a more individualized basis, we were not able to conclude that AQ score or AQ subscale scores have any major effect when predicting a child's ability to learn novel phonetic sounds on a short-term basis. While there may be a slight neurodivergent advantage in short-term L2 phonetic learning, the findings from the present study are not strong enough to confidently draw this conclusion.

The last question we aimed to address in Experiment 1 was the following: if short-term learning is predicted by autistic-like traits, is it due to enhanced perceptual abilities? We now know that short-term learning is not predicted by autistic-like traits; however, we can still provide our insights regarding the perceptual abilities of our participants. First, we did not notice any specific trend of higher or lower AQ scores resulting in stronger perceptual boundaries, rather it seemed that our participants' perceptual abilities were comparable across AQ scores. As a group, our participants performed the strongest when deciphering the voiced and voiceless bilabial and dental stimuli, with the bilabial stimuli being the most straightforward and easiest to comprehend for the majority of our participants (Section 2.3.1.1). This is evidenced by the clear inverse S-shape seen in the group perception results and the high percentage of participants that showed distinct categorical perception, with an allowance of just one mistake. For both the bilabial and dental graphs, the perceptual boundary becomes more clear, and the inverse S-shape becomes more defined, from the first to second task; in the case of the bilabial stimuli, the slope becomes steeper by perception task #2 and, in the case of the dental stimuli, the boundary shifts to the left and becomes steeper by perception task #2, and this task effect was proven to be significant via logistic regression analysis. It seems as though the participants already categorized the bilabial stimuli accurately in the first task and, therefore, did not have to shift their boundary significantly by the second task. Because the dental stimuli were slightly more difficult to categorize than the bilabial stimuli, there was a greater change from perception task #1 to perception task #2, or pre- and post-exposure, resulting in a significant difference in responses across these two tasks. This may suggest that, as a group, the participants' bilabial and dental perceptual boundaries shift and improve after exposure to the model speaker's speech.

The velar perception tokens were more difficult for our participants, as a group, especially the tokens that were presented in perception task #2. Even after removing the data of the participants that frequently guessed which velar stop consonant they heard, it was evident our remaining participants also struggled to distinguish the voiced and voiceless velar stop consonants. While it is possible there was an error in the creation of the velar tokens, it is known that children tend to acquire language in the order of the easiest to the most difficult speech sounds (Fabiano-Smith & Goldstein, 2010; Jakobson, 1968; Locke, 1983). According to Jakobson (1968), voiceless sounds are acquired before voiced sounds, and front consonants are acquired before back consonants, meaning bilabial sounds are acquired before alveolar sounds, which are acquired before velar sounds. Given that Fabiano-Smith and Barlow (2010) determined that bilingual children acquire both phonetic inventories simultaneously and at comparable levels to monolingual children, it seems that the same order of acquisition holds true for second language acquisition for children learning a novel language. Thus, we can interpret the results of the present study in the context of existing literature which acknowledges the more complex nature of velar consonants and, thus, the potentially delayed acquisition of these sounds compared to bilabial or dental sounds. Perhaps our participants would show improved velar categorical perception over time or with additional exposure to Spanish.

The same patterns of performing better when deciphering the bilabial and dental tokens, compared to the velar tokens, and performance improving from the first to second task, persisted in the individual analysis (Section 2.3.2.1). That being said, several of the participants, individually, showed a strong ability to shift their perceptual boundaries in the appropriate direction after hearing the model speaker's speech in the immediate imitation task. For example, P209 and P214 shifted both their bilabial and dental perceptual boundaries appropriately to the left by the second perception task. Six of the other participants (P210, P212, P217, P219, P220, P221), whose data was included in the perception analysis, appropriately shifted their boundary to the left for one place of articulation. P211 and P219 did not need to shift their bilabial perceptual boundaries from perception task #1 to perception task #2 and were consistent in their responses. P211's dental perceptual boundary was also appropriately at step 8.5 for both perception tasks and, thus, did not require any shifting. Several participants shifted their perceptual boundaries to the right, in the English-like direction; this finding was unexpected, though it may be indicative of the influence of the participants' established, English-like VOT boundary. It is unclear why this rightward shift would occur post-exposure as opposed to pre-exposure, but it may be related to the difference in syllables presented in perception task #2 compared to perception task #1. Of the participants highlighted above, who showed strong perceptual abilities, four are TD and six are neurodiverse, four of whom are autistic and two of whom have ADHD. Their AQ scores, as a group, range from four to 39. Thus, we still see substantial variation in the number of autistic-like traits of the top performers, even in terms of perceptual abilities. That being said, of the seven participants whose perceptual boundary was at the expected step by perception task #2, six of the participants were neurodiverse. This suggests the potential of a neurodivergent advantage in achieving appropriate categorical perception following exposure to a model speaker. We would like to explore the potential of this advantage further. However, given the variation in perceptual categorization abilities by AQ score, as of right now, we conclude that the answer to the third question above, again, appears to be no - autistic-like traits do not seem to be predictive of perceptual abilities. The findings from the present experiment may contradict the SLM: perception may not always precede production, as our participants did not demonstrate enhanced perceptual abilities that then resulted in stronger short-term L2 learning (Flege, 1987a).

In conclusion, Experiment 1 allowed us to determine that the number of autistic-like traits an individual possesses does not affect their L2 phonetic imitation, short-term L2 phonetic learning, or L2 perceptual abilities. Thus, regardless of a child's neurodivergent diagnosis, or lack thereof, they are capable of performing the tasks related to SLA outlined in this experiment and should not be discouraged from learning, rather encouraged to learn, an L2. Crucially, the results from the present study do not indicate any difference in a child's strength in learning an L2 whether they are neurodivergent or TD; thus, all children would benefit from increased social interaction to learn the novel L2. The L2 classroom is the ideal space to foster these social interactions in an environment in which TD children and neurodiverse children are starting at a similar baseline and possess the same ability to learn the novel L2. Not only will the children improve in their ability to speak their L2, but they will expand their communication skills and benefit from the increased prosocial interactions with their peers.

Chapter 3: Experiment 2

3.1. Introduction

Experiment 2 took place four to seven days after Experiment 1, as this delay allowed us to better understand the longevity of any L2 phonetic learning that occurred in Experiment 1. The goal of Experiment 2 is to evaluate the same research questions outlined in Section 1.3, from the perspective of long-term learning and in more of a case study approach. While group results are informative, we recognize that each autistic child is unique and experiences language differently, making a case study approach more relevant and applicable for the present analysis. We also acknowledge the small sample size of Experiment 2 and hope to be able to contribute more in-depth findings to the field through a more individualized analysis approach. The more individualized approach is not uncommon in the field of autism research; we will utilize a similar approach to that which was used by Macken and Barton (1980a). The primary questions of interest in Experiment 2 are the following:

3. Do autistic-like traits predict long-term L2 phonetic learning?
    a. Is long-term L2 phonetic learning generalizable to new words?
4. If long-term learning is predicted by autistic-like traits, is it due to enhanced perceptual abilities?

In this chapter, we will discuss both the methodology and results of Experiment 2 and discuss these findings prior to the conclusion presented in Chapter 4.

3.2. Methodology

3.2.1. Recruitment

As a prerequisite for participation in Experiment 2, the participants had to have participated in Experiment 1; we only recruited participants for Experiment 2 by invitation, after successful completion of Experiment 1. This invitation was sent directly to parents via email, using the contact information they provided when they scheduled their child's first session via Calendly (Appendix J; Calendly, 2024). Again, parents were invited to book the Calendly time slot that worked the best for their family and they received a confirmation email upon booking.

All of the participants for Experiment 2 met the same inclusion and exclusion criteria required for Experiment 1. As a reminder, the participants we recruited in Experiment 1 were English-dominant children ages seven to 12. We recruited both TD children and neurodiverse children with either an autism or ADHD diagnosis, given the frequency these diagnoses co-occur or have overlapping symptoms (Antshel et al., 2013; Antshel & Russo, 2019; Craig et al., 2015). In the recruitment materials that we sent to clinics for Experiment 1, we specified that we were recruiting children that did not have any other neurological, hearing, or language challenges (Appendix C). The recruitment flyer and messages, the research information sheet, and the assent sheet, all shared the format of the experimental tasks such that parents and clinicians could decide if each child would be able to successfully complete the tasks (Appendices C, E, and F).

As we explained in Section 2.2.1, we isolated our recruitment efforts to New South Wales, Australia, in an effort to find children that predominantly spoke English and were not exposed to Spanish. Although we did not exclude multilingual participants, we did confirm that these participants were not familiar with any voicing languages (Jensen, 1993). As in Experiment 1, we drove recruitment efforts by advertising the compensation rate of $25 USD, or approximately $37 AUD, per parent-child pair following completion of Experiment 2 (Appendix C). By maintaining the same payment amount across the two experiments, we hoped to incentivize families to return to the clinic for Experiment 2. Although all 18 participants that participated in Experiment 1 were invited to return for Experiment 2, only five participants accepted that invitation.

3.2.2. Participants

Because the participants were the same across Experiment 1 and Experiment 2, for Experiment 2 we utilized the same demographic and linguistic information gathered via the Qualtrics survey in Experiment 1 (Qualtrics XM, 2024). The five participants of Experiment 2 spoke English as their primary language and they were ages seven to 10. All five participants were based in Sydney, New South Wales, Australia. One of the participants was exposed to Cantonese, two of the participants were bilingual English-Mandarin speakers, and two of the participants were monolingual English speakers. Four of the five participants were neurodiverse, while the remaining single participant was TD. There were two male participants, two female participants, and one participant who preferred not to specify their biological sex. Their overall AQ scores ranged from 13 to 43. The single TD participant had an AQ score of 13, while the other four neurodiverse participants all scored between 24 and 43; thus, their AQ scores were validated by their diagnoses. All of this information, as well as their relevant AQ subscale scores, is summarized in Table 13 below.

Participant Code	Biological Sex	Age	Diagnosis	Languages	AQ Score	Attention Switch	Attention to Detail
P202	female	9	autism	English, Cantonese *exposure*	43	10	9
P208	prefer not to say	8	autism, ADHD	English	31	8	6
P204	male	10	autism	English, Mandarin	27	6	5
P203	male	10	autism	English	24	6	9
P205	female	7	TD	English, Mandarin	13	4	6

Table 13. Experiment 2 participant profiles.

Given the nature of Experiment 2, we will elaborate on each of our participant's profiles further in the results (Section 3.3) as we take a case study approach to analysis.

3.2.3. Research Materials

Experiment 2 only consisted of two tasks total: one perception task and one production task. The parents were not asked to complete anything in advance of Experiment 2, since they already provided their child's AQ test responses, demographic information, and linguistic profile, as well as their consent to allow their child to participate in both experiments. All of the perception and production tasks for Experiment 2 were consolidated in a Qualtrics survey (Qualtrics XM, 2024). The research information was presented again, such that the child participant could provide their assent prior to beginning Experiment 2 (Appendix E and Appendix F).

3.2.3.1. Perception Task Materials

The perception task materials for Experiment 2 were the same as those which were used in Experiment 1 (Section 2.2.3.1). Again, the perception task was used as a means by which to evaluate each participant's perceptual boundary at different points across the two experiments. In Experiment 1, we looked at each participant's perceptual boundary before and after exposure to the model speaker's speech. In Experiment 2, the perception task was used to understand the permanence of any potential boundary shift that occurred after exposure to the model speaker's speech in Experiment 1. Thus, the goal of the perception task in Experiment 2 was to determine each participant's perceptual boundary even after a four to seven day delay.

Experiment 1 utilized 12 stop-initial syllables, which began with either a voiceless stop consonant /p, t, k/ or voiced stop consonant /b, d, g/, followed by /o/ or /a/: /pa/, /ba/, /po/, /bo/, /ta/, /da/, /to/, /do/, /ka/, /ga/, /ko/, and /go/. These 12 syllables were distributed across the two perception tasks presented in Experiment 1. For Experiment 2, we included six of the syllables; while we included one minimal pair per place of articulation, we randomly selected whether /o/ or /a/ followed the consonant in each pair. By ensuring that the six syllables covered all three of the minimal pairs (/p/ and /b/, /t/ and /d/, and /k/ and /g/), we were able to evaluate the participants' perceptual boundaries of interest. The syllables tested in the perception task presented in Experiment 2 were the following: /po/, /bo/, /to/, /do/, /ka/, and /ga/.

Our native speaker, a Spanish-English bilingual female in her twenties, produced all of the syllables with both Spanish-like and English-like pronunciation, which allowed us to gather a full range of VOT values, covering those that are typical of both Spanish and English (Kellogg & Chang, 2023; Lozano-Argüelles et al., 2021). We manipulated the spectrum of VOT values to provide stimuli, or steps, in 10 ms increments. This resulted in 42 steps across the three places of articulation: the VOT range for /b/ and /p/ was -70 ms to 60 ms with 13 steps between the two, the VOT range for /d/ and /t/ was -80 ms to 70 ms with 15 steps between the two, and the VOT range for /g/ and /k/ was -60 ms and 80 ms with 14 steps between the two. Thus, Experiment 2 consisted of 42 perception stimuli total, all of which were manipulated such that they matched in pitch, vowel duration, and intensity, like in Experiment 1 (Winn, 2020a). We did not specify the language of the perception task, rather presented the stimuli in a kid-friendly manner, stating in the directions that they would select the sound they heard our "robot friend" saying per audio file (Figure 13, Section 2.2.3.1). As in Experiment 1, the participants' perception responses in Experiment 2 were collected and stored within Qualtrics (Qualtrics XM, 2024). Section 2.2.3.1 above discusses the creation and presentation of the perception stimuli in more detail.

3.2.3.2. Production Task Materials

The production task in Experiment 2 included 48 stimuli total, all of which were in Spanish. These 48 stimuli consisted of 12 of the original target words and 12 of the original filler words from Experiment 1, as well as 12 novel target words and 12 novel filler words, all presented in a randomized order (Table 14).

Original Target Words	**Novel Target Words**	**Original Filler Words**	**Novel Filler Words**
pato	pata	mano	gama
palo	paso	lago	lado
pomo	polo	vaca	bata
poro	poso	gafa	mara
talo	tata	bala	mago
tapa	taca	rana	masa
toro	tono	mono	doce

topo	toca	lobo	nota
cara	cana	ropa	bolo
casa	capa	goma	foro
codo	cono	moto	gota
coro	coco	boca	foca

Table 14. Original and novel tokens for Spanish production task.

The novel target words, like the original target words, were all disyllabic nouns with CVCV structure, controlled for penultimate stress. Each target word began with one of the voiceless stop consonants /p, t, k/, four per place of articulation for the 12 original target words and four per place of articulation for the 12 novel target words. For each set of four target words, two of the words consisted of the vowel "o" following the stop consonant and two of the words consisted of "a" following the stop consonant. The novel filler words, like the original filler words, were all disyllabic nouns with CVCV structure, controlled for penultimate stress, which began with a consonant other than one of the target voiceless stop consonants. Half of the original filler words and half of the novel filler words contained the vowel "o" following the consonant, while the remaining half of the filler words consisted of the vowel "a" following the consonant. All 48 of the Spanish target and filler words were presented in the carrier sentence "di _____" ("*say*___"), with an instructional slide following the directions slide, but preceding the experimental tasks (Figures 14 and 16).

The production task stimuli were presented and recorded following the same procedure outlined in Experiment 1 (Section 2.2.3.2). Again, we relied on Qualtrics and Audacity to present and record, respectively, the production tasks completed in Experiment 2 (Audacity Team, 2023; Qualtrics XM, 2024).

3.2.4. Procedure

Experiment 2 was scheduled four to seven days after Experiment 1. Data collection for Experiment 2 took place at the Bright Eyes Early Intervention clinic location in Pymble, New South Wales, Australia. We collected the data in a quiet, furnished room in order to minimize background noise. Parents were invited to either accompany their child in the room while they completed the experimental tasks or to wait for their child in the lobby, depending on their child's preference. Prior to beginning Experiment 2, each participant had another opportunity to read the same information sheet provided in Experiment 1, ask any questions, and assent to participating by continuing with the Qualtrics survey (Appendix E and Appendix F; Qualtrics XM, 2024). The child's parent did not have to offer their consent again, as their original consent to participate covered both the first and second experiments.

Experiment 2 consisted of one perception task and one production task. Because of the reduced task load, we anticipated this session would take about 15 minutes total to complete, with one break offered between the two tasks. The recommended break time was two minutes, though we allowed participants to take an extended break, as needed. Again, we provided coloring supplies to the child participants during the break periods.

The research team allowed the child to put on and adjust the headset microphone independently, before checking that the mouthpiece was appropriately placed within a couple inches of the child's mouth. We began recording in Audacity at the start of Experiment 2 to ensure the recording was already running when the child continued on to the production task (Audacity Team, 2023). For all of the experimental tasks, the child was permitted to choose whether they wanted to read the directions aloud or hear the directions read to them by the research team. Once the research team confirmed the child's comprehension of the task at hand, the child was allowed to begin the experiment.

The first task was the perception task, presented in the format of a two-alternative forced-choice task. After playing each audio file, the participant selected which sound they heard; they were allowed to replay the audio file, as needed. We allowed the child to choose whether they wanted to navigate through the perception tasks independently, using the trackpad, or if they wanted the research team member to scroll and record their responses for them. If the child preferred to have the research team member navigating the task for them, the research team member asked the child if they heard the first or second option listed, to avoid reading the syllables shown. Similarly, each participant was allowed to choose whether they wanted to click through the Google Slides presenting the production task stimuli or if they preferred to have a research team member move through the slides on their behalf.

For the second experimental session, the child completed the following two tasks:

1. Perception task #3
2. Spanish delayed post-test: read-aloud task (*production task #5*)

As indicated above, the tasks are numbered corresponding to their position in the lineup of tasks across Experiment 1 and Experiment 2. Because we are comparing the participants' performance in Experiment 2 to their performance in Experiment 1, we decided to label the tasks as such.

The research team monitored the flow of the experiment, as well as the breaks. We also ensured all of the tasks were recorded properly and addressed any technical errors that arose. We thanked the parent and child for their participation in both Experiment 1 and Experiment 2 and referred them to our contact information to address any questions that came up after completion of the experiment. Each parent-child pair received $25 USD, or $37 AUD, in the form of an Amazon gift card as compensation for successfully completing Experiment 2.

3.2.5. Variables

With the following research questions at the forefront of Experiment 2, there is one dependent variable of interest, which is the same as that of Experiment 1: VOT.

3. Do autistic-like traits predict long-term L2 phonetic learning?
    a. Is long-term L2 phonetic learning generalizable to new words?
4. If long-term learning is predicted by autistic-like traits, is it due to enhanced perceptual abilities?

To answer question (3), we were interested in comparing the VOT values of the Spanish delayed post-test (production task #5) and both the Spanish immediate post-test (production task #4 from Experiment 1) and the Spanish pre-test (production task #2 from Experiment 1). In comparing the VOT values of the Spanish delayed post-test to the Spanish immediate post-test, we are able to understand whether or not any short-term learning that was achieved in Experiment 1 persisted in the long-term, in Experiment 2. By comparing the Spanish immediate post-test to the Spanish pre-test, we are able to evaluate each participant's long-term learning overall, from the start of Experiment 1 to the end of Experiment 2. In order to answer part (a) of question (3), we also compared each participant's VOT values for the novel target words to their VOT values for the original target words, within the delayed post-test. This comparison allowed us to understand the extent by which each individual did or did not generalize their learned VOT to the novel words.

The independent variables in this study included AQ score or specific AQ subscale score (e.g., attention switching), task (pre-test vs. immediate post-test vs. delayed post-test), and target word (novel vs. original). AQ score and AQ subscale score were treated as continuous variables, since the sum AQ test scores range from zero to 50 and the AQ subscale scores range from zero to 10. We were interested in two of the AQ subscales in particular: attention switching and attention to detail. Both task and target word were categorical variables, with three levels for task and two levels for target word, as outlined above. Lastly, the covariates for Experiment 2 were participant and experimental item, and the fixed effects were the voiceless stop consonant that each experimental item began with (/p, t, k/) and the vowel (/o, a/) following the stop consonant.

In order to answer question (4) listed above, we compared and graphed the perception task responses per place of articulation, per task, and inspected each visualization in order to estimate the boundary, or 50% crossover point, and the boundary shift.

Using the same coding scheme outlined in Experiment 1 (Section 2.2.6.1), we analyzed our participants' perceptual abilities by graphing their responses to the perception task, as we did in Experiment 1. We graphed their responses per place of articulation, per task, and inspected the resulting categorical boundaries. For the purpose of Experiment 2, the perceptual boundary visualizations include the data from the first two perception tasks in Experiment 1. This is discussed further in the analysis section (Section 3.2.6) that follows.

3.2.6. Analysis

Given the small sample size of our participants, we will present a brief report of the group trends, but will prioritize a case study approach for Experiment 2. In addition to the small sample size, each autistic child uses and acquires language differently, further supporting a case study approach. We also acknowledge potential limitations of the present study. Our analysis, as well as the results reported in the following section (Section 3.3) begins with the perception task prior to the production task.

While we did not run any statistical analyses on the data from Experiment 2, we used similar analysis techniques to aid in our interpretation of both the perception and production tasks in Experiment 2. For the perception analysis, participants' responses to the two-alternative forced-choice task were graphed in order to evaluate the point at which their perceptual categories transitioned from voiced to voiceless responses for each place of articulation. The participants' responses were coded as "1" for a voiced response and "0" for a voiceless response. Participants were exposed to the full VOT spectrum per place of articulation, but heard only one stimulus per step. We graphed their responses across the three perception tasks, two in Experiment 1 and one in Experiment 2, in order to visualize any shift in their perceptual boundaries across tasks. This allowed us to determine how they categorized the voiced and voiceless stop consonants prior to exposure to the model speaker's speech (perception task #1), immediately after this exposure (perception task #2), and after a four to seven day delay post-exposure (perception task #3). As shown in Appendix G, we used "ggplot2" in R to create the visualizations, after summarizing and manipulating the data using the "dplyr" package (v4.2.3; R Core Team, 2023; Wickham, 2016; Wickham et al., 2023). Using these visualizations, we were able to comment on the participants' boundary, or 50% crossover point, and boundary shift per place of articulation, per task, which aided in our interpretation of their results. We will include these data points, as applicable, in the results below (Section 3.3.1).

For the production analysis, we uploaded the data and labeled the target stop consonants, /p, t, k/, of both the novel and original target words, using the same analysis procedure outlined in Section 2.2.6.2. We extracted the VOT values of these labeled segments, using the same procedure as Experiment 1.

As mentioned in Section 2.2.6.2, we confirmed the reliability of our data, from both Experiment 1 and Experiment 2, using the test-retest approach across both datasets at once. To reiterate, we randomly selected 20 /p/-initial, 20 /t/-initial/, and 20 /k/-initial words from all of the data collected and repeated our measurement of VOT. The cor.test() function in R revealed a significant, strong, positive correlation ($r = 0.990$, $p < 0.001$) between the original VOT values and the repeated measures, confirming that our results are reliable (v4.2.3; R Core Team, 2023).

In Section 2.3.1.2, we established the Spanish pre-test read-aloud task (production task #2) as a useful baseline for measuring our participants' VOT values. Furthermore, in Section 2.3.2.2 of Experiment 1, the individual and group average VOT values per place of articulation, per task, are reported. For the purpose of Experiment 2, we will summarize each participant's VOT values on the delayed post-test, as well as compare these values to their baseline and immediate post-test VOT values presented in Experiment 1. These results are summarized in tables, one per participant, such that we can better visualize the change in their VOT across the Spanish pre-test, immediate post-test, and delayed post-test. These tables will also divide each participant's VOT values in the delayed post-test according to whether the target words were novel or part of the original list of stimuli from Experiment 1. Additionally, the tables will include the group average VOT values per task.

3.3. Results

The results for Experiment 2 are presented such that the perception results precede the production results. We will begin by commenting on the general trends of the group, before providing a more in-depth analysis via case studies.

3.3.1. Perception Results

As a group, it seems the same order of ease of comprehension followed for perception task #3; in other words, the bilabial stimuli were the easiest to categorize, followed by the dental stimuli, and, finally, the velar stimuli. We based our interpretation of ease of comprehension on the number of participants that showed the expected inverse S-shape in their voiced versus voiceless responses, their accuracy in achieving a VOT boundary closer to the Spanish-like 0 ms boundary, and reduced mistakes or deviations from the anticipated inverse S-curve. As evidenced by the graphs included in the case study analysis below, it seems that, as a group, participants performed as well or better on the delayed perception task than on perception tasks #1 and #2. The graphs below show the group mean percent of "voiced" (/b, d, g/) responses per step for each of the respective places of articulation, across the three perception tasks completed during Experiments 1 and 2 (perception task #1: purple line; perception task #2: green line; perception task #3: bold black line). Again, this improvement is evidenced by the same or fewer deviations from the expected inverse S-curve, as well as a VOT boundary that is closer to the Spanish-like, 0 ms boundary. Thus, it seems as though participants benefited from exposure to the native Spanish speaker, such that their performance on the perception tasks improved from pre-exposure (perception task #1) to post-exposure (perception task #2), and this improvement persisted even in the delayed post-exposure task (perception task #3). As discussed in more detail in the case study analysis below, all of the participants that completed perception task #3 showed either retained improvement or improved further in their categorical perception in this delayed perception task, regardless of AQ score.

As we transition to the case study results, we will provide a brief description of each participant's profile prior to presenting their respective perception results. This overview of each of the Experiment 2 participant profiles offers a strong foundation for our individualized analysis of each participant's perception, as well as production (Section 3.3.2), results. For each of the five participants of Experiment 2, we have included three graphs to aid in interpreting their perceptual abilities. We have highlighted the results of Task 3 specifically for the case study analysis; please refer to Appendix I to see the participants' performance on Tasks 1 and 2 alone. Again, we anticipate the responses will result in an inverse S-shape if participants heard voiced stop consonants for the lower, negative VOT values and voiceless stop consonants for the higher, positive VOT values. Each of the graphs below shows the number of steps there were between the two minimal pairs, per place of articulation. For the purpose of the present case study analysis, we will provide a more in-depth analysis of each participant's performance on perception tasks #1 and #2 in Experiment 1 before discussing their performance on perception task #3 in Experiment 2.

P202 is a nine-year-old female who was diagnosed with autism at the age of three. She has never been diagnosed with any other psychiatric, nor motor disabilities, and does not have any known hearing or neurological challenges. Lastly, P202 does not have any diagnosed speech or language challenges. P202 is a monolingual English speaker, though she has been exposed to Cantonese. Her exposure is not to the point of fluency, but she is exposed to Cantonese for more than 10% of her wake time per week via her grandmothers. P202's parents report that their daughter has heard Cantoneses since birth, but has never been encouraged to learn the language properly, so she can only say a few words in the language and can understand simple sentences. P202 scored a 43 overall on the AQ test, with an attention switching score of 10 out of 10, and an attention to detail score of nine out of 10.

In Experiment 1, it seems perception task #1 was slightly more difficult than perception task #2 for P202, with respect to the bilabial stop consonant stimuli; while the line for the first task shows a spike that deviates from the S-curve, the line for the second task follows the expected inverse S-shape (Figure 33). In the first perception task, P202's bilabial perceptual boundary was either at step 6.5, between -20 ms and -10 ms, or step 8.5, between 10 ms and 20 ms. However, by the second task, P202's bilabial perceptual boundary shifted appropriately by one step such that her boundary was at 7.5, or between -10 ms and 10 ms. Although it is unclear whether this was an English-like rightward shift or a Spanish-like leftward shift, this finding may suggest that P202 learned to shift her perceptual boundary to the appropriate crossover point, at 0 ms, after exposure to the model speaker's speech. Based on Figure 33 below, it seems as though P202 was better able to decipher the bilabial perception stimuli in the third task than the first task, but her perceptual boundary is shifted further to the left (step 5.5) than expected in this third task. Perhaps P202 recalled the need to shift her bilabial perceptual boundary to the left, but over-accommodated this shift such that her perceptual boundary was at step 5.5, or between -30 ms and -20 ms. Regardless, her performance on the third perception task follows the expected inverse S-shape pattern, which is an improvement from her performance on the first perception task.

With respect to both the dental and velar stimuli, it seems P202 struggled slightly more with the second task compared to the first task, as we see more spikes back and forth between reporting voiced and voiceless stop consonants; we see more of these spikes for the velar stimuli presented in perception task #2 (Figures 34 and 35). The initial crossover point between voiced and voiceless for both the dental and velar stimuli, across perception tasks #1 and #2, occurs at step 7.5; we expected to see this transition one step to the right for the dental stimuli, at step 8.5, and one step to the left for the velar stimuli, at step 6.5. P202 seemed to provide more consistent voiced and voiceless responses for the dental and velar stimuli, such that the expected inverse S-shape occurred, in the first task than in the second task. By perception task #3, P202's dental perceptual boundary was exactly between -10 ms and 10 ms, at step 8.5; this is an improvement from her performance on both the first and second perception tasks in terms of boundary accuracy, despite the rightward shift (Figure 34). P202's categorization of the dental stimuli resulted in a perfect inverse S-curve in perception task #3, which is an improvement from the single mistake she made in classifying the dental stimuli in perception task #2. Lastly, Figure 35 shows that P202's main transition from voiced to voiceless velar responses occurred at the exact same point across all three perception tasks: between steps seven and eight (step 7.5), or 10 ms and 20 ms. Her velar perceptual boundary follows the expected inverse S-shape in perception task #1, but not in perception tasks #2 or #3. That being said, P202 only made one mistake in categorizing the velar stimuli, resulting in a deviation from the anticipated inverse S-curve, by perception task #3; in perception task #2, she had three responses that deviated from the inverse S-curve.

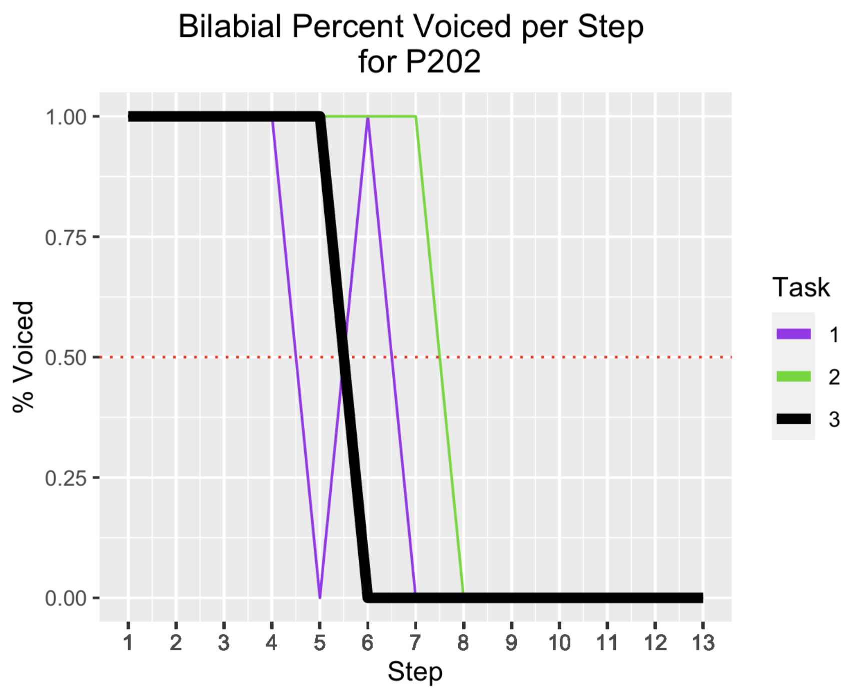


Figure 33. P202 percent bilabial voiced responses per step per task.

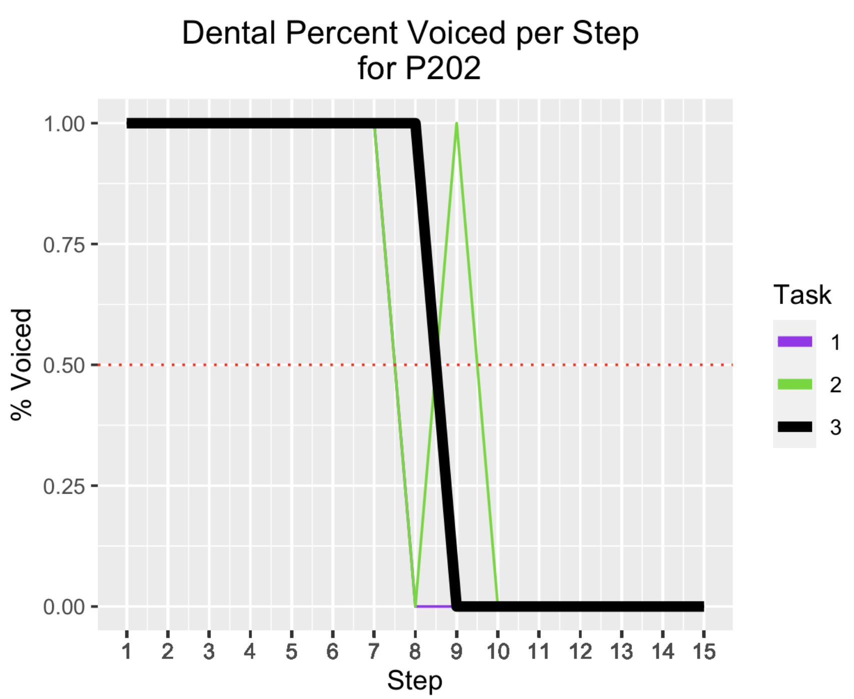


Figure 34. P202 percent dental voiced responses per step per task.

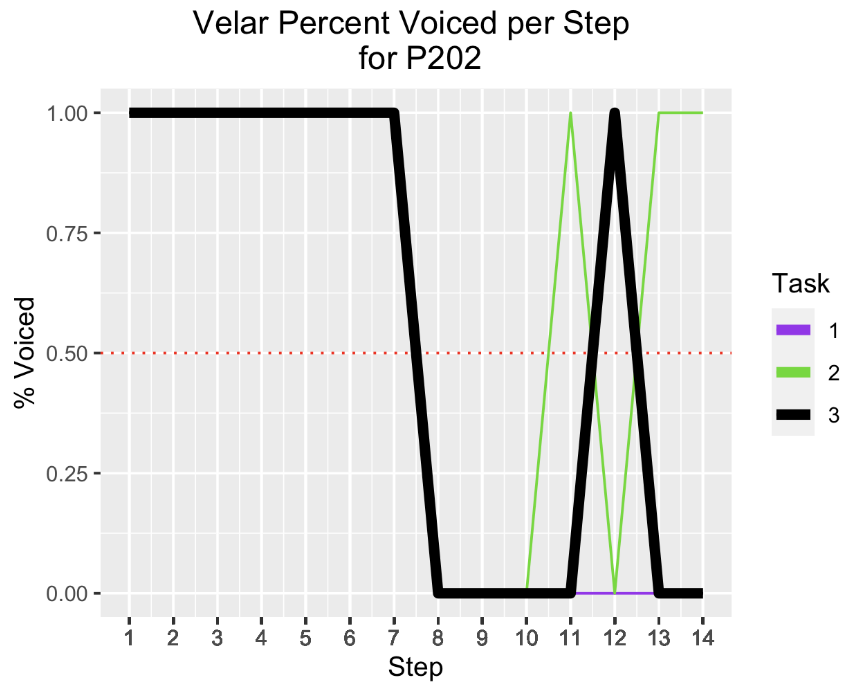


Figure 35. P202 percent velar voiced responses per step per task.

P208 is an eight-year-old autistic child who was diagnosed at age six. P208 also has an ADHD diagnosis, but does not have any other motor or neurological disabilities. P208 does not have any known hearing or language challenges either. They are fluent in English and do not have exposure to any other languages on a regular basis. P208 received an AQ score of 31, with subscale scores of eight and six on the attention switching and attention to detail subscales, respectively.

With respect to Experiment 1, P208's responses to the perception stimuli were the most consistent for the bilabial stimuli in the first task, the dental stimuli in the second task, and the velar stimuli in the first task. It does not seem like exposure to the model speaker's speech was particularly influential on P208's perceptual boundary. P208's responses to the bilabial stimuli in the first task showed a clear inverse S-shape, as expected (Figure 36). Their transition between voiced and voiceless responses occurred at step 8.5, so their perceptual boundary was shifted to the right by one step. While the last transition between voiced and voiceless bilabial stop consonants occurred between the same steps for the second task, their responses spike back and forth, due to a single mistake at step three, before reaching this transition in the second task. P208's bilabial perceptual categorization in the third perception task also showed one deviation from the inverse S-curve (Figure 36). That being said, the last transition from voiced to voiceless responses occurred between steps six and seven (-20 ms and -10 ms) in the third perception task, which is a shift in the appropriate, leftward direction from the first and second perception tasks, in which this crossover point occurred at step 8.5.

P208's responses to the dental stimuli in the first task seem unpredictable, with multiple deviations from the expected inverse S-curve. There was only one spike in P208's categorization of voiced and voiceless dental stop consonants before the final transition to voiceless responses at step 10.5, in the second task (Figure 37). This transition occurred at a point that was slightly shifted to the right, between 20 ms and 30 ms, compared to where we would expect to see it. Similar to P208's performance on the second task, their categorization of the dental voiced and voiceless stops on the third task nearly followed the inverse S-shape pattern, with one spike in the opposite direction in the lower VOT values (Figure 37). While the boundary between voiced and voiceless dental stop categorization occurred between steps 10 and 11 (20 ms and 30 ms) in the second task, this boundary shifted appropriately to the left such that it is between steps nine and 10 (10 ms and 20 ms) by the third task.

Lastly, with respect to Figure 38, P208 showed the most consistent velar voiced responses between steps three and nine for the first task and between steps one and eight for the second task. The boundary for either task line was slightly shifted to the right from where we would expect to see the boundary for velar stop consonants, which is between steps six and seven (step 6.5). Beyond step three, the line for the first task appears to follow the inverse

S-shape we expect to see when looking at perceptual boundaries. Figure 38 shows the extreme inconsistency in P208's categorization of voiced and voiceless velar stop consonants in the third perception task. While there is a slight resemblance of the inverse S-shape pattern in their responses to the greater VOT values in the first task, and, less so, in the lower VOT values in the second task, we do not see any kind of consistency in their responses in the third perception task. Across all three perception tasks, it is very likely P208 was simply guessing which sound they heard, or answering randomly, perhaps in an attempt to complete the task at a faster pace.

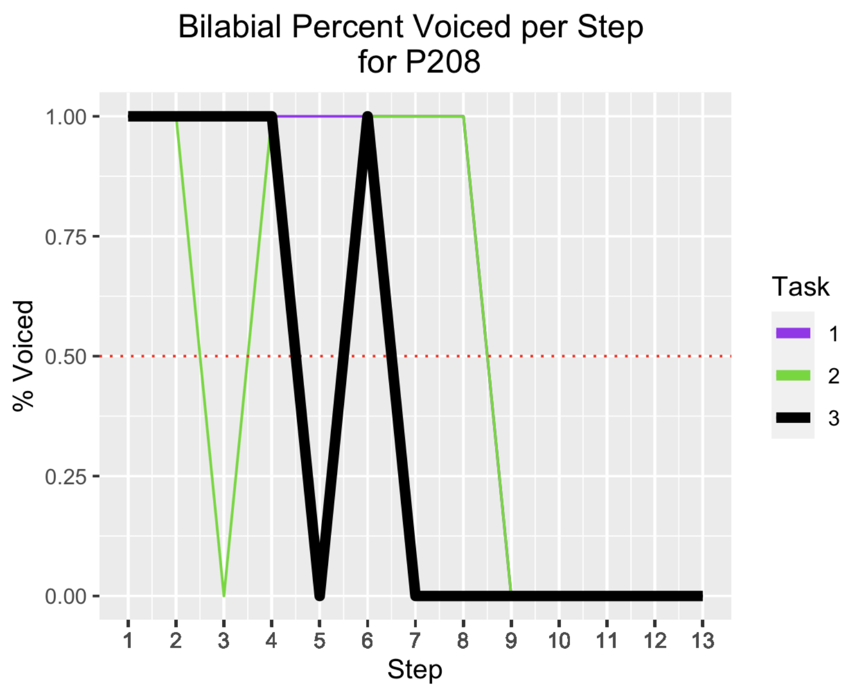


Figure 36. P208 percent bilabial voiced responses per step per task.

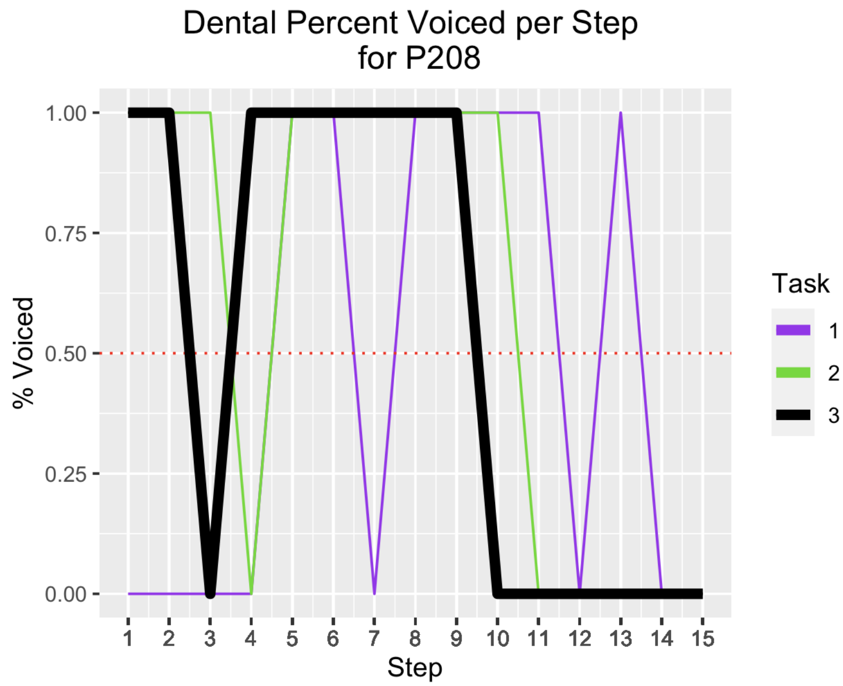


Figure 37. P208 percent dental voiced responses per step per task.

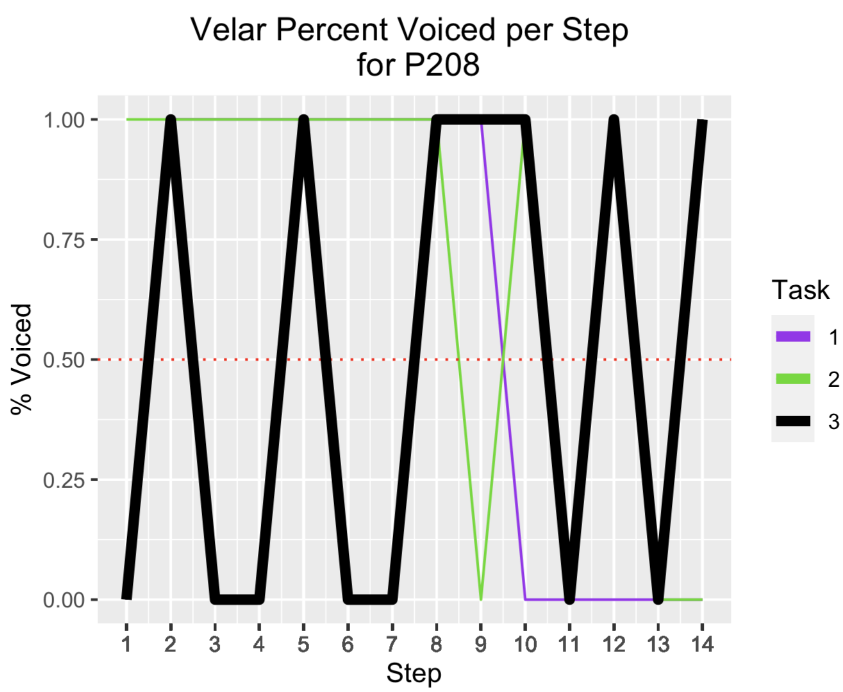


Figure 38. P208 percent velar voiced responses per step per task.

P204 is a 10-year-old male diagnosed with autism at the age of nine. P204 does not have any other psychiatric or motor disabilities, and he has not received a hearing, neurological, nor speech or language impairment diagnosis. He is bilingual and fluent in both English and Mandarin. Other than those two languages, he does not have weekly exposure to any other languages. P204 received an overall AQ score of 27, with an attention switching subscale score of six and an attention to detail subscale score of five.

P204 seemed to have the easiest time deciphering the voiced and voiceless bilabial stimuli across all three perception tasks (Figure 39). For both the first and second perception tasks, P204's responses to the bilabial stimuli followed the expected inverse S-shape pattern. His perceptual boundary was between four and five in the first task, but it shifted to the right, appropriately, such that his boundary was between six and seven, or -20 ms and -10 ms, by the second task. By the third task, P204's bilabial perceptual boundary, occurred between steps six and seven, or 10 ms and 20 ms, as it did in the second perception task (Figure 39). That being said, he still reported hearing a voiceless response at step five, or -30 ms, resulting in a nearly perfect inverse S-shape pattern, with one spike in the opposite direction at step five.

While we see several spikes in the opposite direction for P204's performance on the first task, we see fewer spikes by the second task for the dental stimuli. In fact, there are not any spikes in the opposite direction for the dental stimuli in the second task, rather we see the expected inverse S-shape, just shifted to the right two steps from where we expected to see this boundary (Figure 40). By the third perception task, P204's responses to the dental stimuli followed the expected inverse S-shape pattern, with the transition from voiced to voiceless responses occurring between steps nine and 10, or 10 ms and 20 ms (Figure 40). This is an improvement, and Spanish-like shift, from his perceptual boundary existing at step 10.5, or between 20 ms and 30 ms, in the second perception task.

For the velar stimuli, P204 showed some inconsistent voiced and voiceless responses in the first perception task, but his last crossover point occurred at step 9.5. It appears P204 heard voiced stop consonants for all of the stimuli except for step 10, with a VOT of 40 ms, in the second task (Figure 41). By the third perception task, there is an evident shift in P204's voiced and voiceless responses between steps eight and nine, or 20 ms and 30 ms, but his responses do not show the consistency needed to produce a perfect inverse S-shape (Figure 41). Since P204 seemed to struggle more with perception task #3, and showed less consistency in his responses, it is possible he was less focused or interested in the task by Experiment 2. Perhaps P204 was more engaged in the task when he completed it for the first time in Experiment 1.

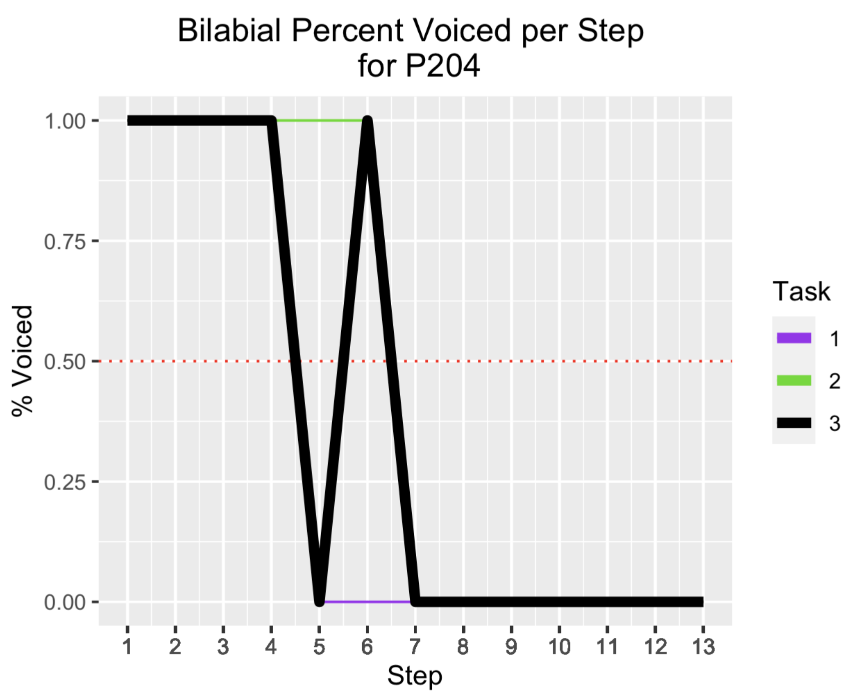


Figure 39. P204 percent bilabial voiced responses per step per task.

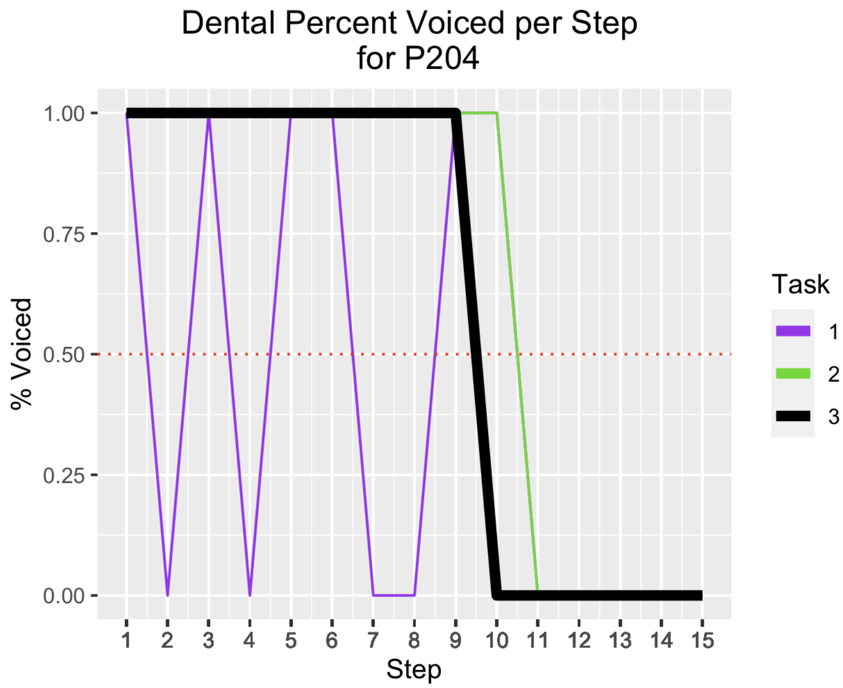


Figure 40. P204 percent dental voiced responses per step per task.

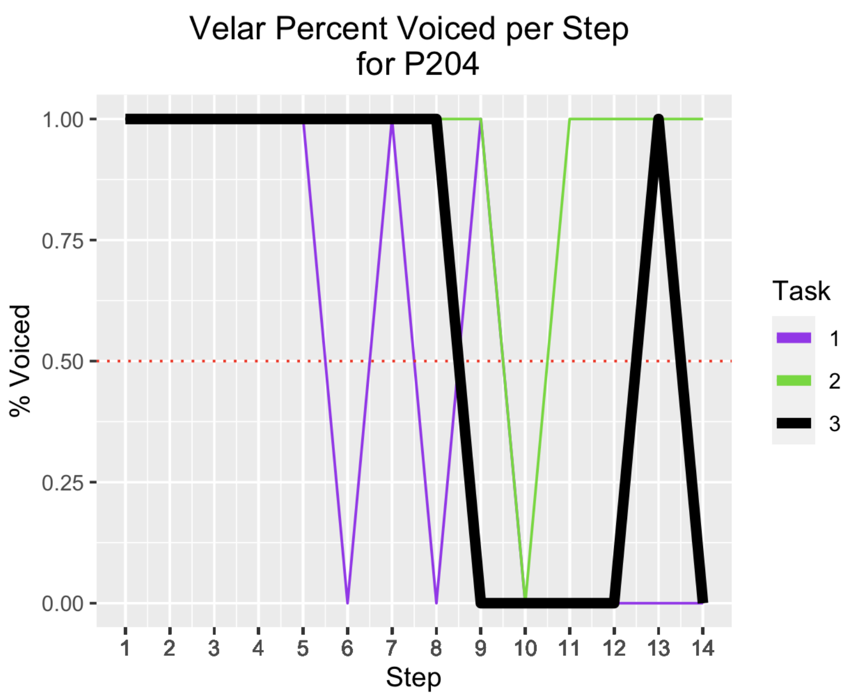


Figure 41. P204 percent velar voiced responses per step per task.

P203 is a 10-year-old autistic male without any other psychiatric or motor disabilities. He was diagnosed with autism at the age of four. P203 has never received a neurological, hearing, nor speech or language impairment diagnosis. He is a monolingual English speaker without any weekly exposure to any other languages. P203 scored a 24 overall on the AQ test, the lowest of any of our neurodiverse participants. His attention switching subscale score was six, while his attention to detail subscale score was nine.

Unfortunately, P203's perception data from Experiment 2 did not save appropriately. His results in Experiment 1 showed that he had difficulties with the perception tasks, across all three places of articulation, as we see inconsistency in his voiced versus voiceless responses. For all three places of articulation, we see several spikes in the opposite direction. While this pattern of bouncing back and forth persists across the full range of steps for the velar stimuli, P203's responses were more consistent after steps seven/eight for the bilabial stimuli and after steps nine/10 for the dental stimuli. For the bilabial stimuli, P203's perceptual boundary appears to be between steps seven and eight in the first task, but it shifts to the left by one step by the second task; P203's perceptual boundary was in the location predicted in the first task, though it shifted such that it was between -20 ms and -10 ms by the second task, perhaps over-accommodating for the shortened VOT values of Spanish (Figure 42). Again, P203's dental perceptual boundary was appropriately further to the left in the first task, but shifted to the right by the second task (Figure 43). The only graph that resembles the anticipated inverse S-shape in this case is P203's dental perceptual boundary as demonstrated by the second perception task. There is no particular pattern for either perception task related to the velar stimuli (Figure 44). Although P203's responses lack consistency, it does seem like he guessed more in the lower VOT values for the bilabial and dental stimuli. Perhaps he had a harder time deciphering and categorizing the syllables in this lower VOT range, but was able to more confidently classify the bilabial and dental voiceless syllables.

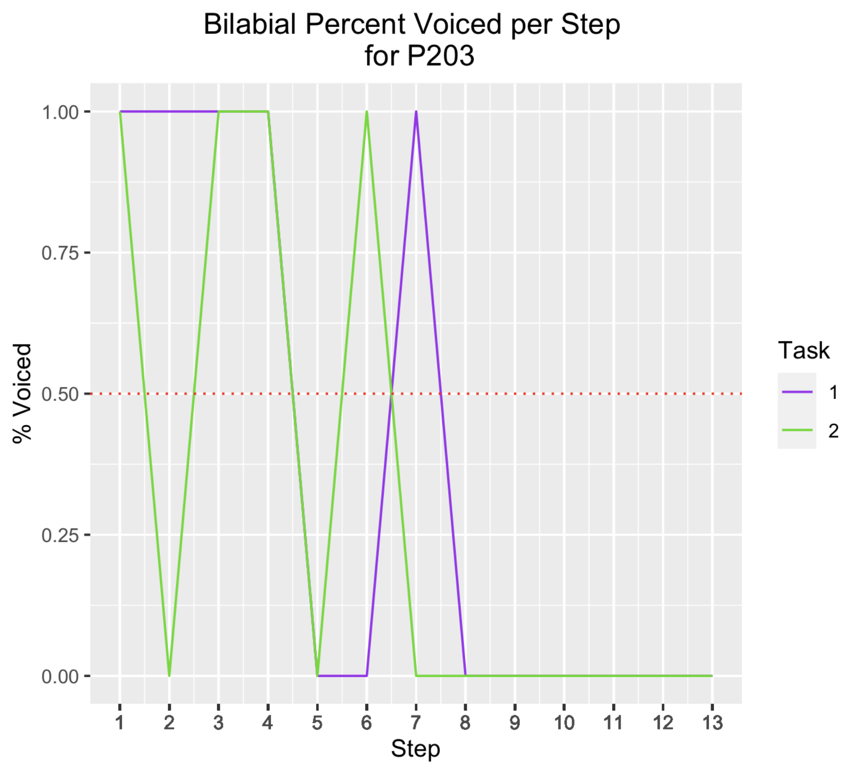


Figure 42. P203 percent bilabial voiced responses per step per task.

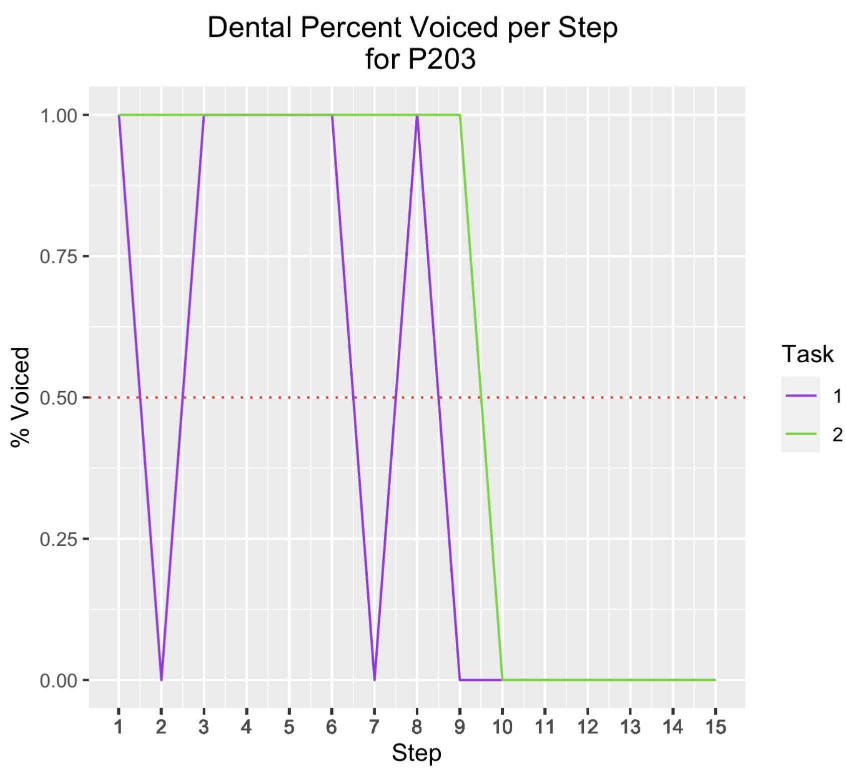


Figure 43. P203 percent dental voiced responses per step per task.

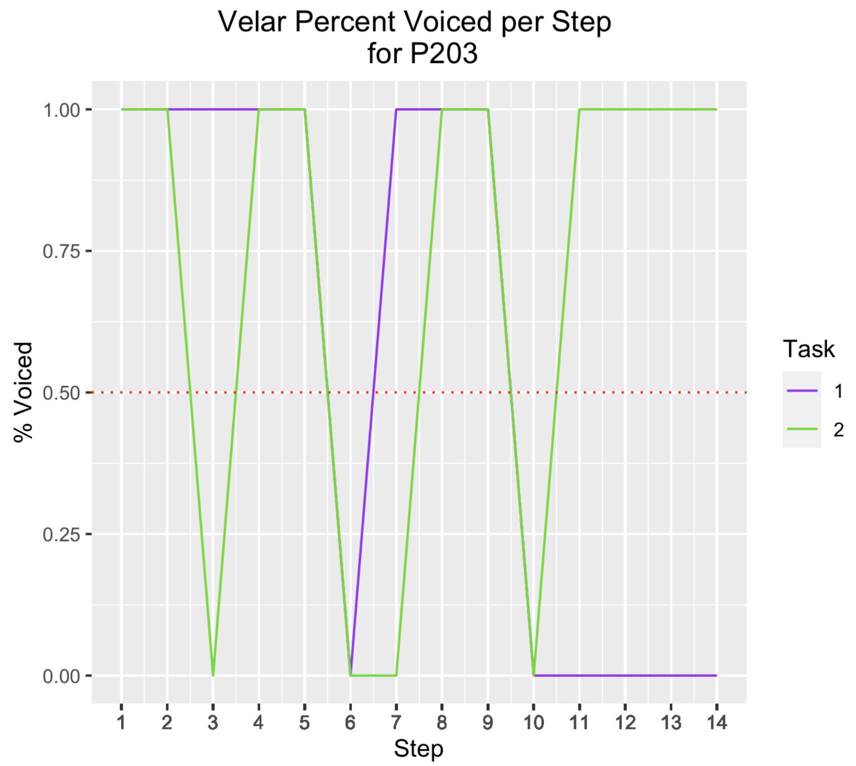


Figure 44. P203 percent velar voiced responses per step per task.

P205 is a seven-year-old, TD female. She does not have any psychiatric, motor, neurological, hearing, nor speech or language challenges. As the sibling of P204, P205 is also bilingual and speaks both English and Mandarín fluently. She is not regularly exposed to any other languages. P205 scored 13 out of 50 on the AQ test, with scores of four and six on the attention switching and attention to detail subscales, respectively.

Based on the graphs below, P205 found the bilabial and dental stimuli slightly easier to distinguish voiced and voiceless stop consonants compared to the velar stimuli (Figures 45 and 46). For both the bilabial and dental stimuli, we see the inverse S-shape we expected by the second perception task; however, there is not a noticeable inverse S-shape to P205's perceptual boundary for the bilabial and dental stimuli in the first task. For the bilabial stimuli, the boundary and last transition towards a voiceless response for the first task occurs at step 9.5, but this transition is shifted appropriately to the left by the second task (Figure 45). Still, P205's perceptual boundary was one step off from the expected boundary, as her boundary between voiced and voiceless responses occurred between 10 ms and 20 ms (step 8.5). Although P205 switched back and forth between hearing voiced and voiceless dental stop consonants inconsistently in the first task, the steepest slope, or the perceptual boundary for both the first and second task were the same: step 10.5, or between VOT values of 20 ms and 30 ms (Figure 46). By the third task, it appears P205 still performed well when categorizing the bilabial and dental stimuli, as we see the expected inverse S-shape pattern for her third perception task performance in both Figures 45 and 46. Furthermore, both her bilabial and dental perceptual boundaries have shifted appropriately to the left between the second and third perception tasks. In Figure 45, we see that her bilabial perceptual boundary shifted from 8.5 in the second task to 6.5 in the third task, shifting left, in the Spanish-like direction. Figure 46 shows that P205's dental perceptual boundary shifted leftward from its position between steps 10 and 11 (20 ms and 30 ms) to between steps nine and 10 (10 ms and 20 ms) from the second to the third perception tasks.

The velar stimuli appeared to be difficult for P205. P205 showed slightly more consistent voiced and voiceless responses in the first task than in the second task, when her responses seemed to bounce back and forth even more (Figure 47). We do not see the expected inverse S-shape in P205's responses to the velar stimuli. Similarly, we see inconsistent voiced and voiceless responses in P205's performance on the third perception task as well.

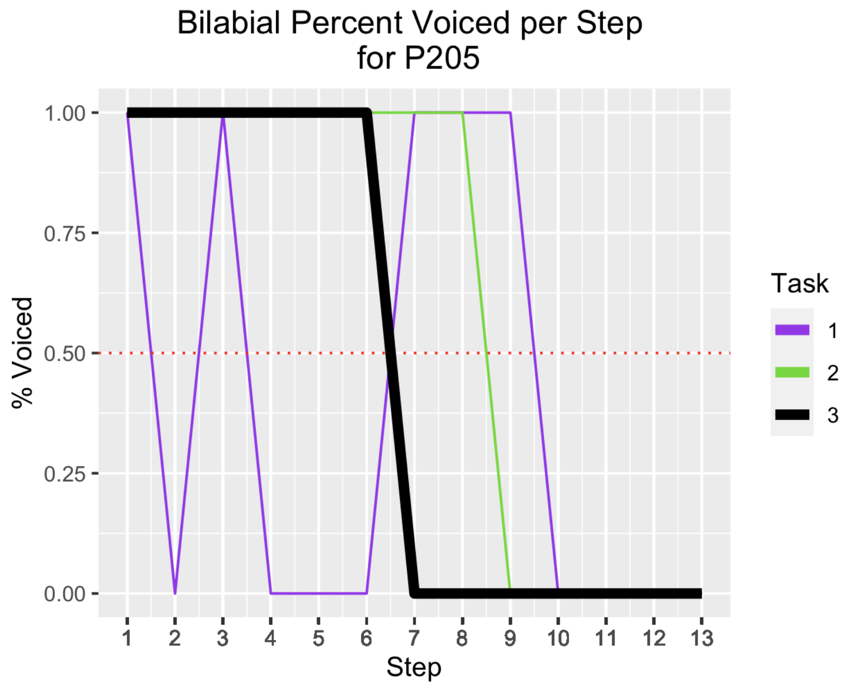


Figure 45. P205 percent bilabial voiced responses per step per task.

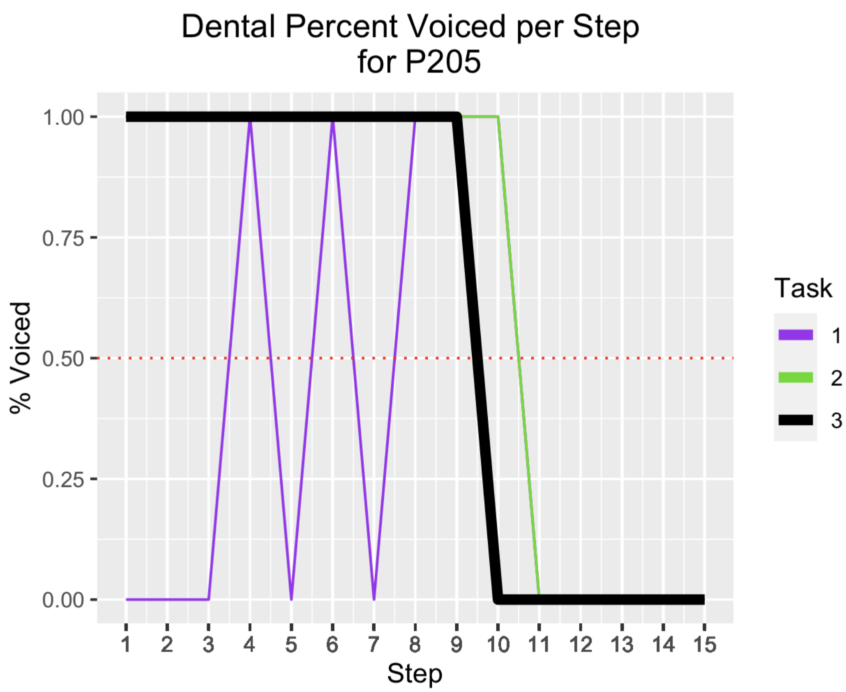


Figure 46. P205 percent dental voiced responses per step per task.

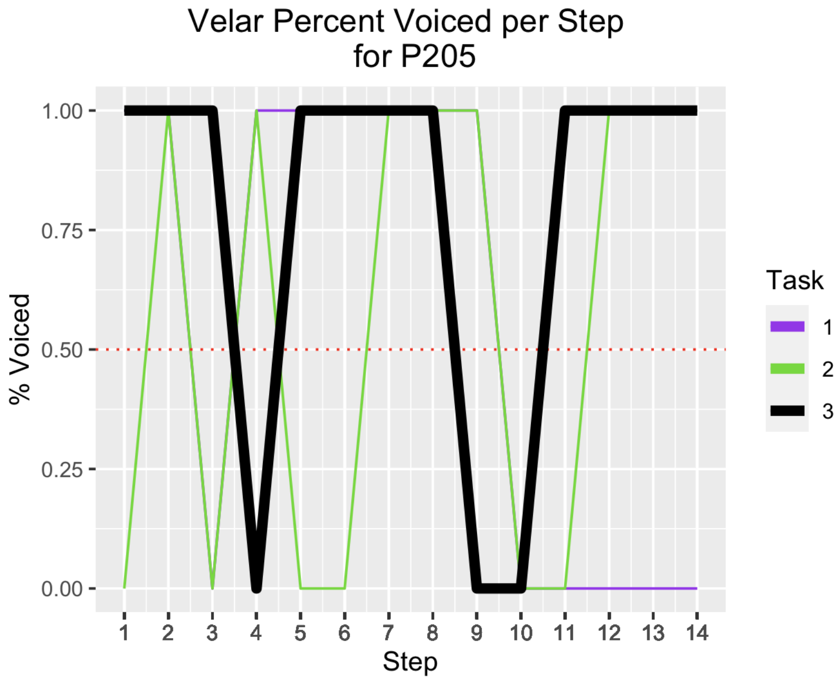


Figure 47. P205 percent velar voiced responses per step per task.

Based on the case studies outlined above, it seems as though the same ease of comprehension order that was discovered in Experiment 1 follows for Experiment 2: bilabial > dental > velar; the case study analysis further supports the group trends we previously reported. To briefly summarize, P202 seemed very consistent in her perception task #3 performance in that she showed perfect inverse S-curves in categorizing the bilabial and dental stimuli, and had only one deviation from the S-curve in her velar data; that being said, her bilabial boundary was shifted to the left, her velar boundary was shifted to the right, and only her dental boundary was in the appropriate location. P208 appeared to be guessing across all three perception tasks and on the stimuli presented across all three places of articulation; they seemed to follow the same ease of comprehension as the rest of the group, but with less consistency across the board. P204 showed some improvement from perception task #1 to perception task #2, at least with respect to the bilabial and dental stimuli; it seemed these improvements persisted into perception task #3. Lastly, P205 improved from perception task #1 to perception task #2 with respect to aligning her bilabial and dental responses with the predicted inverse S-curve; she showed even further improvement by perception task #3 with respect to shifting her boundary leftward, in the Spanish-like direction, for these places of articulation, while maintaining the inverse S-curve. P202 and P205 were the only two participants that showed perfect inverse S-curves for the bilabial and dental stimuli in perception task #3; P202 was the only participant that showed accurate Spanish-like categorization (crossover point at 0 ms) and she only showed this for the dental stimuli. Interestingly, P202 and P205 had both our highest and lowest AQ scores in Experiment 2, respectively. Thus, it does not seem as though there is a neurodivergent advantage with respect to performance on perception tasks delayed four to seven days post-exposure to a native speaker.

3.3.2. Production Results

The graph below summarizes all five participants' production data from Experiment 2. Each bar represents either a single participant's mean VOT value or the group's mean VOT value on the designated task, by which the bars are grouped (Figure 48).

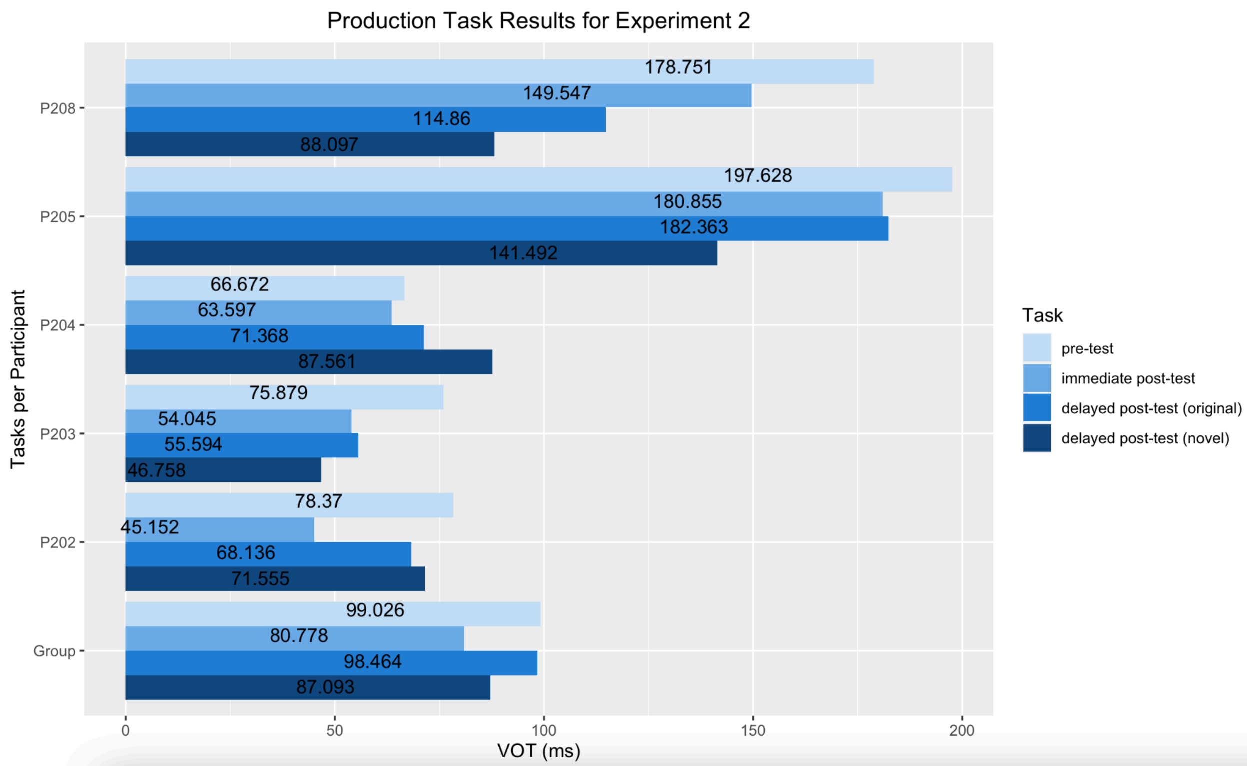


Figure 48. Summarized production task results.

As evidenced by the graph above, of the five participants that completed Experiment 2, two of the participants had consistently higher VOT values than the group, two of the participants had consistently lower VOT values compared to the group, and one participant had lower VOT values than the group on all of the tasks, except when producing the novel target words (Figure 48). The two participants with mean VOT values that were consistently higher than the group, P208 and P205, had AQ scores of 31 and 13, respectively. The two participants with mean VOT values that were consistently lower than the group, P202 and P203, had AQ scores of 43 and 24, respectively. Thus, AQ score was not predictive of participants having more or less Spanish-like VOT values compared to the group. Of the five participants, four participants' (P202, P208, P203, P205) mean delayed post-test VOT values remained lower than their Spanish pre-test mean VOT value, whether they were producing original or novel target words. These same four participants showed the ability to generalize the reduced VOT necessary when producing Spanish words to novel target words; interestingly, three of these participants (P208, P203, P205) reduced their VOT values much more when producing the novel target words compared to the original target words. Only one participant's (P208) mean VOT value on the delayed post-test, including both original and novel target words, was lower than their mean VOT value on the immediate post-test.

Transitioning into the case study results, the tables below compare each participant's delayed post-test VOT values to their pre-test and immediate post-test VOT values in Experiment 1. The delayed post-test VOT values are separated out by the type of stimuli presented, whether it was novel or from the original list of stimuli presented in Experiment 1. The far right column of each table shows the group average VOT values, per task, for comparison sake. Although we have included the mean VOT values per place of articulation, we will focus on the mean VOT values per task in our analysis since place of articulation did not prove to be a meaningful comparison in Experiment 1.

Task	Place of Articulation	PoA Mean VOT	Mean VOT/Task	Group Mean VOT/Task
Spanish pre-test	/p/	51.242	78.370	99.026
Spanish pre-test	/t/	76.160		
Spanish pre-test	/k/	107.707		
Spanish immediate post-test	/p/	47.795	45.152	80.778
Spanish immediate post-test	/t/	28.850		
Spanish immediate post-test	/k/	58.810		
Spanish delayed	/p/	64.130	68.136	98.464

post-test (original)				
Spanish delayed post-test (original)	/t/	49.201		
Spanish delayed post-test (original)	/k/	91.077		
Spanish delayed post-test (novel)	/p/	79.143	71.555	87.093
Spanish delayed post-test (novel)	/t/	69.757		
Spanish delayed post-test (novel)	/k/	65.765		

Table 15. Production task results for P202.

Compared to the group average, P202 started out with a low baseline mean VOT value on the Spanish pre-test (Table 15). Upon exposure to the model speaker's speech, P202's mean VOT value dropped even further, widening the gap between her mean VOT value and the group's mean VOT value on the immediate post-test. While we see an increase in P202's mean VOT value after the four to seven day delay, her Spanish delayed post-test mean VOT values are lower than her Spanish pre-test mean VOT values by 10.234 ms and 6.815 ms on the original and novel target words, respectively. Thus, P202's learned VOT reduction in Experiment 1 persisted in the long-term, in Experiment 2. P202 reduced her VOT slightly more when producing the original target words than the novel target words. It is unclear if this slight difference is meaningful; it could suggest that P202 did not generalize the necessary VOT reduction entirely to the novel words. Across all four tasks, P202's mean VOT values remained below the group's average VOT values on these tasks.

Task	Place of Articulation	PoA Mean VOT	Mean VOT/Task	Group Mean VOT/Task
Spanish pre-test	/p/	206.162	178.751	99.026

Spanish pre-test	/t/	146.269		
Spanish pre-test	/k/	183.821		
Spanish immediate post-test	/p/	113.745	149.547	80.778
Spanish immediate post-test	/t/	185.517		
Spanish immediate post-test	/k/	149.378		
Spanish delayed post-test (original)	/p/	98.439	114.860	98.464
Spanish delayed post-test (original)	/t/	145.935		
Spanish delayed post-test (original)	/k/	100.205		
Spanish delayed post-test (novel)	/p/	92.902	88.097	87.093
Spanish delayed post-test (novel)	/t/	72.638		
Spanish delayed post-test (novel)	/k/	98.751		

Table 16. Production task results for P208.

P208's mean VOT values were consistently higher than the group's mean VOT values (Table 16). That being said, we do see a consistent downwards trend in their VOT values across tasks. P208's baseline mean VOT value was 79.725 ms greater than the group's mean VOT value on the pre-test. By the immediate post-test, following exposure to the model speaker's speech, P208 reduced their mean VOT value by 29.204 ms. Impressively, P208's mean VOT values in the delayed post-test dropped lower than both their pre-test and immediate post-test mean VOT values, when producing both the original and novel target words. Perhaps P208 noticed the decreased VOT produced by the native speaker, but took longer to accommodate this change in their own speech. Even more interestingly, P208's mean VOT value was at its lowest and just over 1 ms greater than the group's mean VOT value when producing the novel target words in the delayed post-test, suggesting that they not only generalized the necessary VOT reduction to the novel words, but were able to reduce their VOT to be even more Spanish-like despite never hearing these novel words produced by a native speaker. It is possible this occurred by chance. Perhaps P208 had a more difficult time applying what they learned about Spanish sounds to words that they had already produced, but could more readily apply the necessary VOT reduction to words that they had yet to produce. We noted that P208 produced /pamo/ instead of /pomo/ for one of the tokens in the production task during the second experiment.

Task	**Place of Articulation**	**PoA Mean VOT**	**Mean VOT/Task**	**Group Mean VOT/Task**
Spanish pre-test	/p/	52.143	66.672	99.026
Spanish pre-test	/t/	58.648		
Spanish pre-test	/k/	89.224		
Spanish immediate post-test	/p/	48.124	63.597	80.778
Spanish immediate post-test	/t/	56.329		
Spanish immediate post-test	/k/	86.338		
Spanish delayed post-test (original)	/p/	56.376	71.368	98.464
Spanish delayed post-test (original)	/t/	80.119		
Spanish delayed post-test (original)	/k/	77.609		

Spanish delayed post-test (novel)	/p/	86.141		
Spanish delayed post-test (novel)	/t/	82.327	87.561	87.093
Spanish delayed post-test (novel)	/k/	94.217		

Table 17. Production task results for P204.

P204's mean VOT values fell below the group averages for all of the tasks except when producing the novel target words in the delayed post-test (Table 17). Compared to the group, P204's baseline VOT values were well below the group's mean VOT values on the pre-test, as he started with a pre-test mean VOT value that was 32.354 ms lower than the group's average. This trend persisted in the immediate post-test, where his mean VOT value was still 17.181 ms below the group's mean VOT value on this task. After exposure to the model speaker's speech, P204's mean VOT value decreased slightly. By the delayed post-test, P204's mean VOT value remained lower than the group's mean VOT value, but it increased slightly from both his Spanish pre-test and immediate post-test mean VOT values. Interestingly, P204's mean VOT value was at its highest when producing the novel target words in the delayed post-test, with his mean VOT value jumping 20.889 ms higher than his baseline mean VOT value. Because P204's mean VOT value when producing the novel target words jumped above his Spanish pre-test VOT values, it is possible he was thrown off by the introduction of these novel words and, therefore, hesitated slightly in producing these words. P204 did not alter his VOT very much throughout the pre-test and immediate post-test tasks, and it seems as though P204 did not generalize any learned VOT reduction to the novel target words either.

Task	Place of Articulation	PoA Mean VOT	Mean VOT/Task	Group Mean VOT/Task
Spanish pre-test	/p/	67.889	75.879	99.026

Spanish pre-test	/t/	76.564		
Spanish pre-test	/k/	83.185		
Spanish immediate post-test	/p/	49.492	54.045	80.778
Spanish immediate post-test	/t/	50.701		
Spanish immediate post-test	/k/	61.942		
Spanish delayed post-test (original)	/p/	38.199	55.594	98.464
Spanish delayed post-test (original)	/t/	64.433		
Spanish delayed post-test (original)	/k/	64.151		
Spanish delayed post-test (novel)	/p/	34.982	46.758	87.093
Spanish delayed post-test (novel)	/t/	40.516		
Spanish delayed post-test (novel)	/k/	64.776		

Table 18. Production task results for P203.

P203 also started with a mean VOT value on the Spanish pre-test that was lower than the group's mean VOT value on this task (Table 18). Upon exposure to the model speaker's speech, P203 reduced his VOT value appropriately, such that he maintained this gap between his mean VOT value on the Spanish post-test and the group's mean VOT value on this task. Between the Spanish pre-test and Spanish immediate post-test, P203's mean VOT value dropped 21.834 ms, which is impressive given his Spanish pre-test mean VOT value was already 23.147 ms lower than the group's average. Even more impressively, P203's mean VOT value on the Spanish delayed post-test remained very low at 55.495 ms, only a 1.549 ms increase from his mean VOT value on the Spanish immediate post-test, for the original target words. Interestingly, P203 reduced his VOT even further when producing the novel target words presented in the Spanish delayed post-test. Compared to his mean VOT value on the Spanish post-test, P203's mean VOT value when producing the novel target words in the Spanish delayed post-test decreased by 7.287 ms, suggesting that he did successfully generalize the need to reduce his VOT, when producing Spanish words, to novel words. Similar to P208, P203 also showed a greater VOT reduction when producing the novel target words than the original target words. Again, this may have occurred by chance or it may have been easier for P203 to apply the necessary VOT changes to words that he had never produced before. Across all four tasks, P203's mean VOT values were well below the group's mean VOT values.

Task	Place of Articulation	PoA Mean VOT	Mean VOT/Task	Group Mean VOT/Task
Spanish pre-test	/p/	202.085	197.628	99.026
Spanish pre-test	/t/	191.078		
Spanish pre-test	/k/	199.719		
Spanish immediate post-test	/p/	176.973	180.855	80.778
Spanish immediate post-test	/t/	146.278		
Spanish immediate post-test	/k/	219.314		
Spanish delayed post-test (original)	/p/	217.809	182.363	98.464
Spanish delayed post-test (original)	/t/	182.606		

Spanish delayed post-test (original)	/k/	146.674		
Spanish delayed post-test (novel)	/p/	158.141	141.492	87.093
Spanish delayed post-test (novel)	/t/	135.984		
Spanish delayed post-test (novel)	/k/	130.352		

Table 19. Production task results for P205.

P205 consistently demonstrated a much higher mean VOT value than the group, across all four tasks presented here (Table 19). Her baseline mean VOT value started out 98.602 ms higher than the group. While her mean VOT value decreased between the pre-test and immediate post-test, it still remained much higher than the group's mean VOT value, sitting nearly 100 ms higher than the group average. By the delayed post-test, P205's mean VOT value remained lower than her baseline VOT value, but increased by 1.508 ms compared to her immediate post-test mean VOT value. Interestingly, P205's mean VOT value was at its lowest when producing the novel target words presented in the delayed post-test; her mean VOT value was 56.136 ms lower for the delayed post-test novel target words compared to the pre-test. Like P208 and P203, we speculate this extreme generalization may have occurred because P205 noticed the need to reduce her VOT value when producing Spanish and found this easier to do when producing words she had not seen nor produced previously. Again, we'd like to point out that P205 hesitated in producing the Spanish sounds at times, resulting in the occasional exaggerated aspiration. P205 also produced two tokens with an /a/, when the token had /o/ as the following vowel, orthographically.

3.4. Discussion

The goal of Experiment 2 was to answer the research questions presented in Section 1.3 from the perspective of long-term learning. The two questions we will focus on in this discussion include:

3. Do autistic-like traits predict long-term L2 phonetic learning?
    b. Is long-term L2 phonetic learning generalizable to new words?
4. If long-term learning is predicted by autistic-like traits, is it due to enhanced perceptual abilities?

We will summarize the findings from Experiment 2 in an effort to answer the questions outlined above.

In order to address question (3), we looked at our participants' mean VOT values prior to exposure to the model speaker's speech, as well as four to seven days after this exposure. Given the small sample size in Experiment 2, we will summarize all five participants' performance on the pre-test, immediate post-test, and delayed post-test. P202's mean VOT values on the delayed post-test were lower, across both the original and novel target words, than her pre-test mean VOT values, but not lower than her immediate post-test mean VOT values. Both P203's and P205's delayed post-test mean VOT values were lower than their pre-test mean VOT values, but not lower, rather very similar to their immediate post-test mean VOT values, for the original target words, and lower than their immediate post-test mean VOT values for the novel target words. P204's delayed post-test mean VOT value, across both token types, was higher than his pre-test and post-test mean VOT values. P208's mean VOT values on the delayed post-test were lower, across both the original and novel target words, than both their pre-test and post-test mean VOT values. Thus, four out of the five participants for Experiment 2 showed a VOT reduction in the long-term, comparing their pre-test and delayed post-test VOT values. This suggests that, at least in this case study, long-term learning did occur for four of the five participants. If we refer back to Table 13 (Section 3.2.2), we see that our participants with the second highest AQ score, and the only participant with a co-occurring ADHD diagnosis, P208 (AQ score of 31), was the only participant that consistently decreased their VOT values from the pre-test to the immediate post-test to the delayed post-test. P202, P203, and P205 all showed that they learned to reduced their VOT values in the delayed post-test compared to their baseline VOT values, but they did not consistently reduce their VOT values beyond the reduction that occurred between the pre-test and immediate post-test; these participants had the highest and lowest AQ scores of 43, 24, and 13, respectively. P204 was the only participant that did not demonstrate any long-term learning of the appropriate VOT reduction, and he had the mid-range AQ score of 27. P204 had the lowest attention to detail subscale score of five, but both P205 and P208 had scores of six on the same subscale and they demonstrated long-term L2 phonetic learning.

Thus, to answer question (3), it does not seem as though autistic-like traits are predictive of long-term learning, as both our highest-scoring and lowest-scoring participants demonstrated long-term learning of the appropriate VOT reduction in an effort to sound more native-like. That being said, we may have discovered a different advantage of having more autistic-like traits: the participant with the second most autistic-like traits continued to improve their production across the tasks, even after just one exposure to the model speaker. It is surprising P208 was capable of improving further, even without additional exposure to the model speaker's speech. P208 had the second highest attention switching subscale score of eight, though this does not seem to offer them an advantage since P203 has the highest attention switching subscale score of 10 and did not experience the same advantage. It is possible this is more related to their high initial VOT value and growing familiarity with the Spanish words, as opposed to any linguistic advantage. We predicted those with more autistic-like traits would be able to demonstrate long-term learning of the reduced VOT values; thus, we expected a performance similar to that of P202, P203, and P205, in which their delayed post-test and post-test VOT values were very similar, and reduced in comparison to their pre-test VOT values.

To answer part (a) of question (3), we specifically looked at our participants' performance when producing both the original target words and the novel target words, after the four to seven day delay. We wanted to evaluate whether or not our participants generalized their learned VOT to the novel target words. As mentioned above, all four of the participants that reduced their delayed post-test VOT values compared to their pre-test VOT values did so for both the original and novel target words presented in Experiment 2. Furthermore, P203, P205, and P208 actually reduced their VOT values even further when producing the novel target words compared to the original target words in the last production task. It is possible this extreme generalization occurred as a result of these participants recognizing the need to reduce their VOT when producing Spanish words, but they were more readily able to accommodate this change in their speech when producing words they had not seen nor produced, repeatedly, previously. Although P204 did not reduce his delayed post-test VOT compared to his pre-test VOT, it is interesting to note that his highest mean VOT value was for the novel target words presented in the delayed post-test, which had an average VOT value that was 20.889 ms higher than his pre-test mean VOT value.

The results from this case study show that, of the participants that demonstrated long-term learning abilities, all four of them generalized their learned VOT values to the novel target words and, thus, relied on procedural, or implicit, memory in completing the delayed post-test. This is the opposite finding from that of Mielke et al. (2013), and does not match our prediction. Although P203, P205, and P208 showed the ability to reduce their VOT values even further, in the direction of sounding more native-like, for the novel target words in the delayed post-test, this could simply be a result of their increased familiarity with Spanish words and sounds by the delayed post-test. These participants have a range of AQ scores, so we do not attribute this greater generalization ability to the number of autistic-like traits they possess. It is also important to note that both P205 and P208 had incredibly high baseline mean VOT values, so there was a greater need to reduce their VOT values; even their lowest mean VOT values, for the novel target words, were higher than the other three participants' baseline VOT values.

Lastly, we will address the final question that Experiment 2 aimed to answer: question (4). First, we know that long-term learning is not predicted by autistic-like traits, as outlined above. That being said, we will still provide a summary of the perception results from our case study to better understand the perceptual abilities of the participants.

Beginning with P202, we found that her performance on the delayed perception task was similar to her performance on the second perception task in Experiment 1, with respect to the bilabial stimuli. Her bilabial perceptual boundary improved after exposure to the model speaker, but it seems as though she over-accommodated the necessary shift in her perceptual boundary between the second and third tasks. P202's dental perceptual boundary improved by the third perception task, as compared to the first and second tasks, showing her dental perceptual skills remained strong even after the delay. While P202's velar perceptual boundary was the best in the first task, the transition between her voiced and voiceless responses stayed consistent across all three tasks. Thus, we can conclude that P202 not only benefited from exposure to the model speaker in the short-term, but this persisted in the long-term where she showed both improved and consistent perceptual abilities. As a group, the velar stimuli were challenging to decipher, with P202 being the most successful at categorizing the velar stop consonants. This is particularly interesting given P202 having the highest AQ score of all the participants in Experiment 2. P208 showed inconsistency in their classification of the bilabial stop consonants across the second and third perception tasks, but the slope of their boundary shifted appropriately by the third task, showing improvement in the long-term. Their dental perceptual boundary improved by the second task, and this improvement persisted into the third task, with a nearly perfect inverse S-shape and an appropriately shifted slope. P204's dental perceptual abilities seemed to improve across all three tasks, with the slope shifting appropriately by the third task. P204's performance on the velar tasks was inconsistent across all three tasks. For both the bilabial and dental stimuli, P204 seemed to benefit from exposure to the model speaker's speech and, again, this persisted in the long-term. While we cannot comment on P203's long-term perceptual abilities, he did not show any particular perceptual strength in Experiment 1. P204's bilabial perceptual boundary improved from the first to second task, and remained strong with respect to the slope, in the long-term. P205 showed the expected inverse S-shape for both the bilabial and dental stimuli in the third perception task, and even shifted the boundary appropriately to the left from the second to the third task. Thus, P205's perceptual abilities seemed to improve upon exposure to the model speaker and into the long-term. Similar to P204, P205's velar categorization was inconsistent across all three perception tasks. Again, P205's velar perception task performance was inconsistent and uninterpretable, like that of P208 and P204.

Thus, the case study analysis and group analysis allowed us to confirm that all four of the participants whose perception data we successfully gathered in Experiment 2 showed either retained improvement from exposure to the model speaker's speech or additional improvement in the delayed perception task. It is unclear how participants improved further by perception task #3; perhaps these participants benefited from the repetitions of producing Spanish in the production tasks, which in turn aided in their perceptual abilities, if perception does not in fact precede production, as proposed by the SLM (Flege, 1987a). Regardless of AQ score, all of our participants demonstrated their ability to learn perceptually, both in the short-term and long-term through Experiments 1 and 2. It does not seem as though any one participant stood out as having the most heightened perceptual abilities across the board, but our highest-scoring participant, P202, did perform the best in categorizing the velar perceptual stimuli. It is possible her heightened number of autistic-like traits gave her this strength, though her performance on the perception task overall did not seem particularly exceptional compared to the other participants.

In conclusion, Experiment 2 allowed us to address the remaining research questions outlined in Section 1.3, from the perspective of long-term learning. The results of the present case study showed that autistic-like traits do not predict long-term learning ability, rather most of our participants, regardless of diagnosis or AQ score, demonstrated long-term learning of the fine-grained phonetic details of Spanish. Of the four participants that demonstrated long-term learning ability, all four of them generalized the need to reduce their VOT when producing Spanish words to novel words; thus, these participants relied on their procedural memory in completing the delayed post-test, which was unexpected based on the results of Mielke et al. (2013). Long-term learning is not predicted by autistic-like traits, and we did not find any especially enhanced perceptual abilities in our participants, rather consistency and improvement across the board. The biggest takeaway from Experiment 2 is further confirmation that neurodiverse and TD children alike are strong L2 learners, even in the long-term after minimal exposure to the language by the model speaker; thus, we support further L2 teaching and education for *all* children in an effort to expand their communicative outlets.

Chapter 4: Conclusion

4.1. Summary of Findings

The previous three chapters (Chapter 1) introduced the motivation behind and the set up for the present doctoral dissertation before explaining the design and results of both (Chapter 2) Experiment 1 and (Chapter 3) Experiment 2. The present chapter will begin with a summary of the findings across the prior chapters, followed by a discussion of the implications of these findings, remaining limitations and further directions of the present study, and a final conclusion.

To begin the summary of our findings over the course of the present study, we will first review our motivation for conducting this study and designing it in the way that we did. Maenner (2021) acknowledged the increasing prevalence of the autism diagnosis and the need for expanded autism identification and evaluation services across different ethnic, racial, and geographical groups. Expanding on this call to action, the aim of the present study was to better understand how autism presents linguistically across different individuals with such a diagnosis. While past studies have focused on a deficit-based approach in assessing language abilities in autism, the present study prioritized a strength-based approach, utilizing past studies that found autistic children are not at a disadvantage when raised bilingually, as opposed to monolingually, as foundation (Jordaan, 2008; Ohashi et al., 2012; Valicenti-McDermott et al., 2013). We hoped that the results from Experiments 1 and 2 would show the strengths autistic children bring to bilingualism and language learning to further promote second language education and expand the number of communicative outlets available to autistic children. Relying on past studies, we developed the present study in an effort to better understand the L2 phonetic imitation and short-term and long-term L2 phonetic learning abilities of autistic children, with Spanish as the

L2 (Yu et al., 2013). With these goals in mind, the research questions this dissertation aimed to address include:

1. Do autistic-like traits predict L2 phonetic imitation ability?
2. Do autistic-like traits predict short-term L2 phonetic learning?
3. Do autistic-like traits predict long-term L2 phonetic learning?
    a. Is long-term L2 phonetic learning generalizable to new words?
4. If short-term and long-term learning are predicted by autistic-like traits, is it due to enhanced perceptual abilities?

Across all four questions outlined above, we hypothesized that autistic-like traits would be predictive of L2 phonetic imitation ability, short-term L2 phonetic learning, and long-term L2 phonetic learning; our hypothesis was as such because autistic individuals tend to have heightened attention to detail and focus abilities, both of which are measured and confirmed by the AQ test (Baron-Cohen et al., 2001; Ruzich et al., 2015b). We also predicted that heightened detail-orientation would allow autistic individuals to show greater perceptual categorization abilities.

Experiment 1 answered the first two questions, as well as the fourth question, from the perspective of short-term learning. Based on the results (Section 2.3) and discussion (Section 2.4) presented in Chapter 2, we found that autistic-like traits are not predictive of phonetic imitation ability, short-term L2 phonetic learning ability, nor perceptual abilities before and after exposure to a model speaker's speech as demonstrated by our individual and group analyses of the results. That being said, our participants showed a significant VOT reduction after exposure to a model speaker's speech, both immediately after this exposure and in the immediate post-test. Thus, all of our participants, regardless of AQ score, AQ subscale score, or whether or not they had a neurodivergent diagnosis, were strong L2 phonetic imitators and capable learners of the phonetics of an L2 in the short-term. Furthermore, our participants seemed to show improved perceptual abilities when comparing their categorical perception prior to the phonetic imitation task and following this task; this trend was observed irrespective of AQ score or AQ subscale score. There may be a slight neurodivergent advantage with respect to achieving the accurate perceptual boundary post-exposure to a native speaker; we leave this for future exploration.

Experiment 2 answered both parts of the third question, as well as the fourth question, from the perspective of long-term learning, as this experiment took place after a four to seven day delay following Experiment 1. According to the results (Section 3.3) and discussion (Section 3.4) presented in Chapter 3, autistic-like traits are not predictive of long-term L2 phonetic learning, memory used to perform the delayed production tasks, nor long-term perceptual abilities. The case studies discussed in Chapter 3 show that four of the five participants demonstrated long-term L2 phonetic learning, using procedural memory, and they had AQ scores ranging from 13 to 43. In Experiment 2, the participants' perceptual abilities either remained consistent or improved beyond their already-improved performance in the second perception task in Experiment 1; this performance was not related to participants' AQ scores nor AQ subscale scores.

Through both Experiment 1 and Experiment 2, our participants demonstrated the ability to imitate the phonetics of a novel L2, learn these phonetics in both the short-term and long-term, and improve their perceptual abilities in both the short-term and the long-term; all of these trends were observed, regardless of the number of autistic-like traits each participant possessed. While Experiments 1 and 2 were not in line with our initial hypotheses that autistic individuals would show superior strengths across the perception and production tasks, the present study further supports previous findings that support bilingual, or even multilingual, environments for children, autistic or not. The neurodiverse participants proved to be equally as successful at the tasks presented in this study compared to their TD counterparts; thus, autistic and neurodiverse children are as capable L2 learners as their TD peers. The implications of these findings are discussed further in Section 4.2 below.

4.2. Implications

The motivation behind this study relied on filling a gap in the literature: to better understand the linguistic profile of autistic individuals, specifically related to their SLA abilities. We hoped to utilize the results of this study to promote L2 learning in autistic populations, by showing their strengths in SLA compared to their TD peers. Similar to other recent studies, the results of the present study do not show a clear relationship between autistic-like traits and performance on linguistic tasks (Yu, 2021). While the results of the present study do not offer support for an autistic, nor neurodiverse, advantage when acquiring a novel language, the results indicate that neurodiverse children are as capable L2 learners as TD children and should be encouraged to be bilingual, or even multilingual.

Long ago there were concerns that bilingualism would negatively impact one's ability to learn their L1, even for TD children (Darcy, 1953). Researchers later discovered the heightened brain activity and potential cognitive benefits of the bilingual mind and then claimed bilingualism was not a hindrance rather an advantage (Bialystok et al., 2012; Marian & Spivey, 2003; Poarch & Krott, 2019). The present study did not show support for a similar autistic advantage in acquiring an L2; however, given (1) the comparable SLA abilities of autistic and TD children and (2) the many benefits of raising bilingual children, schools and families should support neurodiverse children in learning an L2. Recent studies have shown that parents feel their TD children should learn an L2 because it would afford their child additional career opportunities and competitive advantage in the future and create a sense of belonging to one's family while preserving their family language and culture (Lee et al., 2015; Sims & Ellis, 2014). These same benefits apply to neurodiverse children. While we have focused on many of the strengths of autism throughout this study, we recognize there are several barriers autistic children face in their daily lives as well. By raising autistic children bilingually, or even multilingually, these children will be better set up for their futures. Bilingualism allows individuals to participate in more conversations, both literally and metaphorically, as it opens doors for career opportunities, allows deeper community and cultural connection, and a stronger sense of one's identity. These benefits of bilingualism seem even more significant in the lives of autistic individuals who may struggle with finding the right career path, connecting with others, and understanding themselves.

There are a wide array of intervention methods utilized in autism clinics across the globe, ranging from social stories to Applied Behavior Analysis to music therapy. As far as we know, language is not currently being used as an intervention strategy for autism. Given that autistic children are as strong L2 learners as their peers and many autistic children struggle with communicating and connecting with peers, an L2 classroom is the ideal space to foster these prosocial interactions. L2 classrooms rely on constant communication and conversation practice among peers, while learning the pragmatics of a novel language. In a sense, the L2 classroom levels the playing field for autistic and TD children, as everyone is new to the language and learning it together in a structured setting. L2 learning offers all children an additional communicative outlet, which is socially, and even cognitively, beneficial.

4.3. Limitations and Further Directions

The most obvious limitation of the present study is rooted in the heterogeneity of the autism diagnosis. As we mentioned before, “if you’ve met one person with autism, you’ve met one person with autism” (Shore, 2016). Thus, it is always difficult to offer generalized results in autism research. Furthermore, we recognize the small sample size of the present study as a limitation and an area for improvement in future studies. While the case study approach allowed us to perform a more thorough analysis of five of the participants’ results, both individually and compared to the group, it would be beneficial to repeat this study with more participants, especially for Experiment 2. It would be informative to recruit more participants with both an autism and ADHD diagnosis, as well as more participants with a single one of these diagnoses, to tease out any differences in how these diagnoses present themselves linguistically. We could expand beyond the AQ test to evaluate detail-orientation and focus in other ways, or even focus on testing for ADHD-specific skills. While expanding participant recruitment, it would be beneficial to conduct the same study across the globe, extending beyond Australia, in order to do a cross-cultural comparison. The study could also be repeated with additional languages; we could introduce a novel L2 other than Spanish, or begin with monolingual speakers of a language other than English. Another potential limitation of the present study is the inclusion of participants with exposure to or knowledge of languages other than English. Although these other languages are not voicing languages and, thus, do not have the same voicing contrast as Spanish, it is possible that having knowledge of an L2 affected our participants’ performance; therefore, future studies should intentionally recruit a balanced mix of truly monolingual and multilingual participants to evaluate whether or not a multilingual advantage or disadvantage exists when tasked with completing phonetic imitation and production tasks in a novel language.

More specific to our data, we would like to acknowledge that our attempts to normalize the data were unsuccessful, resulting in a lot of variation in participants' VOT values across the tasks. Perhaps in a larger study, it would be worth excluding any participants that hesitated when speaking or reading the phrases such that their VOT values were beyond a specified limit. We could have excluded all of our younger participants that seemed to struggle more with reading the phrases and sounding out the Spanish words, but we saw similar hesitation across several ages. Lastly, it was evident our velar perceptual stimuli were problematic throughout the experiment, though more so during perception task #2. In repeating and expanding upon the present study, we would consider re-recording the perceptual stimuli and running a screening task to ensure naturalness of the stimuli. We could include additional experimental sessions and exposure tasks to test whether or not the velar stimuli simply require more time and instances of exposure to acquire, given that they are back consonants, acquired later than front consonants (Jakobson, 1968). Studies have shown that individuals, ranging from infants to adults, adapt quickly to speaker variation and are capable of shifting their perceptual categories according to the phonetic input they receive (Bieber & Gordon-Salant, 2024; Kuhl et al., 2003); it is possible our participants needed additional exposure to learn the more difficult velar stimuli.

We utilized VOT as a lens through which we could understand the early steps of acquiring a novel language, beginning with the speech sounds that are then used to form the words and sentences of that language; we recognize, however, that phonetic inventories do not represent language learning as a whole. In the future, it would be informative to incorporate tasks to test how well the participants remember the meaning of the novel words since they were excited to translate the words by relying on the clipart images presented. We used a very simple carrier phrase, as well as words with CVCV structure. It would be interesting to repeat this study with different stimuli, perhaps with consonant clusters, and a longer carrier phrase to understand the relationship between AQ scores and both phonetic imitation and L2 learning when the stimuli are more complex. More immediate next steps could entail analyzing the filler words from the present study.

4.4. Conclusion

Through this doctoral dissertation, we confirmed more recent findings that autistic children are equally as competent learners of novel languages, from a phonetic perspective, as their TD counterparts. The participants across Experiments 1 and 2 demonstrated their strengths in L2 phonetic imitation, short-term L2 phonetic learning, and long-term L2 phonetic learning, with Spanish as the L2. Because of these strengths, and the many benefits of bilingualism, we encourage all children to learn an L2 and we believe schools should promote, and prioritize, SLA. The L2 classroom is the ideal space to host prosocial peer interactions in a natural, productive setting; this opportunity for children to have an additional communicative outlet is beneficial for neurodiverse and TD children, alike.

Appendix A

Autism Quotient Test

For each statement below, choose one response (Definitely Agree, Slightly Agree, Slightly Disagree, Definitely Disagree) that best describes how strongly that statement applies to your child:

1. I prefer to do things with others rather than on my own.
2. I prefer to do things the same way over and over again.
3. If I try to imagine something, I find it very easy to create a picture in my mind.
4. I frequently get so strongly absorbed in one thing that I lose sight of other things.
5. I often notice small sounds when others do not.
6. I usually notice car number plates or similar strings of information.
7. Other people frequently tell me that what I've said is impolite, even though I think it is polite.
8. When I'm reading a story, I can easily imagine what the characters might look like.
9. I am fascinated by dates.
10. In a social group, I can easily keep track of several different people's conversations.
11. I find social situations easy.
12. I tend to notice details that others do not.
13. I would rather go to a library than to a party.
14. I find making up stories easy.
15. I find myself drawn more strongly to people than to things.
16. I tend to have very strong interests, which I get upset about if I can't pursue.
17. I enjoy social chitchat.
18. When I talk, it isn't always easy for others to get a word in edgewise.
19. I am fascinated by numbers.

20. When I’m reading a story, I find it difficult to work out the characters’ intentions.

21. I don’t particularly enjoy reading fiction.

22. I find it hard to make new friends.

23. I notice patterns in things all the time.

24. I would rather go to the theater than to a museum.

25. It does not upset me if my daily routine is disturbed.

26. I frequently find that I don’t know how to keep a conversation going.

27. I find it easy to “read between the lines” when someone is talking to me.

28. I usually concentrate more on the whole picture, rather than on the small details.

29. I am not very good at remembering phone numbers.

30. I don’t usually notice small changes in a situation or a person’s appearance.

31. I know how to tell if someone listening to me is getting bored.

32. I find it easy to do more than one thing at once.

33. When I talk on the phone, I’m not sure when it’s my turn to speak.

34. I enjoy doing things spontaneously.

35. I am often the last to understand the point of a joke.

36. I find it easy to work out what someone is thinking or feeling just by looking at their face.

37. If there is an interruption, I can switch back to what I was doing very quickly.

38. I am good at social chitchat.

39. People often tell me that I keep going on and on about the same thing.

40. When I was young, I used to enjoy playing games involving pretending with other children.

41. I like to collect information about categories of things (e.g., types of cars, birds, trains, plants).

42. I find it difficult to imagine what it would be like to be someone else.

43. I like to carefully plan any activities I participate in.

44. I enjoy social occasions.

45. I find it difficult to work out people's intentions.

46. New situations make me anxious.

47. I enjoy meeting new people.

48. I am a good diplomat.

49. I am not very good at remembering people's date of birth.

50. I find it very easy to play games with children that involve pretending.

Appendix B

How the AQ Test is scored (Baron-Cohen et al., 2001; Buchanan, 2023):

AQ Category	Interpretation	Item Number	Answer that Aligns with Neurodivergent Response
Social Skills	A higher score on this sub-section indicates weaker social skills.	1 11 13 15 22 36 44 45 47 48	Disagree Disagree Disagree Disagree Agree Disagree Disagree Agree Disagree Disagree
Attention switching	A higher score on this sub-section indicates weaker attention switching skills, or heightened focus.	2 4 10 16 25 32 34 37 43 46	Agree Agree Disagree Agree Disagree Disagree Disagree Disagree Agree Agree
Attention to Detail	A higher score on this sub-section indicates stronger attention to detail skills.	5 6 9 12 19 23 28 29 30 49	Agree Agree Agree Agree Agree Agree Disagree Disagree Disagree Disagree
Communication	A higher score on this sub-section indicates weaker communication skills.	7 17 18 26 27 31 33 35 38	Agree Disagree Agree Agree Disagree Disagree Agree Agree Disagree

		39	Agree
Imagination	A higher score on this sub-section indicates weaker imagination skills.	3 8 14 20 21 24 40 41 42 50	Disagree Disagree Disagree Agree Agree Disagree Disagree Agree Agree Disagree

Appendix C

Example of first message sent to clinics:

Hello,

My name is Madie Dunlap and I am a PhD candidate in the Department of Spanish and Portuguese at the University of California, Los Angeles. I am currently working on a voluntary, paid, IRB-approved UCLA research study related to the speech imitation abilities of monolingual, English-speaking children with and without autism and I wanted to reach out to ask if your clinic would be able to support my research efforts by sharing my recruitment message with the parents and families you serve. I have attached a flyer that can be shared electronically (perhaps via a newsletter or mailing list). While the study must be conducted in person, I am also happy to conduct the experiment either at UCLA in the Linguistics Lab on campus, or at your clinic, if that is permitted by you and more convenient for your clients.

I have been studying autism throughout my graduate education and I work with children with autism, as well as a variety of other developmental disorders, each summer at an ABA-based Summer Treatment Program hosted by the University of Washington Autism Center. In exchange for sharing my study details, I would be more than happy to support your clinic as well by volunteering or helping out in any way that I can. Please let me know if you are able to share my study information and how I can be of use!

Best,

Madie Dunlap
mdunlap@ucla.edu

Example of second message sent to clinics who agreed to share the study information:

Hello again,

Thank you so much for agreeing to distribute my recruitment message! I included a message below that can be copied and pasted in an email blast or newsletter - please let me know if you would like me to send the message as a separate Word/PDF file or as a flyer and I would be happy to do so.

UCLA STUDY RECRUITING RESEARCH PARTICIPANTS

Hi there! My name is Madie Dunlap and I am a graduate student in the Department of Spanish and Portuguese at UCLA. I am currently working on a voluntary UCLA research study related to the phonetic imitation abilities of monolingual, English-speaking children and I am reaching out to ask if your child would be interested in participating, given that he or she fits the following criteria:

- Ages 7-12 years old

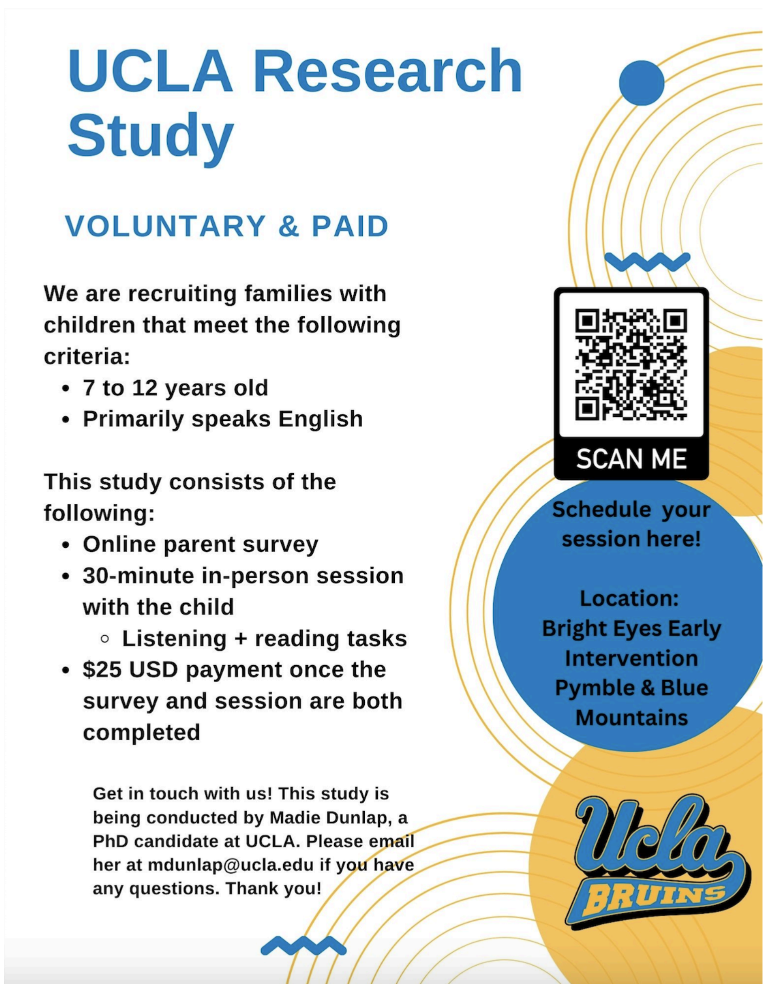
- Speaks English, and primarily English, fluently
- Neurotypical OR diagnosed with autism (but *not* diagnosed with any other neurological, hearing, or language impairments)

By participating in this study, you will be asked to complete a demographic questionnaire and Autism Quotient Test on behalf of your child. Your child will complete two experimental sessions consisting of a variety of English and Spanish production tasks, as well as perception tasks, all of which will be audio-recorded. Together, each parent-child pair will be paid $25 via electronic Amazon gift card after completion of the survey and first experimental session, and $25 via electronic Amazon gift card after completion of the second experimental session. If you are interested in allowing your child to participate in this study, an approximately 45-minute (30 minutes for Session #1 and 15 minutes for Session #2) recorded speech task experiment, **please contact me at mdunlap@ucla.edu.**

Thank you in advance! Please let me know if you have any further questions for me.

Best,

Madie Dunlap
mdunlap@ucla.edu



Appendix D

Demographic and Linguistic Questionnaire

Please answer the following questions about your child to the best of your ability. We ask that you answer each question before moving on to the next one.

1. Gender:
___ Male
___ Female
___ Prefer not to specify
2. Age:
___
3. Has your child ever been diagnosed with autism?
___ Yes
___ No
If yes, please indicate the age at which your child received this diagnosis:
____
You may, optionally, include any other details below:
____
4. Has your child ever been diagnosed with any other psychiatric or motor disabilities?
___ Yes
___ No
If yes, please specify which disabilities below:
____
5. Has your child ever been diagnosed with any speech disabilities or language challenges?
___ Yes
___ No
If yes, please specify which disability below:
____
If yes, please indicate the age at which your child received this diagnosis:
____
6. Does your child have any hearing challenges?
___ Yes
___ No
7. Does your child have any neurological challenges?
___ Yes
___ No
8. Is your child fluent in English?
___ Yes
___ No
9. Is your child fluent in Spanish?
___ Yes
___ No
10. Is your child fluent in any other languages?
___ Yes

___ No
If yes, please specify which language(s) below:

____

11. Has your child been exposed to (not to the point of fluency, but continuously for more than 10% of their awake time per week) any other languages?
___ Yes
___ No
If yes, please specify which language(s) below:

____

If yes, please specify how many years your child has been learning this/these language(s):

____

Appendix E

University of California, Los Angeles

RESEARCH INFORMATION SHEET

Using Autism Quotient Scores to Predict Second Language Phonetic Trait Imitation Ability

INTRODUCTION

Madison Dunlap and Dr. Ji Young Kim, from the Department of Spanish and Portuguese at the University of California, Los Angeles are conducting a research study. Your child was selected as a possible participant in this study because he/she is a monolingual English-speaking child between the ages of 7-12 years old. Your child's participation in this research study is voluntary.

WHAT SHOULD I KNOW ABOUT A RESEARCH STUDY?

- Whether or not you allow your child to take part is up to you.
- You can choose not to allow your child to take part.
- You can agree to allow your child to take part and later change your mind.
- Your decision to allow/not allow your child to participate will not be held against you.
- You can ask any questions before deciding to allow your child to participate.

WHY IS THIS RESEARCH BEING DONE?

This study is designed to learn more about the usefulness of Autism Quotient scores in predicting one's ability to imitate the phonetic traits of a foreign language. The Autism Quotient test is used to assess the number of autism-like traits individuals with and without autism have.

HOW LONG WILL THE RESEARCH LAST AND WHAT WILL I NEED TO DO?

Participation will take a total of about 40-45 minutes.

If you and your child volunteer to participate in this study, the researcher will ask you and your child to do the following:

- First, you will be asked to complete a short questionnaire on behalf of your child

    - This should take about 10 minutes
    - The questions are related to your child's demographics, language experience, and Autism Quotient score
- Next, your child will do the first experimental session with six tasks
    - This should take 20 to 30 minutes
    - Your child will be audio recorded producing both English and Spanish words
    - Your child will also complete two Spanish perception tasks
- Four to seven days later, your child will complete the second experimental session with two tasks
    - This should take about 10 minutes
    - Your child will be audio recorded producing Spanish words
    - Your child will also complete one Spanish perception task
- All of these activities will take place at a school or clinic in which we have permission to collect data or on the UCLA Campus in the Linguistics Lab, depending on which option you and your child choose

Each parent and child pair will receive a $50 electronic gift card as compensation for your participation in the study. You and your child will receive the first $25 electronic gift card following completion of the survey and the first experimental session. You and your child will receive the second $25 electronic gift card after completing the second experimental session.

ARE THERE ANY RISKS IF I PARTICIPATE?

There are not any risks in participating in the research.

ARE THERE ANY BENEFITS IF I PARTICIPATE?

You will not directly benefit from your participation in the research.

The results of the research may contribute to our understanding of how autism and autism-like traits affect second language learning.

HOW WILL INFORMATION ABOUT ME AND MY PARTICIPATION BE KEPT CONFIDENTIAL?

The researchers will do their best to make sure that your private information is kept confidential. Information about you will be handled as confidentially as possible, but participating in research may involve a loss of privacy and the potential for a breach in confidentiality. Study data will be physically and electronically secured. As with any use of electronic means to store data, there is a risk of breach of data security.

Use of personal information that can identify you:

Your name will not be linked to your data.

How information about you will be stored:

All voice recordings will be stored on a password-protected laptop.

People and agencies that will have access to your information: The research team and authorized UCLA personnel may have access to study data and records to monitor the study. Research records provided to authorized, non-UCLA personnel will not contain identifiable information about you. Publications and/or presentations that result from this study will not identify you by name.

Employees of the University may have access to identifiable information as part of routine processing of your information, such as lab work or processing payment. However, University employees are bound by strict rules of confidentiality.

USE OF DATA FOR FUTURE RESEARCH

Your data, including de-identified data may be kept for use in future research.

WHO CAN I CONTACT IF I HAVE QUESTIONS ABOUT THIS STUDY?

The research team:

If you have any questions, comments or concerns about the research, you can talk to one of the researchers. Please contact: Madison Dunlap at mdunlap@ucla.edu or Dr. Ji Young Kim at jiyoungkim@ucla.edu.

UCLA Office of the Human Research Protection Program (OHRPP):

If you have questions about your rights as a research subject, or you have concerns or suggestions and you want to talk to someone other than the researchers, you may contact the UCLA OHRPP by phone: (310) 206-2040; by email:

participants@research.ucla.edu or by mail: Box 951406, Los Angeles, CA 90095-1406.

WHAT ARE MY RIGHTS IF I TAKE PART IN THIS STUDY?

You can choose whether or not you want to be in this study, and you may withdraw your consent and discontinue participation at any time.

Whatever decision you make, there will be no penalty to you, and no loss of benefits to which you were otherwise entitled.

You may refuse to answer any questions that you do not want to answer and still remain in the study.

You will be given a copy of this information to keep for your records.

By clicking the arrow, you are acknowledging the above information and you are agreeing to participate and to allow your child to participate in this study.

Appendix F

UNIVERSITY OF CALIFORNIA LOS ANGELES

ASSENT TO PARTICIPATE IN RESEARCH

Using Autism Quotient Scores to Predict Second Language Phonetic Trait Imitation Ability

1. My name is Madie Dunlap and I am a graduate student researcher at the University of California, Los Angeles.

2. We are asking you to take part in a research study because we are trying to learn more about how well different kids are able to imitate the sounds of a language that they don't speak.

3. If you agree to be in this study, your parent will fill out a survey that asks questions about you, your language background, and different traits that you do/do not have. Then, you'll complete two experimental sessions with several speech tasks, all of which will be recorded via microphone. The activities include reading tasks, a task in which you will repeat the words that you hear, and tasks where you will select the option you heard afterwards. These tasks will take about 30-45 minutes total to complete, across the two sessions.

4. There are not any risks involved in participating.

5. If you choose to participate in this study, you will help us learn more about how well kids with different traits are able to imitate the sounds of a language that is new to them. There are no other benefits in participating in this study.

6. Each child and parent pair will receive a $25 electronic gift card after completing the survey and first experimental session, and a second $25 electronic gift card after completing the second experimental session. Your parent will receive the gift card via email.

7. Please talk this over with your parents before you decide whether or not to participate. We will also ask your parents to give their permission for you to take part in this study. But even if your parents say "yes" you can still decide not to do this.

8. If you don't want to be in this study, you don't have to participate. Remember, being in this study is up to you and no one will be upset if you don't want to participate or even if you change your mind later and want to stop.

9. You can ask any questions that you have about the study. If you have a question later that you didn't think of now, you can call me at 530-519-2866 or email me at mdunlap@ucla.edu. You can also ask your parents to contact me at any time.

10. By continuing with the experiment and completing the tasks, you are agreeing to participate in this study.

Appendix G

Experiment 1 - Code for Perception Data Graphs

Individualized Analysis - Comparing Performance across Tasks per AQ Score

For the individualized analysis, we've included the code for creating the graphs for two of the participants, per place of articulation, as an example. The remaining graphs can be recreated using the same code, but changing out the participant number to match the desired participant.

```
data = read.csv(file.choose(), header = T)
summary(data)
data = subset(data, Task != "3") #Experiment 1 only

data.bilabial = subset(data, POA == "bilabial")
data.dental = subset(data, POA == "dental")
data.velar = subset(data, POA == "velar")
library(dplyr)
library(ggplot2)

new.bilabial = group_by(data.bilabial, Participant)
mean.new.bilabial = new.bilabial %>% group_by(Step, Task, Participant) %>%
summarize(mean = mean(Voiced))

ggplot(subset(mean.new.bilabial, Participant == "202"), aes(x=Step, y=mean, group=Task,
colour=as.factor(Task))) + geom_line() + labs(y = "% Voiced") + scale_x_continuous(breaks =
mean.new.bilabial$Step) + ggtitle("Bilabial Percent Voiced per Step \n for P202") +
theme(plot.title = element_text(hjust = 0.5)) + geom_hline(yintercept = 0.50, colour = "red",
linetype = 3) + scale_color_manual(values=c("#A020F0","#53db00")) +
scale_linetype_manual(values=c("longdash", "solid")) + labs(color  = "Task", linetype = "Task",
shape = "Task")

ggplot(subset(mean.new.bilabial, Participant == "203"), aes(x=Step, y=mean, group=Task,
colour=as.factor(Task))) + geom_line() + labs(y = "% Voiced") + scale_x_continuous(breaks =
mean.new.bilabial$Step) + ggtitle("Bilabial Percent Voiced per Step \n for P203") +
theme(plot.title = element_text(hjust = 0.5)) + geom_hline(yintercept = 0.50, colour = "red",
linetype = 3) + scale_color_manual(values=c("#A020F0","#53db00")) +
scale_linetype_manual(values=c("longdash", "solid")) + labs(color  = "Task", linetype = "Task",
shape = "Task")

new.dental = group_by(data.dental, Participant)
```

```
mean.new.dental = new.dental %>% group_by(Step, Task, Participant) %>% summarize(mean =
mean(Voiced))

ggplot(subset(mean.new.dental, Participant == "202"), aes(x=Step, y=mean, group=Task,
colour=as.factor(Task))) + geom_line() + labs(y = "% Voiced") + scale_x_continuous(breaks =
mean.new.dental$Step) + ggtitle("Dental Percent Voiced per Step \n for P202") + theme(plot.title
= element_text(hjust = 0.5)) + geom_hline(yintercept = 0.50, colour = "red", linetype = 3) +
scale_color_manual(values=c("#A020F0","#53db00")) +
scale_linetype_manual(values=c("longdash", "solid")) + labs(color  = "Task", linetype = "Task",
shape = "Task")

ggplot(subset(mean.new.dental, Participant == "203"), aes(x=Step, y=mean, group=Task,
colour=as.factor(Task))) + geom_line() + labs(y = "% Voiced") + scale_x_continuous(breaks =
mean.new.dental$Step) + ggtitle("Dental Percent Voiced per Step \n for P203") + theme(plot.title
= element_text(hjust = 0.5)) + geom_hline(yintercept = 0.50, colour = "red", linetype = 3) +
scale_color_manual(values=c("#A020F0","#53db00")) +
scale_linetype_manual(values=c("longdash", "solid")) + labs(color  = "Task", linetype = "Task",
shape = "Task")

new.velar = group_by(data.velar, Participant)
mean.new.velar = new.velar %>% group_by(Step, Task, Participant) %>% summarize(mean =
mean(Voiced))

ggplot(subset(mean.new.velar, Participant == "202"), aes(x=Step, y=mean, group=Task,
colour=as.factor(Task))) + geom_line() + labs(y = "% Voiced") + scale_x_continuous(breaks =
mean.new.velar$Step) + ggtitle("Velar Percent Voiced per Step \n for P202") + theme(plot.title =
element_text(hjust = 0.5)) + geom_hline(yintercept = 0.50, colour = "red", linetype = 3) +
scale_color_manual(values=c("#A020F0","#53db00")) +
scale_linetype_manual(values=c("longdash", "solid")) + labs(color  = "Task", linetype = "Task",
shape = "Task")

ggplot(subset(mean.new.velar, Participant == "203"), aes(x=Step, y=mean, group=Task,
colour=as.factor(Task))) + geom_line() + labs(y = "% Voiced") + scale_x_continuous(breaks =
mean.new.velar$Step) + ggtitle("Velar Percent Voiced per Step \n for P203") + theme(plot.title =
element_text(hjust = 0.5)) + geom_hline(yintercept = 0.50, colour = "red", linetype = 3) +
scale_color_manual(values=c("#A020F0","#53db00")) +
scale_linetype_manual(values=c("longdash", "solid")) + labs(color  = "Task", linetype = "Task",
shape = "Task")
```

Case Study Analysis - Including Task #3

```
data = read.csv(file.choose(), header = T)
summary(data)

data.bilabial = subset(data, POA == "bilabial")
data.dental = subset(data, POA == "dental")
data.velar = subset(data, POA == "velar")
library(dplyr)
library(ggplot2)

new.bilabial = group_by(data.bilabial, Participant)
mean.new.bilabial = new.bilabial %>% group_by(Step, Task, Participant) %>%
summarize(mean = mean(Voiced))

ggplot(subset(mean.new.bilabial, Participant == "202"), aes(x=Step, y=mean, group=Task,
colour=as.factor(Task))) + geom_line() + labs(y = "% Voiced") + scale_x_continuous(breaks =
mean.new.bilabial$Step) + ggtitle("Bilabial Percent Voiced per Step \n for P202") +
theme(plot.title = element_text(hjust = 0.5)) + geom_hline(yintercept = 0.50, colour = "red",
linetype = 3) + scale_color_manual(values=c("#A020F0","#53db00","#000000")) +
scale_linetype_manual(values=c("longdash", "solid", "solid")) + labs(color  = "Task", linetype =
"Task", shape = "Task") + geom_line(data = filter(mean.new.bilabial, Participant == "202", Task
== "3"), size = 1)

new.dental = group_by(data.dental, Participant)
mean.new.dental = new.dental %>% group_by(Step, Task, Participant) %>% summarize(mean =
mean(Voiced))

ggplot(subset(mean.new.dental, Participant == "202"), aes(x=Step, y=mean, group=Task,
colour=as.factor(Task))) + geom_line() + labs(y = "% Voiced") + scale_x_continuous(breaks =
mean.new.dental$Step) + ggtitle("Dental Percent Voiced per Step \n for P202") + theme(plot.title
= element_text(hjust = 0.5)) + geom_hline(yintercept = 0.50, colour = "red", linetype = 3) +
scale_color_manual(values=c("#A020F0","#53db00","#000000")) +
scale_linetype_manual(values=c("longdash", "solid", "solid")) + labs(color  = "Task", linetype =
"Task", shape = "Task") + geom_line(data = filter(mean.new.dental, Participant == "202", Task
== "3"), size = 1)

new.velar = group_by(data.velar, Participant)
mean.new.velar = new.velar %>% group_by(Step, Task, Participant) %>% summarize(mean =
mean(Voiced))
```

```
ggplot(subset(mean.new.velar, Participant == "202"), aes(x=Step, y=mean, group=Task,
colour=as.factor(Task))) + geom_line() + labs(y = "% Voiced") + scale_x_continuous(breaks =
mean.new.velar$Step) + ggtitle("Velar Percent Voiced per Step \n for P202") + theme(plot.title =
element_text(hjust = 0.5)) + geom_hline(yintercept = 0.50, colour = "red", linetype = 3) +
scale_color_manual(values=c("#A020F0","#53db00","#000000")) +
scale_linetype_manual(values=c("longdash", "solid", "solid")) + labs(color  = "Task", linetype =
"Task", shape = "Task") + geom_line(data = filter(mean.new.velar, Participant == "202", Task ==
"3"), size = 1)
```

Group Analysis - Task Effect

```
data = read.csv(file.choose(), header = T)
summary(data)

data = subset(data, Task != "3")
data.bilabial = subset(data, POA == "bilabial")
data.dental = subset(data, POA == "dental")
data.velar = subset(data, POA == "velar")

install.packages("dplyr")
library(dplyr)
mean.voiced.bilabial = data.bilabial %>% group_by(Step, Task) %>% summarize(mean =
mean(Voiced))

mean.voiced.dental = data.dental %>% group_by(Step, Task) %>% summarize(mean =
mean(Voiced))

mean.voiced.velar = data.velar %>% group_by(Step, Task) %>% summarize(mean =
mean(Voiced))

library(ggplot2)

ggplot(mean.voiced.bilabial, aes(x=Step, y=mean, group = factor(Task), colour = factor(Task),
linetype = factor(Task))) +  geom_line() + labs(y = "% Voiced") + scale_x_continuous(breaks =
mean.voiced.bilabial$Step) + ggtitle("Bilabial Percent Voiced per Step") + theme(plot.title =
element_text(hjust = 0.5)) + geom_hline(yintercept = 0.50, colour = "red", linetype = 3) +
scale_color_manual(values=c("#A020F0","#53db00")) +
scale_linetype_manual(values=c("longdash", "solid")) + labs(color  = "Task", linetype = "Task",
shape = "Task")
```

```
ggplot(mean.voiced.dental, aes(x=Step, y=mean, group = Task, colour = Task)) +  geom_line() +
labs(y = "% Voiced") + scale_x_continuous(breaks = mean.voiced.dental$Step) + ggtitle("Dental
Percent Voiced per Step") + theme(plot.title = element_text(hjust = 0.5)) +
geom_hline(yintercept = 0.50, colour = "red", linetype = 3) +
scale_color_manual(values=c("#A020F0","#53db00")) +
scale_linetype_manual(values=c("longdash", "solid")) + labs(color  = "Task", linetype = "Task",
shape = "Task")

ggplot(mean.voiced.velar, aes(x=Step, y=mean, group = Task, colour = Task)) +  geom_line() +
labs(y = "% Voiced") + scale_x_continuous(breaks = mean.voiced.velar$Step) + ggtitle("Velar
Percent Voiced per Step") + theme(plot.title = element_text(hjust = 0.5)) +
geom_hline(yintercept = 0.50, colour = "red", linetype = 3) +
scale_color_manual(values=c("#A020F0","#53db00")) +
scale_linetype_manual(values=c("longdash", "solid")) + labs(color  = "Task", linetype = "Task",
shape = "Task")
```

Logistic Regression - Code and Output

```
 library(lme4)

total = rbind(data.bilabial, data.dental, data.velar)
new.total = subset(total, Task != "3")

logit.new.total = glm(Voiced ~ AQ + POA + factor(Task) * Step, family = binomial(link =
"logit"), data = new.total)
summary(logit.new.total)

Call:
glm(formula = Voiced ~ AQ + POA + factor(Task) * Step, family = binomial(link = "logit"),
    data = new.total)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.5969  -0.3095   0.0546   0.3176   3.5338

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      5.19185    0.63606   8.163 3.28e-16 ***
AQ               0.00197    0.01056   0.187   0.8520
POAdental        2.02411    0.32800   6.171 6.78e-10 ***
```

```
POAvelar          0.89145    0.43325   2.058   0.0396 *
factor(Task)2      2.28971    1.05311   2.174   0.0297 *
Step             -0.76841    0.07589 -10.125  < 2e-16 ***
factor(Task)2:Step -0.27944    0.12801  -2.183   0.0290 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 955.40  on 691  degrees of freedom
Residual deviance: 364.92  on 685  degrees of freedom
AIC: 378.92

Number of Fisher Scoring iterations: 7

 mylogit.bilabial = glm(Voiced ~ AQ + factor(Task) * Step, family = binomial(link = "logit"),
 data = data.bilabial)
 summary(mylogit.bilabial)

Call:
glm(formula = Voiced ~ AQ + factor(Task) * Step, family = binomial(link = "logit"),
    data = data.bilabial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.4878  -0.2030  -0.0038   0.2434   2.1122

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)         6.311207   1.082166   5.832 5.48e-09 ***
AQ                  0.001434   0.016938   0.085   0.9325
factor(Task)2       3.969147   2.164611   1.834   0.0667 .
Step               -0.942694   0.145006  -6.501 7.97e-11 ***
factor(Task)2:Step -0.476034   0.296993  -1.603   0.1090
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 468.57  on 337  degrees of freedom
```

```
Residual deviance: 150.01  on 333  degrees of freedom
AIC: 160.01

Number of Fisher Scoring iterations: 7

 mylogit.dental = glm(Voiced ~ AQ + factor(Task) * Step, family = binomial(link = "logit"),
 data = data.dental)
 summary(mylogit.dental)

Call:
glm(formula = Voiced ~ AQ + factor(Task) * Step, family = binomial(link = "logit"),
    data = data.dental)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.3284  -0.2847   0.0252   0.3113   3.9193

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)        6.254782   1.111062   5.630 1.81e-08 ***
AQ                -0.007349   0.016422  -0.447   0.6545
factor(Task)2      5.938344   2.758951   2.152   0.0314 *
Step              -0.645954   0.106240  -6.080 1.20e-09 ***
factor(Task)2:Step -0.665209   0.296109  -2.247   0.0247 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 366.42  on 269  degrees of freedom
Residual deviance: 132.94  on 265  degrees of freedom
AIC: 142.94

Number of Fisher Scoring iterations: 7

 mylogit.velar = glm(Voiced ~ AQ + factor(Task) * Step, family = binomial(link = "logit"),
 data = data.velar)
 summary(mylogit.velar)

Call:
```

```
glm(formula = Voiced ~ AQ + factor(Task) * Step, family = binomial(link = "logit"),
    data = data.velar)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.9693  -0.5126   0.1103   0.5089   2.4559

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        5.61557    1.93314   2.905 0.003674 **
AQ                 0.03133    0.02854   1.098 0.272357
factor(Task)2     -2.66685    2.19680  -1.214 0.224760
Step              -0.74063    0.21777  -3.401 0.000671 ***
factor(Task)2:Step  0.25553    0.25619   0.997 0.318567
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 116.020  on 83  degrees of freedom
Residual deviance:  61.979  on 79  degrees of freedom
AIC: 71.979

Number of Fisher Scoring iterations: 6
```

Group Analysis - AQ Effect

```
data = read.csv(file.choose(), header = T)
summary(data)

data = subset(data, Task != "3")
data1 = subset(data, Task != "2")
data2 = subset(data, Task != "1")

data1.bilabial = subset(data1, POA == "bilabial")
data2.bilabial = subset(data2, POA == "bilabial")
data1.dental = subset(data1, POA == "dental")
data2.dental = subset(data2, POA == "dental")
data1.velar = subset(data1, POA == "velar")
data2.velar = subset(data2, POA == "velar")
```

```
install.packages("dplyr")
library(dplyr)
library(ggplot2)

mean.voiced.bilabial.1 = data1.bilabial %>% group_by(Step, AQ) %>% summarize(mean =
mean(Voiced))
ggplot(mean.voiced.bilabial.1, aes(x=Step, y=mean, group = AQ, colour = as.factor(AQ))) +
geom_line() + labs(y = "% Voiced") + scale_x_continuous(breaks =
mean.voiced.bilabial.1$Step) + ggtitle("Bilabial Percent Voiced per Step in Task 1") +
theme(plot.title = element_text(hjust = 0.5)) + labs(colour = "AQ") + theme_bw()
mean.voiced.bilabial.2 = data2.bilabial %>% group_by(Step, AQ) %>% summarize(mean =
mean(Voiced))
ggplot(mean.voiced.bilabial.2, aes(x=Step, y=mean, group = AQ, colour = as.factor(AQ))) +
geom_line() + labs(y = "% Voiced") + scale_x_continuous(breaks =
mean.voiced.bilabial.2$Step) + ggtitle("Bilabial Percent Voiced per Step in Task 2") +
theme(plot.title = element_text(hjust = 0.5)) + labs(colour = "AQ") + theme_bw()

mean.voiced.dental.1 = data1.dental %>% group_by(Step, AQ) %>% summarize(mean =
mean(Voiced))
ggplot(mean.voiced.dental.1, aes(x=Step, y=mean, group = AQ, colour = as.factor(AQ))) +
geom_line() + labs(y = "% Voiced") + scale_x_continuous(breaks = mean.voiced.dental.1$Step)
+ ggtitle("Dental Percent Voiced per Step in Task 1") + theme(plot.title = element_text(hjust =
0.5)) + labs(colour = "AQ") + theme_bw()
mean.voiced.dental.2 = data2.dental %>% group_by(Step, AQ) %>% summarize(mean =
mean(Voiced))
ggplot(mean.voiced.dental.2, aes(x=Step, y=mean, group = AQ, colour = as.factor(AQ))) +
geom_line() + labs(y = "% Voiced") + scale_x_continuous(breaks = mean.voiced.dental.2$Step)
+ ggtitle("Dental Percent Voiced per Step in Task 2") + theme(plot.title = element_text(hjust =
0.5)) + labs(colour = "AQ") + theme_bw()

mean.voiced.velar.1 = data1.velar %>% group_by(Step, AQ) %>% summarize(mean =
mean(Voiced))
ggplot(mean.voiced.velar.1, aes(x=Step, y=mean, group = AQ, colour = as.factor(AQ))) +
geom_line() + labs(y = "% Voiced") + scale_x_continuous(breaks = mean.voiced.velar.1$Step) +
ggtitle("Velar Percent Voiced per Step in Task 1") + theme(plot.title = element_text(hjust = 0.5))
+ labs(colour = "AQ") + theme_bw()
mean.voiced.velar.2 = data2.velar %>% group_by(Step, AQ) %>% summarize(mean =
mean(Voiced))
```

```
ggplot(mean.voiced.velar.2, aes(x=Step, y=mean, group = AQ, colour = as.factor(AQ))) +
geom_line() + labs(y = "% Voiced") + scale_x_continuous(breaks = mean.voiced.velar.2$Step) +
ggtitle("Velar Percent Voiced per Step in Task 2") + theme(plot.title = element_text(hjust = 0.5))
+ labs(colour = "AQ") + theme_bw()
```

Appendix H

Experiment 1 - Code and Output for Production Data

English Baseline vs. Spanish Pre-test

```
data = read.csv(file.choose(), header = T)

install.packages("dplyr")
library(dplyr)

data = data %>% mutate_at(c(1:4, 8), factor)

library(lme4)
library(lmerTest)

data.pre = subset(data, Time == "english" | Time == "spanish1")
data.pre$Time = factor(data.pre$Time)

fit.pre = lmer(VOT ~ AQ + POA * Time + (1|Speaker) + (1|Word), data = data.pre)

summary(fit.pre)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: VOT ~ AQ + POA * Time + (1 | Speaker) + (1 | Word)
   Data: data.pre

REML criterion at convergence: 4425.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.0027 -0.5568 -0.0725  0.4319  4.8340

Random effects:
 Groups   Name        Variance Std.Dev.
 Word     (Intercept)   84.5    9.192
 Speaker  (Intercept)  960.6   30.993
 Residual             1569.0   39.610
Number of obs: 432, groups:  Word, 24; Speaker, 18

Fixed effects:
```

```
                 Estimate Std. Error       df t value Pr(>|t|)
(Intercept)      101.43682   17.39300  26.58000   5.832  3.5e-06 ***
AQ                -0.06471    0.61342  15.99999  -0.105    0.917
POAp              -8.97967   13.10197  18.00000  -0.685    0.502
POAt             -14.58586   13.10197  18.00000  -1.113    0.280
Timespanish1       7.31984   10.69771  18.00000   0.684    0.503
POAp:Timespanish1  -7.99466   15.12885  18.00000  -0.528    0.604
POAt:Timespanish1   1.16111   15.12885  18.00000   0.077    0.940
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) AQ     POAp   POAt   Tmspn1 POAp:T1
AQ          -0.735
POAp        -0.377  0.000
POAt        -0.377  0.000  0.500
Timespansh1 -0.461  0.000  0.612  0.612
POAp:Tmspn1  0.326  0.000 -0.866 -0.433 -0.707
POAt:Tmspn1  0.326  0.000 -0.433 -0.866 -0.707  0.500

fit.pre.2 = lmer(VOT ~ Sub_switch + POA  * Time + (1|Speaker) + (1|Word), data = data.pre)

summary(fit.pre.2)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: VOT ~ Sub_switch + POA * Time + (1 | Speaker) + (1 | Word)
   Data: data.pre

REML criterion at convergence: 4422.6

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.0039 -0.5550 -0.0724  0.4319  4.8362

Random effects:
 Groups   Name        Variance Std.Dev.
 Word     (Intercept)   84.5    9.192
 Speaker  (Intercept)  961.3   31.005
 Residual             1569.0   39.610
Number of obs: 432, groups:  Word, 24; Speaker, 18
```

```
Fixed effects:
                   Estimate Std. Error       df t value Pr(>|t|)
(Intercept)        99.97030   18.03479  25.79277   5.543 8.28e-06 ***
Sub_switch          0.02197    2.53091  15.99995   0.009    0.993
POAp               -8.97967   13.10202  17.99987  -0.685    0.502
POAt              -14.58586   13.10202  17.99987  -1.113    0.280
Timespanish1        7.31984   10.69775  17.99987   0.684    0.503
POAp:Timespanish1  -7.99466   15.12891  17.99987  -0.528    0.604
POAt:Timespanish1   1.16111   15.12891  17.99987   0.077    0.940
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) Sb_swt POAp   POAt   Tmspn1 POAp:T1
Sub_switch  -0.756
POAp        -0.363  0.000
POAt        -0.363  0.000  0.500
Timespansh1 -0.445  0.000  0.612  0.612
POAp:Tmspn1  0.315  0.000 -0.866 -0.433 -0.707
POAt:Tmspn1  0.315  0.000 -0.433 -0.866 -0.707  0.500

fit.pre.3 = lmer(VOT ~ Sub_attention + POA  * Time + (1|Speaker) + (1|Word), data = data.pre)

summary(fit.pre.3)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: VOT ~ Sub_attention + POA * Time + (1 | Speaker) + (1 | Word)
   Data: data.pre

REML criterion at convergence: 4422

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.0038 -0.5488 -0.0709  0.4303  4.8442

Random effects:
 Groups   Name        Variance Std.Dev.
 Word     (Intercept)   84.5    9.192
 Speaker  (Intercept)  954.2   30.890
 Residual             1569.0   39.610
Number of obs: 432, groups:  Word, 24; Speaker, 18
```

```
Fixed effects:
                    Estimate Std. Error     df t value Pr(>|t|)
(Intercept)           93.855     22.061 22.272   4.254 0.000317 ***
Sub_attention          1.111      3.324 16.000   0.334 0.742565
POAp                  -8.980     13.102 18.000  -0.685 0.501844
POAt                 -14.586     13.102 18.000  -1.113 0.280244
Timespanish1           7.320     10.698 18.000   0.684 0.502536
POAp:Timespanish1     -7.995     15.129 18.000  -0.528 0.603650
POAt:Timespanish1      1.161     15.129 18.000   0.077 0.939671
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
             (Intr) Sb_ttn POAp   POAt   Tmspn1 POAp:T1
Sub_attentn -0.845
POAp        -0.297  0.000
POAt        -0.297  0.000  0.500
Timespansh1 -0.364  0.000  0.612  0.612
POAp:Tmspn1  0.257  0.000 -0.866 -0.433 -0.707
POAt:Tmspn1  0.257  0.000 -0.433 -0.866 -0.707  0.500
```

Spanish Immediate Imitation and Immediate Post-Test vs. Spanish Pre-test

```
data = read.csv(file.choose(), header = T)

install.packages("dplyr")
library(dplyr)

data = data %>% mutate_at(c(1:4, 8), factor)

library(lme4)
library(lmerTest)

data.sp = subset(data, Time != "english")
data.sp$Time = factor(data.sp$Time)

fit.sp = lmer(VOT ~ AQ  * Time + (1|Speaker) + (1|Word), data = data.sp)

summary(fit.sp)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: VOT ~ AQ * Time + (1 | Speaker) + (1 | Word)
```

```
  Data: data.sp

REML criterion at convergence: 9835.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.5166 -0.5577 -0.0788  0.4204  7.4244

Random effects:
 Groups   Name        Variance Std.Dev.
 Speaker  (Intercept)  981.9   31.33
 Word     (Intercept)  163.6   12.79
 Residual             1332.6   36.50
Number of obs: 972, groups:  Speaker, 18; Word, 18

Fixed effects:
                Estimate Std. Error        df t value Pr(>|t|)
(Intercept)     101.48183   15.35697  19.00503   6.608 2.53e-06 ***
AQ               -0.11787    0.62231  17.60671  -0.189 0.851935
Timespanish2    -25.32875    5.63877 933.00003  -4.492 7.94e-06 ***
Timespanish3    -21.01072    5.63877 933.00003  -3.726 0.000206 ***
AQ:Timespanish2   0.01609    0.23304 933.00003   0.069 0.944962
AQ:Timespanish3   0.13259    0.23304 933.00003   0.569 0.569518
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) AQ     Tmspn2 Tmspn3 AQ:Tm2
AQ          -0.844
Timespansh2 -0.184  0.161
Timespansh3 -0.184  0.161  0.500
AQ:Tmspnsh2  0.158 -0.187 -0.861 -0.430
AQ:Tmspnsh3  0.158 -0.187 -0.430 -0.861  0.500

install.packages("emmeans")
library(emmeans)

emmeans(fit.sp, pairwise ~ Time)
NOTE: Results may be misleading due to involvement in interactions
$emmeans
 Time     emmean   SE   df lower.CL upper.CL
```

```
 spanish1   99.0 8.23 22.7     82.0    116.1
 spanish2   74.0 8.23 22.7     57.0     91.1
 spanish3   80.8 8.23 22.7     63.7     97.8

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95

$contrasts
 contrast            estimate   SE  df t.ratio p.value
 spanish1 - spanish2    24.99 2.87 933   8.714  <.0001
 spanish1 - spanish3    18.25 2.87 933   6.363  <.0001
 spanish2 - spanish3    -6.75 2.87 933  -2.352  0.0495

Degrees-of-freedom method: kenward-roger
P value adjustment: tukey method for comparing a family of 3 estimates
```

```
fit.sp.2 = lmer(VOT ~ Sub_switch  * Time + (1|Speaker) + (1|Word), data = data.sp)

summary(fit.sp.2)
```

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: VOT ~ Sub_switch * Time + (1 | Speaker) + (1 | Word)
   Data: data.sp

REML criterion at convergence: 9825.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.5199 -0.5580 -0.0747  0.4306  7.4385

Random effects:
 Groups   Name        Variance Std.Dev.
 Speaker  (Intercept)  981.5   31.33
 Word     (Intercept)  163.7   12.79
 Residual             1330.8   36.48
Number of obs: 972, groups:  Speaker, 18; Word, 18

Fixed effects:
                     Estimate Std. Error       df t value Pr(>|t|)
(Intercept)           99.6962    16.0925  18.8755   6.195 6.12e-06 ***
Sub_switch            -0.1243     2.5662  17.6051  -0.048    0.962
Timespanish2         -26.2277     5.9166 933.0000  -4.433 1.04e-05 ***
```

```
Timespanish3          -24.4779     5.9166 933.0000  -4.137 3.83e-05 ***
Sub_switch:Timespanish2   0.2290     0.9605 933.0000   0.238    0.812
Sub_switch:Timespanish3   1.1560     0.9605 933.0000   1.204    0.229
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) Sb_swt Tmspn2 Tmspn3 Sb_:T2
Sub_switch  -0.859
Timespansh2 -0.184  0.164
Timespansh3 -0.184  0.164  0.500
Sb_swtch:T2  0.161 -0.187 -0.875 -0.437
Sb_swtch:T3  0.161 -0.187 -0.437 -0.875  0.500
```

```
fit.sp.3 = lmer(VOT ~ Sub_attention  * Time + (1|Speaker) + (1|Word), data = data.sp)

summary(fit.sp.3)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: VOT ~ Sub_attention * Time + (1 | Speaker) + (1 | Word)
   Data: data.sp

REML criterion at convergence: 9826.6

Scaled residuals: 
    Min      1Q  Median      3Q     Max 
-3.5134 -0.5446 -0.0797  0.4172  7.4058 

Random effects:
 Groups   Name        Variance Std.Dev.
 Speaker  (Intercept)  980.4   31.31  
 Word     (Intercept)  163.6   12.79  
 Residual             1332.6   36.51  
Number of obs: 972, groups:  Speaker, 18; Word, 18

Fixed effects:
                        Estimate Std. Error       df t value Pr(>|t|)    
(Intercept)              95.0623    16.9568  18.7505   5.606  2.2e-05 ***
Sub_attention             0.6995     2.6167  17.6092   0.267 0.792325    
Timespanish2            -21.9065     6.2530 933.0000  -3.503 0.000481 ***
```

```
Timespanish3            -17.6585     6.2530 933.0000  -2.824 0.004844 **
Sub_attention:Timespanish2  -0.5448     0.9806 933.0000  -0.556 0.578634
Sub_attention:Timespanish3  -0.1041     0.9806 933.0000  -0.106 0.915462
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) Sb_ttn Tmspn2 Tmspn3 Sb_:T2
Sub_attentn -0.874
Timespansh2 -0.184  0.166
Timespansh3 -0.184  0.166  0.500
Sb_ttntn:T2  0.164 -0.187 -0.889 -0.444
Sb_ttntn:T3  0.164 -0.187 -0.444 -0.889  0.500
```

Comparing Place of Articulation

```
data.sp = subset(data, Time != "english")
data.sp$Time = factor(data.sp$Time)
fit.sp = lmer(VOT ~ AQ + POA * Time + (1|Speaker) + (1|Word), data = data.sp)
summary(fit.sp)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: VOT ~ AQ + POA * Time + (1 | Speaker) + (1 | Word)
   Data: data.sp

REML criterion at convergence: 9768.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.6128 -0.5448 -0.0638  0.4326  7.5193

Random effects:
 Groups   Name        Variance Std.Dev.
 Speaker  (Intercept)  938     30.63
 Word     (Intercept)  118     10.86
 Residual             1302     36.08
Number of obs: 972, groups:  Speaker, 18; Word, 18

Fixed effects:
                    Estimate Std. Error        df t value Pr(>|t|)
(Intercept)       107.268952  15.394692  20.760030   6.968 7.46e-07 ***
```

```
AQ               0.006701   0.594041  15.999914   0.011  0.99114
POAp           -16.974324   7.965132  26.855494  -2.131  0.04239 *
POAt           -13.424750   7.965132  26.855494  -1.685  0.10349
Timespanish2     -25.220120   4.910047 931.000044  -5.136 3.41e-07 ***
Timespanish3     -15.439130   4.910047 931.000044  -3.144  0.00172 **
POAp:Timespanish2  -2.036296   6.943855 931.000045  -0.293  0.76940
POAt:Timespanish2   5.642639   6.943855 931.000045   0.813  0.41665
POAp:Timespanish3  -4.193398   6.943855 931.000045  -0.604  0.54606
POAt:Timespanish3   1.017148   6.943855 931.000045   0.146  0.88357
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
          (Intr) AQ     POAp   POAt   Tmspn2 Tmspn3 POAp:T2 POAt:T2 POAp:T3
AQ         -0.804
POAp        -0.259  0.000
POAt        -0.259  0.000  0.500
Timespansh2 -0.159  0.000  0.308  0.308
Timespansh3 -0.159  0.000  0.308  0.308  0.500
POAp:Tmspn2  0.113  0.000 -0.436 -0.218 -0.707 -0.354
POAt:Tmspn2  0.113  0.000 -0.218 -0.436 -0.707 -0.354  0.500
POAp:Tmspn3  0.113  0.000 -0.436 -0.218 -0.354 -0.707  0.500   0.250
POAt:Tmspn3  0.113  0.000 -0.218 -0.436 -0.354 -0.707  0.250   0.500   0.500

library(emmeans)
emmeans(fit.sp, pairwise ~ POA)
Cannot use mode = "kenward-roger" because *pbkrtest* package is not installed
NOTE: Results may be misleading due to involvement in interactions
$emmeans
 POA emmean   SE   df lower.CL upper.CL
 k     93.9 8.71 26.6     76.0    111.7
 p     74.8 8.71 26.6     56.9     92.7
 t     82.7 8.71 26.6     64.8    100.5

Results are averaged over the levels of: Time
Degrees-of-freedom method: satterthwaite
Confidence level used: 0.95

$contrasts
 contrast estimate   SE df t.ratio p.value
 k - p       19.05 6.88 15   2.768  0.0361
 k - t       11.20 6.88 15   1.628  0.2648
 p - t       -7.85 6.88 15  -1.140  0.5054
```

Results are averaged over the levels of: Time
Degrees-of-freedom method: satterthwaite
P value adjustment: tukey method for comparing a family of 3 estimates

Spanish Immediate Imitation vs. Native Speaker

```
t.test(new.data.sp$VOT, new.data.sp$Native.VOT)

	Welch Two Sample t-test

data:  new.data.sp$VOT and new.data.sp$Native.VOT
t = 20.824, df = 413.34, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 40.06368 48.41583
sample estimates:
mean of x mean of y
 73.25753  29.01778

new.data.sp.p = subset(new.data.sp, POA == "p")
t.test(new.data.sp.p$VOT, new.data.sp.p$Native.VOT)

	Welch Two Sample t-test

data:  new.data.sp.p$VOT and new.data.sp.p$Native.VOT
t = 13.431, df = 110.91, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 37.97937 51.12627
sample estimates:
mean of x mean of y
 63.17782  18.62500

new.data.sp.t = subset(new.data.sp, POA == "t")
t.test(new.data.sp.t$VOT, new.data.sp.t$Native.VOT)

	Welch Two Sample t-test

data:  new.data.sp.t$VOT and new.data.sp.t$Native.VOT
t = 14.042, df = 108.99, p-value < 2.2e-16
```

```
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 43.44690 57.72743
sample estimates:
mean of x mean of y
 74.40633  23.81917

 new.data.sp.k = subset(new.data.sp, POA == "k")
 t.test(new.data.sp.k$VOT, new.data.sp.k$Native.VOT)

	Welch Two Sample t-test

data:  new.data.sp.k$VOT and new.data.sp.k$Native.VOT
t = 11.011, df = 132.43, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 30.82826 44.33030
sample estimates:
mean of x mean of y
 82.18844  44.60917
```

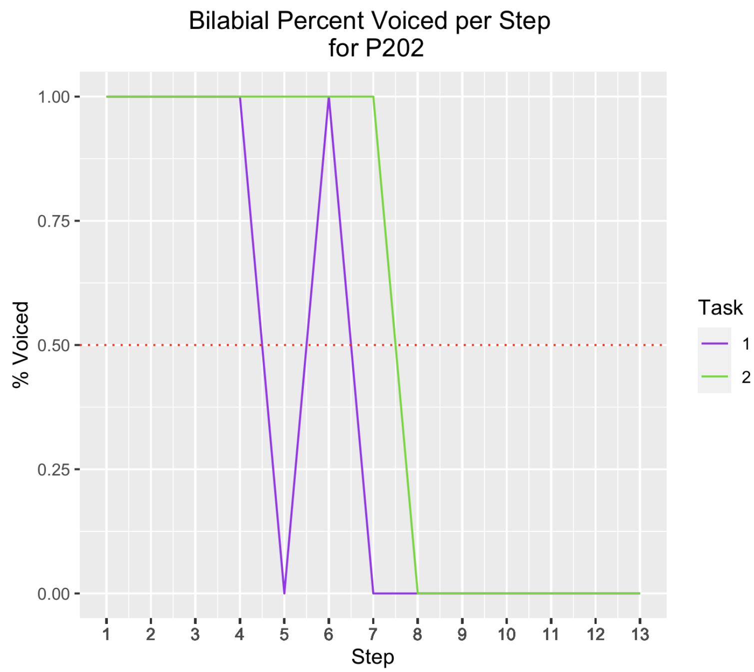
Case Study Analysis - Barplot

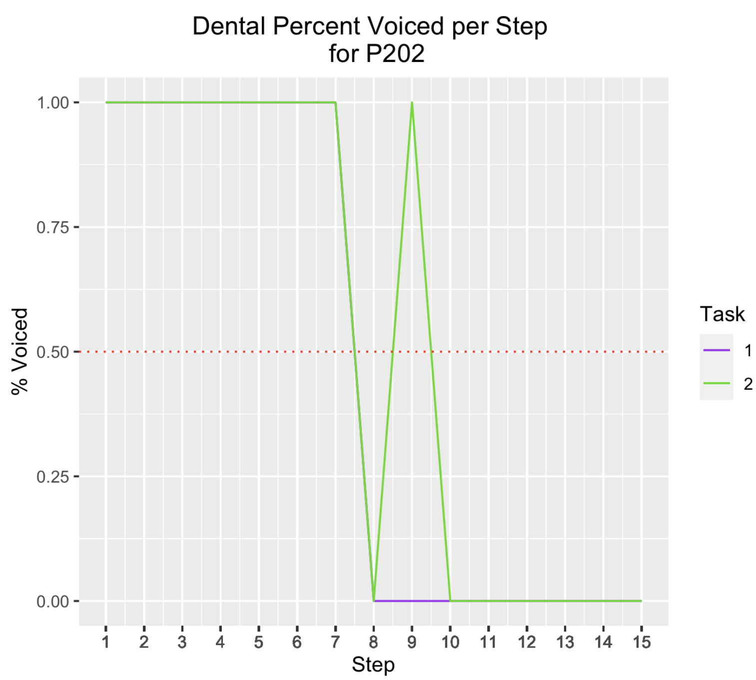
```
data = read.csv(file.choose(), header = T)
summary(data)

library(dplyr)
library(ggplot2)

aggregated = dplyr::group_by(data, Participant, Task)
aggregated = dplyr::mutate(aggregated, Task = base::factor(Task, levels = c("pre-test", "immediate post-test", "delayed post-test(original)", "delayed post-test(novel)")), Participant = base::factor(Participant, levels = "P202", "P208", "P204", "P203", "P205", "Group"))
chart = ggplot2::ggplot(data = aggregated, aes(x=Participant, y=VOT, fill=Task))

chart + ggplot2::geom_bar(position = "dodge", stat = "identity") + ggplot2::geom_text(aes(label = VOT), position = position_dodge2(width = 0.95), vjust = 0.3, hjust = 3.3) + coord_flip() + ggplot2::labs(x="Tasks per Participant", y = "VOT (ms)") + ggtitle("Production Task Results for Experiment 2") + scale_fill_manual(values=c("#134a7f", "#1f7cd5", "#6dace9", "#c2dcf6")) + theme(plot.title = element_text(hjust = 0.5)) + guides(fill=guide_legend(reverse=TRUE))
```

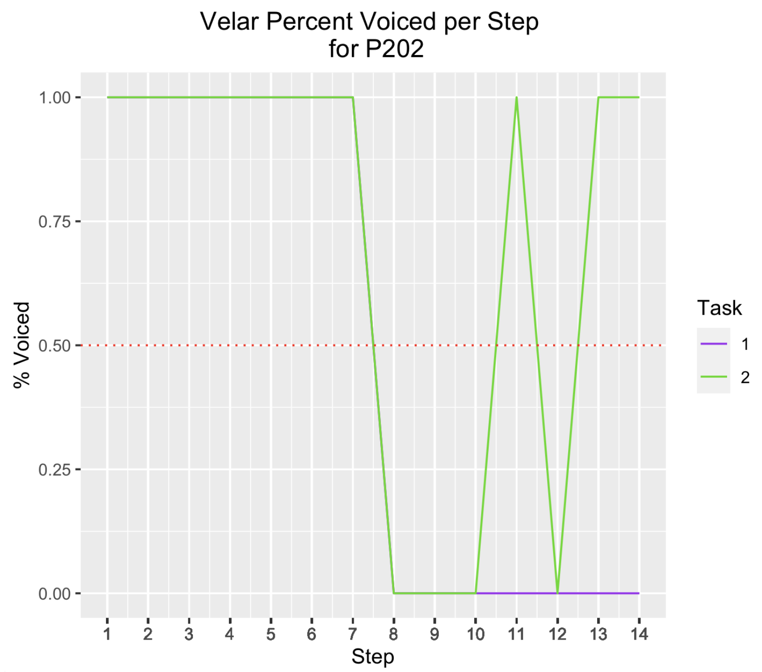
Appendix I

Individual Perception Graphs

P202

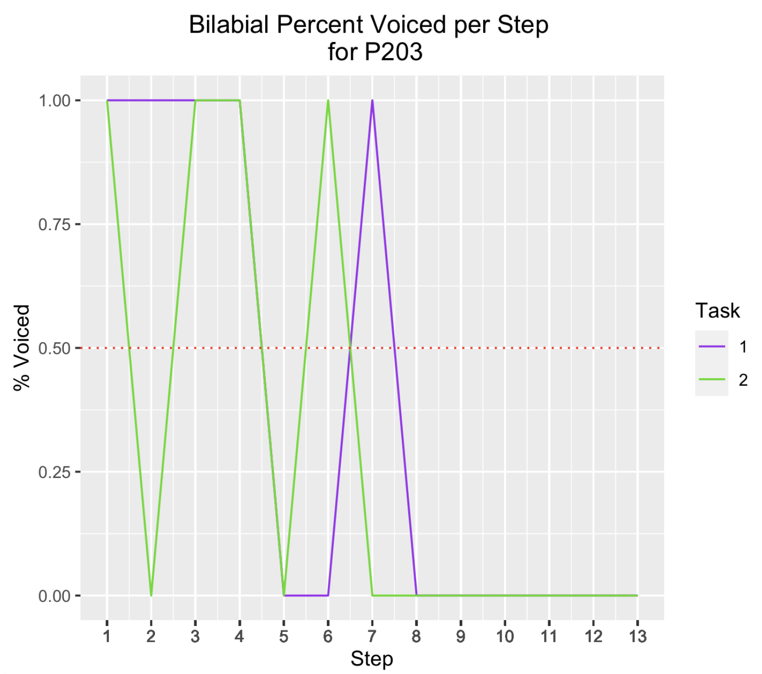




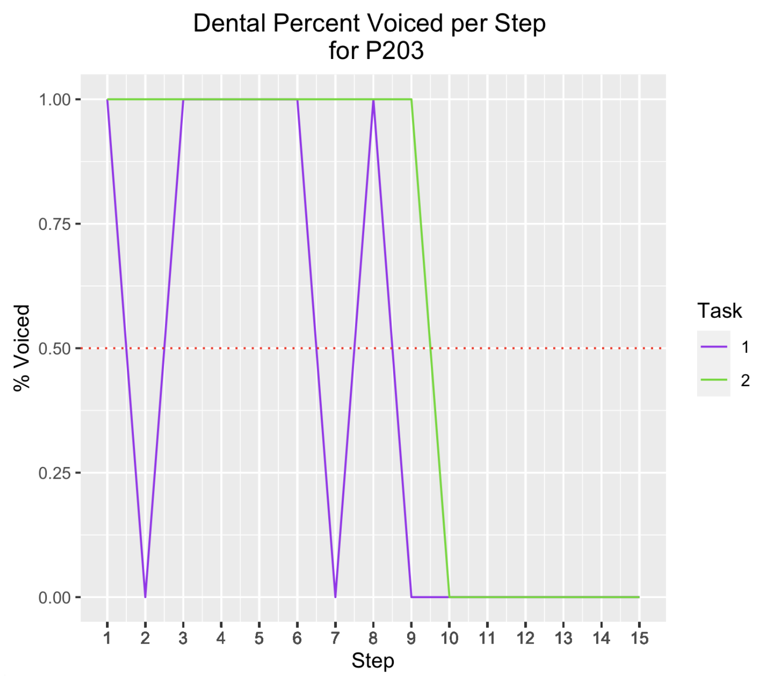


Velar Percent Voiced per Step
for P202
% Voiced
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Step
Task
1
2

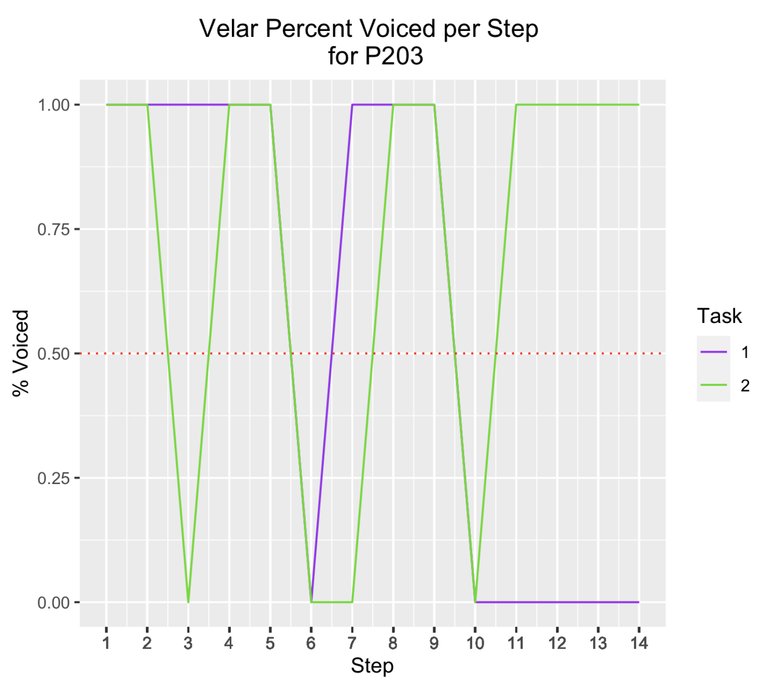
P203



Bilabial Percent Voiced per Step
for P203
% Voiced
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10
11
12
13
Step
Task
1
2

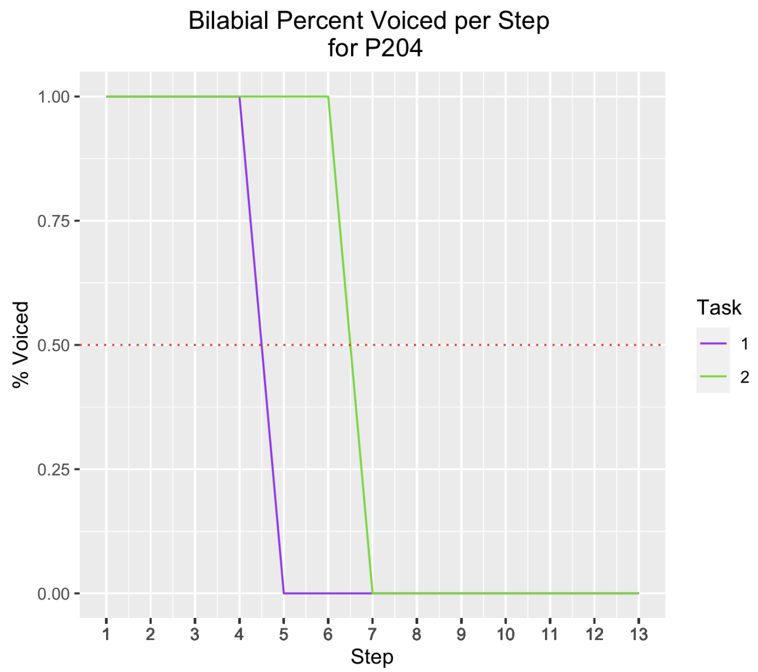


Dental Percent Voiced per Step
for P203
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Step
Task
1
2

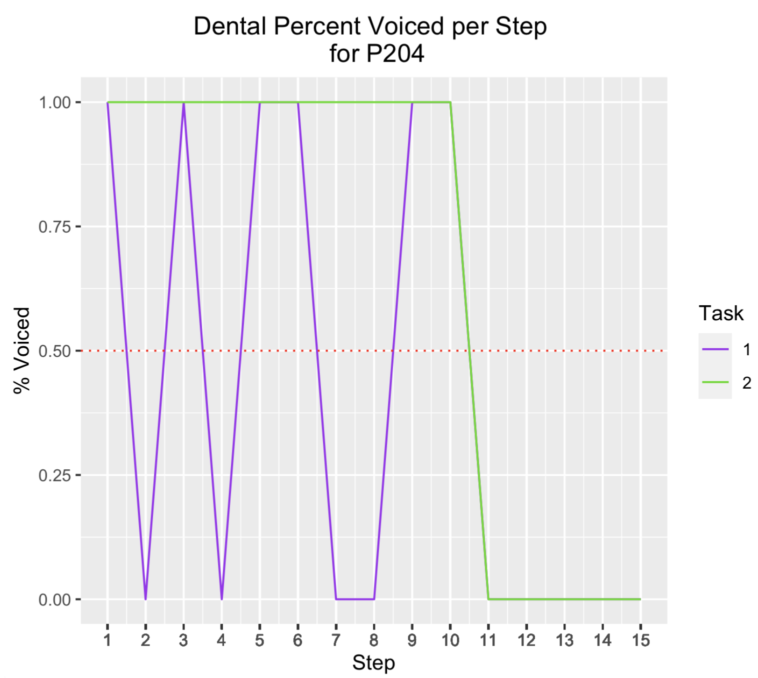


Velar Percent Voiced per Step
for P203
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Step
Task
1
2

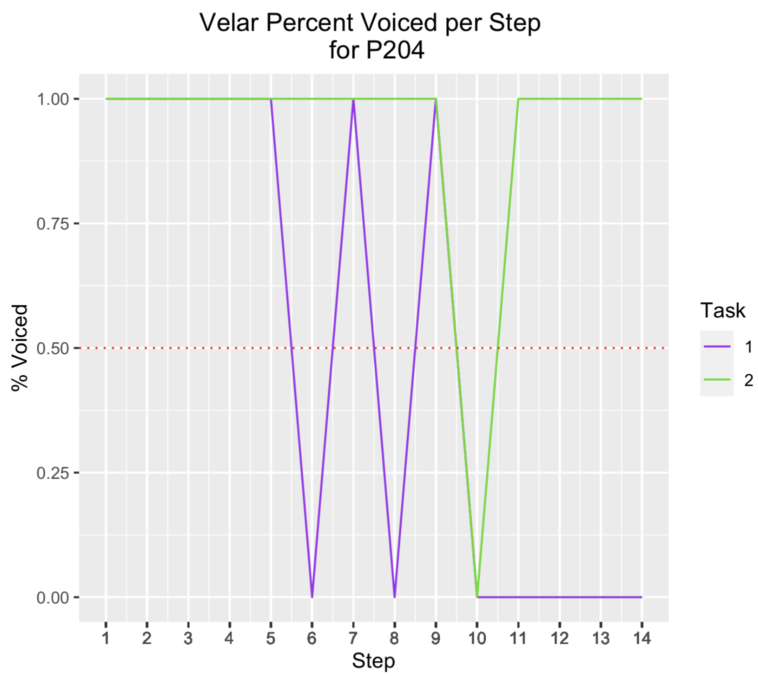
P204



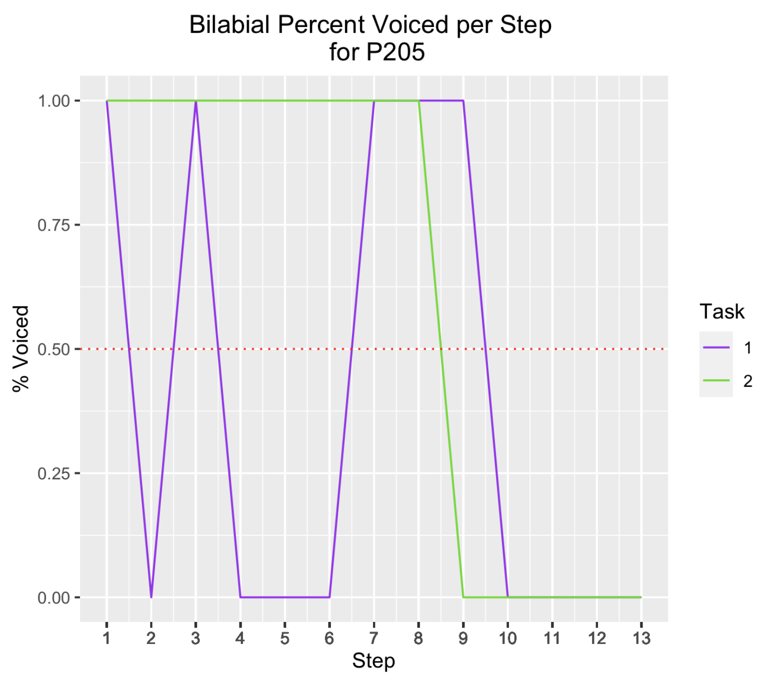
Bilabial Percent Voiced per Step
for P204
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13
Step
Task
1
2

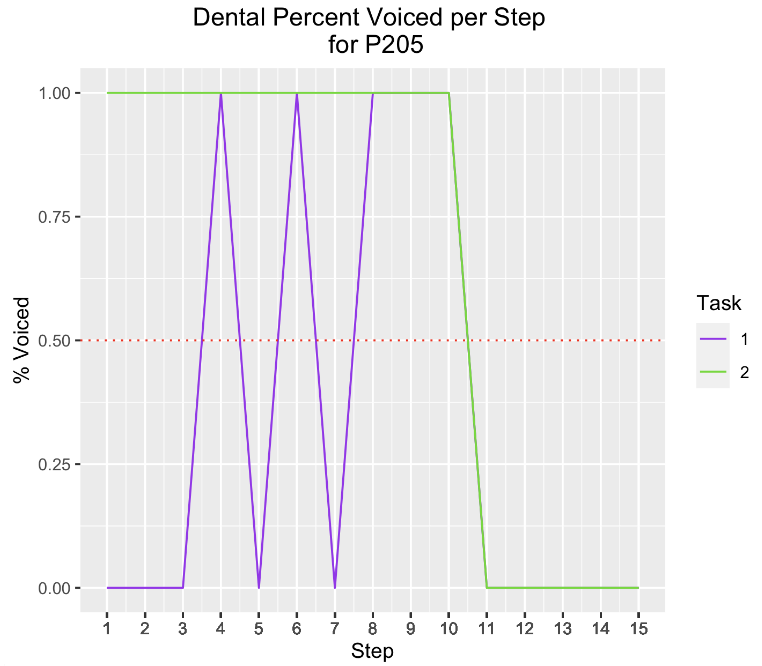


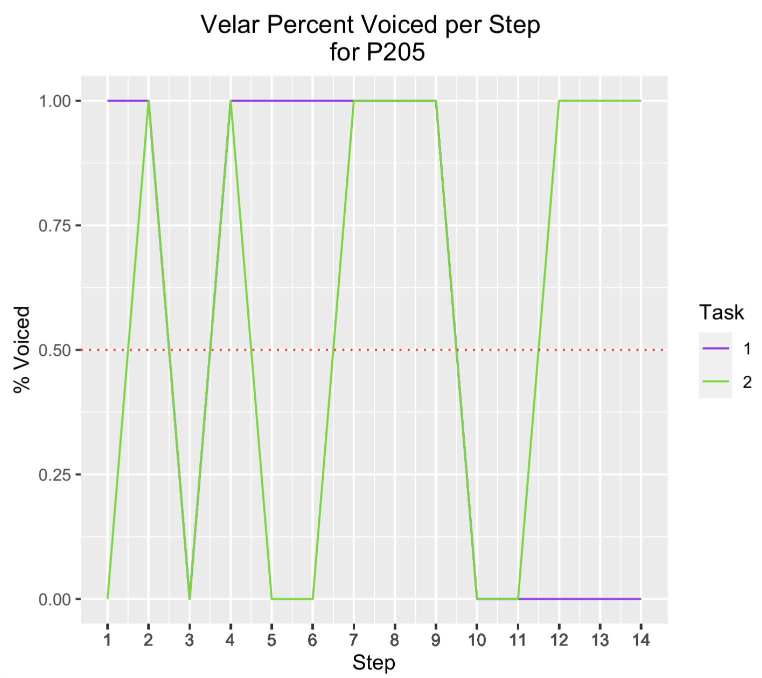
Dental Percent Voiced per Step
for P204
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Step
Task
1
2


Velar Percent Voiced per Step
for P204
% Voiced
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Step
Task
1
2

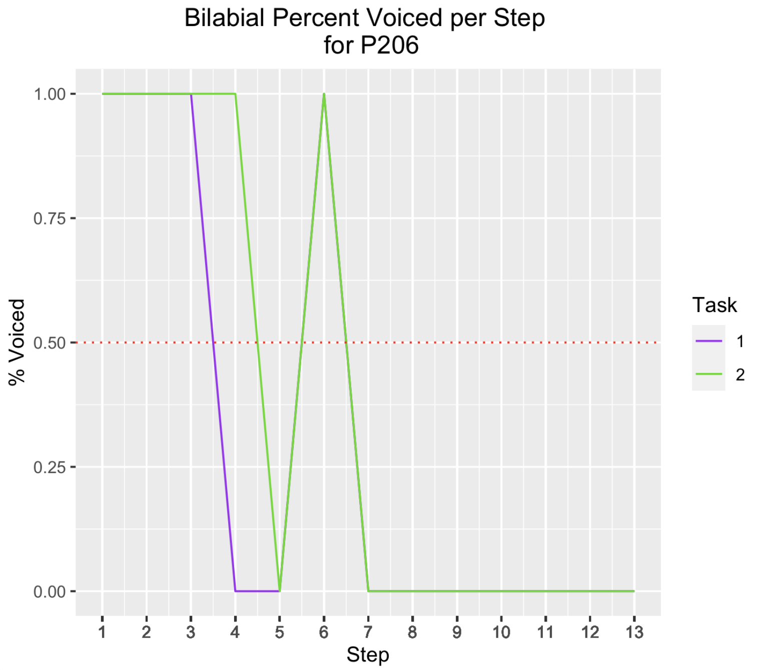
P205


Bilabial Percent Voiced per Step
for P205
% Voiced
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10
11
12
13
Step
Task
1
2

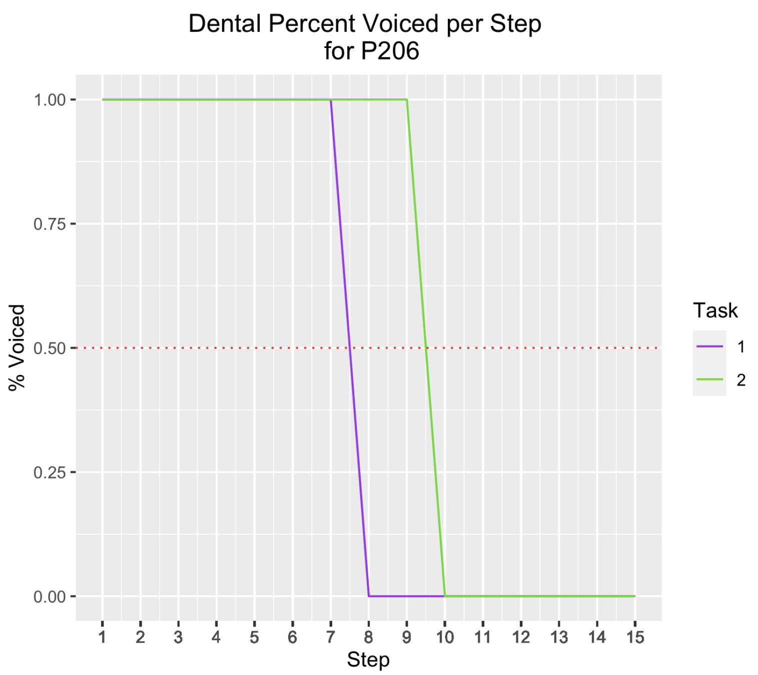

Dental Percent Voiced per Step
for P205
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Step
Task
1
2


Velar Percent Voiced per Step
for P205
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Step
Task
1
2

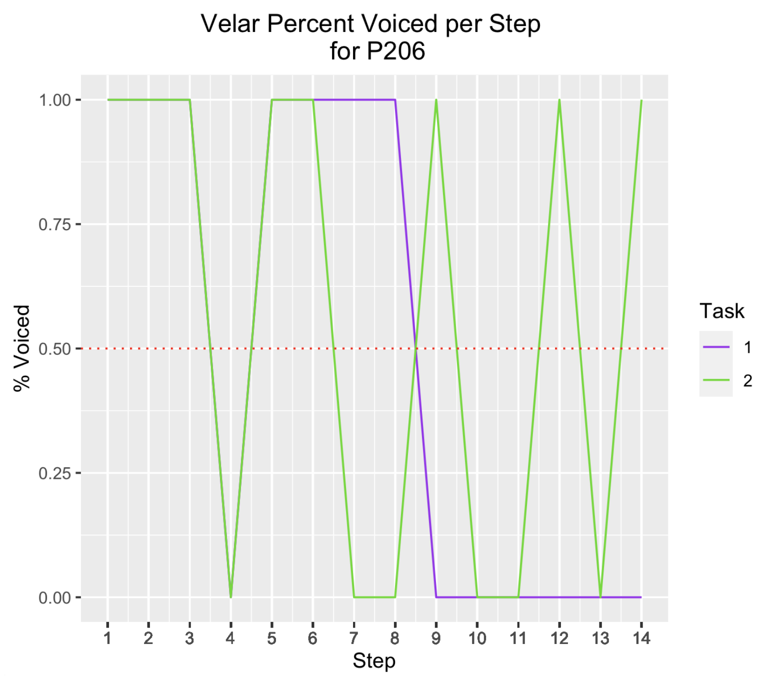
P206



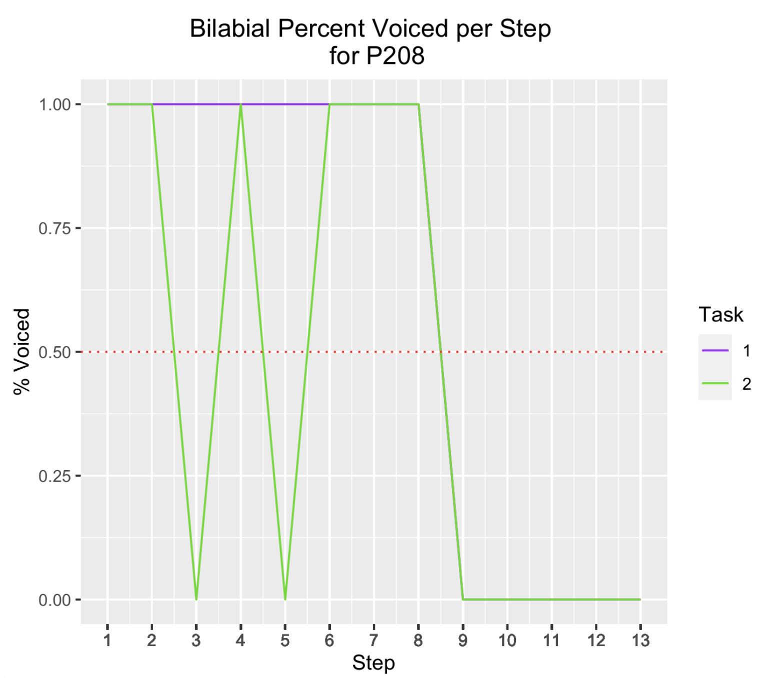
Bilabial Percent Voiced per Step
for P206
1.00
0.75
0.50
0.25
0.00
% Voiced
1 2 3 4 5 6 7 8 9 10 11 12 13
Step
Task
1
2

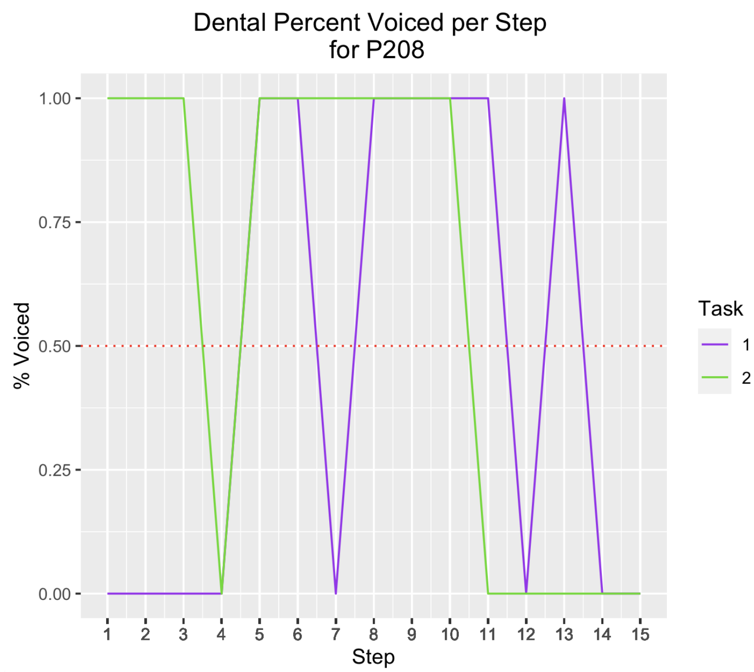


Dental Percent Voiced per Step
for P206
1.00
0.75
0.50
0.25
0.00
% Voiced
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Step
Task
1
2

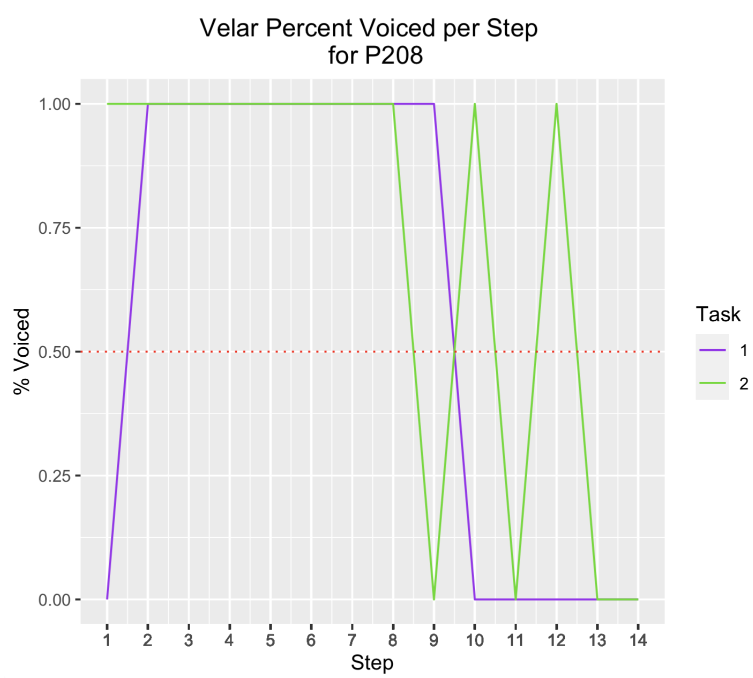

Velar Percent Voiced per Step
for P206
1.00
0.75
0.50
0.25
0.00
% Voiced
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Step
Task
1
2

P208


Bilabial Percent Voiced per Step
for P208
1.00
0.75
0.50
0.25
0.00
% Voiced
1
2
3
4
5
6
7
8
9
10
11
12
13
Step
Task
1
2

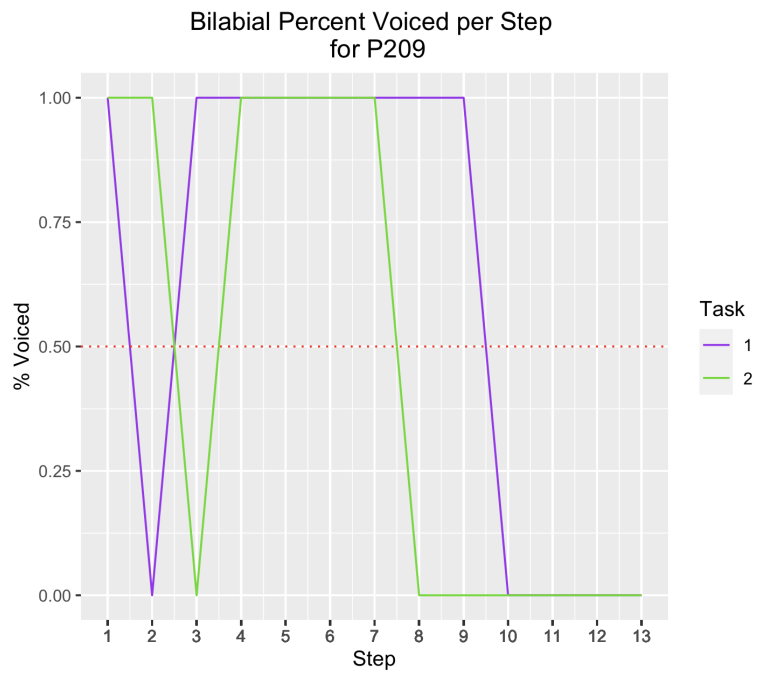


Dental Percent Voiced per Step
for P208
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Step
Task
1
2

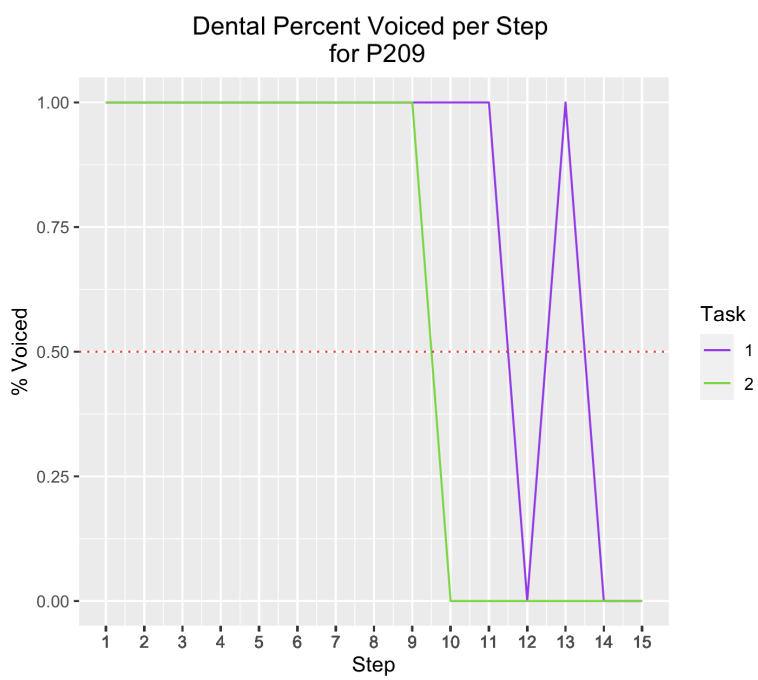


Velar Percent Voiced per Step
for P208
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Step
Task
1
2

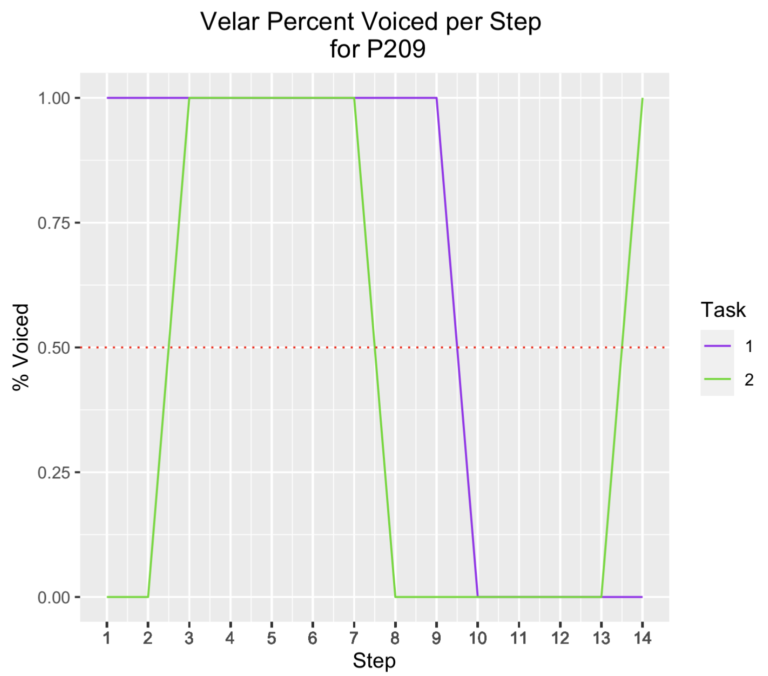
P209



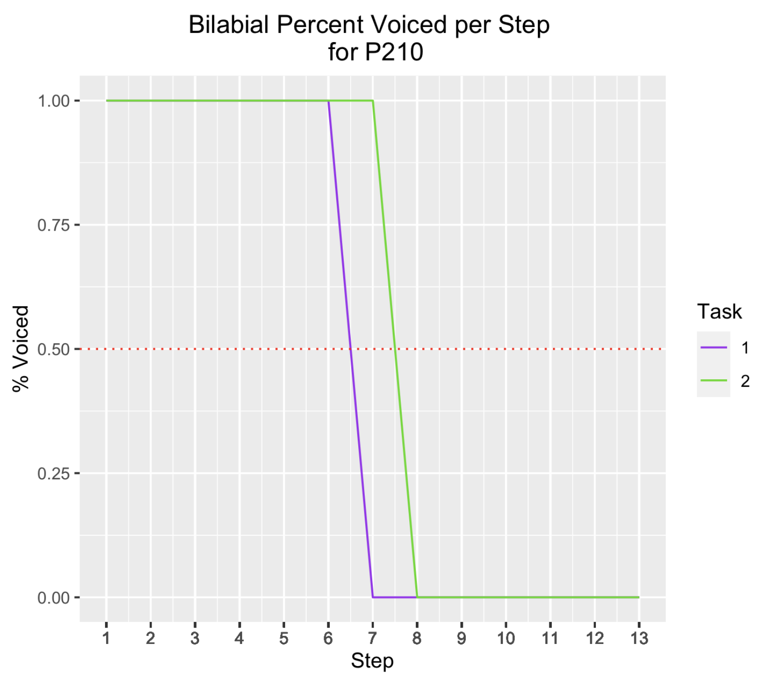
Bilabial Percent Voiced per Step
for P209
% Voiced
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10
11
12
13
Step
Task
1
2

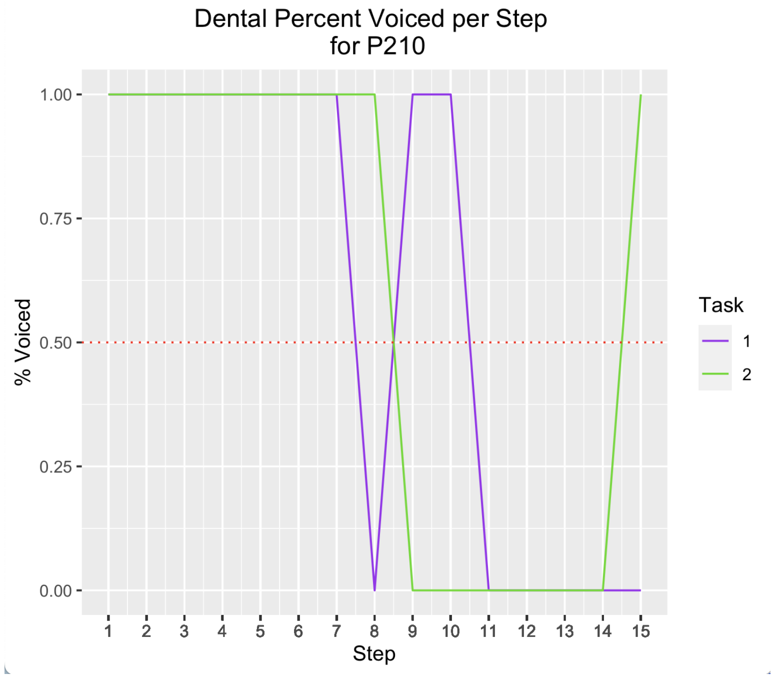


Dental Percent Voiced per Step
for P209
% Voiced
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Step
Task
1
2

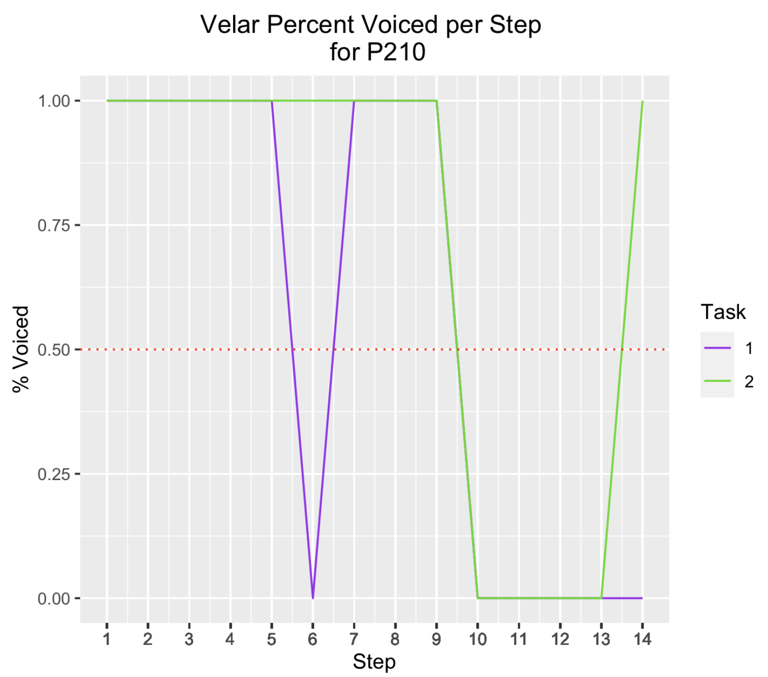

Velar Percent Voiced per Step
for P209
1.00
0.75
0.50
0.25
0.00
% Voiced
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Step
Task
1
2

P210


Bilabial Percent Voiced per Step
for P210
1.00
0.75
0.50
0.25
0.00
% Voiced
1 2 3 4 5 6 7 8 9 10 11 12 13
Step
Task
1
2

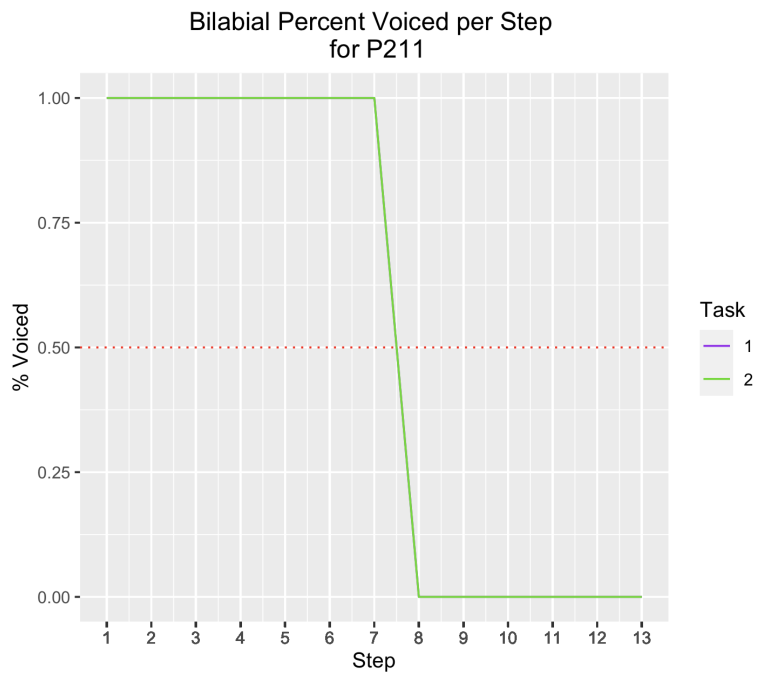


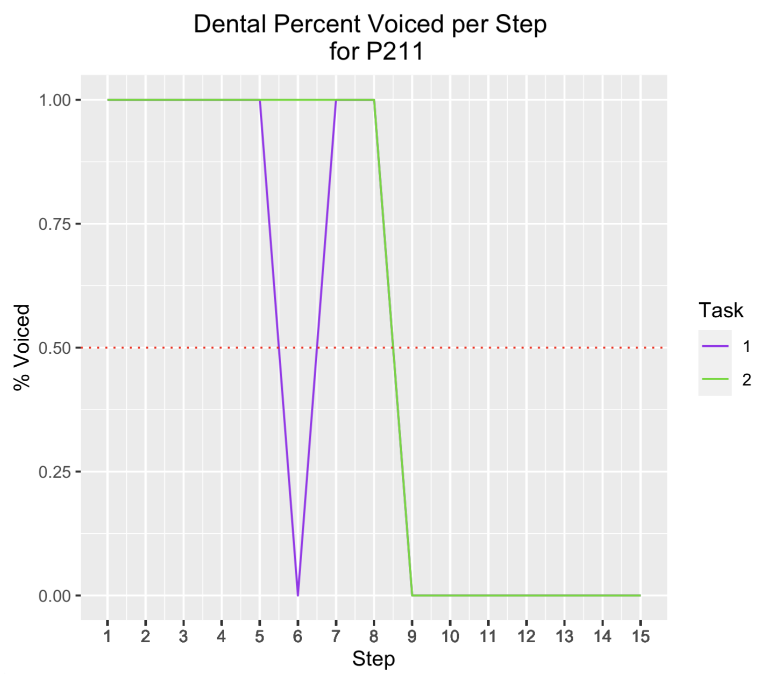
Dental Percent Voiced per Step
for P210
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Step
Task
1
2

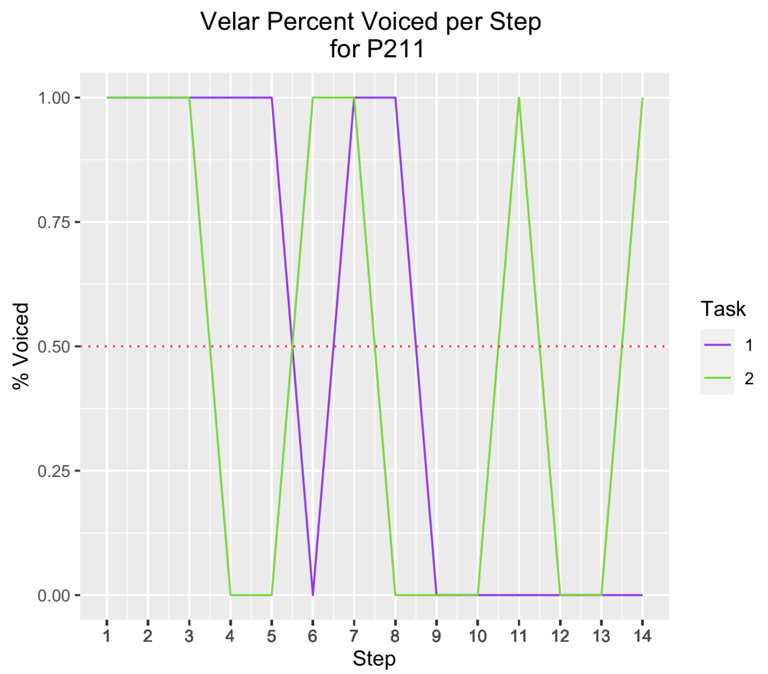


Velar Percent Voiced per Step
for P210
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Step
Task
1
2

P211

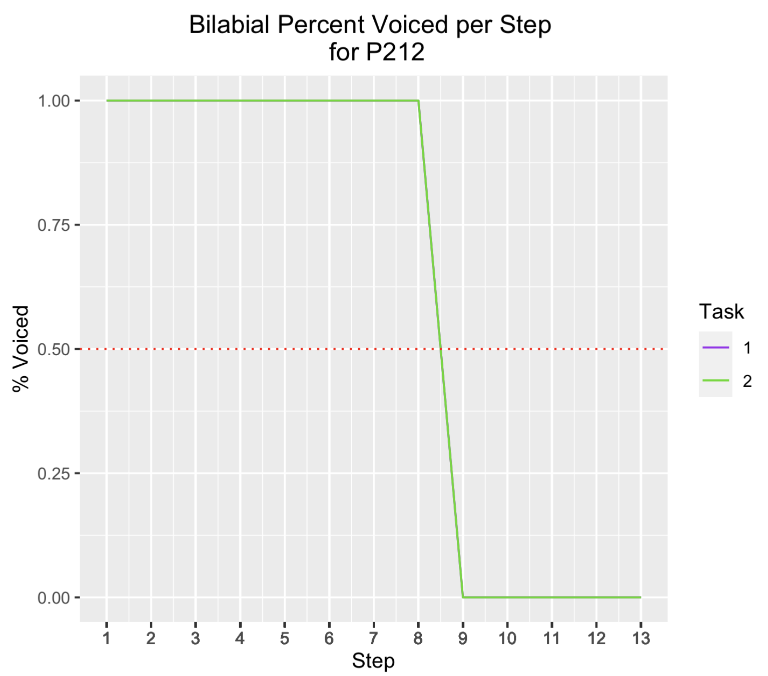

Bilabial Percent Voiced per Step
for P211
1.00
0.75
0.50
0.25
0.00
% Voiced
1
2
3
4
5
6
7
8
9
10
11
12
13
Step
Task
1
2


Dental Percent Voiced per Step
for P211
1.00
0.75
0.50
0.25
0.00
% Voiced
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Step
Task
1
2

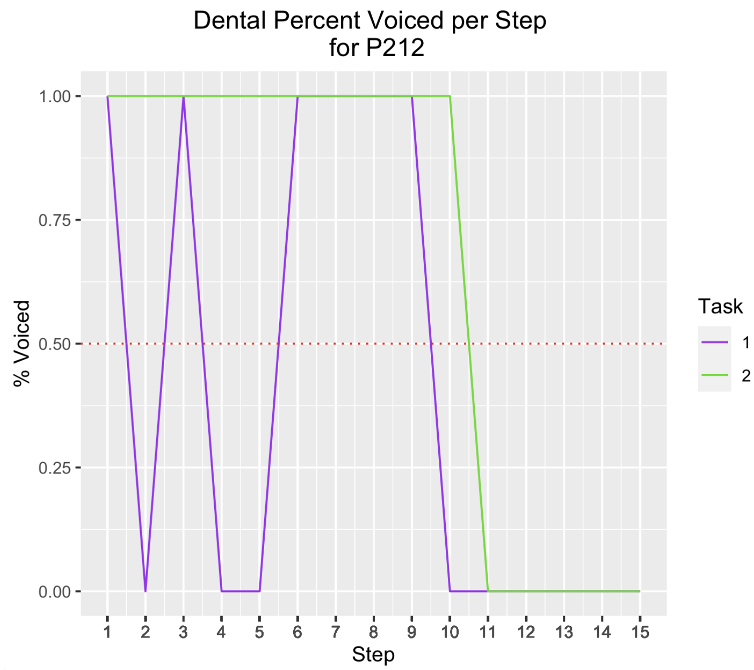


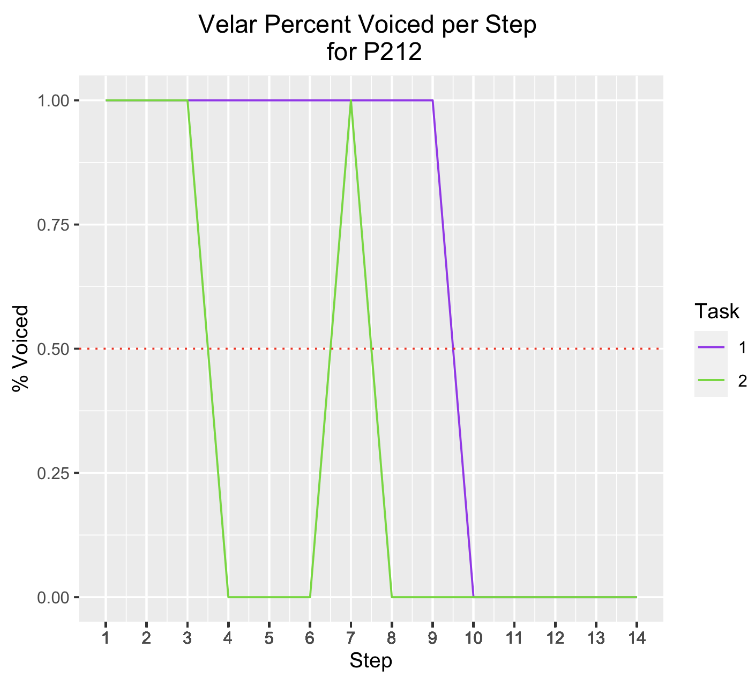
Velar Percent Voiced per Step
for P211
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Step
Task
1
2

P212

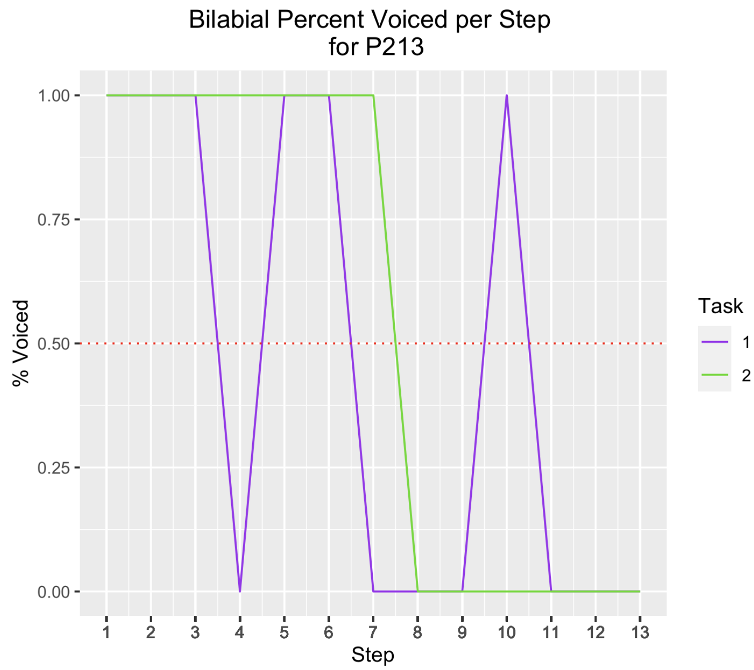


Bilabial Percent Voiced per Step
for P212
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13
Step
Task
1
2

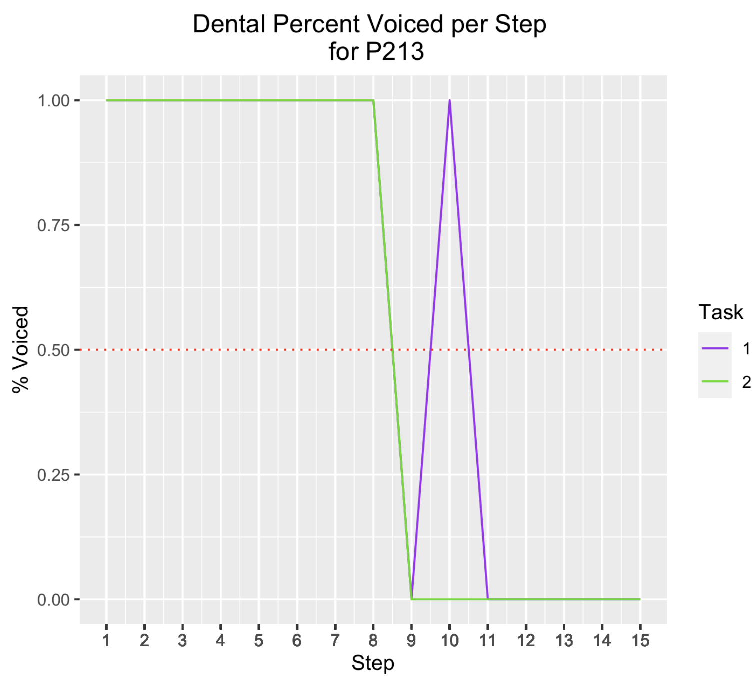

Dental Percent Voiced per Step
for P212
1.00
0.75
0.50
0.25
0.00
% Voiced
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Step
Task
1
2


Velar Percent Voiced per Step
for P212
1.00
0.75
0.50
0.25
0.00
% Voiced
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Step
Task
1
2

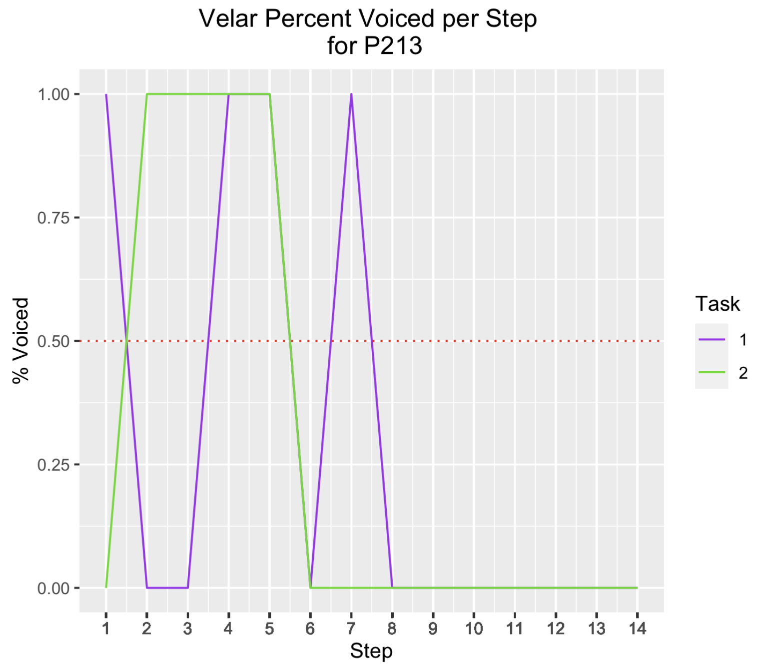
P213



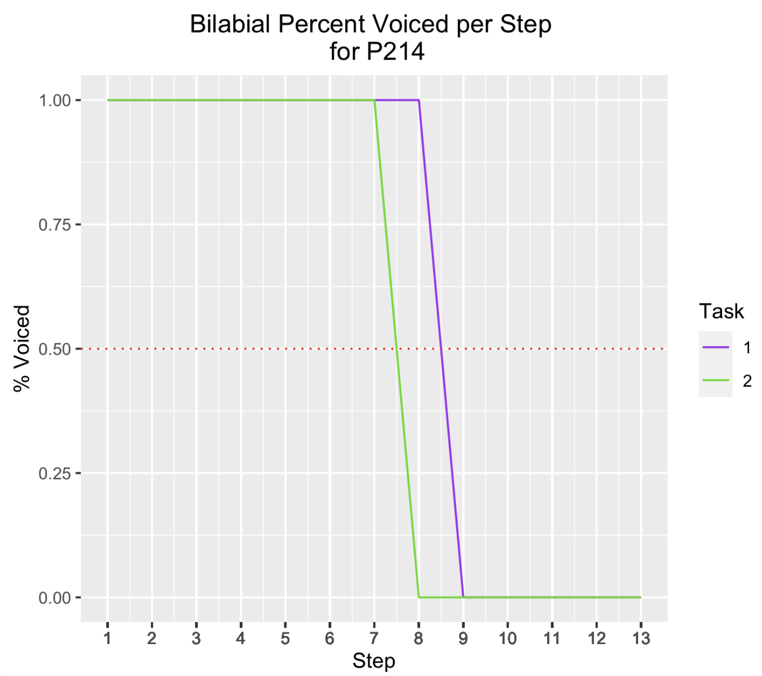
Bilabial Percent Voiced per Step
for P213
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13
Step
Task
1
2

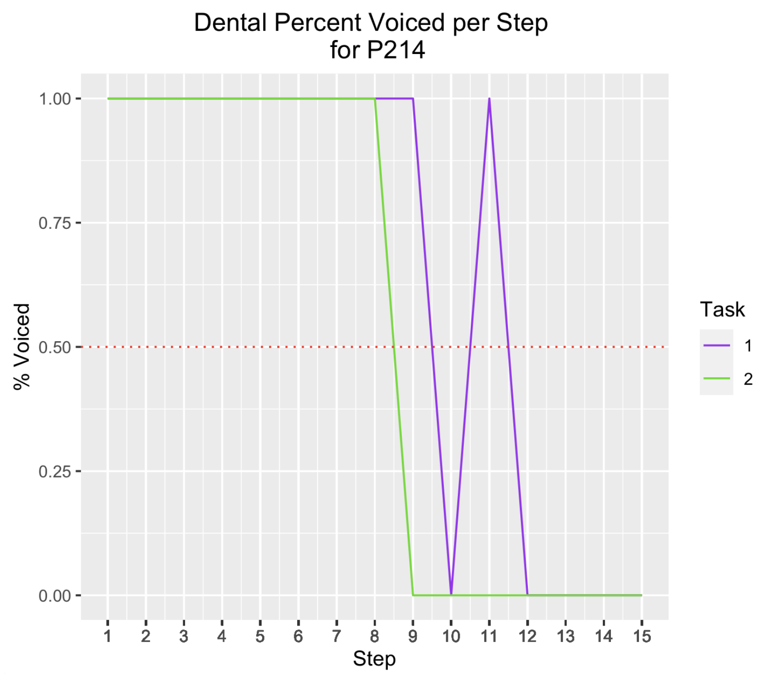


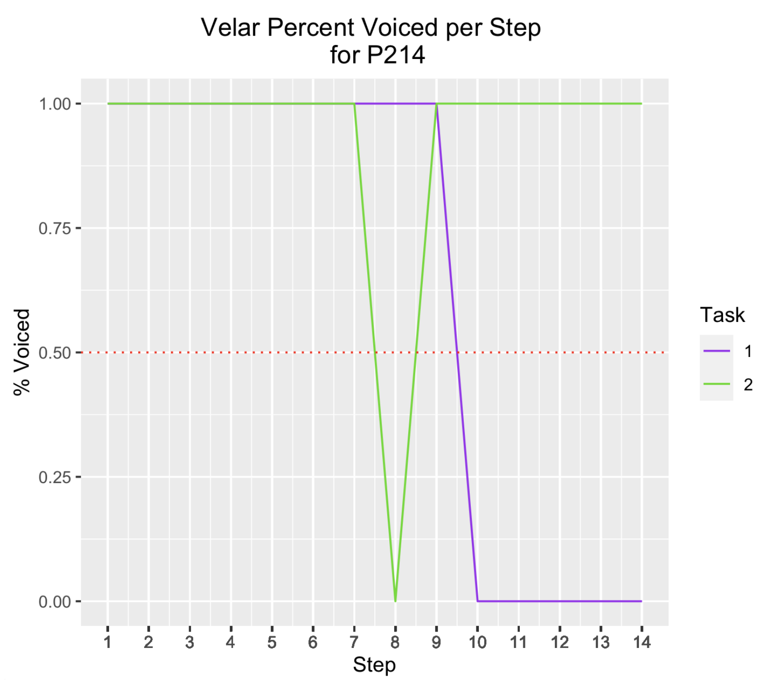
Dental Percent Voiced per Step
for P213
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Step
Task
1
2


Velar Percent Voiced per Step
for P213
% Voiced
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Step
Task
1
2

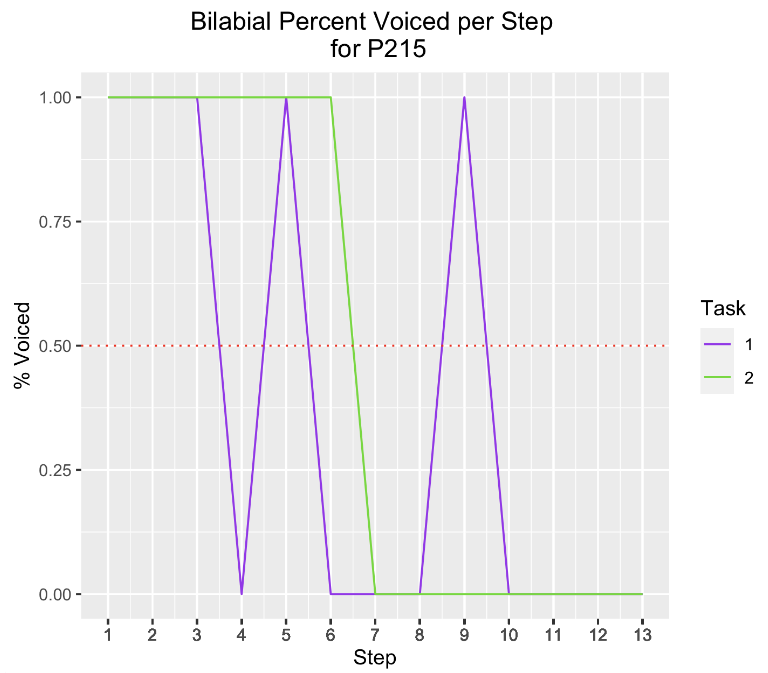
P214

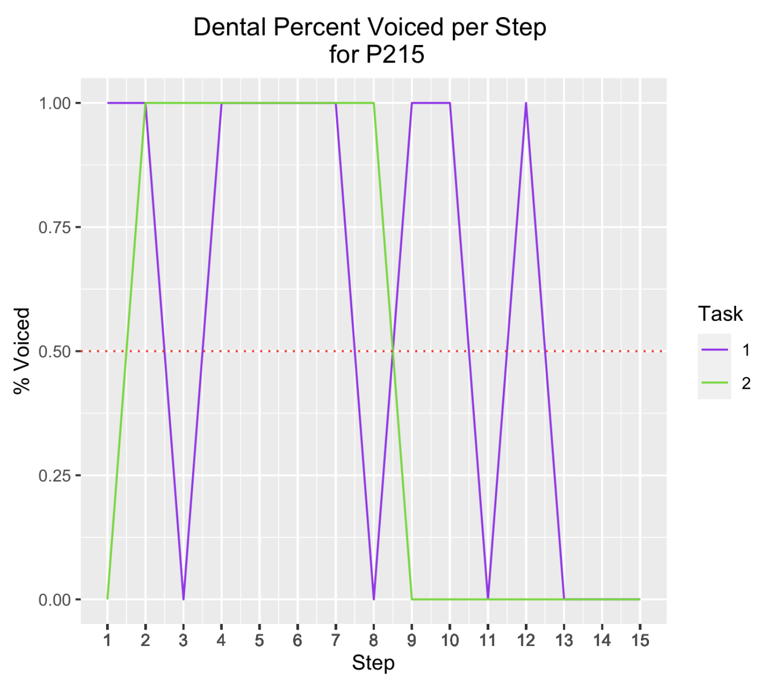

Bilabial Percent Voiced per Step
for P214
% Voiced
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10
11
12
13
Step
Task
1
2

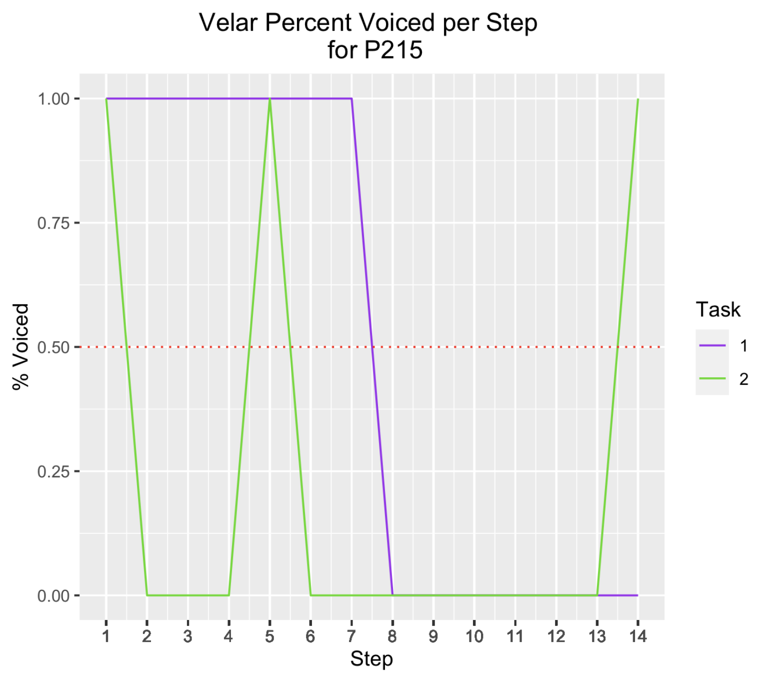

Dental Percent Voiced per Step
for P214
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Step
Task
1
2


Velar Percent Voiced per Step
for P214
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Step
Task
1
2

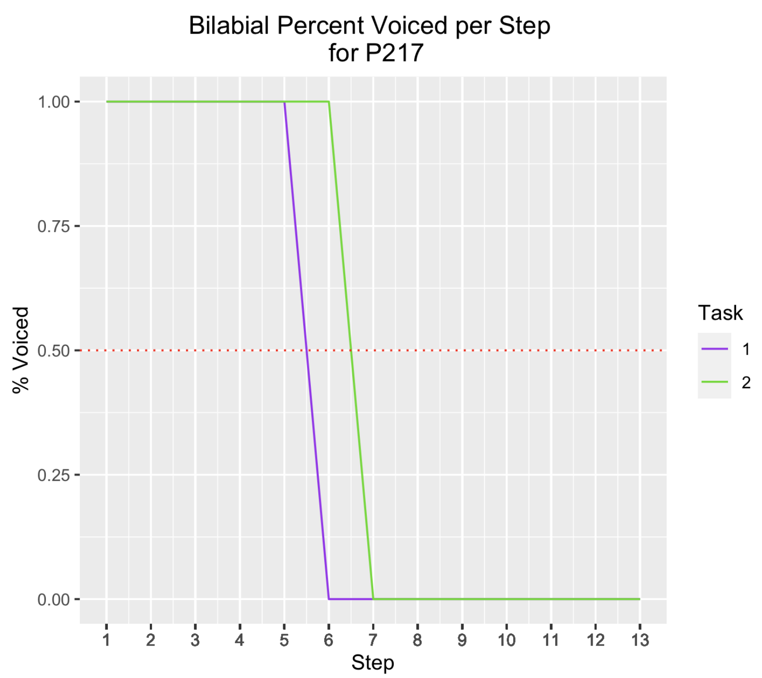
P215

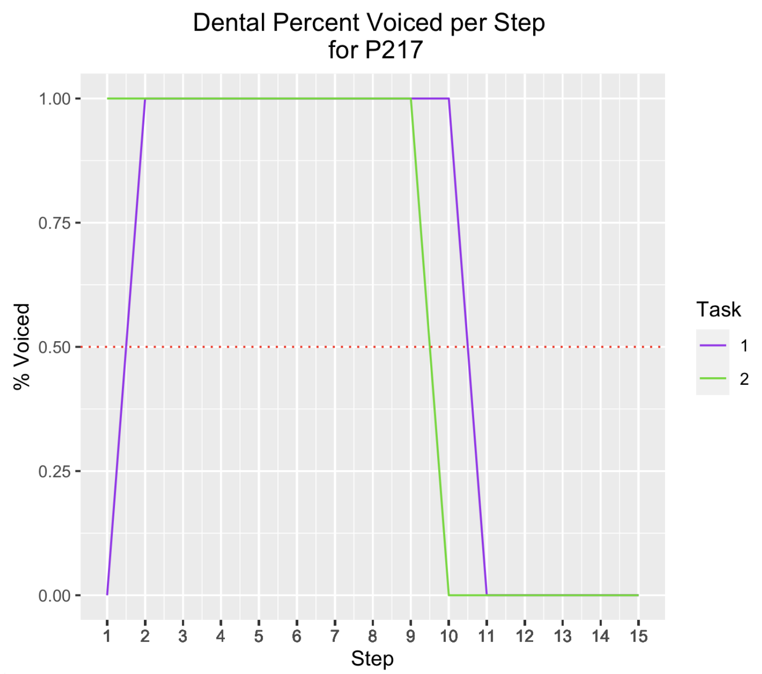

Bilabial Percent Voiced per Step
for P215
% Voiced
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10
11
12
13
Step
Task
1
2


Dental Percent Voiced per Step
for P215
% Voiced
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Step
Task
1
2

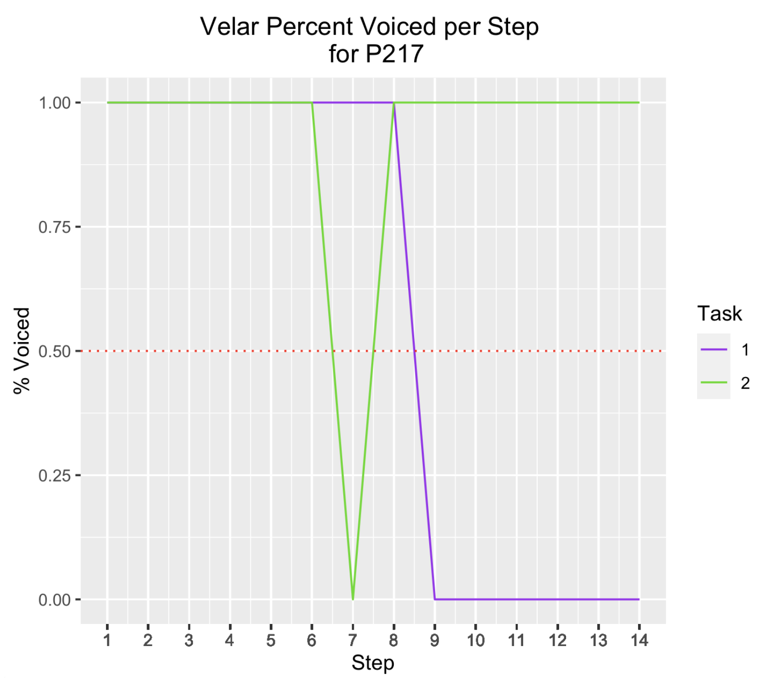

Velar Percent Voiced per Step
for P215
% Voiced
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Step
Task
1
2

P217


Bilabial Percent Voiced per Step
for P217
% Voiced
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10
11
12
13
Step
Task
1
2

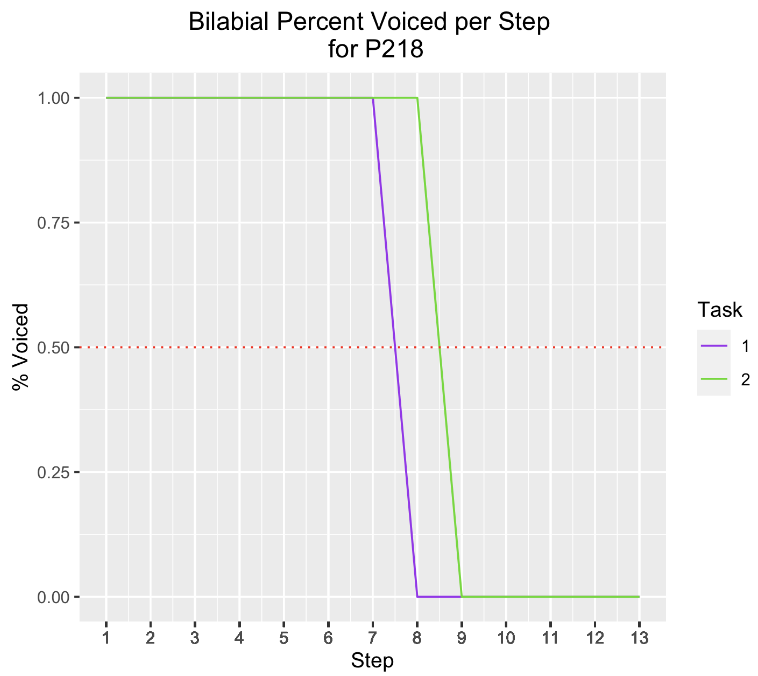


Dental Percent Voiced per Step
for P217
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Step
Task
1
2

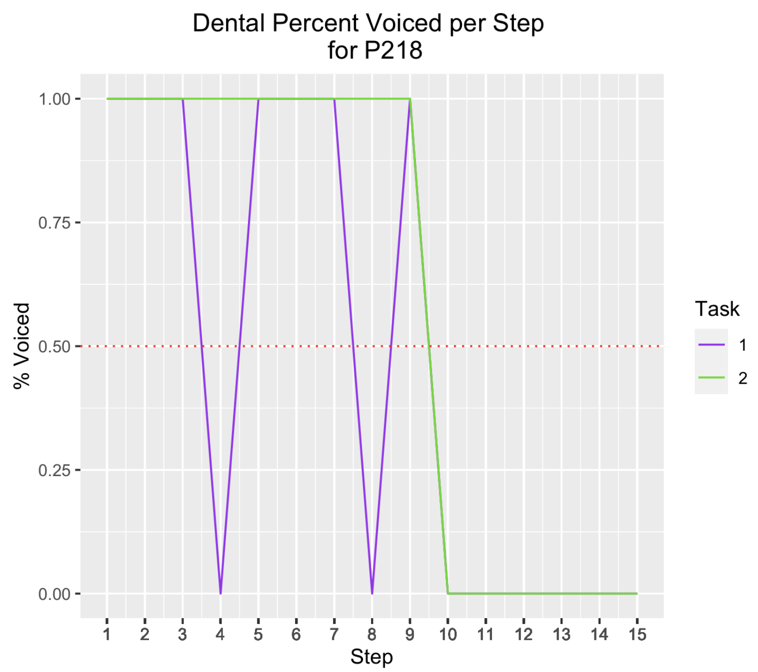


Velar Percent Voiced per Step
for P217
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Step
Task
1
2

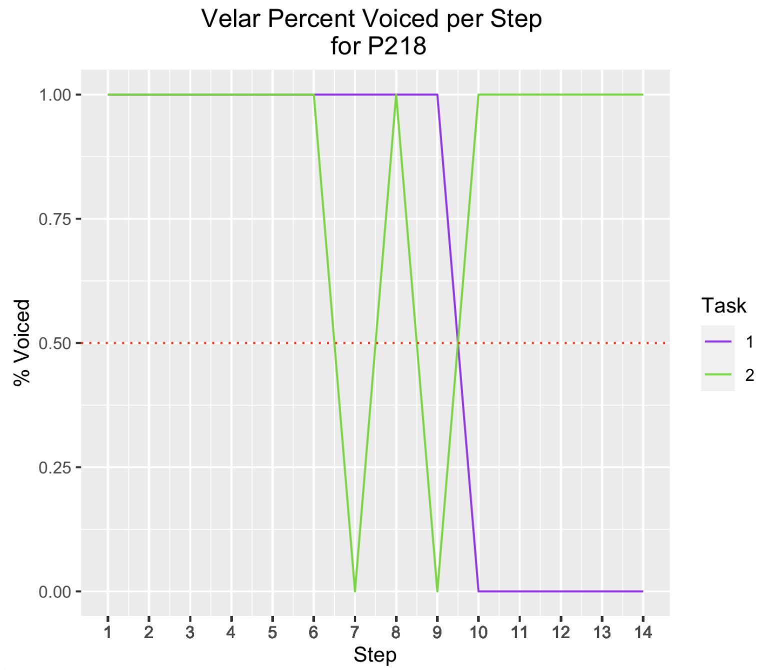
P218



Bilabial Percent Voiced per Step
for P218
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13
Step
Task
1
2

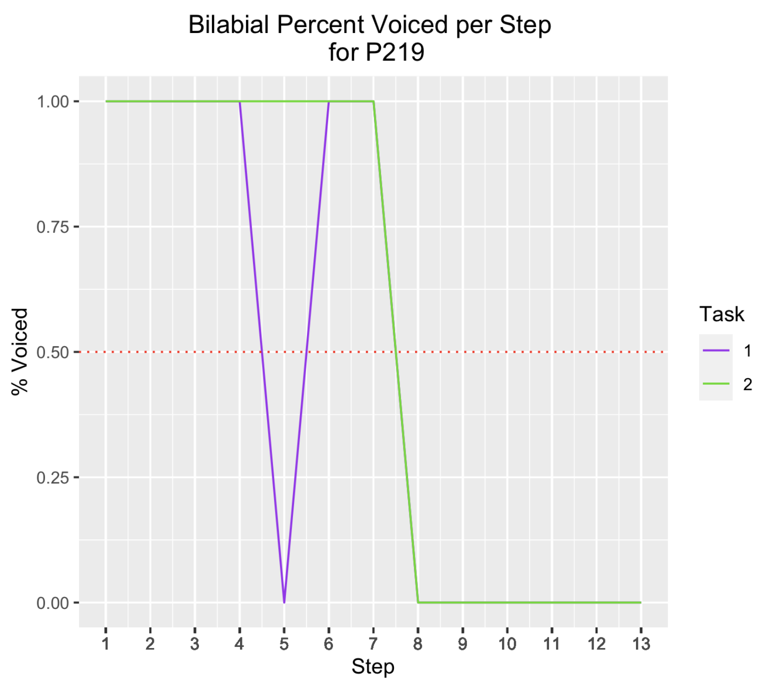


Dental Percent Voiced per Step
for P218
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Step
Task
1
2

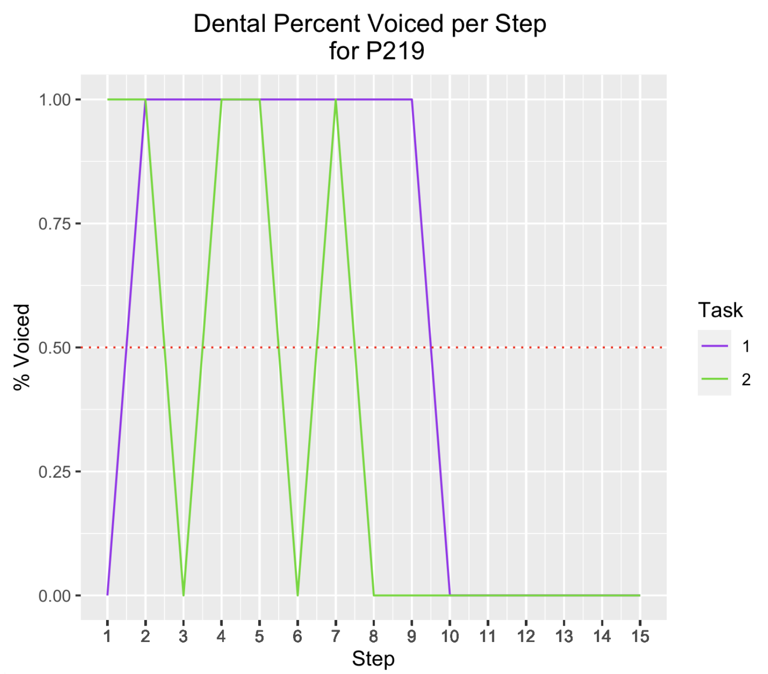


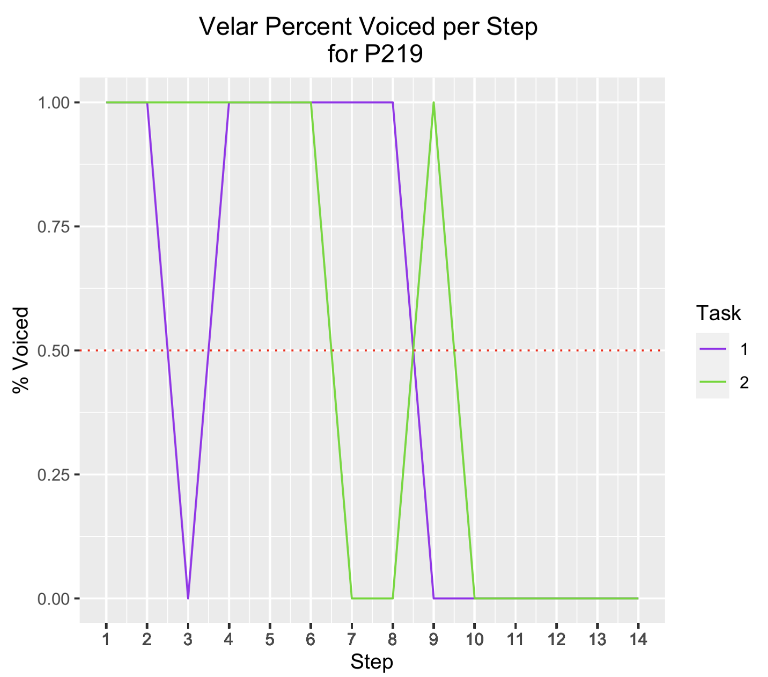
Velar Percent Voiced per Step
for P218
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Step
Task
1
2

P219

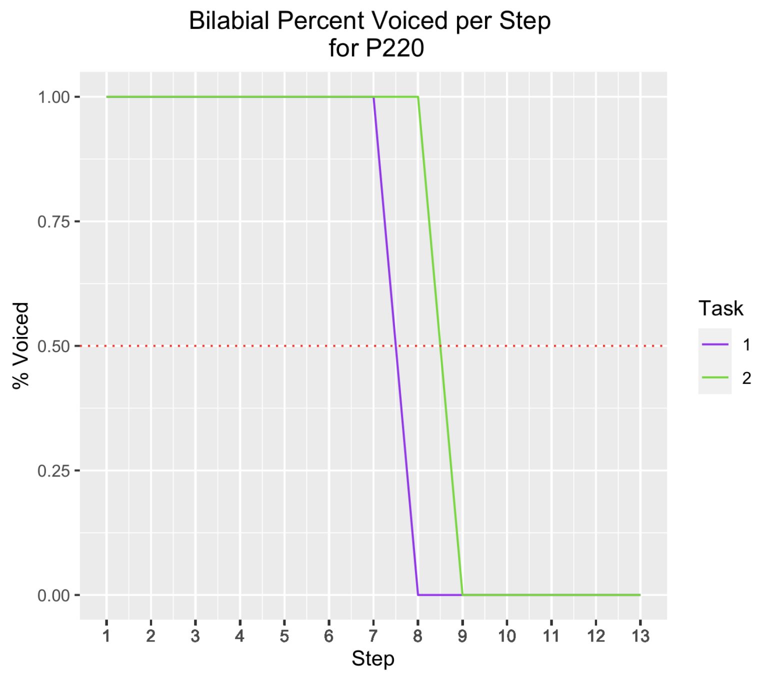


Bilabial Percent Voiced per Step
for P219
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13
Step
Task
1
2

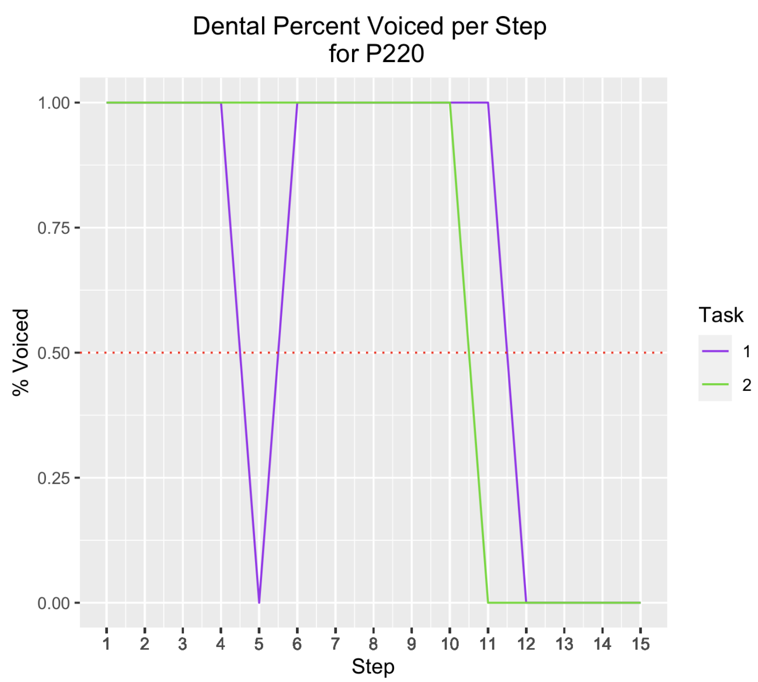

Dental Percent Voiced per Step
for P219
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Step
Task
1
2


Velar Percent Voiced per Step
for P219
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Step
Task
1
2

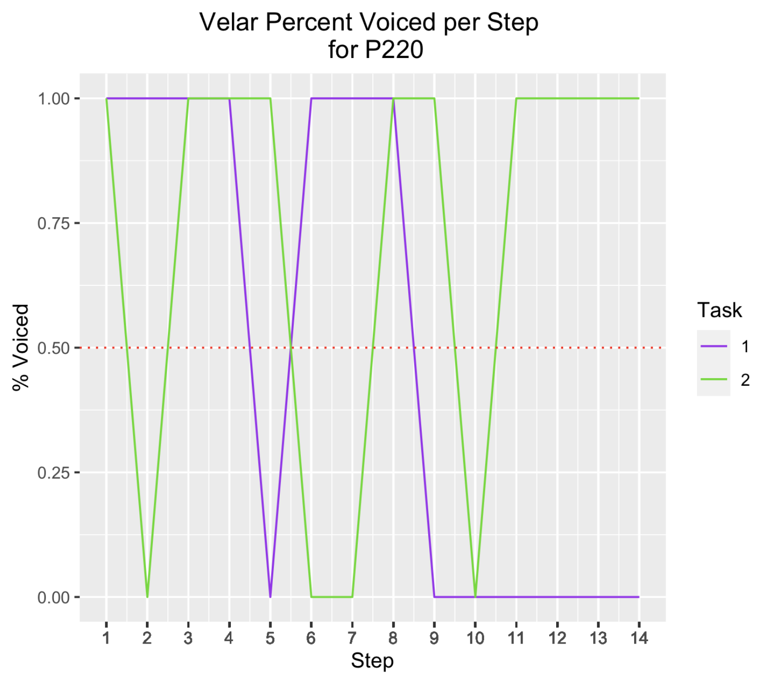
P220



Bilabial Percent Voiced per Step
for P220
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13
Step
Task
1
2

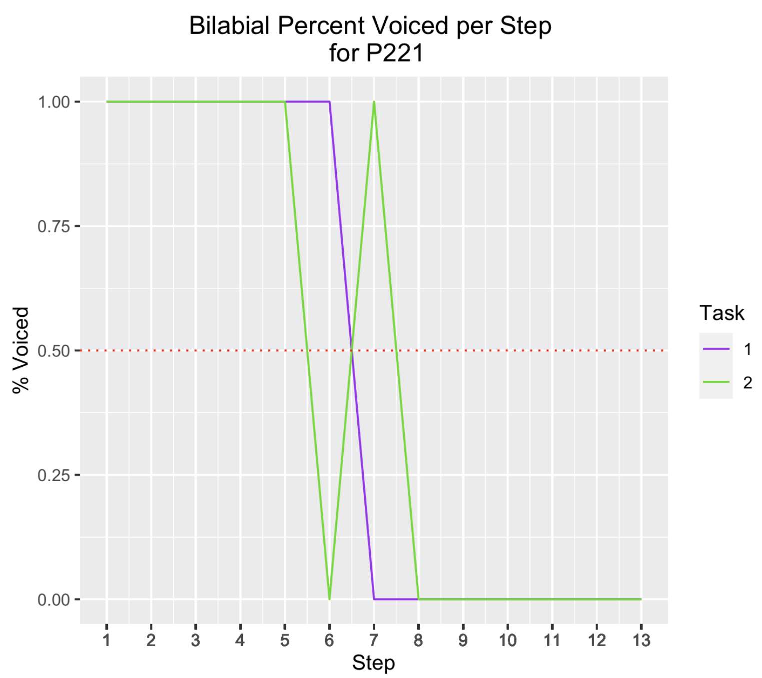


Dental Percent Voiced per Step
for P220
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Step
Task
1
2

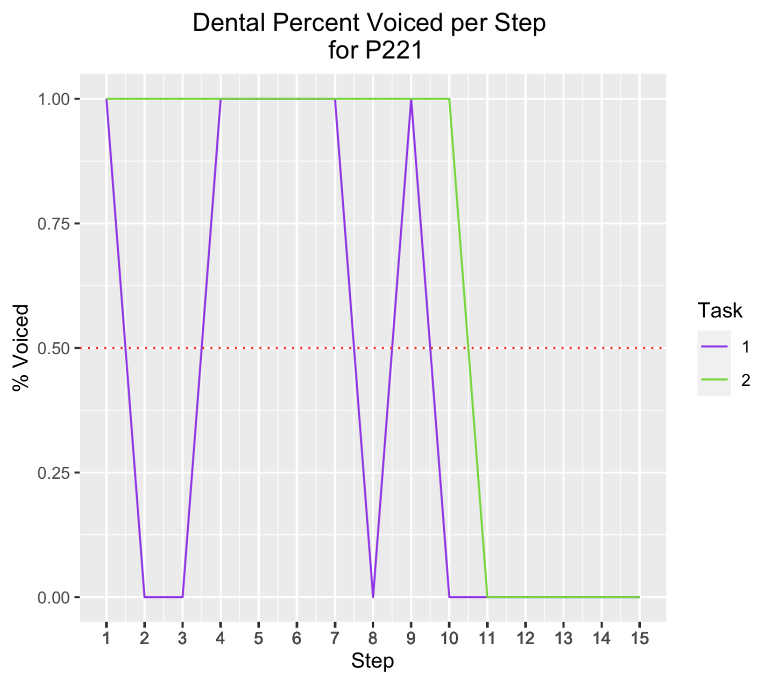


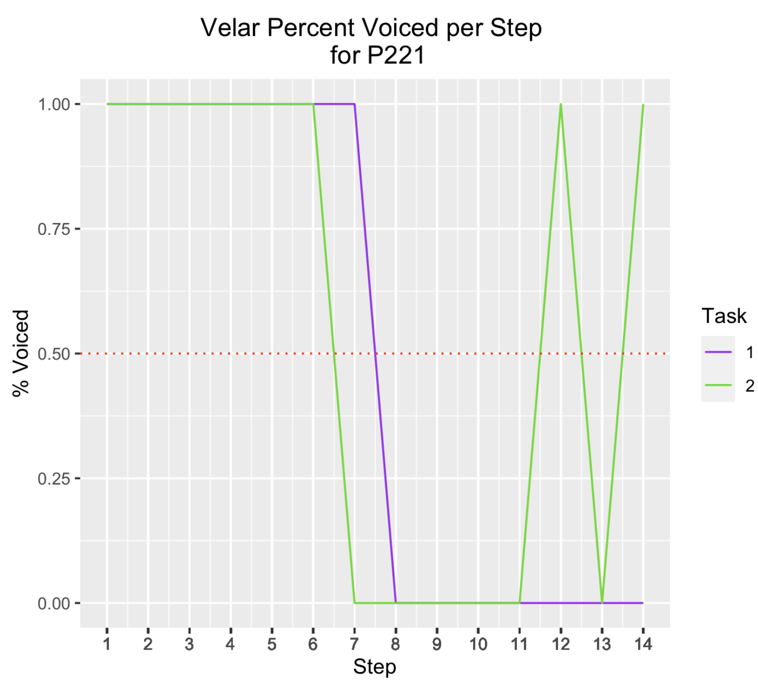
Velar Percent Voiced per Step
for P220
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Step
Task
1
2

P221



Bilabial Percent Voiced per Step
for P221
% Voiced
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13
Step
Task
1
2


Dental Percent Voiced per Step
for P221
1.00
0.75
0.50
0.25
0.00
% Voiced
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Step
Task
1
2


Velar Percent Voiced per Step
for P221
1.00
0.75
0.50
0.25
0.00
% Voiced
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Step
Task
1
2

Appendix J

Hello,

Thank you both for completing the survey and experimental session #1! Here is the link to your $25 Amazon gift card (add link).

If you are interested in scheduling the second experimental session, please follow this link to access Calendly. As a reminder, the second session should be scheduled for 4 to 7 days after the first session. Feel free to reach out to me if you have any questions. We hope to see you again soon!

Thanks again,

Madie Dunlap
mdunlap@ucla.edu

References

Albert, M. L., Goodglass, H., Helm, N. A., Rubens, A. B., & Alexander, M. P. (2013). *Clinical aspects of dysphasia* (Vol. 2). Springer Science & Business Media.

Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America*, *113*(1), 544-552.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Anthony, A., Bogle, D., Ingram, T. T. S., & McIsaac, M. W. (1971). 1971: Edinburgh articulation test. Edinburgh: Churchill Livingstone.

Antonarakis, S. E., Lyle, R., Dermitzakis, E. T., Reymond, A., & Deutsch, S. (2004). Chromosome 21 and down syndrome: from genomics to pathophysiology. *Nature reviews genetics*, *5*(10), 725-738.

Antshel, K. M., & Russo, N. (2019). Autism spectrum disorders and ADHD: Overlapping phenomenology, diagnostic issues, and treatment considerations. *Current psychiatry reports*, *21*, 1-11.

Antshel, K. M., Zhang-James, Y., & Faraone, S. V. (2013). The comorbidity of ADHD and autism spectrum disorder. *Expert review of neurotherapeutics*, *13*(10), 1117-1128.

Audacity Team (2023). Audacity(R): Free Audio Editor and Recorder [Computer program]. Version 3.3.3 retrieved October 19th 2023 from https://www.audacityteam.org/download/

Australian Bureau of Statistics (2021). 2021 Census data [*extracted via Table Builder*], Canberra.

Barletta, N. P. (2018). High-functioning autism and second language development: A case study. *Revista Interamericana de Psicologia/Interamerican Journal of Psychology (IJP)*, *52*(2), 183-193.

Baron-Cohen, S. (1988). Without a theory of mind one cannot participate in a conversation. *Cognition*, *29*(1), 83-84.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of autism and developmental disorders*, *31*(1), 5-17.

Bartak, L., Rutter, M., & Cox, A. (1975). A comparative study of infantile autism and specific developmental receptive language disorder: I. The children. *The British Journal of Psychiatry*, *126*(2), 127-145.

Bartolucci, G., Pierce, S. J., & Streiner, D. (1980). Cross-sectional studies of grammatical morphemes in autistic and mentally retarded children. *Journal of Autism and Developmental Disorders*, *10*(1), 39-50.

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Berry, L. (2022). “Hispanic or Latino Population in the US (Current ACS)” [Web Map]. ArcGIS Living Atlas. https://www.arcgis.com/home/webmap/viewer.html?webmap=bd753e29018f449891057f459db99b18

Berry, L. (2018). “Spanish Speakers in the USA” [Web Map]. ArcGIS Living Atlas. https://hub.arcgis.com/maps/arcgis-content::spanish-speakers-in-the-usa/explore?location=47.381365,-122.168217,9.00

Best, C. T. (1995). A direct realist view of cross-language speech perception. *Speech perception and linguistic experience*, 171-206.

Bialystok, E., Craik, F. I., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in cognitive sciences*, *16*(4), 240-250.

Bieber, R. E., & Gordon-Salant, S. (2024). Influence of Talker and Accent Variability on Rapid Adaptation and Generalization to Non-Native Accented Speech in Younger and Older Adults. *Auditory Perception & Cognition*, 1-30.

Bishop, J. (2012). WPP, No. 111: Focus, prosody, and individual differences in "autistic" traits: Evidence from cross-modal semantic priming.

Boersma, P. & Weenink, D. (2023). Praat: doing phonetics by computer [Computer program]. Version 6.3.18, retrieved 8 October 2023 from http://www.praat.org/

Bourdeau, L. (2009). Categorical Perception Of Stop Consonants In Children With Autism. [Masters Thesis]. STARS Electronic Theses and Dissertations. 4067. https://stars.library.ucf.edu/etd/4067

Bottema-Beutel, K., Kapp, S. K., Lester, J. N., Sasson, N. J., & Hand, B. N. (2021). Avoiding ableist language: Suggestions for autism researchers. *Autism in adulthood*, *3*(1), 18-29.

Boucher, J. (1976). Articulation in early childhood autism. *Journal of Autism and Childhood Schizophrenia*, *6*(4), 297-302.

Bradley, M. (2019). Case study: Second language acquisition with asperger syndrome in a university setting. *Research in pedagogy*, *9*(2), 167-180.

Buchanan, B. (2023). *Autism spectrum quotient (AQ)*. NovoPsych. https://novopsych.com.au/assessments/diagnosis/autism-spectrum-quotient/

Calendly. (2024). Calendly: Scheduling Automation Platform. [Scheduling software]. https://calendly.com/

Cantwell, D. P., Baker, L., & Rutter, M. (1978). A comparative study of infantile autism and specific developmental receptive language disorder: IV. Analysis of syntax and language function. *Child Psychology & Psychiatry & Allied Disciplines*.

Castañeda, M. L. (1986). El VOT de las oclusivas sordas y sonoras españolas. *Estudios de fonética experimental II*, 91-110.

Champely, S. (2020). _pwr: Basic Functions for Power Analysis_. R package version 1.3-0, <https://CRAN.R-project.org/package=pwr>.

Chenausky, K., & Tager‐Flusberg, H. (2017). Acquisition of voice onset time in toddlers at high and low risk for autism spectrum disorder. *Autism Research*, *10*(7), 1269-1279.

Chen, F., & Peng, G. (2021). Categorical perception of pitch contours and voice onset time in Mandarin-speaking adolescents with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research*, *64*(11), 4468-4484.

Chi, N. A., Washington, P., Kline, A., Husic, A., Hou, C., He, C., Dunlap, K., & Wall, D. (2022). Classifying Autism from Crowdsourced Semi-Structured Speech Recordings: A Machine Learning Approach. *arXiv preprint arXiv:2201.00927*.

Chládková, K., Boersma, P., & Escudero, P. (2022). Unattended distributional training can shift phoneme boundaries. *Bilingualism: Language and cognition*, *25*(5), 827-840.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of phonetics*, *27*(2), 207-229.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Cohen, N. J., Poldrack, R. A., & Eichenbaum, H. (1997). Memory for items and memory for relations in the procedural/declarative memory framework. *Memory*, *5*(1-2), 131-178.

Craig, F., Lamanna, A. L., Margari, F., Matera, E., Simone, M., & Margari, L. (2015).

Overlap between autism spectrum disorders and attention deficit hyperactivity disorder: searching for distinctive/common clinical features. *Autism research*, *8*(3), 328-337.

Darcy, N. T. (1953). A review of the literature on the effects of bilingualism upon the measurement of intelligence. *The Pedagogical Seminary and Journal of Genetic Psychology*, *82*(1), 21-57.

De Giacomo, A., & Fombonne, E. (1998). Parental recognition of developmental abnormalities in autism. European child & adolescent psychiatry, 7(3), 131-136.

DePape, A. M. R., Hall, G. B., Tillmann, B., & Trainor, L. J. (2012). Auditory processing in high-functioning adolescents with autism spectrum disorder.

Diehl, J. J., Bennetto, L., Watson, D., Gunlogson, C., & McDonough, J. (2008). Resolving ambiguity: A psycholinguistic approach to understanding prosody processing in high-functioning autism. *Brain and language*, *106*(2), 144-152.

Dietrich, S., & Hernandez, E. (2022). *Nearly 68 million people spoke a language other than English at home in 2019*. Census.gov. https://www.census.gov/library/stories/2022/12/languages-we-speak-in-united-states.html

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, *49*, 77-95.

Dwyer, P., Ryan, J. G., Williams, Z. J., & Gassner, D. L. (2022). First do no harm: Suggestions regarding respectful autism language. *Pediatrics*, *149*(Supplement 4).

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language learning*, *27*(2), 315-330.

Eigsti, I. M., de Marchena, A. B., Schuh, J. M., & Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders*, *5*(2), 681-691.

Eilers, R. E., Gavin, W., & Wilson, W. R. (1979). Linguistic experience and phonemic perception in infancy: A crosslinguistic study. *Child development*, 14-18.

Engelbrecht, N. (2020). *Autism Spectrum Quotient*. Embrace Autism. Retrieved June 8, 2022, from https://embrace-autism.com/tags/autism-spectrum-quotient/

Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of America*, *130*(4), EL206-EL212.

Fabiano-Smith, L., & Barlow, J. A. (2010). Interaction in bilingual phonological acquisition: Evidence from phonetic inventories. *International journal of bilingual education and bilingualism*, *13*(1), 81-97.

Fabiano-Smith, L., & Bunta, F. (2012). Voice onset time of voiceless bilabial and velar stops in 3-year-old bilingual children and their age-matched monolingual peers. *Clinical linguistics & phonetics*, *26*(2), 148-163.

Fabiano-Smith, L., & Goldstein, B. A. (2010). Early-, middle-, and late-developing sounds in monolingual and bilingual children: An exploratory investigation.

Fellbaum, M. L. (1996). The acquisition of voiceless stops in the interlanguage of second language learners of English and Spanish. In *Proceedings of Fourth International Conference on Spoken Language Processing. ICSLP'96* (Vol. 3, pp. 1648-1651). IEEE.

Flege, J. E. (1987a). A critical period for learning to pronounce foreign languages?. *Applied linguistics*, *8*(2), 162-177.

Flege, J. E. (1987b). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of phonetics*, *15*(1), 47-65.

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second language research*, *29*(3), 311-343.

Hambly, C., & Fombonne, E. (2011). The Impact of Bilingual Environments on Language Development in Children with Autism Spectrum Disorders. Journal of Autism and Developmental Disorders, 42(7), 1342–1352. doi: 10.1007/s10803-011-1365-z

Happé, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of autism and developmental disorders*, *36*, 5-25.

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. Cognition, 177, 263-277.

Harvey, I., Bolgan, S., Mosca, D., Mclean, C., & Rusconi, E. (2016). Systemizers Are Better Code-Breakers: Self-Reported Systemizing Predicts Code-Breaking Performance in Expert Hackers and Naïve Participants. Frontiers in Human Neuroscience, 10. Doi: 10.3389/fnhum.2016.00229

Hay, J. S. F. (2005). *How auditory discontinuities and linguistic experience affect the perception of speech and non-speech in English-and Spanish-speaking listeners*. The University of Texas at Austin.

Hoekstra, R. A., Bartels, M., Cath, D. C., & Boomsma, D. I. (2008). Factor structure, reliability and criterion validity of the Autism-Spectrum Quotient (AQ): a study in Dutch population and patient groups. *Journal of autism and developmental disorders*, *38*(8), 1555-1566.

Hoonhorst, I., Colin, C., Markessis, E., Radeau, M., Deltenre, P., & Serniclaes, W. (2009). French native speakers in the making: From language-general to language-specific voicing boundaries. *Journal of experimental child psychology*, *104*(4), 353-366.

Howlin, P. (2003). Outcome in high-functioning adults with autism with and without early language delays: Implications for the differentiation between autism and Asperger syndrome. *Journal of autism and developmental disorders*, *33*(1), 3-13.

Hualde, J. I. (2014). *Los sonidos del español: Spanish Language edition*. Cambridge University Press.

Jakobson, R. (1968). *Child language: aphasia and phonological universals* (No. 72). Walter de Gruyter.

Jensen, J. T. (1993). English phonology. Amsterdam: John Benjamins.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. Cognitive psychology, 21(1), 60-99.

Jordaan, H. (2008). Clinical intervention for bilingual children: An international survey. *Folia Phoniatrica et Logopaedica*, *60*(2), 97-105.

Kamio, Y., Robins, D., Kelley, E., Swainson, B., & Fein, D. (2007). Atypical lexical/semantic processing in high-functioning autism spectrum disorders without early language delay. *Journal of Autism and Developmental Disorders*, *37*(6), 1116-1122.

Kehoe, M., & Kannathasan, K. (2021). Development of voice onset time in monolingual and bilingual French-speaking children. *Lingua*, *252*, 102937.

Kellogg, J., & Chang, C. B. (2023). Exploring the onset of phonetic drift in voice onset time perception. *Languages*, *8*(1), 78.

Kissling, E. M. (2015). Phonetics instruction improves learners' perception of L2 sounds. *Language teaching research*, *19*(3), 254-275.

Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and cognitive processes*, *16*(2-3), 287-308.

Koizumi, M., Saito, Y., & Kojima, M. (2019). Syntactic development in children with intellectual disabilities–using structured assessment of syntax. *Journal of Intellectual Disability Research*, *63*(12), 1428-1440.

Kremer-Sadlik, Tamar. (2005) "To be or not to be bilingual: Autistic children from multilingual families." Proceedings of the 4th International Symposium on Bilingualism.

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *The Journal of the Acoustical Society of America*, *63*(3), 905-917.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental science*, *9*(2), F13-F21.

Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096-9101.

Kuznetsova A., Brockhoff P.B., Christensen R.H.B. (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." _Journal of Statistical Software_, *82*(13), 1-26. doi:10.18637/jss.v082.i13 <https://doi.org/10.18637/jss.v082.i13>.

Lee, M., Shetgiri, R., Barina, A., Tillitski, J., & Flores, G. (2015). Raising bilingual children: A qualitative study of parental attitudes, beliefs, and intended behaviors. *Hispanic journal of behavioral sciences*, *37*(4), 503-521.

Lenneberg, E. H. (1967). The biological foundations of language. New York: John Wiley.

Lenth, R. (2023). _emmeans: Estimated Marginal Means, aka Least-Squares Means_. R package version 1.9.0, <https://CRAN.R-project.org/package=emmeans>.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384-422.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th international congress of phonetic sciences* (Vol. 563, pp. 563-567).

Locke, J. L., & Studdert-Kennedy, M. (1983). *Phonological acquisition and change* (p. xix263). New York: Academic Press.

Lord, C., & Paul, R. (1997). Language and communication in autism. DJ Cohen & FR Volkmar (Eds.), Handbook of autism and pervasive development disorders.

Lozano-Argüelles, C., Arroyo, L. F., Rodríguez, N., López, E. M. D., Pozú, J. J. G., Markovits, J., Varela, J. P., de Rocafiguera, N., & Casillas, J. V. (2021). Conceptually cued perceptual categorization in adult L2 learners. *Studies in second language acquisition*, *43*(1), 204-219.

Macken, M., & Barton, D. (1980a). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language, 7*(1), 41-74. doi:10.1017/S0305000900007029

Macken, M. A., & Barton, D. (1980b). The acquisition of the voicing contrast in Spanish: A phonetic and phonological study of word-initial stop consonants. *Journal of Child*

Language, *7*(3), 433-458.

MacMillan, K. A. R. (2015). Phonological acquisition by children with autism: a case study.

Maenner, M. J., Shaw, K. A., Bakian, A. V., Bilder, D. A., Durkin, M. S., Esler, A., ... & Cogswell, M. E. (2021). Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2018. *MMWR Surveillance Summaries*, *70*(11), 1.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism: Language and cognition*, *6*(2), 97-115.

McCann, J., Peppé, S., Gibbon, F. E., O'Hare, A., & Rutherford, M. (2007). Prosody and its relationship to language in school‐aged children with high‐functioning autism. *International Journal of Language & Communication Disorders*, *42*(6), 682-702.

Michnowicz, J., & Carpenter, L. (2013). Voiceless stop aspiration in Yucatan Spanish: A sociolinguistic analysis. *Spanish in Context*, *10*(3), 410-437.

Mielke, J., Nielsen, K., & Magloughlin, L. V. (2013, June). Phonetic imitation by individuals with autism spectrum disorders: Investigating the role of procedural and declarative memory. In *Proceedings of Meetings on Acoustics ICA2013* (Vol. 19, No. 1, p. 060142). Acoustical Society of America.

Moffitt, A. R. (1971). Consonant cue perception by twenty-to twenty-four-week-old infants. *Child Development*, 717-731.

Nation, K., Clarke, P., Wright, B., & Williams, C. (2006). Patterns of reading ability in children with autism spectrum disorder. *Journal of autism and developmental disorders*, *36*(7), 911-919.

Nordgren, P. M. (2015). Phonological development in a child with autism spectrum condition: Case study of an intervention. *Journal of Interactional Research in Communication Disorders*, *6*(1), 25.

Ohashi, J. K., Mirenda, P., Marinova-Todd, S., Hambly, C., Fombonne, E., Szatmari, P., Bryson, S., Roberts, W., Smith, I., Vaillancourt, T., Volden, J., Waddell, C., Zwaigenbaum, L., Georgiades, S., Duku, E., Thompson, A., & Pathways in ASD Study Team. (2012). Comparing early language development in monolingual-and bilingual-exposed young children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *6*(2), 890-897.

Paradis, J., Crago, M., Genesee, F., & Rice, M. (2003). French-English bilingual children with SLI.

Peña, E. D., Bedore, L. M., Shivabasappa, P., & Niu, L. (2020). Effects of divided input on bilingual children with language impairment. *International journal of bilingualism*, *24*(1), 62-78.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of child psychology and psychiatry*, *37*(1), 51-87.

Poarch, G. J., & Krott, A. (2019). A bilingual advantage? An appeal for a change in perspective and recommendations for future research. *Behavioral Sciences*, *9*(9), 95.

Priester, G. H., Post, W. J., & Goorhuis-Brouwer, S. M. (2011). Phonetic and phonemic acquisition: Normative data in English and Dutch speech sound development. *International journal of pediatric otorhinolaryngology*, *75*(4), 592-596.

Qualtrics XM. (2024). Qualtrics XM: The Leading Experience Management Software (Version January 2024) [Survey software]. https://www.qualtrics.com/

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ricks, D. M., & Wing, L. (1975). Language, communication, and the use of symbols in normal and autistic children. *Journal of autism and childhood schizophrenia*, *5*(3), 191-221.

Rogers, S. J., Hayden, D., Hepburn, S., Charlifue-Smith, R., Hall, T., & Hayes, A. (2006). Teaching young nonverbal children with autism useful speech: A pilot study of the Denver model and PROMPT interventions. *Journal of autism and developmental disorders*, *36*(8), 1007-1024.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. URL http://www.rstudio.com/.

Ruzich, E., Allison, C., Chakrabarti, B., Smith, P., Musto, H., Ring, H., & Baron-Cohen, S. (2015a). Sex and STEM Occupation Predict Autism-Spectrum Quotient (AQ) Scores in Half a Million People. Plos One, 10(10). doi: 10.1371/journal.pone.0141229

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., & Baron-Cohen, S. (2015b). Measuring autistic traits in the general population: a systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Molecular Autism*, *6*(1), 2. doi: 10.1186/2040-2392-6-2.

Saljughian, F. S. (2012). A contrastive study of Persian and English sounds. *NUML Journal of Critical Inquiry*, *10*(2), 32.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of phonetics*, *52*, 183-204.

Schoen, E., Paul, R., & Chawarska, K. (2011). Phonology and vocal behavior in toddlers with autism spectrum disorders. Autism Research, 4(3), 177-188.

Seigfried-Spellar, K. C., Oquinn, C. L., & Treadway, K. N. (2015). Assessing the relationship between autistic traits and cyber deviancy in a sample of college students. Behaviour & Information Technology, 34(5), 533–542. doi: 10.1080/0144929x.2014.978377

Shapiro, T., & Kapit, R. (1978). Linguistic negation in autistic and normal children. *Journal of Psycholinguistic Research*, *7*(5), 337-351.

Shore, S. (2016). If you've met one person with autism, you've met one person with autism.

Sims, M., & Ellis, E. M. (2014). Raising children bilingually is hard: Why bother?. *Babel*, *49*(2), 28-35.

Snyder, C., Cohn, M., & Zellou, G. (2019). Individual Variation in Cognitive Processing Style Predicts Differences in Phonetic Imitation of Device and Human Voices. In *Interspeech* (pp. 116-120).

Stölten, K., Abrahamsson, N., & Hyltenstam, K. (2014). Effects of age of learning on voice onset time: Categorical perception of Swedish stops by near-native L2 speakers. *Language and Speech*, *57*(4), 425-450.

Tager-Flusberg, H. (1985). Basic level and superordinate level categorization by autistic, mentally retarded, and normal children. *Journal of Experimental Child Psychology*, *40*(3), 450-469.

Tager-Flusberg, H. (2006). Defining language phenotypes in autism. *Clinical Neuroscience Research*, *6*(3-4), 219-224.

Tager-Flusberg, H., Calkins, S., Nolin, T., Baumberger, T., Anderson, M., & Chadwick-Dias, A. (1990). A longitudinal study of language acquisition in autistic and Down syndrome children. *Journal of autism and developmental disorders*, *20*(1), 1-21.

Tek, S., Jaffery, G., Fein, D., & Naigles, L. R. (2008). Do children with autism spectrum disorders show a shape bias in word learning?. *Autism Research*, *1*(4), 208-222.

Valicenti-McDermott, M., Tarshis, N., Schouls, M., Galdston, M., Hottinger, K., Seijo, R., Shulman, L., & Shinnar, S. (2013). Language differences between monolingual English and bilingual English-Spanish young children with autism spectrum disorders. *Journal of child neurology*, *28*(7), 945-948.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D. (2023). _dplyr: A Grammar of Data Manipulation_. R package version 1.1.3, <https://CRAN.R-project.org/package=dplyr>.

Winn, M. B. (2020a). Manipulation of voice onset time in speech stimuli: A tutorial and flexible Praat script. *The Journal of the Acoustical Society of America*, *147*(2), 852-866.

Winn, M. B. (2020b). Praat: Scale intensity of all sounds in a Directory. [Computer program]. https://raw.githubusercontent.com/ListenLab/Praat/master/Scale_intensity_all_sounds_in_folder_v1.txt.

Wolk, L., & Edwards, M. L. (1993). The emerging phonological system of an autistic child. *Journal of Communication Disorders*, *26*(3), 161-177.

Wolk, L., & Giesen, J. (2000). A phonological investigation of four siblings with childhood autism. *Journal of Communication Disorders*, *33*(5), 371-389.

Wu, H., Lu, F., Yu, B., & Liu, Q. (2020). Phonological acquisition and development in Putonghua-speaking children with autism spectrum disorders. *Clinical Linguistics & Phonetics*, *34*(9), 844-860.

You, R. S., Serniclaes, W., Rider, D., & Chabane, N. (2017). On the nature of the speech perception deficits in children with autism spectrum disorders. *Research in developmental disabilities*, *61*, 158-171.

Yu, A. C., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and “autistic” traits. *PloS one*, *8*(9), e74746.

Yu, G. (2021). *The influence of sentence context on the categorical perception of English stop VOT in individuals with a high Autism Spectrum Quotient (AQ)*. [PLIN0025 Long Essay] University College London. https://www.lagb.org.uk/resources/Documents/Yu%20UCL%20OUDIL_web.pdf

Zampini, M. L. (2013). Voice Onset Time in Second Language Spanish. *The handbook of Spanish second language acquisition*, 113-129.

Zhang, L., Tsung, L., & Qi, X. (2023). Home language use and shift in Australia: Trends in the new millennium. *Frontiers in Psychology*, *14*, 1096147.