The Emergence of Gendered Phonetic Variation in Preschool Children: Findings from a Longitudinal Study

A THESIS

SUBMITTED TO THE FACULTY OF

UNIVERSITY OF MINNESOTA TWIN CITIES

BY

Kiana Koeppe

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF ARTS IN SPEECH LANGUAGE PATHOLOGY

Benjamin Munson, Ph.D.

June 2021

Acknowledgements

I would like to thank my advisor, Dr. Benjamin Munson for guiding me, encouraging me, and pushing me to my fullest potential throughout both my full collegiate career, and this thesis journey. Thank you to my committee members, Dr. Kerry Ebert and Dr. Erin Durban, for your expertise, feedback, and flexibility in uncertain times. I would also like to thank Natasha Lackas for not only helping pave the way for my thesis, but also for providing me with much needed support and advice. Finally, thank you to the research assistants in the Learning to Talk Lab, Gisela Smith and Katherine Dougherty, for the hours of tedious transcribing and coding work you put in. Without the work of all these individuals, this thesis would not have been possible; Thank you all for your contributions.

Abstract

Gendered speech variation has been found in adults and children. In adults, sexual dimorphism is an important component of this variation, but prepubertal children lack this anatomical differentiation. Research has shown that adults also use learned behaviors to perform their gender, and a growing body of research has suggested that gendered speech variation in children is also due to learning. One of those learned sociophonetic variations is seen in the production of /s/. In this study, the development and variation of /s/ and /ʃ/ production between 55 children assigned male at birth (AMAB) and 55 children assigned female at birth (AFAB) was analyzed. A systematic comparison of /s/ and /ʃ/ accuracy and spectral properties at 28-39 months old and at 53-66 months old suggested that /s/ variation is a possible gender marker that is learned early in life.

Table of Contents

List of Tables

List of Figures

1. Introduction

Speech is a powerful tool that humans use to vocally communicate thoughts, ideas, wants, needs, and other messages. Through phonetic variation alone, speech has also been shown to convey social information about the speaker. The study of these social-cueing, phonetic variables is called *sociophonetics*. Sociophonetics examines phonetic variation that is associated with (and, potentially, capable of revealing) social information about a speaker, including, among others, macrosociological information like ethnicity, race, age, and locally defined information, such as memberships in different social groups (Foulkes & Docherty, 2006; Babel & Munson, 2014). Numerous lines of research (e.g., Stuart-Smith, 2007) have shown that gender is one of those sociophonetic pieces of information shared through speech, and is the primary focus of this study.

Imagine you receive a phone call from an unknown number. Upon answering the call, you hear the speaker's voice on the other line. Within the first 170-230 ms, well before they introduce themselves, it is likely that you have already made a judgement about the speaker's gender based on acoustics alone (Latinus & Taylor, 2012). This is because the human brain is excellent at receiving information, picking up on patterns, and making connections. When you answer that phone call, your brain quickly gathers acoustic information about the speaker, and compares these data points to previously gathered experiences, references, and cultural stereotypes. If a low pitched voice is presented, it would be common in Western societies to determine that the speaker is a male. If you hear a high pitched voice on the other line, the common assumption in Western culture is that you are speaking to a female. There are a number of anatomical and social reasons, both of which will be discussed later, for those pitch to gender associations to be made.

Now imagine you are answering a call from a 4 year old child whose gender you do not know. Just like the females in the adult example, the child will likely have a relatively high pitched voice. This child does not have any articulation errors, and unlike the adults, there are no anatomical reasons for a male child to sound different than a female child. Even still, by perceiving learned sociophonetic variation in the child's speech, it is possible that you will make a guess of the gender of the child. While that guess is not likely to be as accurate as your guess of an adult's gender, it is likely to be better than chance. This thesis investigates those learned, gender-cueing, phonetic variations that children use in their everyday speech.

1.1 Sex and Gender: Operational Definitions

Sex and gender are two extremely complex and multifaceted concepts, so much so that there are entire fields of study dedicated to them. Sex and gender are hugely important ideas that are constantly shaping individual identities, politics, human rights, and society as a whole, but explaning the full breadth of them is outside the scope of this paper. While a simple definition of each would not be sufficient in depicting the entirety of these notions, differentiating between sex and gender, as well as determining operational definitions for both are important to establish for this study. Readers interested in more detailed information on sex and gender, are referred to the works of Judith Bulter, and Penelope Eckert and Sally McConnell-Ginet, among others. A few key points are especially relevant here. First, as emphasized by Butler (1990), gender represents a performance, that is, the construction of an identity through acts, including linguistic acts. This performance need not be solely about masculinity or femininity. As emphasized by Eckert and McConnell-Ginet (1992), the performance of gender occurs in different communities of practice that people engage in. Hence, the construction of gender is best understood in the context of the different communities of practice that people interact in.

Even with the abundance of literature on sex and gender, many people still use and believe that the terms can be interchanged, despite the fact that they refer to two unique concepts. For the sake of this paper, *sex* refers to biological traits, such as chromosomes, gene expression, hormone levels, sex organs, and other physical features. At birth, and sometimes even earlier, when the fetus is in the womb, children are often assigned a sex, either male or female based on the above listed biological markers. It is important to recognize that while male and female are the dominating categories of sex, they are not all-encompassing and not as straight forward as many believe them to be. Individuals whose biological markers do not align with the traditional groups of male or female are generally referred to as an additional category, intersex. For example, an individual with ovaries, XX chromosomes (typically categorized as female), but male external genitalia would be considered intersex. There are a wide variety of possible presentations of sex, making the binary default of male/female problematic and exclusionary to people outside the socially constructed norm. This results in limited respresentation for intersex individuals, even though this presentation of sex is more common than most people probably realize. The exact number is difficult to determine, but experts believe that up to 1.7% of the population display intersex traits (Blackless et al., 2000; Fausto-Sterling, 2000). When intersex babies are born, it is common for the parents and medical practitioners to decide on whether the child should be assigned as male or female, oftentimes resulting in surgeries and raising the child with cisnormative expectations of the child's gender. For all children, not just intersex, being reared and socialized as a pre-determined sex and gender does not necessarily mean that the child's identity, future physical expression, and gender performance will align with the socially constructed ideals of those assigned labels.

The second important term to characterize for this study is *gender*. As stated broadly by Munson and Babel (2019), "gender is a set of practices that individuals engage in that reinforce a social division between men and women." It often refers to social roles (e.g., earning money/raising children), behaviors, and performances (Butler, 1990). In some cases, these performances are of masculinity and femininity, though it goes substantially beyond just that. Gender, in many Western cultures, has traditionally been thought of as a binary (man/woman). The binary expectation, as well as gender presentations that are deemed appropriate, have been shaped by heteronormativity and the assumption that there are two opposite genders who are sexually attracted to each other (heterosexuality). This is not reality and cannot be assumed as the default as sexuality, like gender, is broad and multifaceted. The gender binary has proven inadequate given the complexity and intersectional relationship that exists between gender and other social categories such as class, age, race, and sexuality. As Eckert (2014) pointed out, the oversimplification of gender as a binary has resulted in the idea that individuals with the same gender have the same experience. This is untrue, and proves one of many fatal flaws in the gender binary. For example, the experience of white women is vastly different from that of women of color (Crenshaw, 1991), demonstrating how the universal of ‘woman’ is insufficient in accounting for the variety of existences. Additionally, gender is not determined by biology. A child assigned male at birth (AMAB) is not inherently a boy, just as a child assigned female at birth (AFAB) is not inherently a girl. For the remainder of this paper, AMAB and AFAB will be used for these terms. As put nicely by Simone de Beauvoir (1973), “one is not born, but rather, becomes a woman” (p. 301). That is self-identifying and performing as a woman makes you a woman. Although specific genders are often associated with a specific assigned sex in the West, a perfect one-to-one correlate does not exist. Western society has reinforced the belief that individuals AMAB have to be masculine and present as boys or men, and individuals AFAB have to be feminine and perform as girls or women. In circumstances like this, where an individual’s assigned sex does align with society’s gender-correlate, the term used is cisgender. Today, there are numerous recognized gender identities that non-cisgender individuals use to describe themselves. Some common labels include nonbinary, transgender, gender queer, polygender, omnigender, genderfluid, bigender, and agender.

1.2 Sex and Gender in Adult Speech

As stated earlier, speech can communicate information about the speaker beyond the meaning of the words said. In adult speakers, speech has been shown to relay information about whether a person is male or female. This research has been conducted in the context of cisnormative institutions and practices. Hence, while previous research uses terms like ‘sex’, ‘gender’, ‘male’, and ‘female’, research is sparse on whether this refers to cisgender individuals only. Given the historical context, we interpret this research as being about characteristics of cisgender individuals. A number of studies have shown that there are a variety of acoustic and perceptual differences between adult men and adult women's speech (e.g., Klatt and Klatt, 1990; Iseli et al., 2007; Chen, 1997; Peterson and Barney, 1952; Jongman, et al., 2000; Podesva and Van Hofwegen, 2014; Whiteside and Irving, 1998; and Foulkes and Docherty, 2006).

There are biological/anatomical explanations for much of this variation seen between men and women's speech. The major biological driving forces for these differences are related to the concept of sexual dimorphism. Sexual dimorphism refers to significant physical differences that exist between males and females of the same species, beyond the sex-specific sex organs that are part of sex assigned at birth (SAB). This can include differences in shape, color, structure, and, in the instance of gendered speech in adult humans, size. Adult men have been shown to have longer vocal tracts, and longer, more massive vocal folds than women (Simpson, 2009; Titze, 1989). According to Simpson (2009), the distance from the vocal folds to the lips for the average male is between 17cm and 18cm, whereas that distance is between only 14cm to 14.5cm in females. As for the vocal folds, the membranous length in adult females is about 10mm and 16mm in males. With that, the vocal folds of males are between 20% and 30% thicker than females (Titze, 1989).

Sexually dimorphic articulatory structures in adult vocal tracts have a predictable impact on speech acoustics. The effects of anatomical variation on the acoustic characteristics of speech are described by source-filter theory (Chiba & Kajiyama, 1941). This theory describes sounds' acoustic properties as being the product of a sound source and a filter. One of the most common sound sources for speech is the waveform generated by the larynx, specifically vocal-fold vibration. The sound that is generated by vocal-fold vibration is modified by a filter, the vocal tract. Variations in the larynx and in the vocal tract influence the acoustics of speech. These variations can be due to anatomical differences between individuals (i.e., one individual has longer, thicker vocal folds and a shorter vocal tract than another). They can also be influenced by articulatory maneuvers that a single person produces, as, for example, when a person makes their vocal folds vibrate more tense and hence vibrate more quickly, or when a person raises or lowers their larynx to change the size of their vocal tract.

One way that sexual dimorphism impacts speech acoustics is by shaping fundamental frequency. Fundamental frequency is a measure of vocal fold vibration, and is directly impacted by the measurements of the vocal folds. Longer, more massive vocal folds, like those seen in cisgender males, typically result in a slower vibration rate, which produces a lower fundamental frequency. In conversational speech, the fundamental frequency for cis males is between 100 to 150 Hz and between 180 and 250 Hz for cis females. The perceptual correlate of fundamental frequency is pitch, which means that the sexual dimorphism in adult vocal folds plays a role in the experience of a lower voice in cis men and a higher voice in cis women. Variations in vocal fold thickness and closure are also responsible for the breathy female voice described in Klatt and Klatt (1990).

Formant frequencies are also impacted by sexual dimorphism. Formants are concentrations of acoustic energy at a particular frequency, and are directly related to the shape (i.e., configuration of the articulators) and size of the vocal tract. Every speech sound results in different formant frequencies, as each is uniquely filtered by the vocal tract, resulting in variable resonance. A longer vocal tract, like those seen in cis males, means a longer resonating tube for the sound to pass through. The result of this is a longer wavelength and, again, lower frequencies. The formant frequencies of cis female are approximately 20% higher and more spaced out than cis male's (Hillenbrand et al., 1995). In gender rating studies, fundamental frequency, formant frequencies, and vocal tract resonance have all been shown to differentiate between cis men and cis women, and to affect listeners' judgments of whether a person is a man or a woman (Hillenbrand & Clark, 2009; Skuk & Schweinberger, 2014). While these measures are constrained by anatomy, there is still a range that can be used.

Even though sexual dimorphism is an important component of male and female speech, within the constraints of the vocal mechanism, people do still have agency when it comes to how their speech is presented, as they can select where within their range to speak. That is, people volitionally modify the shape of their vocal folds and tracts to perform gender. Just as gender is a social construct, the ideals of how each gender should speak is also culturally constructed. In many Western societies, low pitched voices are masculine and belonging to men, whereas higher frequency voices are associated with women or femininity (Avery & Liss, 1996). In fact, within American college students, lower fundamental frequencies result in increased perceptions of men's dominance (Puts et al., 2007) and attractiveness (Feinberg, et al., 2005). These two socially constructed ideals for men can be enhanced through manipulation of the vocal mechanism.

Van Bezooijen (1995) showed how cultural values might predispose people to speaking in a particular area of their fundamental frequency range. This study revealed that in Japanese culture, high pitched voices were viewed as more feminine and ideal, whereas, in the Netherlands, medium or low pitched voices in women were viewed more positively. Reflecting those social ideals, the habitual pitch for women in Japan is higher than that of Dutch women. This variation cannot be explained by anatomy, and suggests that fundamental frequency in adults must be partially learned, as the women were modifying their pitch to align with social ideals of gender from their own culture.

Pitch modification is not the only learned gender variation in adult speech. Bradlow et al. (1996) showed that females speak with more precise consonant articulation and more contrasted vowels than males, feeding into the social belief that women speak clearly, while men mumble. Heffernan (2010) then went on to further investigate phonological contrast by studying the speech of eight male radio DJs and listeners' perceptions of them. Results from this study showed that males who produced the least phonetically distinct speech (whom Heffernan called 'mumblers'), were perceived as the most masculine. While one study has suggested that the larger tongues and reduced jaw opening ability of men can explain why women are more clear and men tend to mumble (Weirich et al., 2016), there is still a learned component to this speech presentation. Looking through the research, it is clear that there is evidence that male-female speech differences are at least grounded in sexual dimorphism. Additionally, there is also evidence that these differences are learned. One explanation for this is that the speech differences seen between males and females have evolved as a culture-specific way of enhancing or resisting biology. Similar behaviors are observed in other primates: Diehl et al. (1996) and Puts et al. (2016) point out that other primates exaggerate sexual dimorphism through vocalizations, though humans demonstrate the most sexual dimorphism of all primates when it comes to fundamental frequency. Could this be due to the added awareness of culture? In humans, nature or sexual dimorphism, provides the foundation for speech differences, while nurture, learning, allows for opportunity to exert individual agency. The question then becomes how is that agency learned? To answer this question, we look towards children's speech.

1.3 Sex and Gender in Child Speech

Just as listeners can associate adult speech with gender, a number of studies have also shown that listeners can reliably identify the gender of prepubertal children (Sachs et al., 1973; Perry et al., 2001; Weinberg & Bennett, 1971), possibly as young as 2.5 years of age (Fung et al., 2021). According to a gender perception study where adults listened to sentences and isolated vowels of children at a variety of ages, for children between the ages of 8 and 10 years old (prepubertal), on average, adults identified child gender with 81% accuracy for sentences, and 66% accuracy for vowels (Amir et al., 2012).

The problem with the successful gender recognition of prepubertal children is that it conflicts with the previously established importance of sexual dimorphism shown in adults. The biological changes that result in pitch variation, including the lowering of the larynx and the lengthening and bulking of the vocal folds in males does not occur until puberty. It is generally agreed that the vocal mechanisms of pre-pubertal males and females, when controlled for body size, are indistinguishable and do not show evidence of sexual dimorphism (Perry et al., 2001). It should be acknowledged that two studies by Vorperian et al. (2005) and Vorperian et al. (2011) did find prepubertal sexual dimorphism in the vocal tracts of children between the ages of 3 years to 7 years old. In conflict with these findings is a longitudinal study evaluating the development of the human vocal tract by Barbier et al. (2015). In this study, fetal development was analyzed, as well as subjects between the ages of 1 month old and 25 years old. The findings of Barbier et al. (2015), like many before, showed no evidence of pre-pubertal sexual dimorphism of the vocal tract (Goldstein, 1980; Lieberman et al., 2001; Fitch & Giedd, 1999). Barbier argued that Vorperian's findings were due to statistical anomalies in her data.

Being that there is no significant size difference in the vocal folds, one would expect that there would be no difference in the fundamental frequencies of prepubertal boys and girls as well. This is exactly what research has shown (Busby & Plant, 1995; Perry et al., 2001; Sussman & Sapienza, 1994). In fact, in a study looking at spectral and temporal measures of 436 children between the ages 5 and 17 years, and 56 adults, Lee et al. (1999) found that there are no significant fundamental frequency differences between males and females before 11 years old, with the average age for pubertal pitch change in males starting between 12 and 13 years old.

In that same study (Lee et al., 1999), analysis showed the formant frequencies of the children did differ statistically as a function of sex, even before puberty. While one would probably expect that formant frequencies would remain constant between children AMAB and AFAB, because of the lack of definitive sexual dimorphism in the vocal tract, that is not the case. A number of studies have found inter-sex variation in formant frequencies, and for good reason (Cartei et al., 2014; Perry et al., 2001; Busby & Plant, 1995; Lee et al., 1999; Whiteside & Hodgson, 2000; Bennett, 1981). As has already been established, the vocal tract is a major factor in the determination of formant frequencies. Described by Fant (1975), different portions of the vocal tract are associated with different formants. The pharyngeal cavity is affiliated with the second formant (F2), and the oral cavity is associated with the third formant (F3). Research has shown that the formant spacing between the first formant (F1) and F3 are greater than the spacing between F1 and F2 in children AMAB (Vorperian & Kent, 2007). There is evidence that within the constraints of anatomy, learned maneuvers are used to manipulate these formants. Described by Lindblom and Sundberg (1971), Bennett and Weinberg (1979), Bennett (1981), and Cartei et al. (2014), modifying jaw movement, laryngeal height, and the extent of lip protrusion can impact vocal tract length and formant frequency. Cartei et al. (2014) suggest that children use these maneuvers to alter the gender presentation of their speech. Another important component of formant frequencies is the shape and manipulation of the articulators. Using articulators to shape speech, while there is an anatomical component, is a learned skill, meaning people use their articulators differently and for different reasons. For example, Fox and Nissen (2005) looked at the voiceless fricative (/f, θ, s, ʃ/) production of 100 speakers between the ages of 6 and 52 years old. The results showed that, within the prepubertal group, production for the males and females differed significantly. For instance, /f/ production of female speakers showed significantly lower spectral skewness (m3 - a measure that will be discussed in more detail later) than male speakers. Having comparable anatomy, these males and females must have used differing articulation to produce their speech sounds. Additionally, just as adult's formant frequencies can be used to perceive gender, the same is true for prepubertal children. Cartei and Reby (2013) discovered this by presenting the speech of four 8 year old children with manipulated formants to adult listeners. As the artificial formant frequencies decreased, the adults were more likely to rate the child as masculine and male.

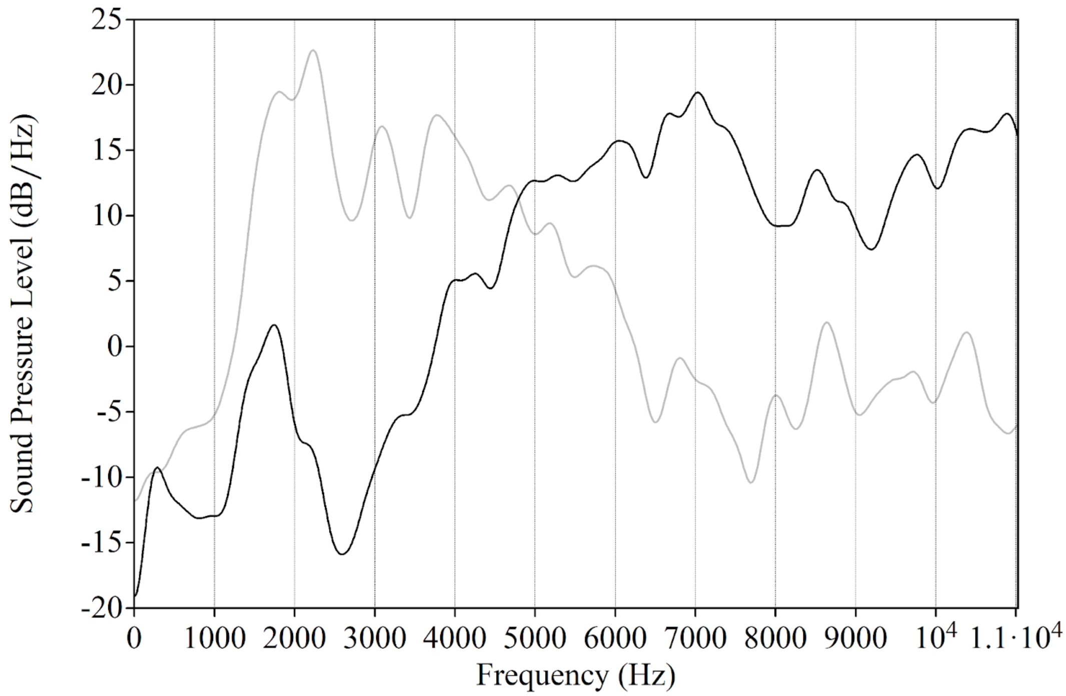
With a lack of definitive evidence of sexual dimorphism, a lack of distinctive fundamental frequency, a driving force in the perception of adult gender, and evidence of a learned formant component, the identification of prepubertal children's gender must rely on the perception of learned behaviors by the child. Children, just like adults, are impacted by cultural ideals of gender and the patterns they see in society, as seen in Martin and Ruble (2004). While it is unclear whether children consciously use learned behaviors to communicate their gender identity specifically, research has shown that children display a learned awareness of gendered speech stereotypes and can intentionally perform them. Cartei et al. (2014) conducted an acoustic analysis of 34 children between the ages of 6 and 9 years old. In the first round of the study, the children were asked to name pictures using their normal speaking voice. The results of this round showed, as expected, no significant difference between the fundamental frequencies of the children. In the following rounds, the children were told to sound more "like a boy" or more "like a girl". The results showed that in the masculinizing probe, girls significantly lowered their fundamental frequency, while boys significantly raised their fundamental frequency in the feminizing probe. For both the boys and girls, formant frequencies were successfully modified to reflect gender patterns seen in adult speech.

Just as adult speakers differentiate their gender with intentional pitch modification, clarity, and contrast, children are learning and performing gender-cueing speech characteristics. Voice onset time (VOT) is one of those characteristics that has been performed differently by prepubertal boys and girls, as girls tend to demonstrate a significantly longer VOT than boys (Whiteside & Marshall, 2001). Another characteristic, which was briefly mentioned earlier as it relates to formant frequencies, is of particular interest in this study. Hypothesized by Barbier et al, (2015) as one of the major factors resulting in the observable differences between male and female children's speech, is learned articulatory behaviors of speech sounds. For this study, articulation of /s/ is the main focus.

1.4 /s/ as a Gender Marker

In trying to better understand how gender is relayed through speech in both adults and children, fricatives, /s/ in particular, have gained a lot of attention. The speech sound /s/ is considered a sibilant fricative and is produced by raising the tongue tip to the alveolar ridge and forcing air past the tongue. /s/ is one of the highest frequency speech sounds, peaking around 6,000 Hz. Figure 1 shows a spectrum for a production of /s/ and a production of /ʃ/, illustrating these differences. This figure is a spectrum of a 40 ms-interval of /s/ (in black) and /ʃ/ (in gray). A spectrum shows the intensity of different frequency regions. As this figure shows, the distribution of energy differs between these sounds, that is, they have a different overall shape.

Figure 1: Spectra of a 40 ms interval of /s/ (in black) and /ʃ/ (in gray) produced by a cisgender adult man, smoothed using cepstral smoothing.

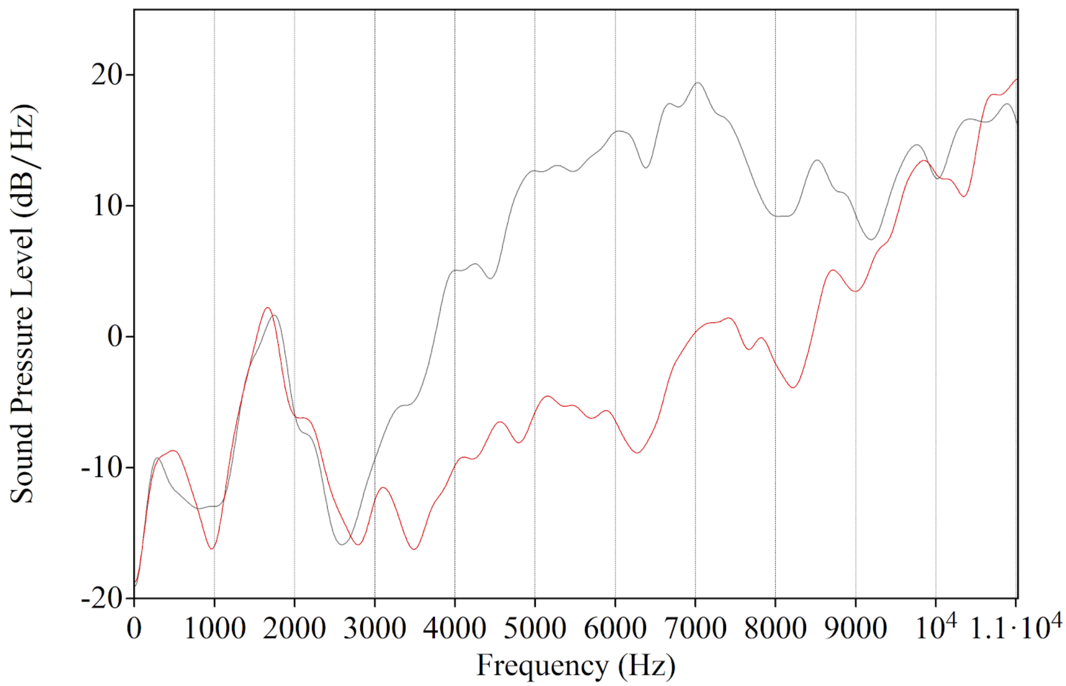


When producing these fricatives, women contrast the peak frequencies to a greater extent than men (Nittrouer et al., 1989; Newman et al., 2001). This increased contrast of similar sounds plays into the concept, discussed earlier, that women articulate with more clarity than men. Munson et al., (2006) also showed that women produce both /s/ and /ʃ/ with higher-frequencies, greater precision, and just like gay men, increased energy above the median frequency.

In order to study variation within fricatives, including sociophonetic variation, very detailed analysis must be done. One challenge to studying fine variation in fricatives is to find a method that can quantify differences like those in Figure 1. One technique that has been used is to treat spectra like those in Figure 1 as random distributions of numbers, and to calculate four summary statistics of this distribution: mean (also called centroid), standard deviation, skewness, and kurtosis. These are calculated the same way that they would for any random distribution of numbers, including a histogram of test scores or heights. These are called spectral moments, and they are usually referred to with numbers: m1 (mean/centroid), which is sensitive to the oral cavity anterior to the constriction (smaller cavities lead to higher m1), and which hence reflects the position of the tongue in the anterior-posterior dimension, m2 (standard deviation), which is sensitive to the size and shape of tongue constriction (broader constrictions lead to higher m2), m3 (skewness), and m4 (kurtosis). Of these, the first three have been used widely to characterize fricatives. In Figure 1, the /s/ has an m1 of 7915 Hz, and the /ʃ/ has an m1 of 2900 Hz.

Spectral moments can be used to characterize fine-grained differences in different productions of the same sound as well. Figure 2 below shows the spectra for two types of /s/. These include a production of a hyper-correct /s/, which is consistent with that of heterosexual women and gay men in Munson et al. (2006), shown in red, and a production of an /s/ that resembles the productions of the heterosexual men in Munson et al. (2006), in gray (the same spectrum as shown in Figure 1).

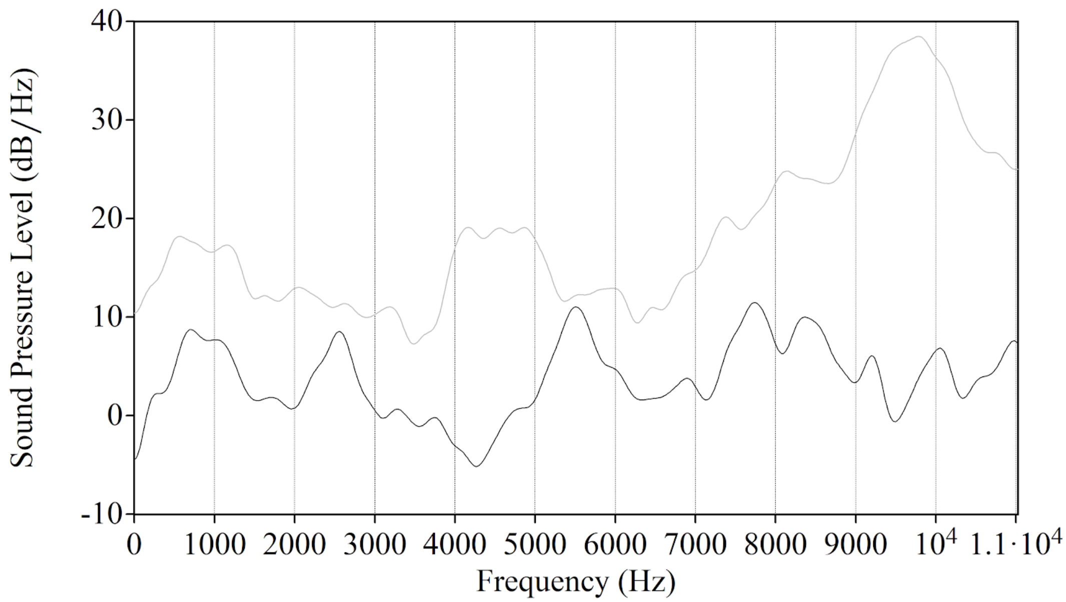
Figure 2: Spectra of a 40 ms interval of the fricative /s/ in two productions by a cisgender adult male, produced intentionally to illustrate different types of /s/, smoothed with cepstral smoothing. The spectrum in red is for an /s/ with an especially high peak frequency.



The two types of /s/ in this figure differ in their m1. The gray spectrum has an m1 of 7915 Hz, and the red spectrum has an m1 of 10442 Hz. They also differ in skewness. The gray spectrum has a skewness of 0.15, indicating that the energy was distributed relatively evenly above and below the median frequency. The red spectrum has a skewness of -4.6, indicating that there was more energy above the median frequency than below it.

Of the other two spectral moments, m2 is relevant to characterizing different types of /s/. Figure 3 below shows spectra for two productions of /s/ by children. Both of these productions were transcribed as accurate, and are included in the analysis in this paper.

Figure 3: Spectra of a 40 ms interval of two different children's productions of /s/ transcribed to be accurate, smoothed using cepstral smoothing



These spectra show two types of correct /s/ tokens, illustrating differences in spectral moments. The gray line is for an /s/ with the following spectral moments: m1=9588 Hz, m2=1391 Hz, m3=-3.24 (dimensionless). The black line is an /s/ with the following spectral moments: m1=6551 Hz, m2=3575 Hz, m3=0.05 (dimensionless). In this figure, the /s/ in gray has a higher centroid frequency than the /s/ in black. However, the /s/ in black is most notably different because it has a high m2, that is, a very diffuse spectrum. It does not have a clear spectral peak the way that the gray spectrum does. Diffuse spectra like those in black in this figure are perceived impressionistically as less accurate tokens of /s/, resembling the sound /θ/. That is, this token of /s/, taken from the word *sick*, might sound somewhat like the *th* in *thick* (though not enough in this case to be transcribed as an error).

According to Schwarts (1968) when presented with isolated /s/-segments, adult listeners can rate speaker sex with 93% accuracy, suggesting that /s/ holds sociophonetic information. With that, a number of studies have shown that acoustic differences in /s/ convey a variety of social messages about the speaker. For instance, Rogers et al. (2000) reviewed /s/ and /z/ production of 25 men (17=gay, 8=straight), which showed that 46 listeners of the men rated speech with longer durations and higher frequencies of /s/ and /z/ as “gayer-sounding”. Levon (2007), also curious about the relationship between /s/ and sexuality, examined the perception of /s/ by digitally manipulating pitch range and duration. Decreasing the pitch of more gay-sounding speakers resulted in an increased likelihood of a straight rating of the speaker. When increasing the pitch of straight-sounding speakers, gayness ratings did not increase, showing that pitch, alone, is not responsible for the perception of gay speech, and that duration of /s/ is also an imporant factor. Munson et al. (2006) also showed that gay men tend to produce /s/ at a higher frequency than heterosexual men. Additionally, this study showed that gay men produce /s/ with an especially high peak frequency and an especially compact spectrum. This is consistent with the characteristics of /s/ produced by heterosexual women in the same study, and the characteristics of /s/ produced in a clear-speech style (Maniwa et al., 2009). Mack and Munson (2012) went on to show that these gay men can be identified via listener perception of their /s/, and that higher frequency, misarticulated /s/, particularly dental and frontal tokens, are rated as being more gay than lower frequency, typical articulation. In these examples of /s/, the speech sound is a social marker of sexual orientation in adult men.

As has already been established, due to anatomy, adult females tend to use higher frequencies than males, so one would likely assume that females also produce /s/ at a higher frequency than males. While this frequency discrepancy holds true for males and females, anatomical differences are not the source. Fuchs and Toda (2010) examined articulatory, acoustic, and morphological data of 12 English and 12 German speakers (6 males and 6 females for each) to determine whether the observable male-female differences in /s/ were rooted in biology or sociophonetics. The results showed no significant biological difference between the sexes, revealing that /s/ production is sociophonetic.

Two studies showed that differences in how adults produce /s/ and /ʃ/ are indicative of learned gender marking. Just like /s/, /ʃ/ is also considered a sibilant fricative. It is also produced by raising the tongue tip and forcing air past the tongue, though the place of articulation is behind the alveolar ridge, with the front of the tongue bunched at the palate. It has been theorized that variation in male and female /s/ production could be due to palatal differences causing unique articulation between the sexes. This idea was studied and debunked by Fuchs and Toda (2010) who used electrode-embedded artificial palates to measure /s/ articulation and anatomy. In an analysis of English and German men and women, there were found to be palatal differences between English and German speakers, but no statistically significant difference between the sexes in terms of palate length. Despite this, women did show a more anterior tongue position when producing /s/ than men, as well as, higher peak frequencies. These results, like many before, point to sociophonetic factors being relayed by /s/.

Sexual dimorphism seen in the vocal mechanisms of adult males and females occurs during puberty. That said, there are clear anatomical differences between adults and children. Despite this, Stuart-Smith (2007) showed a surprising connection between the two groups' production of /s/. In a study of Glaswegian speakers, spectral analyses of /s/ production in 31 speakers separated by age, social class, and gender were conducted. While analysis did find spectral differences between the men and women, as well as, adults and children, one of the most interesting findings from this study was that the /s/ frequencies of young, working class girls was the most similar to the low-frequency production of adult men. This finding further demonstrates the learned articulatory behavior of /s/, as there is no reason, especially not a biological one, for this phenomenon. This interaction between age, class, and gender is not as surprising when considering Eckert's (1989) finding that social categories do not exist independently and interact with each other in important ways. Both Eckert (1989) and Labov (1990) found that gender and class frequently influence speech together. With extreme anatomical differences between the vocal mechanisms of adult cis men and young girls, the spectral similarities and interaction between social categories points to a socially-driven, learned, articulatory explanation.

As shown in Stuart-Smith (2007), /s/ does not only carry social information about adults. With the acceptance of a learned component in the production of /s/, studies have sought out what specific information can be relayed about children through this fricative.

Flipsen et al. (1999) studied /s/ production in children between the ages of 9 and 15 years old. In looking at the acoustics of fricative onset, midpoint, and offset, females showed a higher frequency than males. The most likely explanation for this frequency discrepancy, as presented by Flipsen et al. (1999), is that female speakers could be using a more forward tongue placement than males when producing /s/, just like gay men (Munson et al., 2006) do. This would result in a smaller oral resonating cavity, and consequently, a higher frequency.

In a study explicitly investigating /s/ as a gender marker in children, Li et al. (2016) evaluated 152 children between the ages of 4 and 16 years old. The children were recorded naming /s/-initial objects that were displayed to them over a screen. As an attempt to control for vocal tract length, the children's heights were measured, and as a measure of gender, the parents of each child filled out the 12-item Childhood Gender Identity Questionnaire (CGIQ; Johnson et al., 2004). After collection of the /s/-word samples, an acoustic analysis of the productions was completed. Including all children, the mean frequency (m1) was higher for girls. This higher m1 suggests that the tongue is closer to the teeth, and supports Flipsen et al. (1999)'s theory that girls use a more fronted place of articulation for /s/ than boys. The standard deviation of fricative spectra (m2) was also higher for girls. The one measure where the boys had higher values than the girls was spectral skewness (m3). As discussed earlier in Munson et al. (2006), this higher m3 is associated with more masculine speech, and means there was a greater amount of energy below spectral mean. When isolating the prepubertal children (4-7 years old), there was still a statistically significant difference between boys and girls for m1 and m3, but not m2. Li et al. (2016) also found that within-sex variation of /s/ is highly dependent on gender identity, but only in the boys, suggesting that boys are placing more social weight on /s/ production. Li et al., (2016) postulates that this could be due to the link that exists between men's sexuality and /s/.

The research in this introduction have provided an abundance of evidence of the learning and performance of sociophonetic speech in both adults and children. Further, /s/ has been presented as a major carrier of learned sociophonetic variation. The majority of /s/ research pertains to adult subjects. To further investigate /s/ production as it relates to the gendered speech of children, a series of questions were formed for this thesis. Note that these questions do also require an analysis of /ʃ/. Studies on the development of fricatives often examine the acquisition of the contrast between /s/ and /ʃ/. /ʃ/ is a later-acquired sound than /s/, and [s] for /ʃ/ errors (i.e., saying 'sip' for ship) are common in children acquiring language normally. Moreover, studies of both /s/ and /ʃ/ in adults has shown that females demonstrate more contrast between the peak frequencies of the two fricatives than males (Nittrouer et al., 1989; Newman et al., 2001). By including /ʃ/ we can examine whether any gender marking serves to enhance or attenuate the contrast between these two sounds. Moreover, as has been described earlier, variation in both /s/ and /ʃ/ within and across individuals does not have a clear basis in sex differeniation in anatomy or physiology. This means that if there are differences in the production of these fricatives, there is good reason to believe those features are learned. Finding a difference in /s/ but not /ʃ/ between children AMAB and children AFAB is persuasive evidence that the /s/ differences are learned and not due to a not-yet-discovered anatomical basis. Finally, fricatives like /s/ and /ʃ/ have previously been used to differentiate between adult men and adult women, and are correlated with listeners' ratings of masculinity and femininity in human speech. This makes /s/ and /ʃ/ excellent analysis points for the development of children's gendered speech.

1.5 Research Questions

That said, the primary research question being investigated in this thesis is: **Do children assigned male at birth (AMAB) and those assigned female at birth (AFAB) produce the sibilant fricatives, /s/ and /ʃ/, differently at age ~3 years and ~5 years?** We answer this question through an acoustic analysis of children's fricative productions that were transcribed as correct. Given that our focus is on accurate productions, we also ask **whether children AMAB and AFAB have different accuracy rates for /s/ and /ʃ/, based on closed phonetic transcription, at ~3 years and ~5 years?**

Finally, we capitalize on the fact that the children whose speech we are examining participated in a parallel study which elicited ratings of gender typicality from a group of adults. The ratings from that study confirm that the children varied in the extent to which adults rated their speech to be canonically male or female. We are hence able to ask **whether the summary measures of fricative acoustics in this study are correlated with ratings of gender-typicality of these children's speech collected previously and summarized in Lackas (2019) and Munson et al. (2019)?**

2. Methods

2.1 Participants

This analysis primarily relies on data collected from one group of participants: 1) speech samples collected from 110 Child Talkers. With that, one of our research questions listed above does involve previously collected data from another group: 2) perception ratings from 80 adult listeners. A full description of the Child Talkers will follow, as well as, a brief overview of the Adult Listeners. Readers interested in more detailed information regarding the Adult Listeners are referred to Lackas (2019).

Child Talkers: The first group is composed of 110 children (55 AFAB, 55 AMAB) who were recruited through word of mouth, online advertisements, newspaper bulletins, flyers placed in day-care centers, schools, and other local gathering locations. These children were originally involved in a longitudinal study investigating speech perception, speech production, and word learning with a focus on varying socioeconomic statuses, stages of speech and language development (including typically developing and late talkers), and hearing statuses (cochlear implant users). Summaries of some of the primary research questions of the original study can be found in Munson et al., (2021), Cychosz et al. (2021), and Erskine et al. (2020). While 164 children were included in the longitudinal study, 110 were selected for this study, with each child AMAB being matched to a child AFAB within 3 months of age. Only children who were typically developing, without hearing impairments, and who could be matched in age in AMAB-AFAB pairs were selected from the larger sample of 164. The children included in this study represent a range of socio-economic statuses (SES), and all children were native speakers of either Mainstream American English (MAE, N=114) or African-American English (AAE, N=6). Through parent-child observations and parent interviews, dialect was determined for each child. The children's genders were labeled via parent report. It should be noted that an explicit distinction between sex and gender was not made to these parents, as gender was not a primary focus of the original study. All parents reported their child as male, female, or other. Of the 164 children, one child was being raised without an assigned gender identity, as to avoid a culturally imposed gender and to encourage a self-determined identity by the child. This child was not included in this study. Because the original study from which these speech data were taken did not focus on gender per se, we did not have a detailed measure of gender identity (like the Childhood Gender Identity Questionnaire [CGIQ; Johnson et al., 2004] used by Li et al., 2016). By relying on reports by parents, we imagine that the male/female designations reflect largely the parents' expectations given their child's SAB.

At each of the three time-points in this longitudinal study, each child completed a battery of standardized and non-standardized measures of speech, language, and hearing across two to three sessions. Each session was 90 to 120 minutes long. The first time point 1 (TP1) took place when the children were between the ages of 28-39 months old. The second time-point took place when the children were 40-52 months old. The final time-point, time point 3 (TP3), took place when the children were between the ages of 53-66 months old. For this study, we focus on data collected at TP1and TP3. In hopes of eliminating confusion due to the absence of TP2 in our analyses, TP1 will be referred to as the *earliest time-point* (ETP), and TP3 will be referred to as the *last time-point* (LTP). These terms are consistent with the labels used in Munson et al. (2021) to describe time-points. Mean age for the children at ETP = 32.5 months with a SD of 3.6. Mean age for LTP = 56.7 months with a SD of 3.9. The difference in age between AMAB and AFAB was not statistically significant at either ETP ($t[108]<1$, $p=0.75$) or LTP ($t[108]<1$, $p=0.85$) according to independent-samples t-tests.

At each time point, the children completed standardized and non-standardized speech and language assessments, experimental measures of speech production and perception, as well as, measures of lexical processing. These sessions were led by a team of trained research assistants. Table 1 includes the Child Talkers' standardized test scores that were collected in the previous study. These scores were used as predictors of gender ratings in Lackas (2019). For this study, the scores are displayed only to provide additional descriptive data of the Child Talkers, their language skills, articulatory development, and vocabulary size. Group differences on these measures are described in Lackas (2019). Only one of these differences between AMAB and AFAB children was significant, on the GFTA-3. The AFAB children had slightly lower scores than the AMAB children.

Table 1. Child Talkers' Standardized and Nonstandardized Test Scores

Time-Point	Test	AMAB Mean (SD)	AFAB Mean (SD)
Earliest Time-Point	*Expressive Vocabulary Test-2*[a] Standard Score	116.4 (14.6)[3]	118.3 (16.9)[1]
	Peabody Picture Vocabulary Test-4[b] Standard Score	114.5 (17.1)[3]	113.0 (16.3)[1]
	LENA Conversational Turn Count[c]	55.3 (24.6)[3]	53.5 (23.9)[3]
	LENA Child Vocalization Count[c]	222.8 (89.7)[3]	222.6 (88.5)[3]
	Goldman-Fristoe Test of Articulation-2[d] Standard Score	96.5 (10.6)[6]	90.2 (13.2)[2]
	Inhibitory Control[e] Raw Score	2.2 (0.7)[1]	2.1 (0.7)[2]
	Minimal Pair Perception[f] Percent Correct	73.3% (14.9%)[5]	67.8% (18.8%)[3]
Last Time-Point	*Expressive Vocabulary Test-2*[a] Standard Score	120.1 (15)[1]	117.4 (13.5)[1]
	Goldman Fristoe Test of Articulation-2[d] Standard Score	92.2 (11)[3]	91.9 (12.9)[3]
	Kaufman Brief Intelligence Test-2[g] Matrices Subtest Standard Score	106.7 (14.2)[2]	107.7 (10.9)[2]
	SAILS[h] Percent Correct	71.3% (9.8%)[3]	73.6% (10.9%)[4]

[a]Williams (1997), [b]Dunn and Dunn (1997), [c]Xu et al. (2009), [d]Goldman and Fristoe (2000), [e]Zelazo et al. (2013), [f]Erskine et al., 2020, [g]Kaufman and Kaufman (2014), [h]Rvachew (2009)

[1]N=55, [2]N=54, [3]N=53, [4]N=52, [5]N=50, [6]N=49

AMAB: Assigned Male at Birth
AMAF: Assigned Female at Birth

For participating in the study, the parents of the Child Talkers were monetarily compensated for completing each of the three sessions. When the children successfully completed their final session, the parents received a bonus monetary compensation. Additionally, following completion of each session, the children were gifted with a book to take home. To keep interest in the longitudinal study, and to encourage the Child Talkers to come back for their following sessions, books were also sent to the children on their birthdays.

Adult Listeners: For the gender perception task of children's speech, 80 adults between the ages of 18 and 40 years old acted as listeners/raters. Before participating in the study, adults were asked screening questions to identify their native language, history of hearing status, and history of language/learning disorders. All adults selected were without a history of hearing impairment, language/learning disorders, and identified as native English speakers (defined as having learned English from birth from at least one parent who spoke English natively and who communicated with them in English from birth). These adults were recruited through flyers posted around the University of Minnesota Twin Cities, and recruitment presentations in undergraduate courses at the university. The estimated completion time of the study was 40 minutes to an hour. For participating in the study, the adults were compensated $10 for the first hour, and $2.50 for every 15 minutes past the hour it took them to complete the task.

2.2 Stimuli

The stimuli for this study were taken from a Real Word Repetition (RWR) task at ETP and LTP of the original, longitudinal study. For this task, pictures of familiar objects were presented to the children in conjunction with an auditory prompt naming each picture. Upon viewing and hearing each prompt, the children were then instructed to repeat the name of the object pictured. Taking the children's dialects into account, two sets of auditory prompts were used. The prompts for each child were presented in their native dialect. For the children who spoke MAE, the prompts they were presented with were that of a white woman who was a Minnesota-native and who spoke the local regional dialect. For the children who spoke AAE, the prompts they were presented with were that of an African-American woman who was a Wisconsin-native and who routinely code-switched between AAE and her local regional dialect. Of the 110 children in the study, 106 received the words in MAE and 4 received the words in AAE.

Stimuli words without a strong gender association were selected, as to limit potential influence on gender rating by the Adult Listeners. For example, *truck* and *doll*, typically stereotyped as being for boys and girls respectively (Fine & Rush, 2018), were excluded from the study. At each time-point, the 110 children produced a list of about 100 words, designed to elicit a variety of sound contrasts. Of these 100 words, 16 were /s/-initial and 16 were /ʃ/-initial. The specific words were chosen to be very familiar to children, and to have balanced vowel contexts (i.e., there were the same number of /s/ and /ʃ/ words in a given vowel context). Given these constraints, it was impossible to elicit 16 different /s/- and /ʃ/-initial words at each time-point. Some of the words were elicited two, three, or four times to get the requisite number of 16 productions. The words were selected for each time-point based on phonological complexity that the children were more likely able to produce, hence there is more phonological complexity in the words used at the LTP. The words included are as follows:

ETP: /s/: sun, sad, sandwich, scissors, sick, soap, sock, soup

ETP: /ʃ/: share, sheep, shoe, shovel, shower

LTP: /s/: sun, cereal, sandbox, sandwich, scissors, seven, sidewalk, sink, sister, soup, suit, suitcase, summer

LTP: /ʃ/: shadow, sharing, sheep, shell, ship, shoes, shoulder, shovel, shower, sugar

2.3 Procedure

Trained research assistants listened to the entirety of each child's recordings at each time-point and completed a multi-step analysis using custom-written Praat scripts. The first step in this process is referred to as *segmentation,* where the best, most natural sounding productions at each time-point were selected. Selection criteria to determine the best production included completeness (i.e., no omitted sounds), clarity, and productions that were most likely to be representative of the child's typical mode of production. This means that any words that were sung by the child, produced while shouting or whispering, or said using a 'silly' voice, were excluded from selection. A small number of productions were coded as non-response, when a child refused to produce a word or never produced a usable token. This affected 97 of the possible 3520 productions at ETP, and 1 of the 3520 productions at LTP. Once the best productions for each word were determined, the recordings were cropped to include only the child's production of the word.

The research assistants then annotated the productions. For a small number of responses, the child did not provide a usable response or refused to respond all together.

Each of these types of responses were coded as *missing data.* Occasionally, the child would begin saying the word before the completion of the audio prompt, or the child would produce the target while the experimenter was talking. These instances resulted in overlapping recordings and were coded as *overlapping*, excluding them from selection. The remaining productions were coded as to whether they produced the target word or a *malapropism* (i.e., whole-word errors). If the whole-word error resulted in another word that started with either an /s/ or /ʃ/, they continued to the transcription phase. There were only six malapropisms across all time-points, targets, and children. In four of the the six occurrences, the response was a morphological variant of the target word (i.e., *sharing* for *share*) or a perseveration of responses from an earlier /s/ or /ʃ/ trial. For the final two malapropisms, the non-specific "that one" was produced.

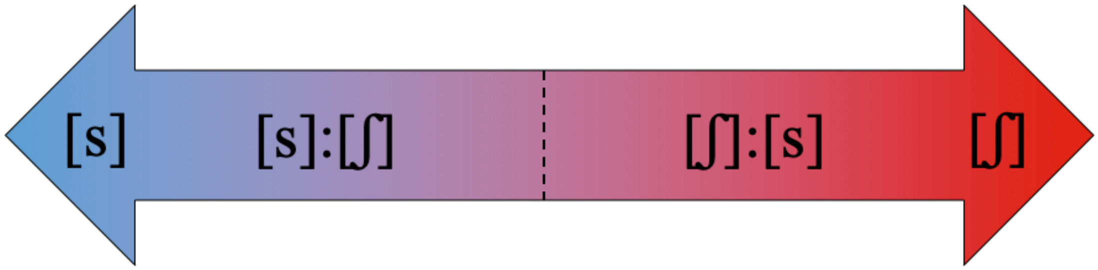
Next, a two-step transcription process of the remaining words was completed. In the first step, the responses were coded for manner of articulation. Each fricative was coded as either a *sibilant fricative* (e.g., /s/, /z/, /ʃ/, /ʒ/) *, sibilant affricate* (e.g., /tʃ/, /dʒ/)*, non-sibilant fricative* (e.g., /θ/, /ð/, /f/, /v/, /h/)*, non-sibilant plosive* (e.g., /t/, /d/, /k/, /g/, /p/, /b/)*,* or *other* (e.g., lateral productions, frontal misarticulations that weren't codable as /θ/)*.* The counts of the productions separated by manner of articulation, time-point, and talker's SAB are displayed in Table 2.

Table 2: Manner of Articulation Counts by Time-Point and Sex Assigned at Birth

Time Point	SAB	Sibilant Fricative	Sibilant Affricate	Non-Sibilant Fricative	Non-Sibilant Plosive	Other	Mala-propism	Missing Data
ETP	AFAB	1252	117	110	203	25	5	48
ETP	AMAB	1235	122	97	189	69	1	47
LTP	AFAB	1523	227	2	4	3	0	1
LTP	AMAB	1442	305	2	6	5	0	0

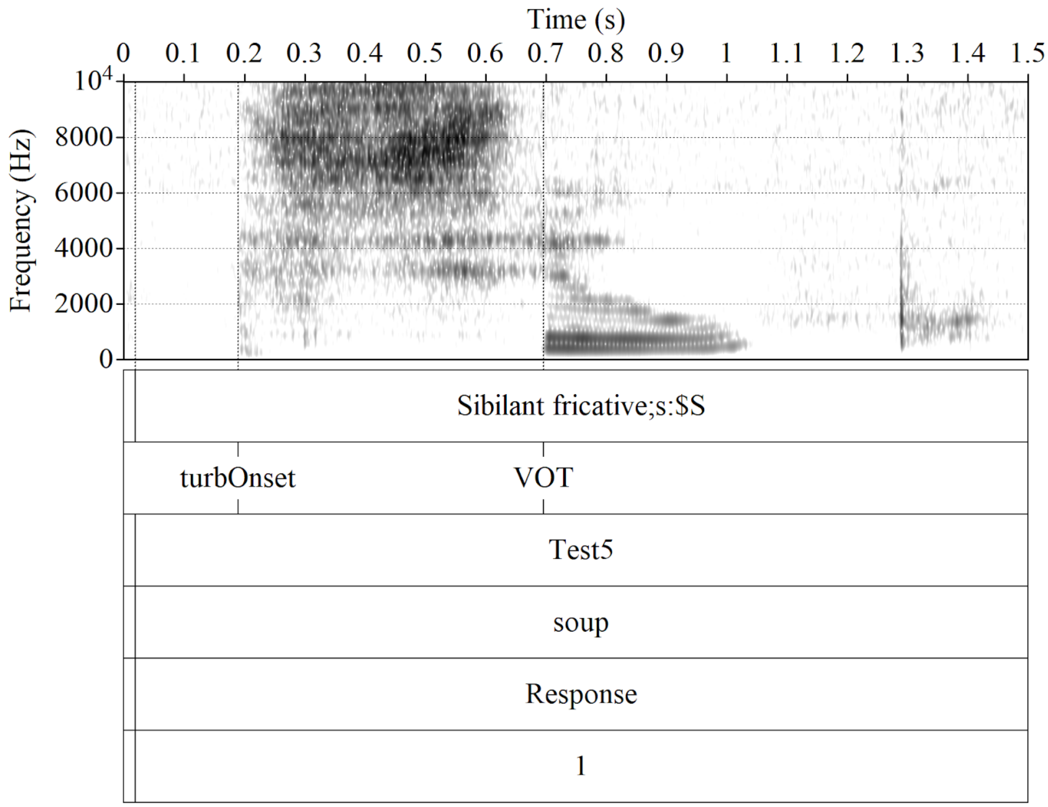
If the production was determined to contain a sibilant fricative, the research assistants then coded the initial sound as an [s], [ʃ], a fricative in between [s] and [ʃ] but more similar to an [s] ([s]:[ʃ]), or a fricative in between [s] and [ʃ] but more similar to an [ʃ] ([ʃ]:[s]). The use of intermediate categories follows the suggestion of Stoel-Gammon (2001). Figure 4 displays a visual representation of this transcription method.

Figure 4: Transcription Coding for /s/ and /ʃ/



The final step in the transcription process was to measure the duration of the fricative turbulence. To do this, the research assistants coded the onset and offset of the fricative. Figure 5 displays an example of the Praat layout used to analyze the Child Talkers' recordings.

Figure 5. Display for the analysis environment for transcription and turbulence tagging, with a production of soup transcribed with an initial [s]:[ʃ], and with the onset of turbulence (turbOnset) and the onset of voicing (VOT) marked.



As stated earlier, the adult perception data of the Child Talkers from Lackas (2019) are utilized in this thesis. That said, the following is a brief description of the Adult Listener study and how those ratings were collected: The Child Talkers were matched for average age at each time-point, test scores, and maternal education. With this demographic balancing, the children were divided into four sets. One child's samples were used in every set to act as a point of reference. The Adult Listeners were brought into a sound-treated booth, fitted with headphones, and guided through the study via computer. One of the three Child Talker sets was randomly selected for each Adult Listener. Each Adult Listener completed two trial prompts before rating their assigned set. This was done to familiarize the listeners with the task and allow for any questions to be addressed. The recordings were presented to the listener while a continuous rating scale with "definitely a boy" on one end and "definitely a girl" on the other was displayed. Upon hearing each recording, the Adult Listeners clicked where along the scale they believed the child's gender to fall.

2.4 Analysis

Following the collection of the Adult Listeners' ratings and coding of the Child Talkers' samples, statistical analyses were completed on the data using R.

2.4.1 Listener Ratings

The listener ratings are analyzed in detail in Lackas (2019). A brief summary of these results is provided here. Listeners rated the AFAB children as more sounding more girl-like than the AMAB children at both time-points. The difference in ratings between the AMAB and AFAB children was greater at LTP than at ETP. This difference was driven by changes in ratings for the AMAB children, who were rated as sounding more boy-like at LTP than they were at ETP. The ratings for the AFAB children were statistically equivalent at the two time-points. At both time-points, the mean ratings for individual children overlapped, and most ratings were in the mid-point of the scale.

2.4.2 Accuracy

The first set of analyses performed was concerned with production accuracy as measured by phonetic transcription. In these analyses, we examined whether the likelihood of producing the correct manner of articulation, sibilant fricative, could be predicted from time-point, target, and SAB. The details of the statistical modeling procedure for this analysis were similar to those used in many subsequent analyses. They are described in detail to serve as a reference for our description of other, similar analyses.

For this analysis, non-responses and malapropisms that were not sibilant-initial words were excluded. The four sibilant-initial malapropisms and the tokens transcribed as sibilant fricatives were coded as correct, while all other productions were coded as incorrect. A logit mixed-effects model was used, as the dependent measure was a binary outcome. Time-point, target (/s/ or /ʃ/), and SAB were fixed effects, and were contrast-coded (AFAB=1, AMAB=-1, ETP=-1, LTP=1, /s/=1, /ʃ/=-1). Word and Child Talker were random effects, as we did not expect every child and word to behave exactly the same way. The model included random slopes for the effect of time-point on Child Talker. This is due to the idea that the amount of growth from ETP to LTP is going to vary randomly across children, as some children will have grown more than others. The random slopes were also applied to the effect of target on talker, as it is to be expected that some children showed a bigger difference between /s/ and /ʃ/ than others. To fit the model, the package lme4 (Bates et al., 2015) was used. The package LMERtest was used to approximate significant values, using Satterwaith's approximation for degrees of freedom.

Model-building proceeded iteratively. We started out with a simple model that had random effects for talker or word (indicating that we expected that some of the variation in accuracy would occur across talkers and words randomly). We then added each of the fixed predictors of *target*, *time-point*, and *sex assigned at birth (SAB),* in that order. Every time we added a new fixed effect, we tested whether the resulting model fit the data better than the next-simplest model (i.e., did adding *target* result in a model with better fit than a model with just the random effects? did a model with *time-point* and *target* result in a model with better fit than with a model with just *target*?). We also evaluated whether the coefficients for a particular model were significant. Hence, our assessment of models involved both the consideration of model fit, and the consideration of whether the coefficients predicted variation in the dependent measure.

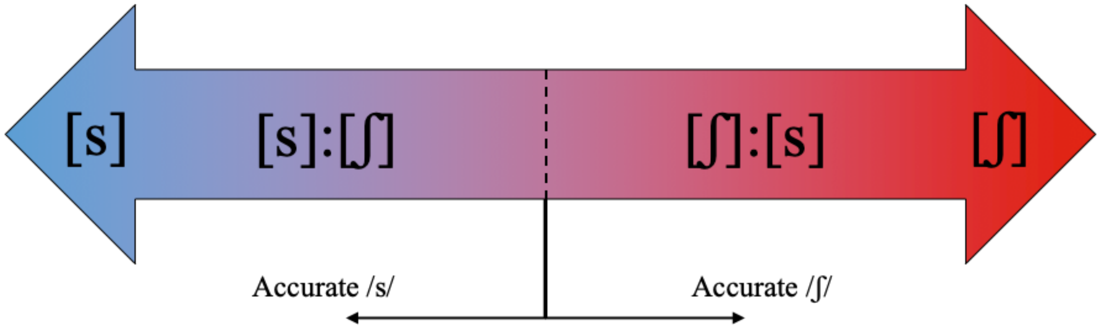
In the analysis of the accuracy of manner of articulation, the most-complex model that fit the data better than the next-simpler model was one in which there were fixed effects of time-point and target. However, an analysis of the coefficients for this model, shown in Table 3, show that only the coefficient for time-point was significant. Critically, adding SAB to this model did not improve model fit. This means that there were fewer manner of articulation errors in LTP than in ETP, but that the manner of articulation errors were similar across the two targets, and across the groups of AMAB and AFAB children.

Table 3: Coefficients of the Logit Mixed-Effects Model Predicting Accuracy of Manner of Articulation from Time-Point and Target.

Estimate	Estimate	Standard Error	z-value	Pr(>|z|)
(Intercept)	1.76334	0.13595	12.971	<0.001
Time-Point (LTP=1, ETP=-1)	0.33091	0.10980	3.014	0.003
Target (/s/=1, /ʃ/=-1)	-0.02718	0.06862	-0.396	0.692

The next analysis examined accuracy in producing place of articulation. That is, for the sounds that were transcribed as sibilant fricatives, did the child produce the sound in a way that the adult transcribers deemed accurate. The dependent measure in this series of models was whether the sounds that were coded as sibilant had an accurate place of articulation. For these models, a *lenient* coding system was used. As stated earlier, when annotating and transcribing the recordings, the research assistants labeled all sibilant fricatives as either [s], [ʃ], [s]:[ʃ], or [ʃ]:[s]. In lenient coding, an /s/ would only be coded as accurate if it was transcribed as either [s] or [s]:[ʃ], whereas an /ʃ/ would only be coded as accurate if it was transcribed as either an [ʃ] or [ʃ]:[s]. For a justification for lenient coding over a stricter coding (in which, for example, [s]:[ʃ] would be counted as an error for /s/), please see Munson et al. (2021). Errored productions, such as lateral /s/ or /ʃ/ would have been coded as *other* in the transcription phase, and were not included in the accuracy analysis. Figure 6 shows a visual representation of the lenient coding system used for this analysis.

Figure 6: Schematic of Lenient Coding for /s/ and /ʃ/



The same procedures as were used in the analysis of manner of articulation accuracy were used in the analysis of place of articulation accuracy. The model that fit the data better than the next-simplest model was one in which there were fixed effects of target and time-point. Adding SAB, or adding an interaction term between target and time-point, did not improve model fit. The coefficients from the model show that there were strong and significant effects of both time-point and target on accuracy. As expected, accuracy was greater at LTP than ETP. Moreover, consistent with much previous research, /ʃ/ was less accurate than /s/.

Table 4: Coefficients for the Logit Mixed-Effects model Predicting Place of Articulation Accuracy from Target and Time-Point.

Estimate	Estimate	Standard Error	z-value	Pr(>|z|)
(Intercept)	2.9591	0.2156	13.727	<0.001
Time-Point (LTP=1, ETP=-1)	1.0049	0.1327	7.573	<0.001
Target (/s/=1, /ʃ/=-1)	0.8265	0.1955	4.227	<0.001

These effects are illustrated graphically in Figure 7 below. As this figure shows, there was a wide range of accuracy for individual children for /ʃ/, and some variation in accuracy for /s/.

Figure 7: Proportion of Productions Transcribed as Accurate, Separated by Time-Point, Sex Assigned at Birth, and Target.

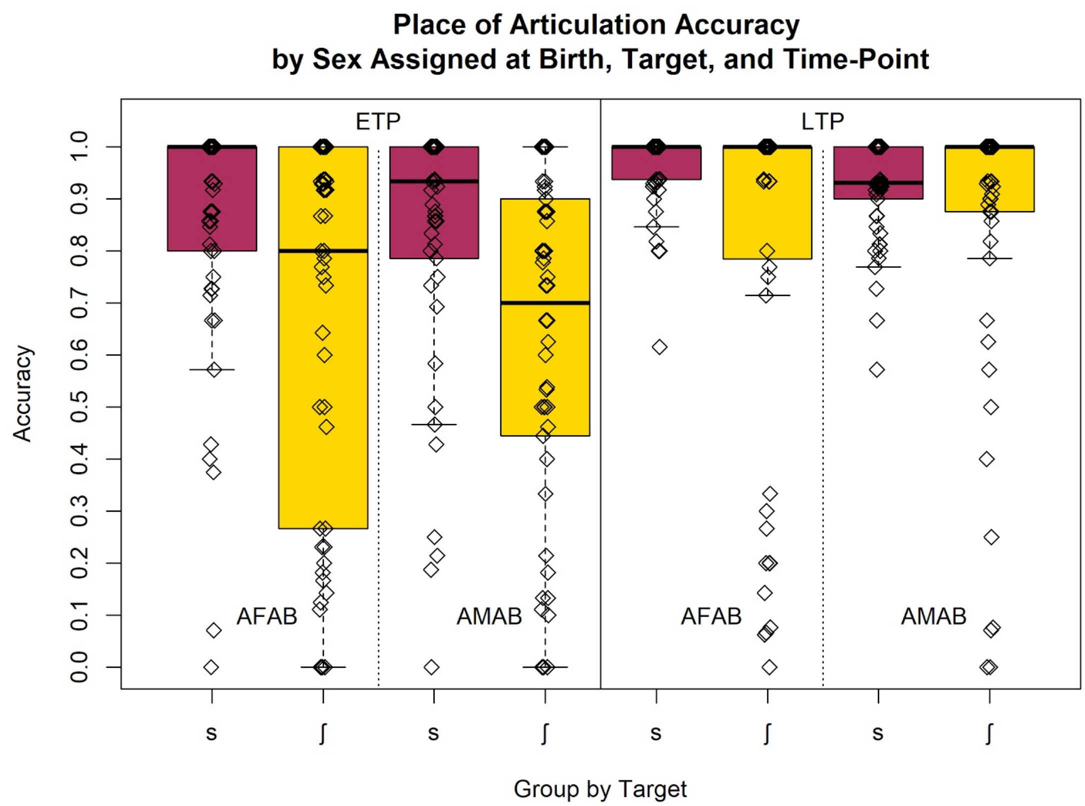


Figure 7 shows the mean proportion correct place of articulation for all 110 children, using the lenient criterion, separated by target, time-point, and SAB. Individual symbols represent the proportion accurate for individual children. This is only calculated for targets that were transcribed as sibilant fricatives, and not as other manners (stops or affricates).

2.4.3 Acoustic Measures

The next set of statistical analyses focused on the acoustics extracted from the Child Speakers' recordings. We examined m1, m2, and m3 separately for tokens coded to be correct. Because these were continuous measures, we used linear mixed-effects models instead of logit models. All other aspects of model-building were the same. For consistency's sake across the three dependent measures, we report the coefficients for the model with all three fixed effects and all possible interactions among those effects. We note the cases where those models did not result in a better model fit than a simpler model, and compare the pattern of coefficients in the full model with those in the simpler model.

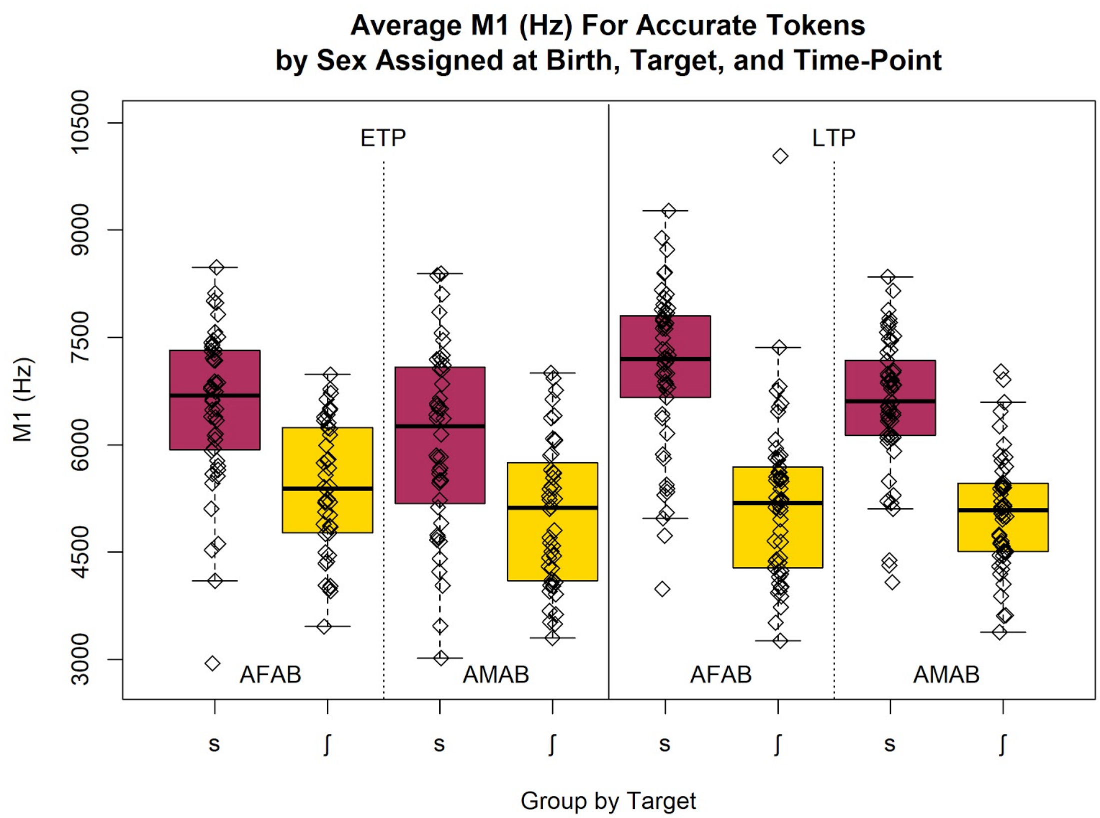
Mixed-effects models are thought to be robust to uneven data sets. Hence, the different numbers of stimuli for each combination of participant by target by time-point should not affect model fit. We return to this issue below.

The first model examined m1. The most complex model that fit the data better than the next-simplest model was the full model with all three fixed effects and all possible interactions. The coefficients for this model, and their associated significance values, are shown in Table 5. This table shows coefficients for the model predicting m1 from fixed effects of target, time-point, and SAB, with all interactions for all 110 children, for tokens of /s/ and /ʃ/ coded as accurate using the lenient criterion. Participant and word were random effects. These data are illustrated in Figure 8, which plots the mean m1 for correct tokens for each participant, separated by target, SAB, and time-point.

Table 5. Coefficients for the Linear Mixed-Effects model Predicting m1 from Target and Time-Point

Estimate	Estimate	Standard Error	Df	t-value	Pr(>|t|)
(Intercept)	5861.78	69.96	113.90	83.784	<0.001
Time-Point (LTP=1, ETP=-1)	37.64	23.64	277.11	1.592	0.113
Target (/s/=1, /ʃ/=-1)	777.80	33.22	27.70	23.412	<0.001
SAB (AFAB=1, AMAB=-1)	166.39	64.34	103.41	2.586	0.011
Time-Point * Target	177.44	23.38	266.18	7.589	<0.001
Time-Point * SAB	-30.09	18.86	4374.65	-1.595	0.111
Target * SAB	79.61	18.79	4364.33	4.237	<0.001
Time-Point * Target * SAB	37.37	18.53	4341.09	2.017	0.044

Figure 8: Average m1 for Productions Transcribed as Accurate, Separated by Time-Point, Sex Assigned at Birth, and Target



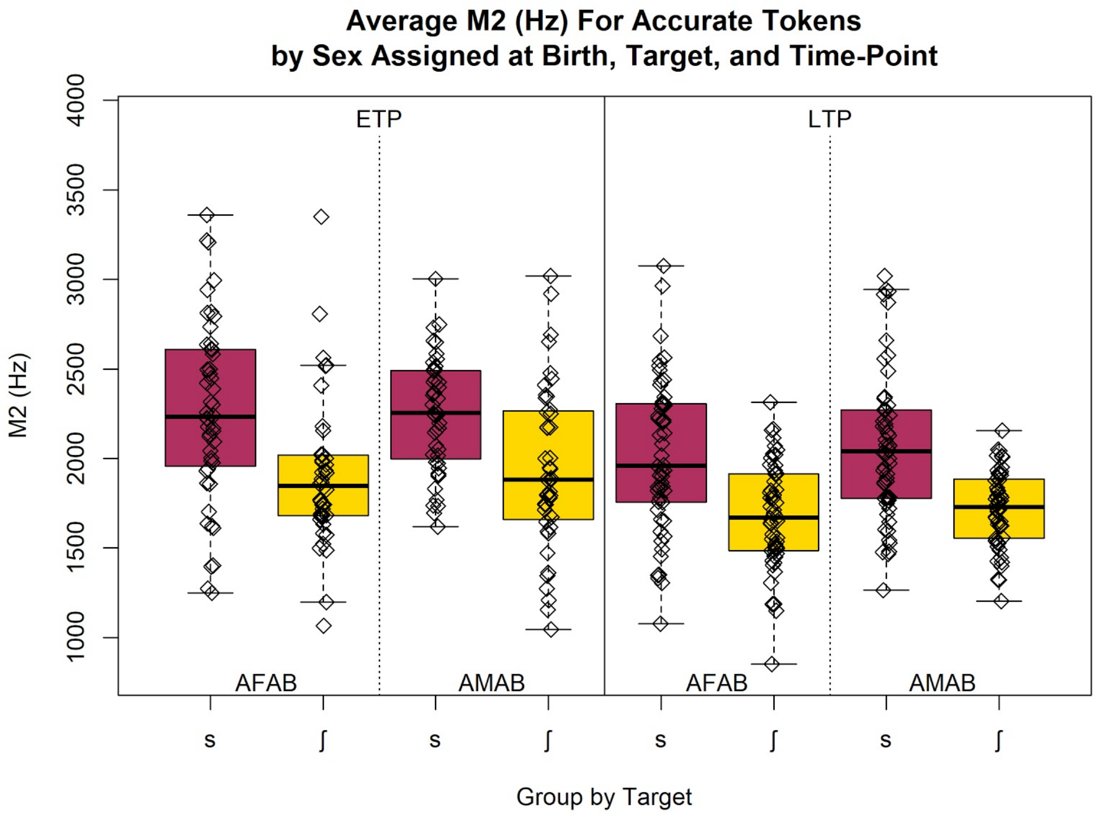
As this figure shows, m1 for /s/ was higher than m1 for /ʃ/, consistent with numerous acoustic studies. Moreover, the m1 for /s/ at LTP was higher than that for /s/ at ETP. The net result of this was that the difference in m1 between /s/ and /ʃ/ was larger at LTP than at ETP. At ETP, the m1 for /ʃ/ produced by AFAB children was higher than that for AMAB children, but at LTP the m1 for /ʃ/ was very similar for both groups. Moreover, the m1 for /s/ was higher for AFAB children than for AMAB children at both time-points. The three-way interaction appears to have been driven by the fact that the difference between /s/ and /ʃ/ was biggest for AFAB children at the LTP. That is, the high m1 of /s/ served to make it most different from /ʃ/ for that combination of SAB and time-point.

The next model examined m2. The most complex model that fit the data better than the next-simplest model was one in which there were fixed effects of SAB and target, and an interaction between them. Adding time-point, either alone or in an interaction with SAB or target, did not improve model fit. The coefficients for the full model including all of those variables are shown in Table 6, and are illustrated graphically in Figure 9. The full model, as opposed to the best fitting, is included for ease of comparison with other models. The coefficients for the simpler model with just SAB, target, and their interaction were very similar: there were significant effects of target, and a significant interaction between target and SAB.

Table 6: Coefficients for the Linear Mixed-Effects model Predicting m2 from Target and Time-Point

Estimate	Estimate	Standard Error	Df	t-value	Pr(>|t|)
(Intercept)	1953.320	23.733	102.224	82.304	<0.001
Time-Point (LTP=1, ETP=-1)	-89.509	9.371	135.694	-9.552	<0.001
Target (/s/=1, /ʃ/=-1)	187.045	10.991	22.820	17.018	<0.001
SAB (AFAB=1, AMAB=-1)	-26.577	22.594	105.496	-1.176	0.242
Time-Point * Target	-14.409	9.253	128.761	-1.557	0.122
Time-Point * SAB	-10.479	8.282	4398.843	-1.265	0.206
Target * SAB	16.738	8.253	4387.031	2.028	0.043
Time-Point * Target * SAB	-4.680	8.146	4357.785	-0.575	0.566

Figure 9:Average m2 for Productions Transcribed as Accurate, Separated by Time-Point, Sex Assigned at Birth, and Target



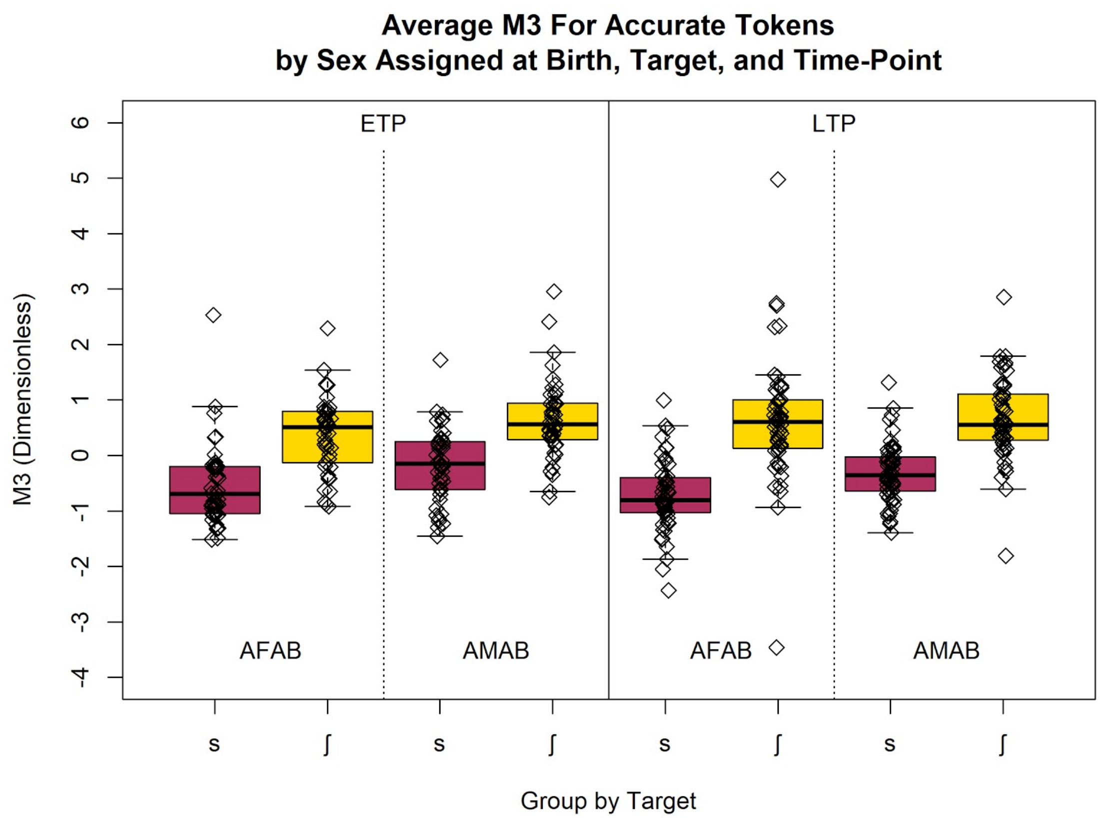
As this figure shows, /s/ productions had higher m2 than /ʃ/ productions for every combination of SAB and time-point. The interaction between SAB and target is difficult to discern visually from this figure, but appears to be due to the AMAB children's /s/ productions having a higher m2 than the AFAB children's /s/ productions, but not for their /ʃ/ productions. The difference in AMAB and AFAB children's /s/ productions is in the direction of the difference between adult men and women's /s/ productions. However, it is notable that this effect is much smaller than the effect of SAB on m1.

The next model examined m3. Once again, m3 indicates whether there is more energy above or below the average frequency. With a higher m3, there is more energy below the average, and this presentation is associated with perceptions of greater masculinity. The most complex model that fit the data better than the next-simplest model was the full model, with all three fixed effects and all possible interactions. The coefficients for this model are shown in Table 7 below, and are illustrated graphically in Figure 10.

Table 7: Coefficients for the Linear Mixed-Effects model Predicting m3 from Target and Time-Point

Estimate	Estimate	Standard Error	Df	t-value	Pr(>|t|)
(Intercept)	0.08141	0.04416	92.02879	1.844	0.068
Time-Point (LTP=1, ETP=-1)	0.02748	0.01883	140.46283	1.460	0.147
Target (/s/=1, /ʃ/=-1)	-0.57891	0.02303	21.38897	-25.140	<0.001
SAB (AFAB=1, AMAB=-1)	-0.14006	0.04098	100.17129	-3.418	<0.001
Time-Point * Target	-0.07771	0.01861	134.11917	-4.175	<0.001
Time-Point * SAB	0.03538	0.01619	4401.75634	2.186	0.029
Target * SAB	-0.07125	0.01613	4389.98615	-4.417	<0.001
Time-Point * Target * SAB	-0.04089	0.01593	4356.80250	-2.567	0.010

Figure 10: Average m3 for Productions Transcribed as Accurate, Separated by Time-Point, Sex Assigned at Birth, and Target



The findings for m3 closely mirror those for m1. There were robust differences between the m3 of target /s/ and target /ʃ/. The AMAB and AFAB participants' m3 for /s/ differed, with the AFAB children having a lower m3 than the AMAB children. This mirrors the differences found between adult men and women. The three-way interaction appears to be driven by the fact that /s/ and /ʃ/ differ the most for the AFAB children at LTP. Given how closely the findings for m3 mirror those for m1, we examined how closely correlated these measures are. While a distribution's mean frequency and its skewness are not statistically dependent, they might be highly correlated within a sample. Indeed, there was a very strong correlation between these measures, $r[4448]=-0.80$, $p<0.001$. Hence, the findings for m1 and m3 likely reflect the same underlying source of articulatory variation in /s/. In contrast, the correlation between m1 and m2 was not significant, $r[4448]=-0.01$, $p=0.49$.

The previous analyses examined all 110 children, some of whom did not have any accurate productions for one or more combinations of target and time-point. We ran an additional set of analyses to examine variation in m1 and m2 in subsets of participants. The first of these additional analyses was a subset of 82 participants who had at least two accurate productions of both of the targets at both of the time-points. The full set of results will not be presented here for brevity's sake, as they closely paralleled those for the full set of participants. The only exception to this was that the three-way interaction among target, time-point, and SAB was not significant for m1. The second of these additional analyses examined variation in /s/ only for the subset of 97 participants who had at least two usable productions of /s/ at both time-points. For m1, there were significant effects of SAB and time-point, but no interaction. For m2, there was a significant effect of time-point but not of SAB.

To summarize thus far, we found robust evidence that AMAB and AFAB children produce /s/ differently. The most consistent finding was that AFAB children produce this sound with a higher mean frequency than AMAB children. There was less consistent evidence that the degree of differentiation between the two groups increases over development. There was also evidence, though less robust, that AMAB children produce /s/ with a more-diffuse spectrum (i.e., a higher m2) than AFAB children. Given that these differences were present for /s/ but not for /ʃ/ supports an interpretation that the differences reflect intentional gender marking, rather than passive effects of sex-dimorphic anatomical structures.

2.4.4 Relationships between Acoustics and Listener Ratings

The final statistical analysis evaluated whether the summary measures of fricative acoustics are correlated with the measures of perceived gender typicality reported in Lackas (2019) and summarized in section 2.4.1. For this analysis, we focused on correlations between mean ratings and average m1, as this was the measure that showed the clearest evidence of gender marking in our analyses thus far. We examined correlations between mean m1 and gender ratings separately for each combination of SAB and time-point.

The correlations between mean m1 of /s/ and ratings at the ETP were not statistically significant for either AFAB children ($r[49]=0.23$, $p=0.11$) or the AMAB children ($r[46]=0.00$, $p=0.98$). This was also true at the LTP, though the correlation for the AFAB children did approach statistical significance at the conventional $\alpha=0.05$ level (AFAB children: $r[52]=0.25$, $p=0.07$, AMAB children: $[52]=0.20$, $p=0.15$). These relationships are shown graphically in the following two figures:

Figure 11: Scatterplot showing the relationship between mean gender ratings from Munson et al. (2019) and the mean m1 values from the current study at ETP. AFAB and AMAB children plotted separately

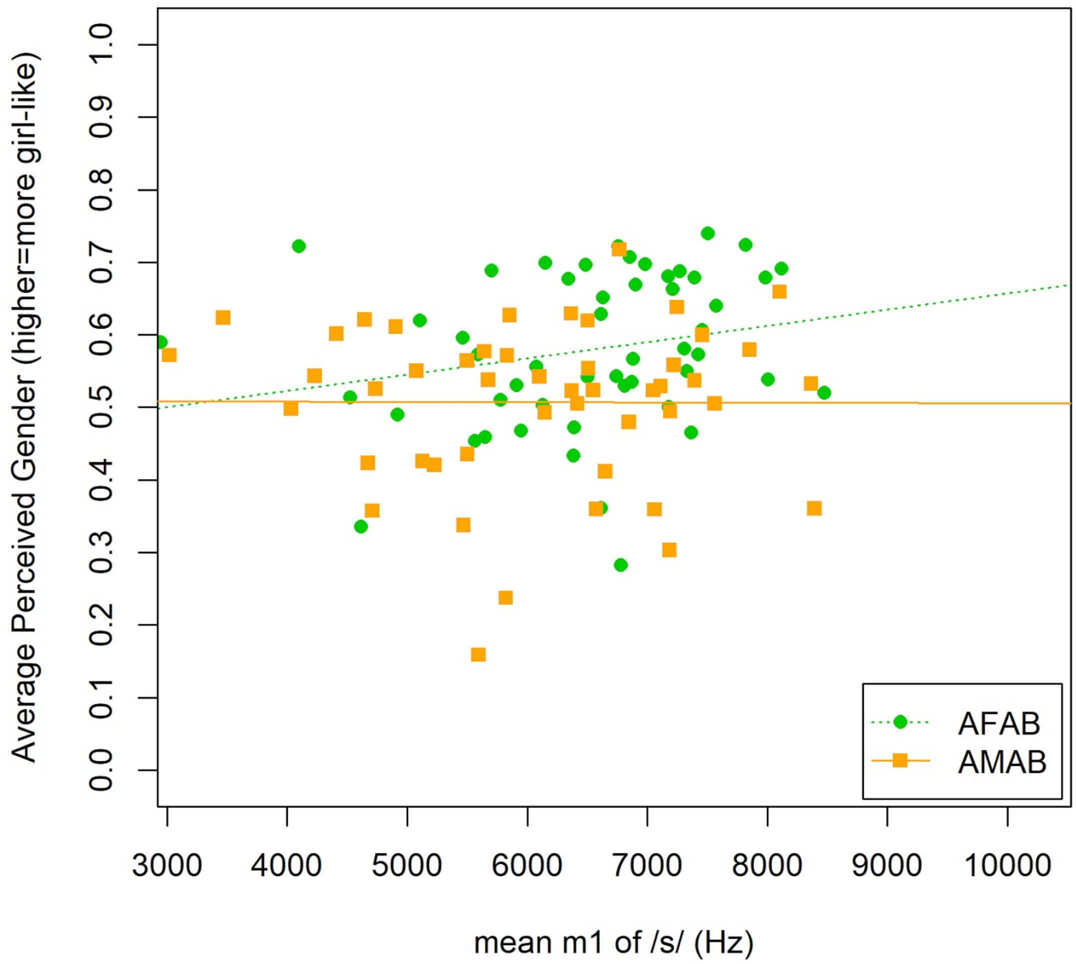
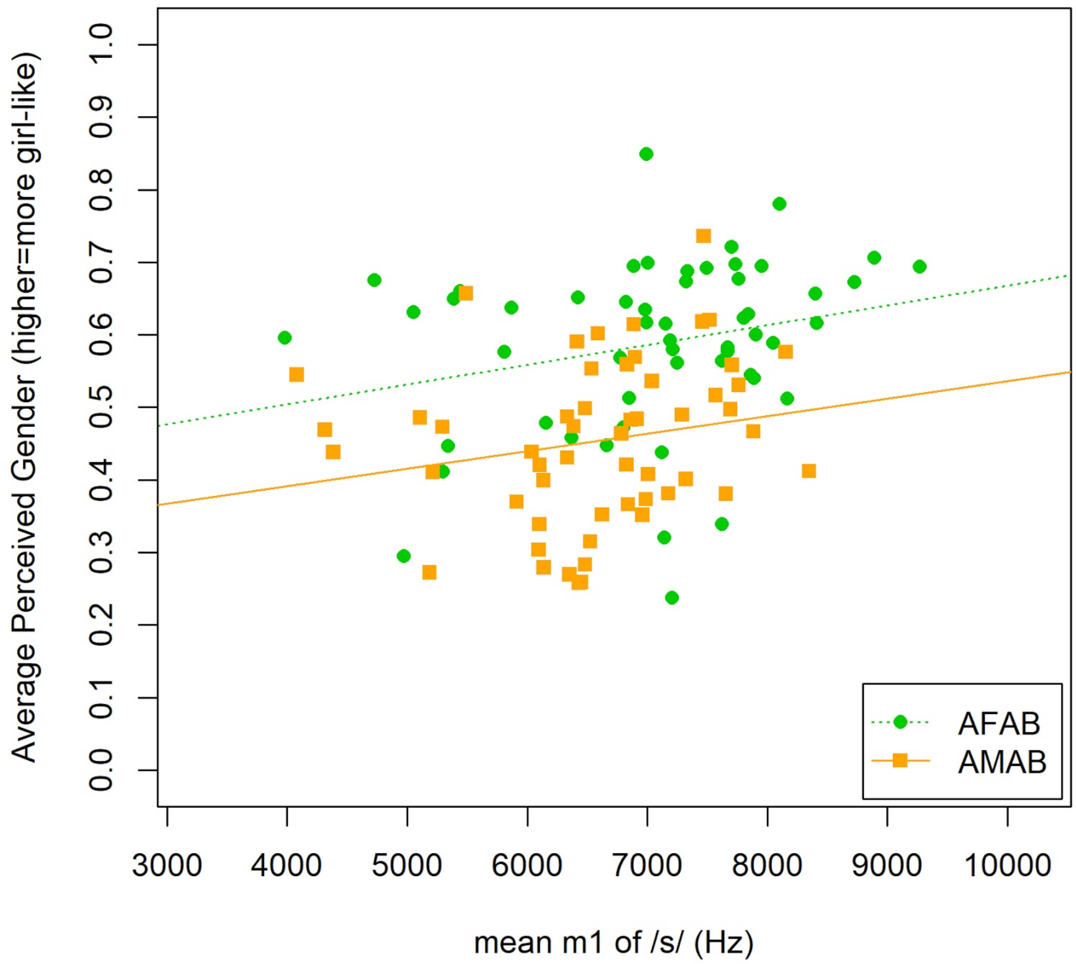


Figure 12: Scatterplot showing the relationship between mean gender ratings from Munson et al. (2019) and the mean m1 values ratings from the current study at LTP. AFAB and AMAB children plotted separately



These results do not provide strong additional support for the interpretation that the observed differences in /s/ across the group of children in this study represent gender marking, in that variation in /s/ was not correlated with gender ratings.

3. Summary

A few results can be concluded from the analyses performed that answer the outlined research questions. The first finding is that children AMAB and AFAB did not differ in their accuracy of /s/ and /ʃ/ at either time-point, though overall accuracy did increase for LTP. While these results do conflict with some previous studies, which show AFAB children having increased speech accuracy over AMAB children, most of those studies examined accuracy on articulation tests, such as the GFTA-3. They also conflict with the results of the standardized test scores for these children, in which the AFAB children had slightly lower scores on the GFTA-3 than the AMAB children. These tests measure the production of a wide range of sounds, as opposed to a targeted analysis of /s/ and /ʃ/. It is possible that children AFAB are more accurate, on average, across all speech sounds than children AMAB, with /s/ and /ʃ/ being speech sounds that are equally as accurate for each group. Being that the accuracy of /s/ and /ʃ/ are similar between AMAB and AFAB children, any acoustic differences that were found are not due to one group simply being more accurate than the other. Additionally, only tokens coded as accurate were used in the acoustic analysis. This means that any acoustic differences found can also not be attributed to there being more tokens in either the AFAB or AMAB group.

The next finding is that children AMAB and children AFAB do differ in the way they produce /s/. Similar to the findings of Li et al. (2016), children AFAB had higher m1 values and lower m3 values than children AMAB for /s/, but that same pattern was not found for /ʃ/ production. Unlike Li et al.'s (2016) prepubertal group findings, this study did find small differences between m2 values of /s/ for children AMAB and AFAB. As explained earlier, there is no anatomical basis for /s/ and /ʃ/ production, so the fact that we find increased differentiation only in the /s/, and across all three spectral measures, further supports the notion that /s/ is an especially important gendering-cueing speech sound, and that children are learning it. These differences in /s/ and /ʃ/ production for AMAB and AFAB children are similar across ETP and LTP, suggesting that the learning of gendered variants takes place early in life, possibly as early as the beginning of speech learning.

Analysis also showed that the m1 distinction between /s/ and /ʃ/ does seem to be driven by AFAB children. This means that the children AFAB were contrasting their /s/ and /ʃ/ to a greater extent than children AMAB. There is no anatomical reason for this differentiation, which suggests learning. Additionally, this pattern of contrast mimics the pattern that is displayed in adults. While this distinction could possibly suggest that children are not actually gender marking, and AFAB children are just more linguistically proficient than AMAB children, the inclusion of the children's standardized language scores suggests otherwise. With there being no significant difference between the language scores of the children who participated in this study, we can say with more confidence that gender marking is taking place.

Finally, despite the robust differences in /s/ production between AMAB and AFAB children, mean /s/ frequency in this study did not correlate with gender ratings from Munson et al. (2019). One possible explanation for this is that the acoustic analysis of /s/ is better suited to assess fine-grained differences than the listeners' ratings. The listeners heard the entire production of each word in order to make their rating. In listening to each word, the raters may have picked up on other gender cues or a combination of cues. As discussed in the introduction, gender is multifaceted and can be presented in many different ways. With these many gender expressions, there are also many ways to perform gender through speech, beyond /s/. We see this in adult speech, and evidence has suggested that children also use multiple different speech tools to convey gender. For example, Whiteside and Marshall (2001) showed VOT is a gender-cue in prepubertal children. Fundamental frequency and /s/ variation can also be used to express gender. This does not mean that all children use every one of these cues to differentiate their gender. Some children may express their gender through /s/ variation. Some may express it through VOT. Other children might express their gender through fundamental frequency, and it's possible that some children use entirely different gender markers. It may even be the case that each of these acoustic variables carry unique meanings about the speaker's gender. If this is true, there would not necessarily be a strong correlation with /s/ variation and gender ratings. For instance, when an AMAB child was rated as sounding more prototypically boy-like despite producing a prototypically girl-like /s/, it is likely that this child is simply expressing their boy-gender through something other than /s/.

4. Discussion and Conclusion

The results of this thesis add to the growing body of evidence that pre-pubertal children AMAB and AFAB are learning and performing gender differences in speech that align with adult speech patterns. The question then becomes how are children learning these patterns? Ladegaard and Bleses (2003) researched gender variation in children's communication and proposed two hypotheses for how children acquire gendered speech. The first hypothesis suggested was the frequency hypothesis, which claims that children's communication is shaped by speech and language that they are more often exposed to. While this is a potential factor in the acquisition of gendered communication of children, it does not account for children who are most often exposed to one speech style, but communicate in another. For instance, a child AMAB being raised and cared for by his single-mom and all-AFAB daycare providers, and who has three sisters, but still gender marks like other individuals AMAB. Inevitably, this child would be receiving almost all speech and language input from individuals AFAB, meaning there has to be another reason that this child developed the gender patterns that they did. This is where Ladegaard and Bleses's (2003) second hypothesis comes in, the role-model hypothesis. This hypothesis claims that children acquire gendered communication by modeling their speech and language after individuals that they identify with. Different speech styles are often used to affiliate oneself with a particular group (Pierrehumbert et al., 2004). As suggested by Li et. al. (2016), the gendered /s/ variants seen in children are used to align with certain peer groups. This is also likely the case for the children analyzed in this study. Stated by Li et al. (2016), children selectively attend to and learn from speech input from individuals they identify with, as to affiliate themselves as similar. In many cis-cases, this model is often a child's same-sex parent or same-sex peers. In the case of the child AMAB with majority of exposure to individuals AFAB, this hypothesis would suggest that this child learned gendered communication possibly through peers at daycare. Additionally, the role-model hypothesis could also allow for counter-identification. That is, if a child sees someone as an anti-role model, and does not identify with them, they could reject the gendered communication of that individual. While the role-model hypothesis is more plausible than the frequency hypothesis, Ladegaard and Blese (2003) acknowledge that it is still not complete enough to explain all the gendered variation in children's communication.

Another proposed hypothesis explaining how gendered variation is learned in children's speech is related to the concept of child-directed speech (CDS). It is known that adults modify their speech, language, and even pitch and cadence when talking to infants. Research has shown that CDS is beneficial for young children, especially for vocabulary growth and speech and language development. That said, it seems plausible that the way adults are speaking to children is developing gendered variation in speech as well. A number of researchers have investigated this very hypothesis. Serbin et al. (1993) found that parental input to children is not only integral to the child's gender development, but is also shaped by the child's SAB. That is, how adults speak to their child is largely influenced by whether the child was AMAB or AFAM. With that, Foulkes et al. (2005) found that Newcastle England mothers actually used nonstandard variations of the speech sound /t/ when talking to boys, as opposed to using standard forms for girls. This differing /t/-input was claimed to be due to the mismatch in social expectations that girls and women were to use more "formal" and socially-"prestigious" styles than boys or men. Additionally, Eccles et al. (1990) showed that adults also modified their communication input to children depending on how masculine or feminine the child performed, though parents, especially heterosexual cis-fathers, are more likely to accomadate for cis-children and non-cis girls (Kane, 2006). The more stereotypically feminine a girl presents, the more likely adults will use stereotypically feminine communication, such as soft-spoken language. This finding by Eccles et al. (1990) is particularly interesting because it suggests that the learning of gendered variation for children is iterative and reinforcing.

While these hypotheses have revealed valuable information in the study of the acquisition of children's gendered variation, there are still many unanswered questions regarding this topic. Some of these questions are due to the nature of this study as a secondary analysis of existing data that were collected for a different reason. Our analyses assume that all 110 children in this study were cisgender, and this is likely not true. The variation in ratings that we see for each of the groups likely represents a mix of differences in the extent to which children have learned the gendered phonetic variants associated with their gender, and the extent to which their gender can be assumed from their SAB. This is an unavoidable weakness of this secondary analysis. Studies like Li et al. provide methods for how a future prospective longitudinal study on the development of gendered speech could do a more-robust assessment of gender.

As this study, and others before it have shown, there is definite speech-gender learning taking place in children. Additional research should be done to better understand how this learning occurs, and where else it can be seen. While it would be difficult, it would also be interesting to investigate just how much intentionality and individual agency children have in their gendered speech performance. As Eccles et al. (1990) showed, gendered presentations of speech are somewhat of a self-fulfilling prophecy, in that, once a gendered variant is performed, it is reinforced by the adults in the child's life.

In moving away from the cis-ideal, and as society becomes more aware and accepting of the variety of gender indentities, will Eccles et al's (1990) findings still hold true? Will this allow for more gendered variation or less?

As Zimman (2017) pointed out, "the voice acts as a critical mediator of gender perceptions" (p. 340), making it an essential gender-performance tool for individuals. That said, it is hugely important to research the implications of speech and voice on gender perception. Additionally, when working with individuals in the field of speech-language pathology, acknowledging and accepting preferred gender identity is of the utmost importance in helping the individual learn, explore, express themselves, as well as, feel validated and supported in determining how they want to be perceived. This is just one of many reasons researching sex, gender, and the relationships they have with speech and language is valuable, necessary, and worthwhile. The importance of understanding these relationships is true for many fields, including speech-language pathology. Munson et al. (2015) showed that children AMAB who were identified by their parents as displaying behaviors that are not consistent with cisgender males (i.e., AFAB peer preferences, preferences for toys that are designed for cis girls) produce a variety of /s/ that is different from AMAB children whose behaviors were consistent with those of cis males. Specifically, they found an /s/ that was hyper-articulated, consistent with the /s/ produced by cis adult heterosexual women, and cis adult gay men. Munson et al. (2015) reasoned that the distinct /s/ production by AMAB children might reflect their performance of gender. Without knowing this phonetic variation was socially meaningful, a speech-language pathologist could identify this distinctive /s/ as an error and correct it, all while unknowingly invalidating the child's identity and perpetuating the cis ideal. Just as dialect is considered a difference, as opposed to a disorder, so should gendered variation. Researching gendered speech development in children can help normalize non-cis genders, validate identities, and distiguish disorders from differences.

To conclude, for years, research has suggested that like adults, children AFAB and AMAB modify their speech to perform gender. This study supports this notion by revealing acoustic differences in the production of /s/ and /ʃ/ between children AFAB and children AMAB. The most important findings are as follows: AMAB and AFAB children did not differ in their accuracy of /s/ and /ʃ/, and for both time points, AMAB and AFAB children did differ in their /s/ production across all three spectral measure (m1, m2, m3), but did not differ in their /ʃ/ production. These findings add to the research showing the importance of /s/, specifically, as a gender-marking speech sound. Additionally, these results suggest that the learning of gendered speech variation takes place very early in a child's life.

References

Amir, O., Engel, M., Shabtai, E., & Amir, N. (2012). Identification of Children's Gender and Age by Listeners. *Journal of Voice, 26*, 313-321.

Avery, J. D., & Liss, J. M. (1996). Acoustic characteristics of less−masculine−sounding male speech. *The Journal of the Acoustical Society of America*, *99*(6), 3738-3748.

Babel, M., & Munson, B. (2014). Producing socially meaningful linguistic variation.

Barbier, G., Boë, J., Captier, G., & Laboissière R. (2016). Human vocal tract growth: A longitudinal study of the development of various anatomical structures. *16th Annual Conference of the International Speech Communication Association (Interspeech 2015)*, International Speech Communication Association, Sep 2015, Dresden, Germany.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-9, https://CRAN.R-project.org/package=lme4.

Beauvoir, S. (1973). *Second Sex*. New York: Vintage.

Bennett, S. (1981). Vowel formant frequency characteristics of preadolescent males and females. *The Journal of the Acoustical Society of America*, 69(1), 231-238.

Bennett, S., & Weinberg, B. (1979). Sexual characteristics of preadolescent childrens' voices. *The Journal of the Acoustical Society of America*, *65*(1), 179-189.

Blackless, M., Charuvastra, A., Derryck, A., Fausto−Sterling, A., Lauzanne, K., & Lee, E. (2000). How sexually dimorphic are we? Review and synthesis. *American Journal of Human Biology: The Official Journal of the Human Biology Association*, *12*(2), 151-166.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech communication*, *20*(3-4), 255-272.

Butler, J., & Taylor & Francis. (1990). *Gender Trouble : Feminism and the Subversion of Identity* (Thinking Gender). New York: Routledge.

Busby, P. A., & Plant, G. L. (1995). Formant frequency values of vowels produced by preadolescent boys and girls. The Journal of the Acoustical Society of America, 97(4), 2603-2606.

Cartei, V., Cowles, W., Banerjee, R. and Reby, D. (2014), Control of voice gender in pre‐pubertal children. Br J Dev Psychol, 32: 100-106.

Cartei, V., & Reby, D. (2013). Effect of formant frequency spacing on perceived gender in pre-pubertal children's voices. *PLoS One*, *8*(12).

Chen, M. Y. (1997). Acoustic correlates of English and French nasalized vowels. *The Journal of the Acoustical Society of America*, 102(4), pp. 2360–2370.

Crenshaw, K. (1991). Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. *Stanford Law Review, 43*(6), 1241-1299.

Cychosz, M., Munson, B., & Edwards, J. R. (2021). Practice and experience predict coarticulation in child speech. *Language Learning and Development*, 1-31.

Diehl, R. L., Lindblom, B., Hoemeke, K. A., & Fahey, R. P. (1996). On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics, 24*, 187-208.

Eccles, J. S., Jacobs, J. E., & Harold, R. D. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of social issues*, *46*(2), 183-201.

Eckert, P. (1989). The whole woman: Sex and gender differences in variation. *Language variation and change*, *1*(3), 245-267.

Eckert, P. (2014). The Problem with Binaries: Coding for Gender and Sexuality. *Language and Linguistics Compass 8(11)*, 529-535.

Eckert, P., & McConnell-Ginet, S. (1992). Think practically and look locally: Language and gender as community-based practice. *Annual review of anthropology*, *21*(1), 461-488.

Erskine, M. E., Munson, B., & Edwards, J. R. (2020). Relationship between early phonological processing and later phonological awareness: Evidence from nonword repetition. *Applied Psycholinguistics*, 41(2), 319-346.

Fant, G. (1975). Non-uniform vowel normalization. *STL-QPSR*, *16*(2-3), 1-19.

Fausto-Sterling, A. (2000). *Sexing the body: Gender politics and the construction of sexuality*. Basic Books.

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal behaviour*, *69*(3), 561-568.

Fine, C., & Rush, E. (2018). “Why does all the girls have to buy pink stuff?” The ethics and science of the gendered toy marketing debate. *Journal of Business Ethics*, 149(4), 769-784.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *106*(3), 1511-1522.

Flipsen Jr, P., Shriberg, L., Weismer, G., Karlsson, H., & McSweeny, J. (1999). Acoustic characteristics of /s /in adolescents. *Journal of Speech, Language, and Hearing Research*, *42*(3), 663-677.

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics 34*, 409-438.

Foulkes, P., Docherty, G., & Watt, D. (2005). Phonological Variation in Child-Directed Speech. *Language, 81(1)*, 177-206.

Fox, R. A., & Nissen, S. L. (2005). Sex-Related Acoustic Changes in Voiceless English Fricatives. *Journal of Speech, Language, and Hearing Research*, *48,* 753-765.

Fuchs, S., & Toda, M. (2010). Do differences in male versus female /s/ reflect biological or sociophonetic factors? In *Turbulent sounds: An interdisciplinary guide* (pp. 281-302). Mouton de Gruyter.

Fung, P., Schertz, J., & Johnson, E. K. (2021). The development of gendered speech in children: Insights from adult L1 and L2 perceptions. *JASA Express Letters*, *1*(1), 014407.

Goldstein, U. G. (1980). *An articulatory model for the vocal tracts of growing children* (Doctoral dissertation, Massachusetts Institute of Technology).

Heffernan, K. (2010). Mumbling is Macho: Phonetic Distinctiveness in the Speech of American Radio DJs. *American Speech, 85(1)*, 67-09.

Hillenbrand, J. M., & Clark, M. J. (2009). The role of f 0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, & Psychophysics*, *71*(5), 1150-1166.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America*, *97*(5), 3099-3111.

Iseli, M., Shue, Y. L., & Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of America*, *121*(4), 2283-2295.

Johnson, L. L., Bradley, S. J., Birkenfeld-Adams, A. S., Kuksis, M. A. R., Maing, D. M., Mitchell, J. N., & Zucker, K. J. (2004). A parent-report gender identity questionnaire for children. *Archives of sexual behavior*, 33(2), 105-116.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, *108*(3), 1252-1263.

Kane, E. W. (2006). "No way my boys are going to be like that!" Parents' responses to children's gender nonconformity. *Gender & Society*, *20*(2), 149-176.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *the Journal of the Acoustical Society of America*, *87*(2), 820-857.

Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language variation and change*, *2*(2), 205-254.

Lackas, N. (2019). The Development of Gender-Typical Speech in Children: A Longitudinal Analysis.

Ladegaard, H. J., & Bleses, D. (2003). Gender differences in children's speech: the acquisition of sociolinguistic competence. *International Journal of Applied Linguistics, 13(2)*, 222-233.

Latinus, M., & Taylor, M. J. (2012). Discriminating male and female voices: differentiating pitch and gender. *Brain topography*, *25*(2), 194-204.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America, 105*, 1455-1468.

Levon, E. (2007). Sexuality in context: Variation and the sociolinguistic perception of identity. *Language in Society*, 533-554.

Li, F., Rendall, D., Vasey, P. L., Kinsman, M., Ward-Sutherland, A., & Diano, G. (2016). The development of sex/gender-specific /s/ and its relationship to gender identity in children and adolescents. *Journal of Phonetics, 57*, 59-70.

Lieberman, D. E., McCarthy, R. C., Hiiemae, K. M., & Palmer, J. B. (2001). Ontogeny of postnatal hyoid and larynx descent in humans. *Archives of oral biology*, *46*(2), 117-128.

Lindblom, B. E., & Sundberg, J. E. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America*, *50*(4B), 1166-1179.

Mack, S., & Munson, B. (2012). The influence of/s/quality on ratings of men's sexual orientation: Explicit and implicit measures of the 'gay lisp'stereotype. *Journal of Phonetics*, *40*(1), 198-212.

Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America*, *125*(6), 3962-3973.

Martin, C. L., & Ruble, D. (2004). Children's Search for Gender Cues: Cognitive Perspectives on Gender Development. *Current Directions in Psychological Science, 13(2)*, 67-70.

Munson, B., & Babel, M. (2019). The phonetics of sex and gender. In W.F. Katz and P.F. Assmann (Eds.), *The Routledge Handbook of Phonetics* (pp. 499-525). New York: Routledge.

Munson, B., Crocker, L., Pierrehumbert, J. B., Owen-Anderson, A., & Zucker, K. J. (2015). Gender typicality in children's speech: A comparison of boys with and without gender identity disorder. *The Journal of the Acoustical Society of America, 137*, 1995-2003.

Munson, B., Lackas, N., & Koeppe, K. (2019). The Longitudinal development of gendered speech production in children: Patterns and predictors. In *Oral Presentation at the Boston University Conference on Language Development*.

Munson, B., Logerquist, M. K., Kim, H., Martell, A., & Edwards, J. (2021). Does Early Phonetic Differentiation Predict Later Phonetic Development? Evidence From a Longitudinal Study of/ɹ/Development in Preschool Children. *Journal of Speech, Language, and Hearing Research*, 1-21.

Munson, B., McDonald, E. C., DeBoe, N. L., & White, A. R. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics*, *34*(2), 202-240.

Newman, R. S., Clouse, S. A., & Burnham, J. L. (2001). The perceptual consequences of within-talker variability in fricative production. *The Journal of the Acoustical Society of America*, *109*(3), 1181-1196.

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech, Language, and Hearing Research*, 32(1), 120-132.

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America, 109*, 2988-2998.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the acoustical society of America*, *24*(2), 175-184.

Pierrehumbert, J. B., Bent, T., Munson, B., Bradlow, A. R., & Bailey, J. M. (2004). The influence of sexual orientation on vowel production (L). *The Journal of the Acoustical Society of America*, *116*(4), 1905-1908.

Podesva, R. J., & Van Hofwegen, J. (2014). How conservatism and normative gender constrain variation in inland California: The case of/s. *University of Pennsylvania Working Papers in Linguistics*, *20*(2), 15.

Puts, D. A., Hill, A. K., Bailey, D. H., Walker, R. S., Rendall, D., Wheatley, J. R., & Ramos-Fernandez, G. (2016). Sexual selection on male vocal fundamental frequency in humans and other anthropoids. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1829), 20152830.

Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. (2007). Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior*, *28*(5), 340-344.

Serbin, L. A., Powlishta, K. K., Gulko, J., Martin, C. L., & Lockheed, M. E. (1993). The development of sex typing in middle childhood. *Monographs of the society for research in child development*, i-95.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass 3(2)*, 621-640.

Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender.

Stoel−Gammon C. (2001). Transcribing the speech of young children. *Topics in Language Disorders, 21,* 12–21

Stuart-Smith, J. (2007). Empirical evidence for gendered speech production: /s/ in Glaswegian. In *Laboratory Phonology 9* (pp. 65-86). New York: Mouton de Gruyter.

Sussman, J. E., & Sapienza, C. (1994). Articulatory, developmental, and gender effects on measures of fundamental frequency and jitter. *Journal of Voice*, 8(2), 145-156.

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America, 85,* 1699-2007.

Van Bezooijen, R. (1995). Sociocultural Aspects of Pitch Differences between Japanese and Dutch Women. *Language and Speech, 38(3)*, 253-265.

Vorperian, H. K., & Kent, R. D. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research*.

Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., & Yandell, B. S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *The Journal of the Acoustical Society of America*, *117,* 338-350.

Vorperian, H. K., Wang, S., Schimek, E. M., Durtschi, R. B., Kent, R. D., Gentry, L. R., & Chung, M. K. (2011). Developmental Sexual Dimorphism of the Oral and Pharyngeal Portions of the Vocal Tract: An Imagining Study. *Journal of Speech, Language, and Hearing Research, 54*, 995-1010.

Weinberg, B., & Bennett, S. (1971). Speaker Sex Recognition of 5- and 6-Year-Old Children's Voices. *The Journal of the Acoustical Society of America, 50*, 1210-1213.

Weirich, M., Fuchs, S., Simpson, A., Winkler, R., & Perrier, P. (2016). Mumbling: Macho or Morphology? *Journal of Speech, Language, and Hearing Research, 59*, 1587-1595.

Whiteside, S. P., & Hodgson, C. (2000). Some acoustic characteristics in the voices of 6-to 10-year-old children and adults: a comparative sex and developmental perspective. *Logopedics Phoniatrics Vocology*, *25*(3), 122-132.

Whiteside, S. P., & Irving, C. J. (1998). Speakers' sex differences in voice onset time: a study of isolated word production. *Perceptual and motor skills*, *86*(2), 651-654.

Whiteside, S. P., & Marshall, J. (2001). Developmental trends in voice onset time: Some evidence for sex differences. *Phonetica*, *58*(3), 196-210.

Zimman, L. (2017). Gender as stylistic bricolage: Transmasculine voices and the relationship between fundamental frequency and /s/. *Language in Society*, 46(3), 339-370.