THE DEVELOPMENT OF GENDERED SPEECH IN CHILDREN AND ADOLESCENTS: A CROSS-LINGUISTIC STUDY OF ENGLISH AND MANDARIN VOICE ONSET TIME

GRAHAM MCKENZIE
Bachelor of Science, **University of Lethbridge, 2017**

A thesis submitted
in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in

PSYCHOLOGY

Department of Psychology
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

THE DEVELOPMENT OF GENDERED SPEECH IN CHILDREN AND ADOLESCENTS: A CROSS-LINGUISTIC STUDY OF ENGLISH AND MANDARIN VOICE ONSET TIME

Graham McKenzie

Date of Defence: May 6th, 2020

Dr. Fangfang Li Thesis Supervisor	Associate Professor	Ph.D
Dr. Paul Vasey Thesis examination committee	Professor	Ph.D
Dr. David Logue Thesis examination committee	Associate Professor	Ph.D
Dr. Scott Allen Thesis Examination Chair	Professor	Ph.D

七転び八起き

"Fall seven times and stand up eight"

– Japanese proverb

DEDICATION

This research is dedicated to my Bāchan Sue Kado, who always encouraged me to put my best foot forward and to prioritize my education. It is also dedicated to my loving wife Morgan, who has unfailingly supported me. It is also dedicated to my daughters, Spencer, Addison, and Millicent.

ABSTRACT

Stop consonant production has been shown to vary with speaker's sex, but cross-language developmental studies are lacking to explain how sex-specific speech patterns come about independent of a specific-language context. In this research, I examined English- and Mandarin-speaking children and adolescents aged 6 to 17 to determine cross-language similarities and differences via measuring the voice onset time (VOT) of their stop productions. VOT of word initial /t/ and /d/ in Mandarin, and /t, d, p, b, k, g/ in English was measured and normalized by vowel and word duration to control for speech rate. An overall sex difference was found for both languages. This difference was much more robust in Mandarin than in English. Developmental trends and cross language comparison suggest that males and females utilize VOT for gender-specific production to varying degrees in different culture.

ACKNOWLEDGEMENTS

I want to formally thank the University of Lethbridge, and the School of Graduate Studies, for providing the opportunity for me to study as a graduate student. My experience has been rich, and I feel that I have grown both in knowledge and as a person. I want to acknowledge that the university is located on traditional Blackfoot territory and that the Blackfoot people, and their traditions, have shaped my experience as a student. I also want to especially acknowledge the Department of Psychology, which has been my home these past few years. The knowledge I've gained and the friendships I have developed while in the department are more than I could have hoped for when I started.

I want to express my sincere gratitude for Dr. Fangfang Li and her guidance throughout my entire graduate studies. It would be nearly impossible for me to count how Dr. Li has gone out of her way to help me achieve my goals. Dr. Li has been an excellent mentor in every aspect of my research. She has always encouraged me to look at big picture ideas and to connect ideas and perspectives. Her patient and firm guidance has pushed me to achieve more than I thought possible when I started my master's degree.

I also want to acknowledge the patient mentorship of my committee members, Dr. Paul Vasey and Dr. David Logue. Dr. Vasey was a constant source of enthusiasm and encouragement throughout my graduate studies. I am incredibly grateful for his vast knowledge and his ability to break down complex topics into simple terms that I could easily understand. Dr. Logue has been a huge source of help to me as well during my studies. Similar to Dr. Vasey, I am appreciative of his patience and mastery in teaching complex ideas simply to me. Dr. Logue's work ethic, both in research and in his life, has been a huge inspiration to me as I have been at the university. He always challenged me to dig a little deeper and think critically about my research, while also being incredibly supportive. I am indebted to both of these professors for how much they have taught me these past few years.

My graduate experience would not have been the same without the constant mentorship and support from Dr. Jennifer Mather. I have learned so much from Dr. Mather about both academic success and life itself. Through my graduate studies, she has gone out of her way to teach me, include me in her research, and show interest in my research. From guiding me when I was a teaching assistant, to bringing me fresh vegetables from her garden, taking me to meetings with the Dean of Arts and Science, and pushing me to follow my research passions, there was never a dull moment with Dr. Mather. Importantly, she always pushed me to tackle problems from multiple angles and to see the big picture in research. It is thanks to Dr. Mather's mentorship that I was able to succeed in graduate school.

A special acknowledgment needs to be given to the two administrative assistants from the Department of Psychology during my studies, Leanne Wehlage-Ellis and Erin McSween. Both Leanne and Erin did so much for me during my studies that were often in the minutia of my research. They kept me sane and organized, and their contributions cannot be overstated.

My thesis project would not have been possible without the support that I received from the Speech Development Lab. I received valuable feedback on my work throughout my graduate experience from several lab members. I would like to particularly acknowledge the contributions of Nicole Van Rootselaar and Mitchell Harris. I am greatly indebted to Nicole for her mentorship throughout my entire graduate journey. Nicole patiently guided me through a variety of projects, from research to presentations. Her experience and friendship have been invaluable, and she has been one of my most significant supports in the lab. Mitchell helped me analyze the acoustic data found and was a companion when the lab was nearly empty. I was greatly supported by his hard work and sense of humour during the first year of my graduate studies. It was the support of these individuals that made me look forward to coming into the lab day after day.

I would also like to acknowledge the support I received from the Birdsong and Vokey labs during my studies. I was able to share ideas with the members of these labs and gain valuable perspective on my research. This was especially true of my fellow graduate students Peter Mower, Lauren Vomberg, Kevin McKillop, and Anthony (AJ) Schoen. Peter always let me bounce ideas off of him and would intelligently play devil's advocate for me when I needed to think through an idea. Lauren's work ethic, creativity, and enthusiasm were always an inspiration for me. Kevin's level-headedness, patience, and sense of humour helped keep my sane through the latter half of my studies. And a special thanks to AJ for helping me to think outside the box, and to never assume anything. The friendship and research expertise of these individuals bolstered me during my studies.

Finally, I want to thank all my family and friends who made it possible for me to complete my graduate studies. It is because of the support and love of these individuals that I was able to conduct this research. I want to thank my loving wife, Morgan, who has always supported me in my schooling, and who is the rock of my life. I also want to thank my three daughters Spencer, Addison, and Millicent. Your enthusiasm for learning and curiosity about the world inspires me.

I want to acknowledge the support and love of my parents, Ronald and Vicki McKenzie. They are my original mentors, and have always encouraged me to push myself and to try to become my best self. I also want to thank my sister Kimiko, whose encouragement, support, and feedback has been an essential part of my university learning and experience. I would also like to thank Neil and Julie Sieben, as well as the whole Sieben family, who have been an unfailing source of encouragement since I became a part of the family. A special thank you goes to Jimmy and Alex Stringam for the support and laughter we have shared. While most of these individuals did not directly contribute to this thesis, this research would not have been possible without the immense support they provided while I studied.

Finally, I want to acknowledge and thank my Bāchan, Shizue "Sue" Kado, who passed away before being able to see me finish my master's degree. She was, and still is, a tremendous source of inspiration. She always encouraged me to pursue education, to aim high, and to work hard. I hope that I can live up to her legacy and embody the lessons she taught me.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AMM	Additive mediation model
PCA	Principal components analysis
VOT	Voice onset time

CHAPTER 1: INTRODUCTION

In addition to linguistic information, human speech also encodes a variety of other personal dimensions of a speaker, including age, educational background and socioeconomic status (Labov, 1972; Peterson & Barney, 1952; Purnell, Idsardi, & Baugh, 1999). However, by far, the most noticeable variation in adult speech are those that cue a speaker's sex (Trudgill, 1983). Differences in female and male speech production have been observed in several different levels of speech. Male and female speakers have been observed to differ in conversational strategies (Roger & Schumacher, 1983; Roger & Nesshoever,1987), sentence structure (Mulic, Wiemann, & Widenmann, 1988), word choice (Xie, 2013), and use of intonation (Jiang, 2011). One of the most interesting of these sex-specific differences is the difference in the production of individual speech sounds.

Sex-specific patterns have been observed in a variety of speech sounds, or phonemes, in both adult and children's speech. For example, male and female speakers are noted for producing vowels differently. In general, women produce vowels with higher voice pitch than men (Hillenbrand et al., 1995; Peterson & Barney, 1952; Sachs, Lieberman, & Erikson, 1973). Similarly, sex-specific production has been observed in consonant production. Female speakers produce the sibilant /s/ (/s/ as in see) with a higher-frequency noise than male speakers (Fox & Nissen, 2005; Kinsmen & Li, 2013; Li et al., 2016). Similarly, sex-specific production differences have also been observed in stop-consonants such as /p/, /t/, and /k/ (ie. Pig, tea, cup respectively; Koenig, 2000; Li, 2013; Robb, Gilbert, & Lerman, 2005; Ryalls, Zipprer, & Baldauff, 2014; Scharf & Masure, 2002; Schwartz, 1992; Whiteside & Irving, 1997; Whiteside & Irving, 1998). Female speakers tend to produce more aspiration, which is the ejection of air from the vocal tract when they produce these stop-consonant sounds than males.

Two main factors contribute to the differences in sex-specific phonetic production (Simpson, 2009). The first factor is differences that arise from biophysical differences between male and female speakers. Differences in the dimensions of the vocal tract can alter speech sound output. The second is socially conditioned behaviours that speakers associate their gender identity. The relationship between these biophysical and learned behaviours, commonly referred to as nature vs. nurture, on speech is complex. The following sections overview how both anatomical and socially motivated behaviour on phonetic output.

1.1 ANATOMICAL FACTORS THAT AFFECT SPEECH

There are several anatomical differences between females and males that could potentially affect speech output. The first is a general difference in size between females and males. Males tend to be larger than females, both in height and weight (LoMauro & Aliverti, 2018). Larger individuals tend to have proportionally larger vocal tract dimensions and vocal cords than smaller individuals regardless of sex (LoMauro & Aliverti, 2018). This difference in vocal tract size affects the acoustic quality of a speaker. For example, taller individuals tend to produce lower pitch during vowel production (Pisanski et al., 2014). A review of the airways and vocal tract dimensions found that some of the differences between males and females, such as lung size, oral cavity size, and airway diameter could mostly be attributed to a difference in physical size, rather than a difference in sex (see LoMauro & Aliverti, 2018). Therefore, size-based differences could not be considered a true sex difference, since it is a function of size rather than sex. We would predict that the taller individual would produce a relatively lower pitch voice than the shorter individual (Pisanski et al., 2014). However, while physical size differences play some role in the production of speech, these differences in physical size are not sufficient to explain the differences between female and male vocal anatomy. This is specifically true of vocal cord size and vocal tract length.

The vocal cords in adult males are both thicker and longer than adult females. Before puberty, there are few differences in vocal cord size between males and females (Fitch & Giedd, 1999; LoMauro & Aliverti, 2018; Markova et al., 2016). During puberty, male vocal cords can grow up to twice the size of female vocal cords (Fitch & Giedd, 1999; Markova et al., 2016). The 'voice drop' in teenage boys is a direct result of this increase in vocal cord size. A speaker's voice pitch is generated as air passes from the lungs into the vocal tract through the vocal cords, which causes the cords to vibrate. Due to the size difference in vocal cords, the larger male vocal cords vibrate at a slower rate than female vocal cords, and therefore males produce a lower pitch. This difference in pitch is a key difference between male and female speech production. Previous research has found that speaker pitch is a reliable cue for speaker sex, with the lower pitch being perceived as more male-like than higher voice pitch (Hillenbrand et al., 1995).

In addition to vocal cord differences, the size of adult male vocal tract is longer than female's and differs in shape (Fitch & Giedd, 1999). In addition to pitch frequencies, as speech sound moves through the vocal tract it produces resonant frequencies. Similar to how the shape of a musical instrument affects the sound quality that is produced, the difference in vocal tract shape affects vocal sound quality. As noted above, before the onset of puberty, female and male vocal tracts are very similar both in size and shape and are primarily correlated with physical size (LoMauro & Aliverti, 2018; Markova et al., 2016). However, after puberty, the correlation between physical size and vocal tract length becomes sex-specific (Fitch & Giedd, 1999; Markova et al., 2016). During puberty, there is an increase in the size of the larynx for both males and females. However, this change is much more prominent in males, which results in proportionally larger larynx than females. The average length of the adult female vocal tract from the lips to the vocal cords is approximately 14 cm in length, while males are approximately 17.5 cm (Markova et al., 2016). The difference in size and shape of the vocal tract results in the production of lower resonant frequencies in males than in females. Relatively lower resonant frequencies have been associated with male voice quality, while higher frequencies are associated with female speech (Hillenbrand et al., 1995).

The differences in vocal tract anatomy account for much of the variation in sex-specific vowel production. Studies looking at the average pitch and formant frequencies of male and female speakers have found that male speakers consistently produce vowels with lower pitch and lower formant frequencies than females (Simpson 2009). While there seem to be some language-specific effects on vowel production, this relative difference between male and female speakers in vowel production frequencies is observed across multiple languages such as English (Takafuta, 1972), French (Boë et al., 1975), and Chinese (Rose, 1991). However, each of these studies found a different average pitch frequency for males and females. This language-specific difference seems to suggest that sex-specific differences in speaker's pitch cannot entirely be explained only vocal anatomy. Social factors also seem to contribute to these production differences.

1.2 THE ROLE OF SOCIAL LEARNING AND GENDER

While anatomical differences can explain some sex-specific output differences in speech, they are not sufficient to explain all the differences and variations seen between male and female speakers. An individual's gender identity, which stems from socially conditioned expectations about behaviour, also affects an individual's speech output. It was previously noted that differences in vowel production seem to be affected by language background. Before the onset of puberty, males and females have similar vocal cord and vocal tract sizes (Fitch & Giedd, 1999) However, when children's vowel pitch and resonant frequencies were tested, there was a male/female difference, with female speakers producing higher pitched vowels than males (Fitch & Giedd, 1999; Lee, Potamianos, & Narayanan, 1999; Perry, Ohde, & Ashmead, 2001). It was observed that children could manipulate both pitch and resonant frequencies when asked to produce speech like a boy or a girl (Cartei, Cowles, Banerjee, & Reby, 2014; Lee, Potamianos, & Narayanan, 1999; Perry, Ohde, & Ashmead, 2001). These studies demonstrate that children can intentionally produce gendered speech to express gender identity as well as identify the intended gender of other speakers (Cartei, Cowles, Banerjee, & Reby, 2014).

The role of social learning is further evident in the production of consonant sounds. These sounds are compelling to look at because the production of consonant sounds less affected by the size of the vocal tract than vowels. This can be seen in the production of the consonant /s/. /s/ is produced by bringing the tongue tip close to the alveolar ridge (the small hard ridge behind the upper front teeth) to create a small opening. Air is passed through this opening, producing turbulent noise that is perceived as /s/. The position of the tongue tip affects the noise frequency, can be used to signal speaker sex. Female speakers tend to produce /s/ with a more anterior tongue tip than males. This results in higher frequency noise components relative to male speakers in the fricative noise spectrum (Fox & Nissen, 2005; Li et al., 2016).

The majority of research conducted on /s/ has found that social or cultural specific influence is largely responsible for its sex-specific production and variation within sex (Fox & Nissen, 2005; Kinsmen & Li, 2013; Li et al., 2016). For example, a study of Glasgean speakers found that male and female speakers' /s/ production differenced significantly. However, additional social factors, such as social class, also seemed to affect /s/ production in women (Stuart-Smith, 2007). This is evidence that sex-specific production of /s/ is influenced by more than just biological sex. Similarly, a study of adolescent boys diagnosed with gender identity disorder (GID) found that these boys produce less gender-typical /s/ than non-GID boys, suggesting that gender identity can play a significant role in speech production (Munson et al., 2015). When this was tested, Li et al. (2016) found that gender identity could explain both the variation between sexes and within sexes for /s/ variation.

These examples seem to suggest that sex-specific production of /s/ is influenced more by social factors than by biological sex. However, the role of biological sex cannot be discounted entirely. Li et al. noted that there were potential anatomical differences that could affect /s/ production (2016). Munson et al. (2015) noted that the gender atypical behaviours that were observed in boys with GID were adopted at an early age before there would have been adequate exposure to either a gay or transgender speech or social network from which participants could observe. Underlying biophysical components may contribute to these differences.

More research will be needed to understand the relationship between social and anatomical factors on sex-specific speech sounds, such as vowels, /s/, and the far less understood stop-consonants. The relationship between biological sex and gender identity is both complex and nuanced. In terms of phonetic output, this can make exploring the underlying origins of sex-specific speech difficult. Two effective methods that have been used to investigate the social/anatomical relationship of speech production are developmental studies and cross-linguistic comparison. Developmental studies look at changes that develop in speech before, during, and after the onset of puberty and across the lifespan. Crosslinguistic studies compare language production of similar sounds across different language contexts. Crosslinguistic studies help elucidate language-specific features of language, and help reveal common language milestones. The following two sections will detail how these methods have been used to study the sex-specific phonetic production.

1.3 DEVELOPMENT

Speech output continually develops throughout the lifespan (Rosselli, Ardila, Matute, & Vélez-Uribe, 2014; Sweeting & Baken, 1982). This development is often slow and gradual. However, there are several significant social and physical milestones that can drastically affect speech output. For sex-specific speech, the most drastic of these milestones is the onset of puberty (Fitch & Giedd, 1999; LoMauro & Aliverti, 2018). As noted above, this change affects both male and female speakers. The developmental changes in the vocal tract are sex-specific, whereas, before puberty, vocal tract anatomy differs very little between sexes.

Phonetic developmental studies take advantage of this change by looking at speech output before and after the onset of puberty (see Ma et al., 2017; Yu et al., 2014). Socially motivated features of speech develop before and persist after puberty. In contrast, anatomically motivated features of speech develop for one sex during puberty.

As noted above, this method has been used to investigate the differences in pitch and formant frequencies differences between males and females. The lowering of male vowel frequency during puberty is a result of physical development in the vocal tract (Fitch &

Giedd, 1999). In contrast, developmental studies for /s/ have found that sex-specific production develops in childhood and persists into adolescence (Flipson et al., 1999; Fox & Nissen, 2005; Li et al., 2016). This suggests that anatomical development is not the main contributing factor to the sex-specific production of /s/. To our knowledge, there is currently only one study that looks specifically at sex differences in stop-consonant production in English (Yu et al., 2014). Yu et al., found that during puberty male speakers produce aspiration comparable to female speakers, reducing the difference in production. This difference does not seem to persist into adulthood, with adult male speakers reducing the relative amount of aspiration when compared to adult females. The cause of this change is currently not understood. As a result, the extent to which stop consonants are affected by puberty is currently not clear and warrants further investigation.

1.4 CROSSLINGUISTIC COMPARISON OF VARIATION

While looking at speech development has several advantages, there are limitations from what can be learned from this method. Developmental studies help make it clear whether anatomical differences are present in speech but is confounded by language-specific influence. One method that has been used to overcome these limitations is to compare sound production across different languages, or cross-linguistic comparison.

One of the advantages of cross-linguistic comparison is that it can be used to investigate how different language contexts can affect speech production and development. This method has been used to analyze different conversational strategies (Ford & Mori, 1994) and differences in language acquisition (Turner et al., 2014). Differences in language input can result in different patterns of production and development of speech sounds.

However, similarities in language development could represent some common developmental pattern in speech.

On a phonetic level, individual sounds that are similar or shared can be compared between languages. As was previously noted, there is a gendered speech difference between males and females in vowel production. It is almost universal that male vowel frequency production is lower than females. However, the average pitch of male and female speakers is not consistent across languages (Simpson, 2009). If vowel frequency was wholly the result of vocal anatomy, we would expect to see similar production frequencies across languages. However, the language-specific input influences the develop of these vowel sounds. Vowel frequency differences demonstrate how both physical and social factors can interact to produce gendered speech.

Similarly, /s/ production is sex-specific in English. However, the differences in /s/ production are not universally found across languages, suggesting that social conditioning may play a significant role in its development and use. As far as we know, there are currently no studies that exist that have a crosslinguistic analysis of stop consonants about gendered speech.

1.5 SEX-SPECIFIC PRODUCTION OF STOP CONSONANTS

Taken together, the study of development and crosslinguistic are effective tools for studying sex-specific phonetic production. They can be useful in elucidating the origins of these differences. This can be particularly true of speech sounds in which sex-specific speech is not well studied, such as stop-consonants. While there has been a large amount of research dedicated to sex-specific production of vowels and /s/, stop-consonant production remains controversial (Morris, Macrea, & Herring, 2008)

By far, most such reports on the sex-specific stop consonant productions are in English. The most common acoustic method of evaluating aspirations in stop consonant production is by measuring a speaker's voice onset time (VOT). During the production of stop-consonants, the airflow is temporarily stopped and then released. In English, there is a slight delay in time between the opening of the vocal tract and when the speaker's vocal cords begin to vibrate. VOT is defined as the time difference between the opening of the vocal tract and the vibration of the vocal cords (see Lisker & Abramson, 1964; Cho & Ladefoged, 1999). Speaker differences in VOT have used to study a variety of social factors that affect language, such as speakers from different language backgrounds, dialect differences, and speaker sex (Robb, Gilbert, & Lerman, 2005; Ryalls, Zipprer, Baldauff, 1997; Scharf & Masure, 2002; Sweeting & Baken, 1982; Swartz, 1992;).

Currently, there exist controversies regarding the magnitude of the VOT sex difference in English. Many previous studies have found that females produce more aspirated stop consonants than males, which results in a longer VOT (Koenig, 2000; Robb, Gilbert, & Lerman, 2005; Ryalls, Zipprer, Baldauff, 1997; Vasishth et al., 2018). This sex-specific difference has been shown to develop in childhood (Whiteside & Irving, 1997; Whiteside & Irving, 1998; Yu et al., 2014). However, this sex-specific difference in production is disputed by some researchers who have found that no sex difference exists for stop-consonants (Morris, Macrea, & Herring, 2008; Yu et al., 2014). These researchers have noted that lack of consistent methodologies may have lead to this controversy. This includes controls for stimulus type and individual speaking rate. When these factors are controlled for, the sex difference seems to be greatly reduced. For example, in the most extensive of these studies, Yu et al. (2014) looked at female and male stop consonant production from the age of 4 to adulthood. They found very little difference between male and female speakers except between the ages of 8 and 11.

Interestingly, the sex-related difference in VOT has also been reported in Mandarin Chinese (Li, 2013; Ma et al., 2017). This is of particular interest since Mandarin stop-consonants shares several similarities to English stop-consonants, which makes a comparison of these two languages ideal. Mandarin speakers have a similar catalogue of stop-consonant sounds to English (ie. /p/, /t/, /k/, /b/, /d/, /g/). Unlike English, the existence of sex-specific stop-consonant production is far less controversial. Adult females generally produce more aspirated speech than males (Li, 2013; Ma et al., 2017). There is also research has found that this sex difference begins to develop before the onset of puberty (Ma et al., 2017). However, similar to English, there are only a few studies that investigate this.

In a Mandarin-language developmental study similar to Yu et al. (2014), Ma et al. looked at the development of VOT in children and adolescents from ages 6 to 18 (2017). Ma et al. observed that female speakers generally produced more aspirated stop-consonants. Similar to Yu et al. findings, Ma et al. observed an increase in aspiration between the ages of 10 and 12. This similarity in the two studies seems to suggest that there biological or anatomical factors that affect the production of male VOT during puberty. However, unlike Yu et al., Ma et al., observed a general sex difference in VOT. This would suggest that language-specific input also seems to affect the sex-specific stop-consonant productions.

Developmental studies of this kind that elucidate the origin of sex-specific stop-consonant patterns in adult speech are scarce in both languages. Most studies on VOT development focus on infants or children before school age (Hitchcock & Koenig, 2013; Kent, 1976; Kewley-Port& Preston, 1974; Macken & Barten, 1980). Among the studies that examine older children, the effect of sex is very often neglected (Hazan, Romeo, &Pettinato, 2013; Lowenstein & Abramson, 1964; Perry, Ohde, & Ashmead, 2001). To my knowledge, there are currently no studies that investigate sex-specific production of stop consonants across languages.

1.6 THE PURPOSE OF THIS RESEARCH

The objective of this research is to investigate the development of sex differences in speech production by investigating sex-specific speech in VOT in English and Mandarin-Chinese. This will be done in two steps. The first step will be to establish whether there is a sex difference in production in each language at different development stages. The second step, if a difference is found, will be to investigate what potential factors may be responsible for this sex difference. Both of these languages have been found to produce sex-specific patterns in VOT in adult speakers. However, this difference has been known to be controversial in at least English (Li, 2013; Yu et al., 2014; Ma et al., 2017; Robb, Gilbert, & Lerman, 2005). There is evidence that these patterns emerge in childhood in each language. However, this pattern is not well understood (Ma et al., 2017; Yu et al., 2014).

Three methods have been employed in this research to investigate the contributions of both developmental and socially learned factors on voice onset time. The first is to investigate the developmental pattern of VOT production from early childhood into late adolescence. As was noted previously, the vocal tract anatomy of males and females before puberty is very similar and develops differently following the onset of puberty. Three age groups have been selected for this research; late childhood (ages 6-8), early adolescence (ages 10-12), and late adolescence (ages 14-16). This makes it possible to observe VOT production before the onset of puberty, during its onset, and following the vocal drop in males during puberty. If vocal tract anatomy contributes to the sex difference in VOT production, I would expect to see similarities across sexes before puberty, and differentiation following the puberty.

The second method used to investigate the contributions to sex differences in VOT production is to use a cross-language comparison between Mandarin and English. Mandarin and English belong to different language families but contain similarities ideal for comparison. Both languages share a similar set of stop-consonant sounds (ie. /p/, /t/, /k/, /b/, /d/, /g/). Because of these similarities between the two languages, comparing similarities and differences in VOT development may reveal common linguistic milestones in VOT production from childhood into adolescence. It may also reveal language-specific patterns of development, which would represent socially learned patterns of development.

Due to differences in research methodologies, this kind of cross-language comparison is challenging to conduct with previous research. For example, Yu et al. (2014) and Ma et al. (2017) both have similar acoustic stimuli. However, the speaker's speech rate was not controlled in either experiment. Speech rate has been shown to significantly affect VOT values, with the faster the speaker's speech rate, the shorter the resulting VOT (Kessinger & Blumstein,1998). Without control of speech rate, it is challenging to compare these two studies, despite their similarities. Other additional factors were also not controlled for, such as vowel-context (Ryalls, Zipprer, & Baldauff 1997; Yu et al., 2014), number of syllables (Yu et al., 2014), and speaker test-environment (Robb, Gilbert, & Lerman, 2005). In order to accurately compare these two languages, this research controls for these factors in both languages by controlling for speech rate, using similar stimuli, and using similar methods for acoustic analysis. How these controls have been implemented will be outlined in the following chapter.

The third method employed was the use of language-specific gendered behaviour questionnaires. Gender identity and behaviour can play a significant role in the production of sex-specific speech. The purpose of these questionnaires was to look at the gendered and social behaviours of the participants in the study. Previous research has found that biological sex and gender both contribute to the production and development of speech (see Li et al., 2016). An investigation into the gendered behaviour of speakers makes it possible to investigate to what extent VOT production differences are socially motivated. The method used for this study will mirror that of Li et al.'s (2016) study of sibilant production. In this study, Li et al. used gendered behaviour questionnaires to look at the variation in /s/ production. They found that gendered behaviour patterns explain sex-specific variation between sexes and variation within each sex. A similar method will be adopted for this analysis. Details on how these questionnaires will be used can be found in chapter 2.

1.7 HYPOTHESIS

Given the previous research for Mandarin VOT and sex differences, it is expected that there will be a male-female sex difference for voiceless stop-consonant VOT, with males producing relatively shorter VOT than females (Li. 2013; Ma et al., 2017). While the sex difference in English remains controversial, many previous studies have found that females generally produce longer VOT than males (Koenig, 2000; Robb, Gilbert, & Lerman, 2005; Ryalls, Zipprer, Baldauff, 1997; Vasishth et al., 2018). Given the findings from the developmental studies in both languages, I expect to observe a difference in stop-consonant production in males during the onset of puberty (Ma et al., 201; Yu et al., 2016).

As far as I know, there are no studies that look at gender-typical behaviour in connection to VOT production. Previous research has found that gender-typical behaviour can mediate the production of speech sounds. Using the same English-speaking corpus as this study, Li et al. (2016), found that production of /s/ was mediated by gender-typical behaviour. VOT has long been shown as a variable that has been used to investigate speaker variation between different dialects of speakers. Additionally, while other languages have found sex/gender differences in VOT production, the effect size varies greatly. This seems to suggest that socially learned behaviours have a strong influence on VOT. Based on this previous research, I would expect that gender-typical behaviour will influence stop-consonant production. Specifically, male-typical behaviour will lead to shorter relative VOT times, while more feminine typical behaviour should lead to longer relative VOT.

1.8 IMPLICATIONS

It has long been established that sex differences in speech production exist, but it remains what the effect size on language is. Developmental studies that elucidate the origin of sex specific VOT patterns in adult speech are scarce in both English and Mandarin. Most studies on VOT development focus on infants or children before school age (Hitchcock & Koenig, 2013; Johnson, 1991; Kent, 1976; Kewley-Port & Preston, 1974; Macken & Barton, 1980). Among the studies that examine older children, the effect of sex is very often neglected (Hazen, Romeo, & Pettinato, 2013; Lowenstein & Nittrouer, 2008; Murry & Singh, 1980; Perry, Ohde, & Ashmead, 2001; Zlatin & Koenigshnekt, 1976).

One of the novel aspects of this research is the comparison of gendered behaviour and speech production. Both sex and gender have been shown to affect the production of speech. In addition to looking at language development, this research will also look at two languages. While there is a large body of research dedicated to the cross-linguistic comparison of VOT, as far the I know, few of these studies focus have either a focus on the developmental trajectory or the effects of sex and gender. Cross-language developmental studies of this sort with controlled methodology could better reveal similarities shared between languages or differences due to specific language context. It also sheds light on the extent to which common biological factors shape the development in a different social context.

CHAPTER 2: METHODS

This chapter outlines the methods used for this study. The first two sections detail the Mandarin and English corpus, respectively. Both corpora were organized similarly, with only minor differences. For this reason, details that are analogous between the two corpora will be described in detail within the Mandarin Corpus section, and described more generally in the English corpus section, while highlighting the differences between the two corpora. The third section describes the acoustic analysis of the speech data. The final section contains the procedures used in the general statistical analysis.

2.1 MANDARIN CORPUS

The Mandarin-language data that was analyzed was taken from an existing corpus put together by Yanjun Shi in 2013 (See Shi, 2014). The corpus contains data collected from school children in Luoyang, Henan, China, in three different grades. For each child participant, the corpus contained the following data: a personal demographic questionnaire, measures of physical size, voice recordings, and a gendered behaviour questionnaire.

The demographic questionnaires contained information regarding the participant age, language and family background information, and information about any speaking or hearing deficiencies, such as stuttering, incorrect pronunciations of individual sounds, or speech delays. Children who spoke a dialect at home rather than Mandarin were not included in this study. Children who reported a speech delay, dyslexic symptoms (e.g. trouble in reading or a significantly smaller vocabulary compared with peers), incorrect pronunciation with individual sounds, or chronic ear infections were also not included in the study. Two hundred sixty children met the selection criteria and used for analysis.

The three school grades were used to generate three age groups (6-8, 10-12, and 14-16) that were used for later analysis. These age groups are ideal for comparison because they represent child development before the onset of puberty, the onset of puberty, and following puberty. In the context of speech output, both male and female vocal anatomy changes during puberty that affect voice quality. A breakdown of the participant's age and sex by age group can be found in table 2.1.

Table 2.1: Age for Mandarin-speaking children. The mean age is reported for each age group and sex.

Age Group	Sex	N	Mean Age (years)	SD (years)
6-8	Female	54	6.87	0.34
	Male	57	6.96	0.42
10-12	Female	39	11.74	0.44
	Male	39	11.74	0.47
14-16	Female	39	15.69	0.47
	Male	33	15.87	0.48

Three measures of physical size were used to assess children's physical development. The corpus provided three measures of physical size: height (reported in centimetres), weight (in kilograms), and head circumference (in centimetres).

2.1.1: SPEECH PRODUCTION EXPERIMENT AND STIMULI

The speech data in the corpus was recorded using a digital recorder (Marantz PMD 661) connected to a Shure SM87A condenser microphone. Recordings were made using a 44.1 kHz sampling rate and 16-bit quantization.

Eighteen unique speech tokens were recorded for each child. These words each started with either /t/ or /d/ followed by vowel /a/, /i/ or /u/. The consonants /t/ and /d/ are ideal for comparison because they are produced using the same place of articulation, the alveolar ridge. What differentiates these two sounds is that /t/ is produced using aspiration, while /d/ is not. The word list included single syllable and multi-syllable words for each combination of sounds. A complete list of the word stimuli is found in Appendix 2.1

2.1.2: GENDER TYPICALITY QUESTIONNAIRE

The guardian-filled gendered behaviour questionnaire assessed the participant's gender-related behaviours. These kinds of questionnaires have been successfully used to evaluate gendered behaviour in children and adolescents, and its relationship with speech output (Li et al., 2016; Yu et al., 2010). Questions include children's toy preferences, play behaviour in the home, the kinds of games played, and whether they enjoy male or female company. For participants under 12 years old, the guardian rated the frequencies of these behaviours on a 5-score Likert Scale of frequency from 1 (Never) to 5 (Always). For adolescent participants, guardians would indicate how often these behaviours occurred while the participants were a child. The original and translated questionnaire can be found in appendix 2.2.

Before analysis, the results of the gendered behaviour questionnaire were converted to gender-typicality scores by converting male scores to be on the same scale as female scores. The resulting gender-typicality scores ranged from 1 to 5, with a higher score indicating a higher degree of female-typicality, and a lower score indicating a higher degree of male-typicality.

2.2 ENGLISH CORPUS

The English language data that was analyzed was taken from an existing corpus put together by Mellissa Kinsmen (See Li et al., 2015). The data collected from the English corpus was collected in Lethbridge, Alberta, Canada. Similar to the Mandarin corpus, the English corpus contained a signed guardian consent form, a filled demographic questionnaire, measures of physical size (height (centimetres) and weight (kilograms)), voice recordings, and a completed gendered behaviour questionnaire.

Based on the demographic questionnaires, one hundred and thirty children were selected for analysis in this study. These participants were placed into three age groups, based on the available number of subjects per age. These age groups are the same three age groups as the Mandarin corpus (6-8, 10-12, 14-16). A breakdown of the participant's age and sex by age group can be found in table 2.2.

Table 2.2: Age for English-speaking children. The mean age is reported for each age group and sex.

Age Group	Sex	N	Mean Age	SD
6-8	Female	38	7.26	0.82
	Male	35	7.09	0.79
10-12	Female	16	11.69	1.30
	Male	19	11.47	1.26
14-16	Female	16	14.31	0.60
	Male	12	15.08	0.90

2.2.1 ENGLISH SPEECH PRODUCTION EXPERIMENT AND STIMULI

Fifty-four words were included as stimuli for the production task. The target words began with one of six stop consonants (/b/, /d/, /g/, /p/, /t/, or /k/) followed by a vowel (/ae/, /i/ or /u/; see appendix 2.3 for a complete list of the words used). Similar to the Mandarin corpus, the set of stop consonant sounds that were analyzed are aspirated vs. non-aspirated pairs of sounds that are articulated at the same places in the oral cavity. A breakdown of these sounds and the places of articulation can be found in table 2.3.

Table 2.3: The consonants, places of closure, and use of aspiration for the stop-consonant sounds found in the English corpus.

Consonant	Place of closure	Aspiration
/p/	Lips	Yes
/b/	Lips	No
/t/	Alveolar Ridge	Yes
/d/	Alveolar Ridge	No
/k/	Velum	Yes
/g/	Velum	No

2.2.2 ENGLISH QUESTIONNAIRE

The English corpus gendered behaviour questionnaire was a 12 question five-point Likert scale of frequency (1 (Never) to 5 (Always)) questionnaire. There were 12 items in total, all of which were either boy- or girl-typical. The gendered behaviour questionnaire used for the current research can be found in Appendix 2.4. This questionnaire has been previously used to assess gender-typical behaviour both in the present (Johnson et al., 2004; Cohen-Kettenis et al., 2006) and retroactively (Bartlett, & Vasey, 1996; VanderLaan, Petterson, & Vasey, 2015). The results of the gendered behaviour questionnaire were converted to gender-typicality scores before analysis using the same method as the Mandarin questionnaire.

2.3 ACOUSTIC ANALYSIS

The goal of the acoustic analysis was to measure the voice onset time (VOT) of the word-initial stop consonants found in the speech data. The acoustic analysis software Praat (Boersma & Weenink, 2005) was used to conduct the acoustic analysis. Each speech recording contained a series of spoken words, uttered one at a time. The individual words were identified and analyzed one at a time. The following measured points were gathered using both the sound waveform and sound spectrogram for each word. All words were collected using a window length of 0.005 seconds.

VOT is a temporal feature of the production of stop-consonants. Word-initial stop consonant production is initiated by creating a complete closure in the vocal tract. During the time of closure, air pressure builds in the vocal tract behind the point of closure. The air pressure is then released before the onset of the following sound. A vowel sound always follows word-initial stop consonants in both Mandarin and English. Vowels in both languages are voiced, meaning that the vocal cords vibrate when they are produced. The difference in time between the stop-consonant release and the vowel onset is known as voice onset time (VOT). The moment of stop-consonant release and the vowel onset can be identified by examining the waveform and spectrogram of a word. The stop-consonant release is seen as a sudden increase in aperiodic noise in both the waveform and the spectrogram (see between 1 and 2 in figure 2.1). For vowels, the fundamental frequency and the vowel formant frequency used to mark vowel onset and conclusion. The fundamental frequency, or 'voicing bar,' is visualized as a low-frequency band, around 150-200 Hz (see figure 2.1, point 5). Acoustic resonance in the vocal tract during vowel production also produces formant frequencies. Formant frequencies are visualized as frequency bands that appear above the fundamental frequency (see figure 2.1, point 6). The stop-consonant release and the vowel onset were marked for each word stimuli using these acoustic cues. The vowel onset was then subtracted from the release point to calculate the VOT.

In addition to the previous two points, the vowel endpoint (see figure 2.1, point 3) and the word endpoint (see figure 2.1, point 4) were also measured. This was done to control for individual differences in speech rate. Previous research has found that the duration of the following vowel can affect VOT duration (Morris et al., 2008). The vowel duration was calculated by finding the difference in time between the vowel onset and the vowel endpoints. As an additional control for speech rate, the duration of the whole word was used. This was calculated by finding the difference in time between the stop-consonant release and the word end.

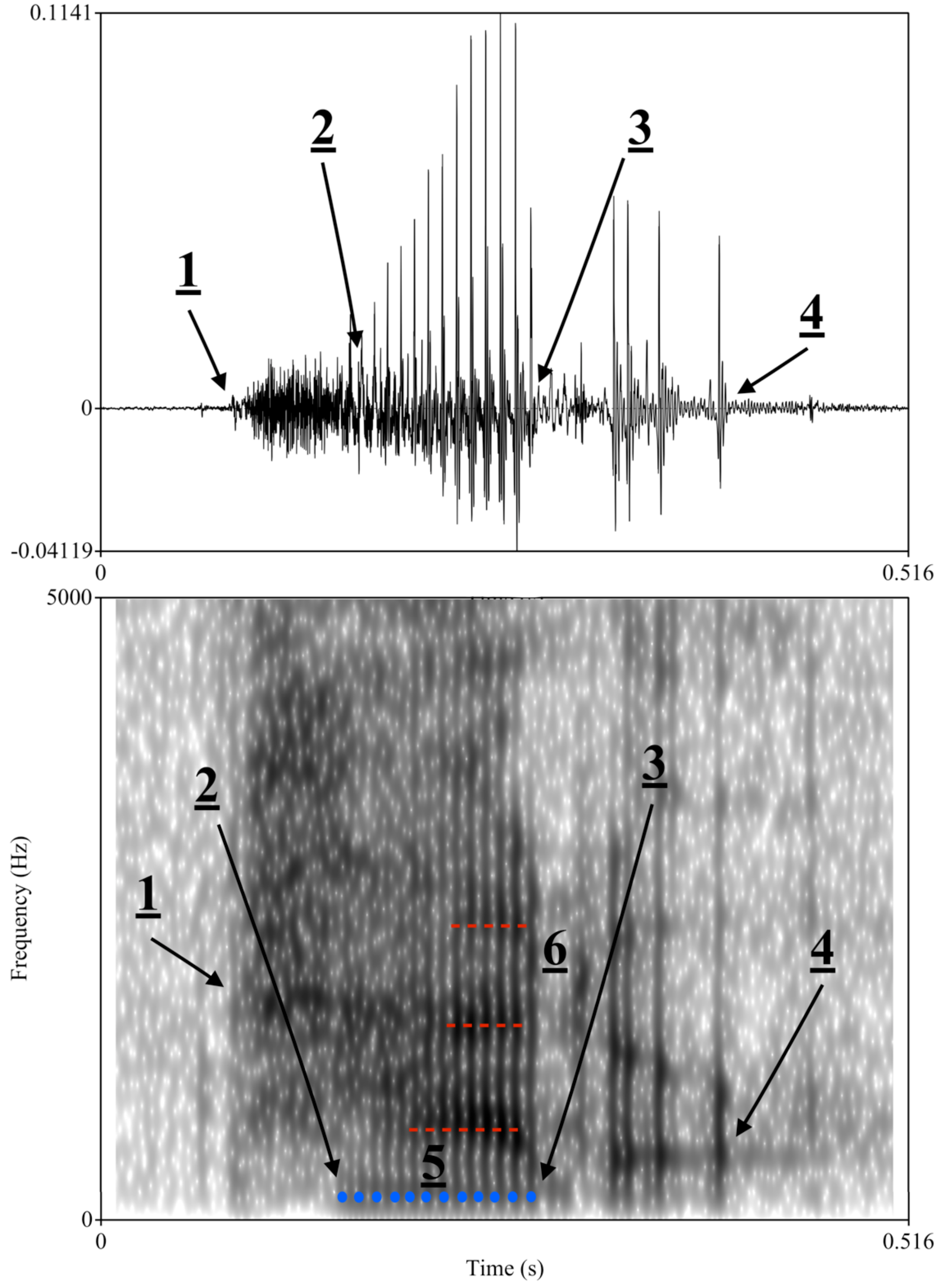

Figure 2.1: The waveform and spectrogram for the word 'paddle.' The four measurement points for VOT are marked on both the waveform and the spectrogram. Points of interest recorded: burst (1), vowel onset (2), vowel end (3), and word end (4). Vowel acoustic markers for the voicing bar (5) has been marked with blue dots. The vowel formants (6) have been marked with red dotted lines.

2.3 STATISTICAL ANALYSIS

The primary goal for this analysis was to identify factors that contribute to the sex difference in production in VOT if such sex difference indeed exists. In order to accomplish this, the analysis was conducted through a two-step process. First, VOT values were summarized and tested to determine the existence of a sex difference. Second, when a confirmed sex difference was identified, factors that could potentially affect speech output were also summarized with a specific emphasis on their variation with speakers' sex. These factors include additional acoustic data such as word duration and vowel duration that were collected to access speech rate, the measures of physical size and measures on children's sex-typed behaviour assessed through a guardian-filled questionnaire.

For the first step, a linear mixed effect model (LME) was used to confirm sex differences in VOT production between males and females. LME analysis was chosen to factor in the clustering effect of repeated samples taken from the same speaker and for the same words produced by different speakers. The analysis was conducted using the LME4 package (Bates et al., 2015) in R (R Development Core Team, 2011). VOT served as the dependent variable, with sex as the independent variable, and with subjects and words set as random effects.

If a sex/gender difference in VOT production existed, the second step of the analysis was conducted in which an additive mediation model (AMM) was used to identify factors that contribute to the sex difference. Acoustic, physical, and gender-typicality measures were successfully entered into the model that initially has sex/gender as the dependent variable and VOT as the independent variable. Measures that reduce the sex/gender gap were assumed to mediate this sex/gender difference.

Before the two-step analysis, gender-typicality measures gathered through the gender-typicality questionnaires were aggregated through the principal component analysis (PCA) to reduce the number of variables associated with this measure. This reduction helped unveil the underlying components that contributed to the participant's gendered behaviour. The analysis was run in SPSS 24 (IMB Corp. 24). Regression scores for each of the underlying components were saved and used as variables for analysis for each participant.

The following two chapters summarize the results from this two-step analysis. The results for the Mandarin corpus are found in chapter 3, and the English corpus results are found in Ch. 4.

CHAPTER 3: MANDARIN RESULTS

This chapter contains the results of the Mandarin corpus analysis. The primary goal of this research was to determine which factors help contribute to the gendered production of stop consonants. In order to address the primary research question, A two-step analysis was conducted. Section 3.1 summarizes the first step in the two-step analysis. It contains a summary of the descriptive statistics for corpus data. Each variable was tested for sex differences. The variables in which sex difference was observed would be considered potential contributing factors to speech production and would be included in the second step in the two-step analysis found in section 3.3. Section 3.2 contains the results of the principal components analysis performed on the gender-typicality questionnaires in order to reduce the number of variables, as well as help uncover the underlying components that motivated responses within the questionnaire. Section 3.3 outlines the second step in the two-step analysis. This second step directly addresses the primary research question. An additive mediative model (AMM) was constructed to elucidate the underlying factors that contributed to the male/female sex difference in VOT production.

It should be noted that Chapter 4 contains the results of the English corpus analysis. The English-language results underwent the same analysis and are organized similarly to the results in this chapter. For this reason, details that are analogous between the two corpora will be described in this chapter and will not be repeated in the next chapter.

3.1 DESCRIPTIVE STATISTICS

The purpose of the following section is to summarize how the corpus data was distributed and to test variables for sex differences. The three variables included in this section are the acoustic data, the measures of physical size, and the gender-typicality questionnaires responses. The acoustic data summarizes the VOT data. Vowel duration and whole word duration, which were used to control for speech rate, were also included. These controls were to normalize the VOT data and control for speaker differences. The measures of physical size include height, weight, and head circumference. Physical measures are useful in looking at the average growth of the corpus from childhood into adolescence, as well as size differences between males and females. This chapter also summarizes the 46-question mandarin language gender-typicality questionnaires. These three variables were used for the additive mediation models in further analysis (see 3.3).

3.1.1 Acoustic measures

The VOT values were calculated for each age group and gender and separated by consonant sounds. Figure 3.1 illustrates the mean VOT for these age groups.

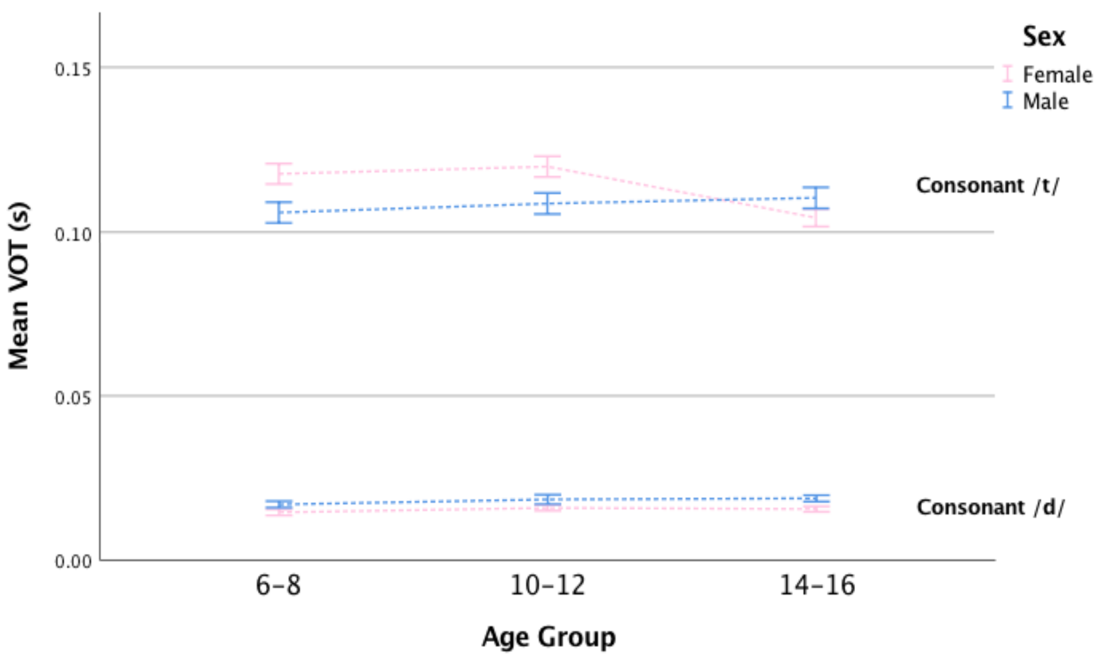

Figure 3.1: A line graph summarizing the mean overall voice onset time by age group. Error bars at a 95% confidence interval.

For the sound /d/, male speakers produced longer VOT than females in all age groups. For /t/, female speakers produced longer VOT than male speakers in the 6-8 age group and the 10-12 age group. However, for the 14-16 age group, the mean VOT for female speakers is relatively shorter than male speakers. Male speakers' VOT, while stable, gradually increased as age increased. Female /d/ production remains relatively stable, while /t/ had a minor increase from 6-8 to 10-12 before shortening in the oldest age group. A breakdown of the mean VOT values for each group can be found in table 3.1.

Table 3.1: The mean voice onset time for /d/ and /t/ for age groups 6-8. 10-12 and 14-16, separated by sex.

Age	Sex	/d/ N	/d/ Mean (ms)	/d/ SD	/t/ N	/t/ Mean (ms)	/t/ SD
6-8	Female	54	14.56	4.04	54	117.68	25.04
	Male	57	16.83	5.84	57	105.48	19.67
10-12	Female	39	15.87	3.84	39	119.73	20.75
	Male	38	18.70	7.76	38	108.25	19.76
14-16	Female	39	15.55	3.37	39	104.34	16.12
	Male	33	18.65	4.14	33	111.052	19.55

Female adolescents produce shorter VOT in the 14-16 age group relative to both earlier age groups and males. The observed pattern in the adolescent group is an unexpected result. Previous research in both adolescents and adults has found that female speakers produce relatively longer VOT across the lifespan (Li, 2013; Ma et al., 2017).

A linear mixed-effect model was used to determine whether a sex difference was present in VOT. The dependent variables were the VOT of /d/ and /t/. The independent variable was sex (see Table 3.2). Due to the different patterns of development for /d/ and /t/, a test was conducted for both consonants.

Table 3.2: Summary of the follow-up VOT analysis. The effect size for each variable is measured in milliseconds. The statistical significance key can be found at the bottom of the table.

Dependant Variable	Independent Variable		Coefficient (ms)
/d/	Sex	male	2.62***
	Age	11-13	1.57
		14-16	1.38
/t/	Sex	male	-6.78**
	Age	11-13	2.50
		14-16	-4.42

* p < 0.05, ** p < 0.01, *** p < 0.001

A significant sex difference was found for both /d/ and /t/, suggesting that there is a difference between male and female production of these two consonants. For /d/, males' VOT was 2.62 milliseconds longer relative to females. For /t/, a significant effect of sex was found, with males' VOT being 6.79 milliseconds shorter than females. This indicates that for /t/ female speakers have a longer lag period. This is possibly due to a greater amount of aspiration.

While the statistical analysis of /d/ revealed some significant sex and age-related effects, figure 3.1 makes it clear that these results should be interpreted with caution. There appears to be minimal variation in both sex and age for /d/. Give the range and mean values of speakers, there seems to be a large overlap. This suggests that the small, but statistically significant, difference may not be a useful marker of speaker sex.

Vowel duration

In order to control for the effects of speech rate on VOT production, vowel duration was measured and summarized for each word token. The vowel duration was measured for three different vowels: /a/, /i/, /u/. The overall mean results can be seen in figure 3.2. A breakdown of the vowel duration by vowel type, age group, and sex can be found in table 3.3.

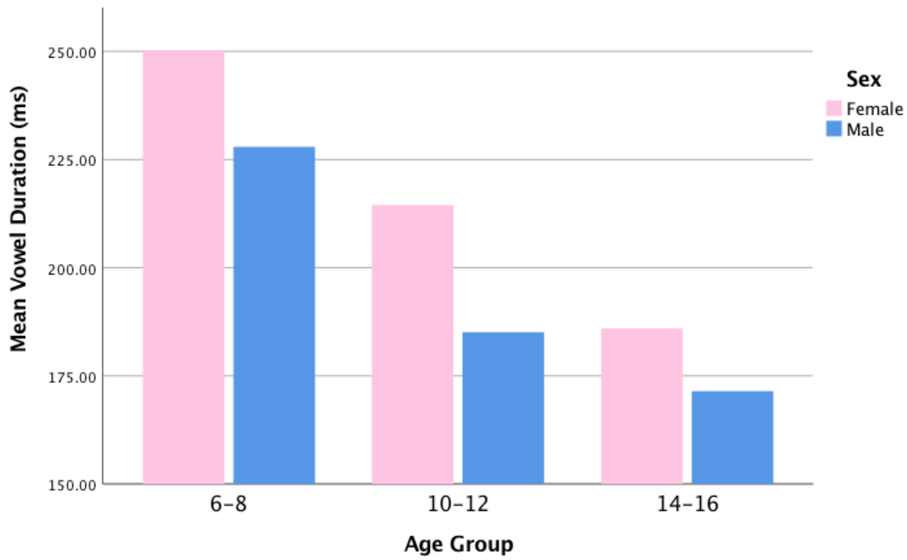

Figure 3.2: A histogram summarizing the mean overall vowel duration separated by age group and sex.

Female and male vowel duration decreased from childhood into adolescence (see figure 3.2). This decrease is an expected part of language development for speakers (Skoog, Eriksson, & Sörqvist,2015). Female speakers had longer vowel duration relative to males, which is also typical of Mandarin speakers (Feldstein, Dohm, & Crown, 1993).

Table 3.3: The mean vowel duration for three vowels: /a/, /i/, and /u/. These values are separated by sex and age group.

Age	Sex	N	/a/ Mean (ms)	/a/ SD	/i/ Mean (ms)	/i/ SD	/u/ Mean (ms)	/u/ SD
6-8	Female	54	293.43	63.34	224.22	56.64	233.80	58.31
	Male	57	271.16	52.65	201.86	44.08	210.03	48.41
10-12	Female	39	246.31	38.25	192.82	36.71	204.96	41.18
	Male	38	211.26	39.08	165.27	32.78	178.71	32.21
14-16	Female	39	213.17	31.92	169.13	35.83	176.12	36.50
	Male	33	194.03	37.54	155.17	33.84	165.39	35.24

A repeated measures ANOVA found a significant main effect of sex on vowel duration ($F(1, 252) = 14.44$, $p < 0.001$). These results indicate that male and female speakers produced average vowel length differently, with female speakers producing longer mean vowel lengths than males. There was also a significant effect of age ($F(2, 252) = 48.88$, $p < 0.001$), indicating that there was a significant change in vowel length as participants aged, decreasing in length with age. There was, however, no significant interaction between sex and age ($F(2, 252) = 0.67$, $p = 0.52$). This seems to indicate that the development of vowel length seen in both sexes was similar from youngest to oldest age group.

Word duration

In addition to vowel duration, word duration was also measured and summarized to help control for speech rate. The overall mean word duration can be seen in figure 3.3. A breakdown of the word duration by syllable, age group, and sex can be found in table 3.4. Syllable number was included as a measure since words syllable number affects word length, with words with more syllables being longer.

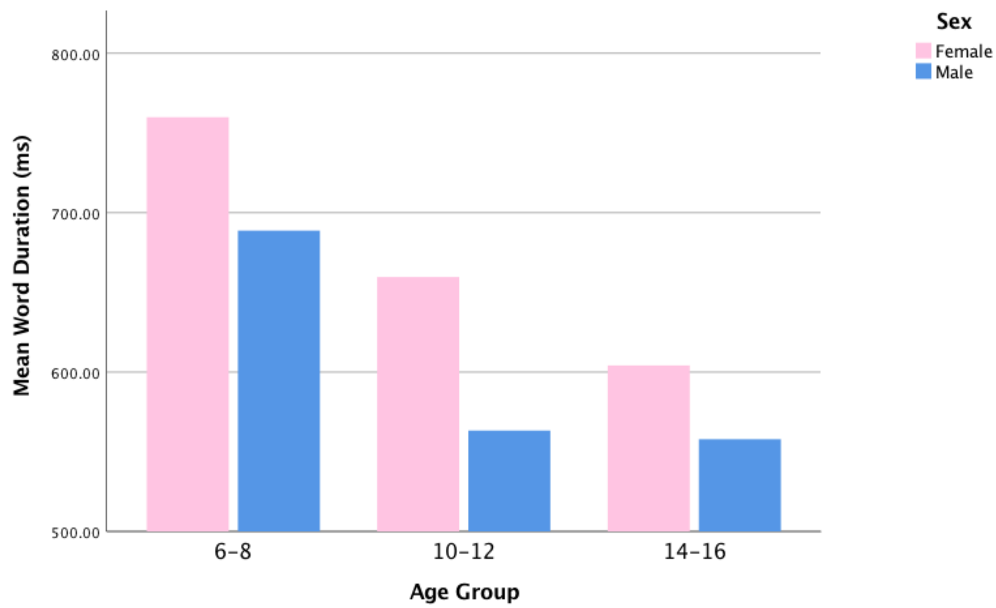

Figure 3.3: A histogram summarizing the mean overall word duration separated by age group and sex.

Similar to the vowel duration, the decrease in word duration from childhood into adolescence is an expected pattern of speech development (Skoog, Eriksson, & Sörqvist, 2015). It is also typical that female word duration is longer than male word duration (Feldstein, Dohm, & Crown, 1993).

Table 3.4: The mean duration of Mandarin words, separated by numbers of syllables syllable, age group and sex.

Age	Sex	1 Syllable N	1 Syllable Mean (ms)	1 Syllable SD	2 Syllable N	2 Syllable Mean (ms)	2 Syllable SD	3 Syllable N	3 Syllable Mean (ms)	3 Syllable SD
6-8	Female	110	589.38	131.66	793	742.00	170.71	117	1041.19	242.48
	Male	122	545.80	101.58	849	678.81	147.48	126	893.82	193.10
10-12	Female	79	488.77	80.09	566	646.09	122.10	78	931.37	165.85
	Male	76	429.82	75.70	546	548.00	113.72	79	797.87	132.90
14-16	Female	81	455.41	80.15	551	590.35	106.50	80	849.40	124.67
	Male	69	434.05	112.83	468	542.96	102.14	68	786.27	124.36

While controlling for syllable number, a repeated measures ANOVA found a significant main effect for sex ($F (1, 252) = 30.56, p < 0.001$) and age group on word duration ($F (2, 252 = 58.31, p < 0.001$). This indicates that female speakers produced longer word durations than male speakers, and that word duration decreased with age. There was no significant interaction between sex and age group ($F (2, 252) = 1.40, p = 0.29$). This suggests that the developmental pattern was not significantly different between males and females.

3.1.2 Physical size

The Mandarin-speaking corpus provided three measures of physical size: height, weight, and head circumference. A breakdown of these physical measures by age group and sex can be found in table 3.5.

Table 3.5: Mean measures of physical size for Mandarin-speaking children. The mean measures of height, weight, and head circumference are reported for each age group and sex.

Age Group	Sex	Height (cm)		Weight (kg)		Head Circumference (cm)	
		Mean	SD	Mean	SD	Mean	SD
6-8	Female	124.73	4.67	24.12	4.52	51.41	1.28
	Male	128.16	5.11	27.29	5.62	51.92	4.25
10-12	Female	156.21	6.41	45.53	9.94	54.31	1.47
	Male	158.58	7.32	49.72	11.55	54.78	1.80
14-16	Female	164.10	4.85	55.07	6.92	55.76	1.61
	Male	175.08	5.65	65.14	9.36	57.14	2.10

The measures of physical size were used to assess physical development. A series of Pearson correlation tests were conducted to rule out collinearity. The first test conducted was between weight and height. Weight and height were highly correlated with each other, $r = 0.911$, $p < 0.01$ (two-tailed). The second correlation tested the relationship between height and head circumference. Height and head circumference were positively correlated with each other, $r = 0.647$, $p < 0.01$. Previous research has found that height is correlated with vocal tract length (Fitch & Giedd, 2002). The taller an individual, the larger their vocal tract. Since all three measures of physical size were correlated, and due to the correlation between height and vocal tract length, the decision was made only to use height in the analysis. Height data separated by age group and sex can be found in figure 3.4.

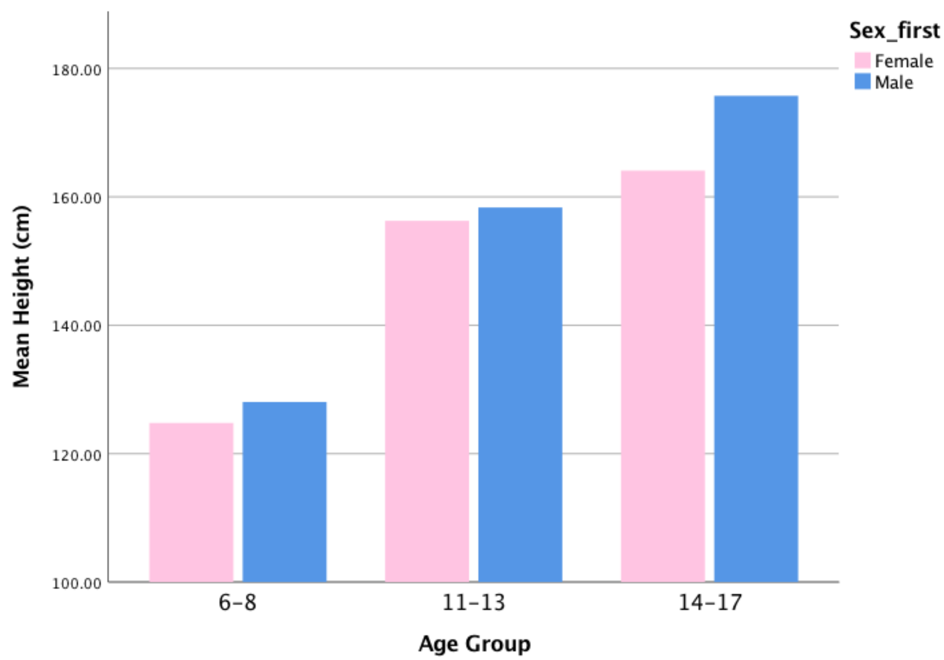

Figure 3.4: The mean height of Mandarin Participants separated by sex and age group.

A two-way ANOVA found a significant effect of sex ($F(1, 254) = 565.62$, $p < 0.001$, partial $\eta2 = 0.205$) and age group ($F(2, 254) = 1510.95$, $p < 0.001$, partial $\eta2 = 0.922$) on height. Follow-up t tests found that there was a significant difference in height between males and females, in both the 6-8 ($t = -3.48$, $p = 0.001$) and the 14-16 groups ($t = -10.40$, $p < 0.001$), with males being taller. There was also a significant interaction between sex and age group ($F (2, 254) = 16.93$, $p < 0.001$, partial $\eta 2 = 0.118$).

3.1.3 Gender-typicality questionnaire

The Mandarin gender-typicality questionnaires contained 46 questions. The scores were first converted to gender-typicality scores. The Likert-scale questionnaires asked specific questions about male and female-typical behaviours, on a scale from 1 (behaviour rarely occurs) to 5 (behaviour occurs frequently). Participants were also able to indicate if the question did not apply to them with a 0. Responses marked with a 0 were removed and left blank for the analysis. Male-typical questions were converted to the same scale as the female-typical questions by multiplying each male-typical response by -1, and then by adding 6. The resulting gender-typicality scores ranged from 1 to 5, with a higher score indicating a higher degree of female-typicality, and a lower score indicating a higher degree of male-typicality.

A summary of the questionnaire response can be found in table 3.6. The mean female response to the questionnaire after was 3.55 (SD = 0.36), while the mean male response was 2.31 (SD = 0.27; see table 3.6). An independent samples t-test found that this difference in gender-typicality questionnaires response was significant ($t = 31.679$, $p < 0.001$).

Table 3.6: Mean gender-typicality scores separated by age group and sex.

Age Group	Sex	Mean response	Std. Dev
6-8	Female	3.69	0.30
	Male	2.34	0.29
10-12	Female	3.47	0.38
	Male	2.23	0.25
14-16	Female	3.44	0.35
	Male	2.32	0.29
Overall	Female	3.55	0.36
	Male	2.31	0.27

A two-way ANOVA test found a significant main effect for sex (F (1, 254) = 1000.02, $p < 0.001$, partial $\eta^2 = 0.797$) and age group (F (2, 254) =727, $p < 0.001$, partial $\eta^2 = 0.054$). Follow-up within-sex ANOVA tests found that there was a significant difference for age group for females (F(2, 129) = 7.47, $p < 0.001$, partial $\eta^2 = 0.104$), but not for males (F(2, 125) = 1.85, p =161, partial $\eta^2 = 0.029$). A post hoc Tukey test found that the 6-8 group females had a significantly higher mean gender-typicality score than the 10-12 and 14-16 groups.

3.2 PRINCIPAL COMPONENTS ANALYSIS

A principal component analysis (PCA) was used to determine the underlying components of the gender-typicality questionnaires. The scores first underwent cleaning and missing data replacement before analysis. Questions that did not apply to the participants were left blank for the analysis and were considered missing data. The Mandarin gender-typicality questionnaire had 46 questions. Due to the large number of questions, it was expected that most participants would indicate that some of the questions were not applicable. The response rate to these questions ranged between 76 and 96%, indicating that some questions seemed to be more applicable to the population than others. The individual response rate to the whole quiz ranged between 41 (19 questions) and 100%, with a mean response rate of 89%. (41 questions).

The principal components exclude subjects with missing data. The decision was made to replace data for participants who had a high questionnaire response rate. Data cleaning and replacement were done in two steps. First, due to the range of response rates for each question, the decision was made to only include questions with the highest response rate. Questions with less than an 85% response rate were excluded from the analysis. This cut-off rate was chosen because it was able to balance the integrity of the questionnaire while preserving individual subjects. Sixteen questions were removed on this basis. The second step was used to determine which participants would be eligible for data replacement. Only participants with a response rate of 80% or higher (missing more than six questions of the new set of 30 questions) were considered eligible for data replacement. Thirty-five individuals were removed on this basis, resulting in 244 available individuals for analysis. Missing values for each of the remaining individuals were replaced by using their mean response to the other questions. No univariate or multivariate outliers were observed.

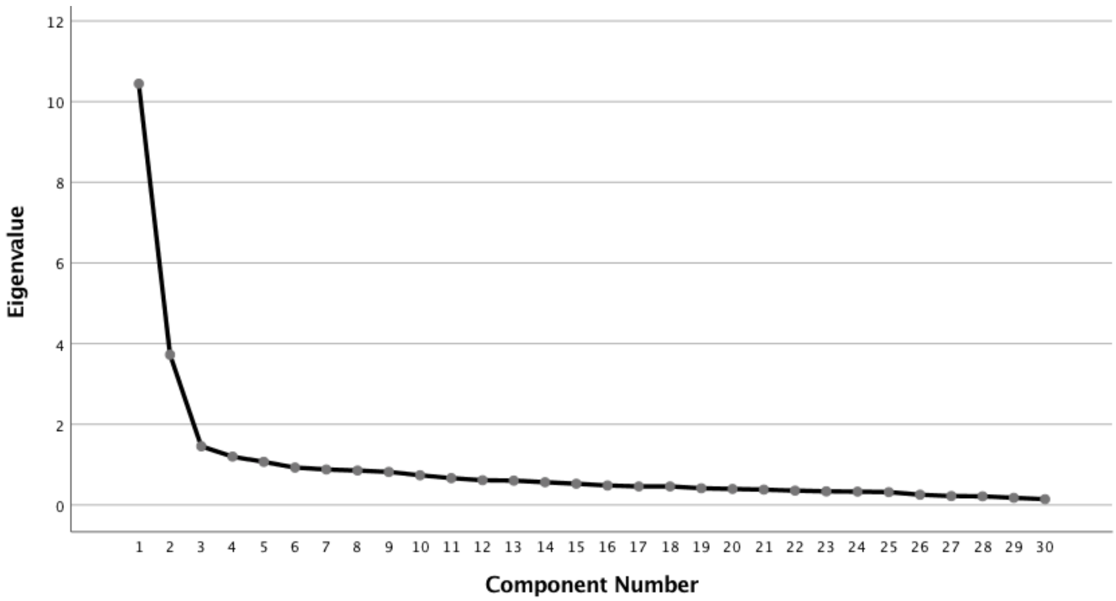

Figure 3.5: The scree plot for the Mandarin Gender-typicality questionnaire.

The principal component analysis revealed two main components. As illustrated in figure 3.5, these two components were found to have eigenvalues higher than 3. Component 1 accounted for 25.3% of the variation, while Component 2 accounted for 21.9% of the variation. The components were identified by comparing which questions were associated with each component. Each question was given a factor score for each component during the PCA analysis, and then these scores were rotated using Varimax rotation. Since the questionnaires were predicted to have two expected components (i.e. male and female-typical behaviour), Varimax rotation was chosen to help simplify the interpretation of the components. The highest rotated factor score was used to identify the component. The component categories were identified by looking at which questions were associated with each component.

Table 3.7: Factor Loadings after Varimax rotation. The highest factor loadings for each question are bolded indicating which factor the question belongs to.

Question	Component		Question
	1	2	
2	**0.579**	0.279	He/she plays football or basketball
3	0.321	**0.619**	He/she plays house
4	**0.728**	0.37	He/she plays toy guns
5	0.31	**0.661**	He/she plays with stuffed animals
6	**0.782**	0.312	He/she plays with mechanical or machine-like toys
7	0.354	**0.703**	He/she likes fairy tales like Snow White
9	0.267	**0.651**	He/she plays teacher or nurse games
10	**0.506**	-0.235	He/she likes building forts in games
11	0.176	**0.606**	He/she likes to play "Put the handkerchief behind you"
13	**0.699**	0.386	He/she likes dressing in male clothing, such as shorts
17	**0.669**	0.26	He/she plays rough and tumble games
18	0.535	**0.654**	At school, he/she plays with boys
19	**0.688**	0.044	At school, he/she plays with girls
20	**0.768**	0.381	He/she likes to use tools, such as hammers, screw drivers, etc
22	**0.686**	0.153	He/she imitates female characters seen on T.V or in movies
24	**0.696**	0.241	He/she wears thinking like towels around the waist as a skirt
26	**0.626**	0.008	He/she likes knitting or sewing
27	0.135	**0.582**	He/she is popular among boys
28	0.082	**0.491**	He/she is popular among girls
31	**0.717**	0.114	He/she is good at imitating females
32	**0.563**	-0.021	He/she is good at imitating males
34	0.204	**0.613**	He/she likes singing
35	-0.007	**0.456**	He/she likes dancing
37	0.174	**0.686**	He/she likes cooking games
38	-0.01	**0.581**	He/she likes to play jump rope
40	0.046	**0.689**	He/she plays hopscotch
41	0.018	**0.584**	He/she likes to play with plasticine
42	-0.165	**0.622**	He/she plays store
44	**0.644**	-0.077	He/she likes to play chess
46	**0.655**	0.189	He/she likes books about dinosaurs and space

Table 3.8 outlines which questions were associated with each component. Component 1 asked about feminine typical behaviour and accounted for 15 questions (see figure 3.8). The question in Component 2 asked about male-typical behaviour, and also accounted for 15 questions. Principle component scores were generated for each subject. These scores are a general score for each component in the PCA model. For example, if one of the identified components were feminine-typical behaviour, with five questions relating to that component, the component score would be generated to represent the responses that make up the component. The generated component scores were used for the additive mediation model analysis.

3.3 ADDITIVE MEDIATION MODEL

In order to address the primary research question, an additive mediation model (AMM) was employed. The purpose of this analysis was to identify which variables contribute to the sex differences in the production of VOT for /d/ and /t/. The AMM was constructed by first establishing a base model, and then constructing successive additive models by adding in additional variables that could contribute to VOT patterning. The method of additive modelling has been previously used to evaluate sex/gender differences in social science research (see Goodwin & Gotlib, 2004; Rogers, Everett, Onge, & Krueger, 2010) and gendered speech production (see Li et al., 2016).

The linear mixed-effects model found in section 3.1 for VOT was used as the base model. The dependent variable was VOT, and the independent fixed effects variable was speaker sex. Speaker and word token were included as random effects. The second variable that was added to the model was age.

Additional variables were then added to the base model successively. As previously shown, there is a significant difference in VOT between males and females. Variables added to the model that can reduce the statistical significance are considered to contribute to the sex-difference in VOT. The variables relating to anatomical measures were first added to the model to help determine whether anatomical differences between sexes could explain the sex/gender difference. After anatomical differences, variables relating to learned behaviours were added to the model. These variables included speaker speech rate and gender-typicality regression scores. The models were evaluated for statistical significance, as well as their effect size. A summary of the overall analysis for /d/ and /t/ can be found in table 3.9.

3.3.1 Additive mediation models for all groups

The base model used for the analysis is the linear mixed effects model that was used in the descriptive statistics section previously. There is a significant effect for sex for the overall models for both /d/ and /t/, with their effect sizes being 2.62ms and -6.62ms, respectively. The first variable added to the base model was age, followed by measures of physical size, speaking rate, and gender typicality scores. The successive models, and the effect sizes can be seen in table 3.9

Table 3.8: Summary for the overall additive mediation models for /d/ and /t/. The effect size for each variable is measured in milliseconds. The statistical significance key can be found at the bottom of the table.

Dependant Variable	Independent Variable		Model 1	2	3	4	5
VOT /d/	Sex		2.62***	2.67***	2.63***	2.51***	1.53
	Age	10-12		1.60*	1.39	1.21	1.14
		14-16		1.40	1.10	0.66	0.96
	Height				0.007	0.01	0.003
	Vowel Duration					-19.68***	-19.98***
	Word Duration					5.23***	6.09***
	Gender-typicality	FT					0.73
		MT					0.15
VOT /t/	Sex		- 6.62*	114.94**	83.09**	-4.80	1.53
	Age	10-12		2.50	-8.14	0.74	4.27
		14-16		-4.43	-15.56	-7.06	-3.35
	Height				0.26	0.23	0.13
	Vowel Duration					34.13***	36.09***
	Word Duration					35.04***	37.33***
	Gender-typicality	FT					2.00
		MT					2.43

* p < 0.05, ** p < 0.01, *** p < 0.001

. For /d/, age, height, and speaking rate were not able to reduce the sex difference. However, the addition of gender-typicality scores reduced the significance. When the significance was reduced the total effect size was 1.53 ms, with males producing longer speech than females. This seems to suggest that gendered behaviour affects the VOT of /d/. While the factors that were identified were able to reduce the significance, the results should be interpreted with caution. For /d/, the effect size was only reduced from 2.62 ms to 1.53 ms. As was noted above, given the range and mean values of /d/ production this difference in effect size is minimal.

For /t/ the significance held in models 2 and 3. However, the addition of age and height increased the effect size to 114.94ms and 83.09ms, respectively. The sex gap was closed in model 4, with the addition of speaking rate, with the mean male VOT being 4.80ms shorter than female speakers. This suggests that sex specific VOT for /t/ is correlated with the speaking rate of individuals.

3.3.2 Additive models for each age group

Previous research has revealed that sex and age can significantly interreact to affect stop-consonant production (Ma et al., 2017). In order to confirm this, additive models separated by age group were constructed for both /d/ (see table 3.10) and /t/ (see table 3.11).

Table 3.9: Summary for each age group's additive mediation models for /d/. The effect size for each variable is measured in milliseconds. The statistical significance key can be found at the bottom of the table.

Dependant Variable - VOT	Independent Variable		Model 1	2	3	4
Age 6-8	Sex		2.29*	2.28*	2.16*	-0.4
	Physical Measures - Height			0.006	0.01	0.02
	Speaking Rate	Vowel Duration			-13.66**	-16.56**
		Word Duration			3.44	4.49*
	gender-typicality	Female				-1.60
		Male				-0.16
Age 10-12	Sex		2.81*	2.98*	2.36	-0.09
	Physical Measures - Height			-0.08	0.06	-0.11
	Speaking Rate	Vowel Duration			-32.96***	-27.24***
		Word Duration			5.04	5.57*
	gender-typicality	Female				-1.39
		Male				-0.37
AGE 14-16	Sex		3.01***	0.96	1.14	1.76
	Physical Measures - Height			0.18	0.17	0.21*
	Speaking Rate	Vowel Duration			-19.08***	-17.50**
		Word Duration			8.33***	9.38***
	gender-typicality	Female				0.85
		Male				-0.46

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

For the voiced consonant /d/, a significant sex effect was observed in every age group. For the youngest group, models 1, 2, 3 found a significant sex difference, with the addition of gender-typicality scores closing the sex gap. In the 10-12 age group, controlling for speaking rate closed the sex gap. The 14-16 group only had a significant effect on sex when nothing was added to the model. Given the small changes in effect size in each of the models while reducing the significance further suggests that the sex difference in /d/ is not robust.

For the production of /t/, a significant difference in sex only existed in the 6-8 group and the 10-12 group. Both the 6-8 and 10-12 age groups follow a similar pattern to the production of /d/, with the 6-8 group closing the sex gap with the addition of gender-typicality scores, and the 10-12 age group closing the sex gap with the addition of speaking rate.

Table 3.10: Summary for each age group's additive mediation models for /t/. The effect size for each variable is measured in milliseconds. The statistical significance key can be found at the bottom of the table.

Dependant Variable - VOT	Independent Variable		Model 1	2	3	4
Age 6-8	Sex		-12.12***	-12.58***	-9.70*	-9.03
	Physical Measures-Height				0.12	0.09
	Speaking Rate	Vowel Duration			25.45**	26.23**
		Word Duration			31.70***	37.42***
	gender-typicality	Female				-1.53
		Male				0.71
Age 10-12	Sex		-11.50*	-11.23*	-5.59	-6.07
	Physical Measures - Height			-0.13	-0.16	-0.12
	Speaking Rate	Vowel Duration			43.54***	49.70***
		Word Duration			46.36***	46.22***
	gender-typicality	Female				0.91
		Male				--0.55
AGE 14-16	Sex		6.78	4.69	5.76	13.15
	Physical Measures-Height			0.18	0.28	0.33
	Speaking Rate	Vowel Duration			56.43***	56.08***
		Word Duration			30.92***	30.45***
	gender-typicality	Female				2.77
		Male				4.71

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results of this section suggest only the weak presence of a sex difference in speakers. There was however a significant difference in participant speaking rates, measures of physical size, and responses to the gender-typicality questionnaires with relation to sex. The sex differences in production do not seem to be the result of physical growth, or gendered behaviour. Individual speech rate seems to significantly affect VOT length. It is possible that gendered differences in speech rate are partly responsible for the sex differences observed.

CHAPTER 4: ENGLISH RESULTS

This chapter contains the results of the English corpus. There are three major sections: 4.1 summarizes descriptive statistics for acoustic data, measures of physical size, and gender-typicality questionnaire data; 4.2 contains the results of the principal components analysis performed on the gender-typicality questionnaire; 4.3 outlines the linear mixed effect model analysis.

4.1 DESCRIPTIVE STATISTICS

The following sections are a descriptive summary of acoustic data, measures of physical size, and gender-typicality questionnaire data. These three variables were used for the linear mixed-effects model analysis (see 4.3). Each variable was tested to determine whether a significant sex difference existed. This was done in order to justify the inclusion of these variables in further analysis. Similarly, each of these variables was also tested to see if a significant age difference existed.

4.1.1 Acoustic summary

The VOT values were calculated for each age group and gender, separated by voiced and voiceless consonant sounds. Figure 4.1 illustrates the mean VOT for these age groups.

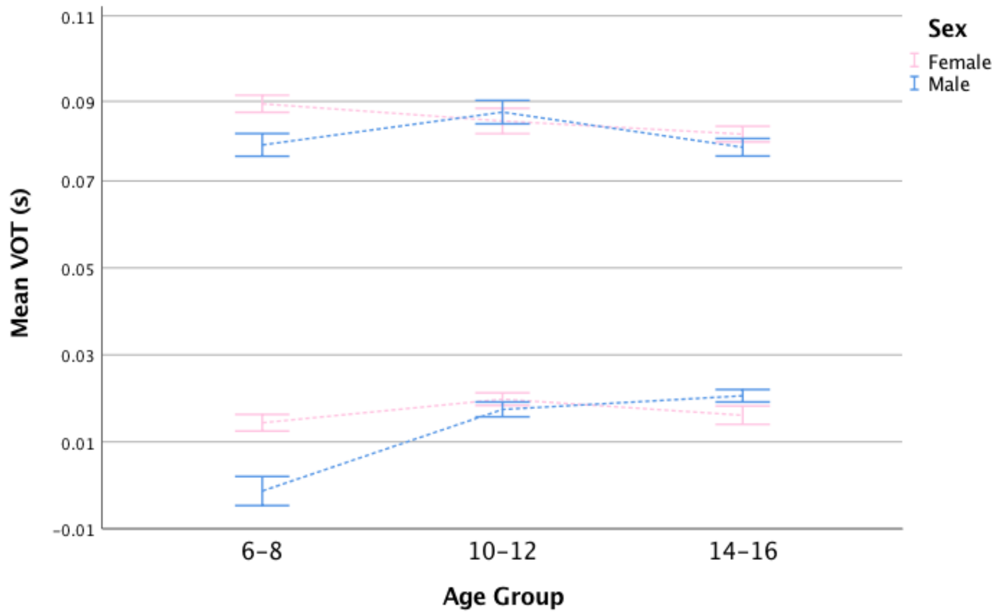

Figure 4.1: A line graph summarizing the mean overall voice onset time by age and sex group. Error bars at a 95% confidence interval.

For voiced consonants, female speakers produced relatively stable VOT between the three age groups. In contrast, male voiced stop consonant VOT gradually increased from the youngest age group to the oldest age group. Notably, male speakers produced negative VOT for both /b/ and /d/, and relatively shorter VOT in the youngest age group, which was an unexpected result. Negative VOT, or prevoicing, is when the vocal cords begin vibrating before the stop consonant release. While this is typically not used in English, it is present in other languages such as French (Netelenbos, Li, & Rosen, 2016). A complete breakdown of the voiced stop consonant VOT production can be found in table 4.1a.

Table 4.1a: The mean voice onset time for voiced consonants for age groups 6-8. 10-12 and 14-16, separated by sex.

Age	Sex	N	/b / Mean (ms)	SD	N	/d/ Mean (ms)	SD	N	/g/ Mean (ms)	SD
6-8	Female	39	8.63	12.23	39	13.12	16.45	39	21.17	17.98
	Male	35	-1.85	30.00	35	-1.30	39.82	35	6.07	39.62
10-12	Female	10	12.38	4.59	10	17.76	2.23	10	29.29	6.76
	Male	13	10.74	6.16	13	17.78	5.47	13	23.65	8.60
14-16	Female	22	10.71	7.29	22	15.41	7.63	22	20.31	16.59
	Male	18	14.32	5.11	18	20.76	7.02	18	26.59	8.00

For voiceless stop consonant production, female speakers overall decreased between the youngest age group and the oldest age group. Male speakers produced relatively shorter VOT than females in the youngest age group. In the 10-12 age group VOT increased , and then decreased in length in the oldest group. This relative increase in male production is consistent with previously observed patterns of English VOT development (Yu et al., 2014). A complete breakdown of the voiceless stop consonant VOT production can be found in table 4.1b.

Table 4.1b: The mean voice onset time for voiceless consonants for age groups 6-8. 10-12 and 14-16, separated by sex.

Age	Sex	N	/p/ Mean (ms)	SD	N	/t/ Mean (ms)	SD	N	/k/ Mean (ms)	SD
6-8	Female	39	79.13	20.34	39	92.33	25.13	39	93.40	20.29
	Male	35	73.35	23.89	35	80.59	24.00	35	89.95	20.49
10-12	Female	10	76.98	10.11	10	86.86	12.56	10	92.39	16.73
	Male	13	79.90	15.41	13	85.72	16.19	13	96.91	15.83
14-16	Female	22	76.63	15.13	22	81.82	15.89	22	88.10	14.02
	Male	18	72.02	17.38	18	80.97	15.96	18	85.02	13.43

A linear mixed-effect model was used to determine whether a sex difference was present in VOT. The dependent variables were the VOT of voiced stop consonants (/b, d, g/) and voiceless stop consonants (/p, t, k/). The independent variables were sex and age group (see Table 4.1). A breakdown of the mean VOT divided into age group and sex can be found in Table 4.2.

Table 4.2: Summary for the follow-up VOT analysis. The effect size for each variable is measured in milliseconds. The statistical significance key can be found at the bottom of the table.

Dependant Variable	Independent Variable		Coefficients (ms)
VOT	Sex	male	-5.801*
	Age	10-12	6.29
		14-16	2.47
	Consonant /b/	/d/	4.16*
		/g/	11.13***
		/k/	84.66***
		/p/	68.68***
		/t/	77.59***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

An overall significant sex effect was found, with males VOT being 5.801 ms shorter than females, suggesting that there is a difference between male and female production of voiced and voiceless stop consonants. There was no overall effect of age. For /d/, males VOT was 2.67 milliseconds longer relative to females. A significant effect was found for the 10-12 age group relative to the youngest age group, with the older group producing 1.60 milliseconds longer VOT. For /t/, a significant effect of sex was found, with males VOT 6.79 milliseconds shorter than females. No significant effect of age group was found.

While the statistical analysis revealed some significant sex effect, figure 4.1 makes it clear that these results should be interpreted with caution. There appears to be minimal variation in both sex for both voiced and voiceless consonants for the 10-12 and the 14-16 age group. Similar to the Mandarin result, while these results may be statistically significant, when compared to mean and range values, the effect size represents a small variation. The exception to this seems to be the youngest age group for both consonant sets. For voiced stop-consonants, males are producing negative VOT for both /b/ and /d/. For voiced consonants, males VOT is relatively shorter than females. This gendered pattern is typical for speakers of English (see Yu et al., 2017).

Vowel duration

Vowel duration was measured for three different vowels: /ae/, /i/, /u/. The overall mean results can be seen in figure 4.1. A breakdown of the vowel duration by vowel type, age group, and sex can be found in figure table 4.3.

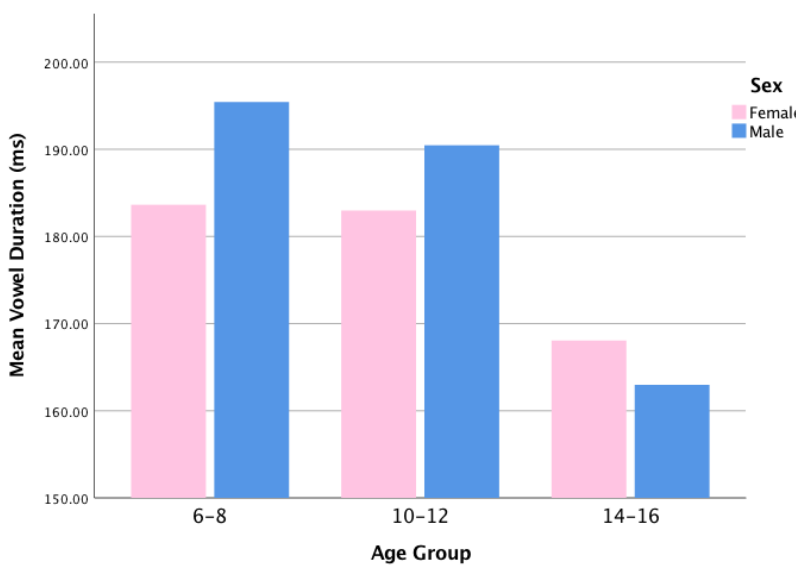

Figure 4.2: A histogram summarizing the mean overall vowel duration separated by age group and sex.

The vowel duration for both females and males were observed to decrease from childhood into adolescence (see figure 3.2). This decrease in vowel duration is an expected part of language development for English speakers (Skoog, Eriksson, & Sörqvist, 2015). Male speakers were observed to produce longer vowel duration relative to males for the first two age groups, while females produced longer vowel duration in the oldest age group. Male speakers produced relatively longer vowel duration in the first two age groups, and then the pattern is reversed in the oldest age group.

Table 4.3: The mean vowel duration for three vowels: /a/, /i/, and /u/. These values are separated by sex and age group.

Age	Sex	/a/ N	/a/ Mean (ms)	/a/ SD	/i/ N	/i/ Mean (ms)	/i/ SD	/u/ N	/u/ Mean (ms)	/u/ SD
6-8	Female	48	182.26	32.18	92	176.20	36.12	94	191.63	31.63
	Male	49	188.60	41.81	78	191.61	50.46	83	203.05	53.78
10-12	Female	14	193.76	33.90	25	180.11	22.82	21	179.16	22.63
	Male	18	206.85	37.19	35	188.06	21.45	25	182.02	32.16
14-16	Female	35	172.86	26.25	51	163.25	26.80	46	169.73	30.75
	Male	22	164.64	32.71	46	160.24	31.47	40	165.19	26.28

A repeated measures ANOVA found no significant main effect of sex on vowel duration ($F(1, 128) = 3.78$, $p = 0.054$). These results indicate that male and female speaker's vowel length do not differ significantly. However, there was also a significant effect of age ($F(2, 128) = 10.28$, $p < 0.001$), indicating that there was a significant change in vowel length as participants aged, decreasing in length with age.

Word duration

In addition to vowel duration, word duration was also measured and summarized to help control for speech rate. The overall mean word duration can be seen in figure 4.3. A breakdown of the word duration by syllable, age group, and sex can be found in table 4.4.

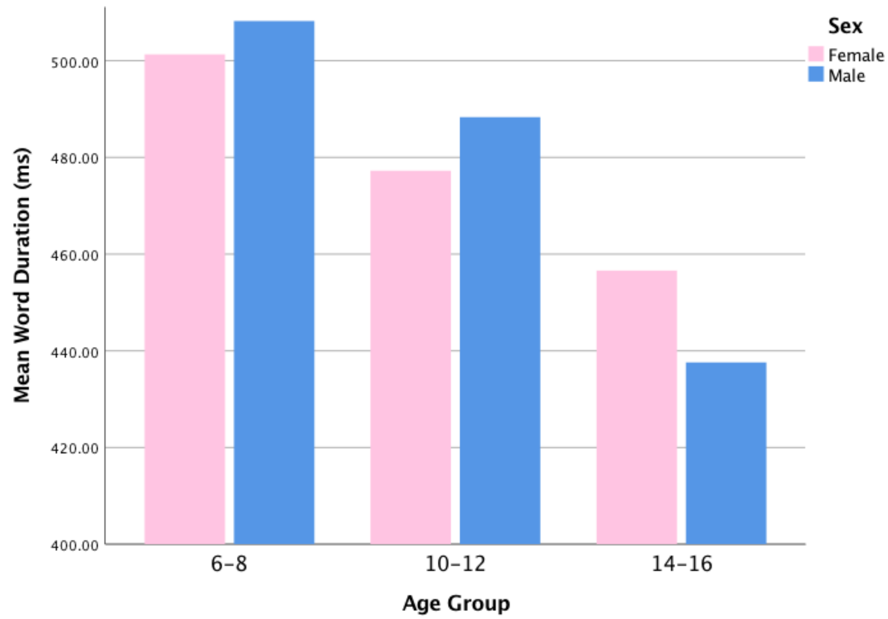

Figure 4.3: A histogram summarizing the mean overall word duration separated by age group and sex.

Similar to the vowel duration, the decrease in word duration from childhood into adolescence is an expected pattern of speech development (Skoog, Eriksson, & Sörqvist, 2015). Similar to vowel duration, male speakers had a slower speech rate in the 6-8 and 10-12 age groups. An opposite pattern was observed in the 16-18 group, which is typical for that age group (Feldstein, Dohm, & Crown, 1993).

Table 4.4: The mean duration of English words, separated by syllable, age group and sex.

Age	Sex	1 Syllable			2 Syllable			3 Syllable		
		N	Mean (ms)	SD	N	Mean (ms)	SD	N	Mean (ms)	SD
6-8	Female	35	443.01	62.06	35	559.83	72.65	33	681.44	72.38
	Male	122	545.80	101.58	849	678.81	147.48	126	893.82	193.10
10-12	Female	10	416.48	46.24	10	537.07	49.45	10	665.35	58.90
	Male	13	430.45	41.69	13	548.66	34.64	13	650.73	67.73
14-16	Female	18	388.11	44.60	18	488.06	55.23	18	578.14	67.54
	Male	69	434.05	112.83	468	542.96	102.14	68	786.27	124.36

While controlling for syllable number, a repeated measures ANOVA found that there was no significant main effect for sex ($F (1, 128) = 0.39$, $p =0.93$). However, there was a significant effect of age group on word duration ($F (2, 128) = 15.26$, $p < 0.001$). This indicates that female speakers produced longer word durations than male speakers, and that word duration decreased with age.

4.1.2 Physical size

The English-speaking corpus provided two measures of physical size: height and weight. A breakdown of these physical measures by age group and sex can be found in table 4.5.

Table 4.5: Mean measures of physical size for English -speaking children. The mean measures of height and weight are reported for each age group and sex.

Age Group	Sex	Height (cm)		Weight (kg)	
		Mean	SD	Mean	SD
6-8	Female	130.01	7.75	27.56	6.37
	Male	129.71	6.46	28.15	6.10
10-12	Female	152.19	13.28	43.62	16.57
	Male	152.76	24.85	52.62	16.93
14-16	Female	164.88	6.75	58.77	9.36
	Male	176.79	5.51	70.65	10.17

A Pearson correlation between weight and height was taken. Weight and height were found to be highly correlated with each other ($r = 0.815$, $p < 0.01$). Similar to the Mandarin corpus, due to height being previously established correlation with vocal tract length (Fitch & Giedd, 2002), the decision was made to only include height in further analysis. Height data separated by age group and sex can be found in figure 4.3.

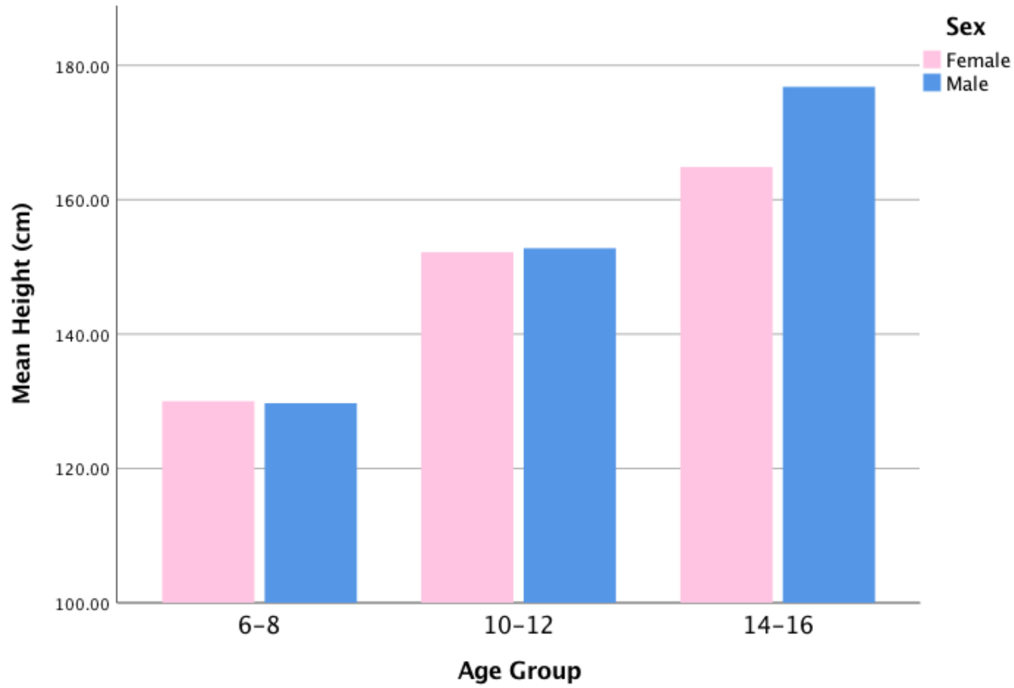

Figure 4.4: The mean height of English Participants separated by sex and age group.

A two-way ANOVA found a significant effect of sex ($F(1,129) = 12.99$, $p < 0.001$, partial $\eta^2 = 0.91$) and age group ($F(2, 129) = 132.37$, $p < 0.001$, partial $\eta^2 = .672$) on height. Follow-up t tests found that there was only a significant difference in height between males and females in the 14-16 age group ($t = -4.99$, $p < 0.001$), with males being taller. There was also a significant interaction between sex and age group ($F(2, 129) = 3.65$, $p < 0.001$, partial $\eta^2 = 0.03$).

4.1.3 Gender-typicality questionnaire

The English gender-typicality questionnaire contained 12 questions. Before they were used for summary statistics, gender typicality scores were converted to gender-typicality scores. Similar to the Mandarin gender typicality questionnaire, this questionnaire was designed to have specific questions that asked participants about both male and female-typical behaviours, on a scale from 1 (behaviour rarely occurs) to 5 (behaviour occurs frequently). Participants were also able to indicate if the question did not apply to them with a 0. Responses marked with a 0 were removed and left blank for the analysis. Male-typical questions were converted to the same scale as the female-typical questions.

The mean female response to the questionnaire after was 3.73 (SD = 0.40), while mean male response was 2.04 (SD = 0.42; see table 4.6). An independent samples t-test found that this difference in gender typicality response was significant ($t = 24.02$, $p < 0.001$).

Table 4.6: Mean Gender-typicality scores separated by age group and sex.

Age Group	Sex	Mean	Std. Dev
6-8	Female	3.76	0.42
	Male	2.00	0.36
10-12	Female	3.64	0.35
	Male	2.05	0.57
14-16	Female	3.72	0.40
	Male	2.16	0.30
Overall	Female	3.73	0.40
	Male	2.04	0.42

A two-way ANOVA test found a significant main effect for sex ($F(1,131) = 450.79$, $p < 0.001$, partial $\eta^2 = 0.775$) and but not for age group ($F(2, 131) = 0.373$, $p = 0.689$, partial $\eta^2 = 0.006$). There was also no significant interaction between sex and age group ($F(2, 131) = 0.898$, $p = 0.41$, partial $\eta^2 = 0.014$).

4.2 PRINCIPLE COMPONENTS ANALYSIS

A principal component analysis (PCA) was used to determine the underlying components of the gender-typicality questionnaires. The scores first underwent cleaning and missing data replacement before analysis. Similar to the Mandarin corpus, questions that were marked as non-applicable were left blank and considered missing data in the anlysis. The English gender-typicality questionnaire had 12 questions. The response rate to these questions ranged between 79 and 100%. The individual response rate to the whole quiz ranged between 83 and 100%. These response rates were due to individuals indicating that the questions did not apply to their circumstances and were left blank.

The principal components analysis requires that subjects with missing data must either be excluded from the analysis or that their data must be replaced. Participants with a response rate of 80% or higher (missing more than two questions) were considered eligible for data replacement. The missing data was replaced with the mean questionnaire response. No individuals were removed based on missing data. No univariate or multivariate outliers were observed.

The principal component analysis revealed two main components. The scree plot can be seen in figure 4.4. Component 1 accounted for 41.0% of the variation, while Component 2 accounted for 26.1% of the variation. The components were identified by comparing which questions were associated with each component. Each question was given a factor score for each component during the PCA analysis, and then these scores were rotated using Varimax rotation. The highest rotated factor score was used to identify which component the question belongs to. The component categories were identified by looking at which questions were associated with each component.

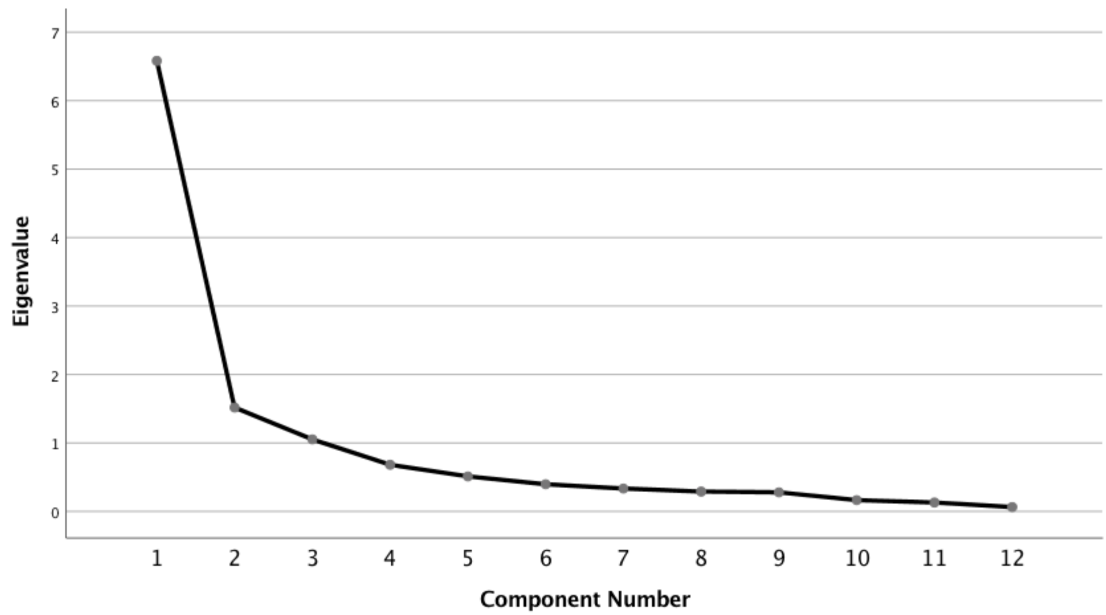

Figure 4.5: The scree plot for the English Gender-typicality questionnaire.

Table 4.8 outlines which questions were associated with each component. Component 1 asked about feminine typical behaviour and accounted for eight questions (see figure 3.8). The question in Component 2 asked about male-typical behaviour, and also accounted for four questions. Component regression scores were generated for each subject for every component used.

Table 4.7: Factor Loadings after Varimax rotation. The highest factor loadings for each question are bolded indicating which factor the question belongs to.

	Factor 1	Factor 2	Questions
Question 1	**0.605**	0.445	Favourite playmates are (1 always boys, 5 – always girls)
Question 2	**0.839**	0.309	Plays with girl type toys, such as "Barbie"
Question 3	-0.06	**0.808**	Plays with boy type toys, such as "Gi Joe"
Question 4	**0.836**	0.268	Experiments with cosmetics and jewelry
Question 5	**0.747**	0.151	Imitates female characters seen on TV or in movies
Question 6	0.168	**0.74**	Imitates male characters seen on TV or in movies
Question 7	0.299	**0.601**	Plays sports with boys
Question 8	**0.575**	-0.156	Plays sports with girls
Question 9	**0.777**	0.515	In playing house, your child takes on the role of a girl or women
Question 10	**0.848**	0.298	Plays girl-type games such as "princess"
Question 11	0.459	**0.708**	Plays boy-type games such as "weapons"
Question 12	**0.772**	0.553	In a dress-up game, your child dresses up as (1 – a girl or women most of the time, 5, a boy or man most of the time

4.3 ADDITIVE MEDIATION MODEL (AMM)

In order to address the primary research question of determining what factors influence the sex difference in VOT production, an additive mediation model was employed. The purpose of this analysis was to identify which variables contribute to the sex differences in the production of VOT for /d/ and /t/. Similar to the Mandarin AMM analysis, the English language analysis was constructed by first establishing a base model. Variables were then added successively to the model. The base model used for analysis can be seen in table 4.9.

Table 4.8: Summary for the overall additive mediation model. The effect size for each variable is measured in milliseconds. The statistical significance key can be found at the bottom of the table.

Dependant Variable	Independent Variable		Coefficient (ms)
VOT Overall	Sex	male	-5.81*
	Age	10-12	6.32
		14-16	2.50
	Target Consonant	d	4.26***
	(base consonant b)	g	11.08***
		k	84.66***
		p	68.81***
		t	77.70***
VOT Voiced	Sex	Male	-6.32
	Age	10-12	11.17*
		14-16	9.83*
	Target Consonant	d	4.19***
		g	10.97***
VOT voiceless	Sex	Male	-5.26
	Age	10-12	1.43
		14-16	4.88
	Target Consonant	p	15.86***
		t	6.97***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

There is a main effect of sex (see table 4.9) in the overall model. However, this result shows a small total effect size of -5.26ms for males when compared to females. This indicates that male speakers generally produce shorter VOT for both voiced and voiceless stop-consonants. Due to the inherent differences in VOT ranges between voiceless and voiced consonants (see tables 4.1), a separate base model was constructed for each sound category. For voiced consonants, the analysis revealed no overall significant effect of sex, with an effect size of -6.31ms for males compared to females (see table 4.9). There was also no significant effect of sex for voiceless consonants, with an overall effect size of -5.26ms for males compared to females (see table 4.9).

These results seem to indicate that the sex difference in VOT is either not very strong or non-existent. Similar to the Mandarin corpus, there was a significant sex difference for the measures of physical size and gender-typicality scores. However, unlike the Mandarin corpus, there was no significant sex-effect for the speech rate measures.

CHAPTER 5: DISCUSSION

Sex-specific production of speech sounds has been observed in several sounds, however how these differences develop is not well understood. For stop-consonant sounds, sex-specific production has been controversial. This thesis investigated the sex-specific production of stop consonants in English and Mandarin. In addition to producing comparable stop-consonant sounds, both languages have been shown to produce sex-specific stop-consonants. The purpose of this thesis was to investigate the origin of the sex-specific production of stop consonants in both languages.

Two steps were taken to explore this production. The first was to determine whether sex-specific production of stop consonants was being used in both languages. Sex-specific speech has been shown to be both language-specific or language-universal (Li et al., 2016). Cross-language comparison is a useful tool for distinguishing the difference between the two. The aspect of stop-consonant production that was measured was VOT, which is a temporal measure of the time between vocal tract closure and vocal cord vibration. The second step was, if there was a difference in production, to determine what potential factors may contribute to this difference.

In order to determine what factors could potentially affect the sex-specific production of stop consonants, two main methods were used. The first was to look at development before, during, and after the onset of puberty. This method was done to look at potential sex differences due to the anatomical differences in male and female vocal tract anatomy during development, and to determine the developmental pattern of VOT for each sex. The second method was to make comparisons across both languages. This method was used to investigate the sex-specific patterns observed in both languages, to determine if the observed patterns were language specific or shared. An additive model was adopted to determine factors which could affect the sex difference.

Three hypotheses were made in the introduction about the findings from this research. The first hypothesis, based on previous development studies of stop consonants (see Ma et al., 2017; see Yu et al., 2014), that there would be a difference in VOT production between males and females, with females producing more aspirated speech generally across both languages. The second hypothesis would be that the difference between male and female speech would be reduced during puberty. This was observed in prior developmental studies in both languages and would suggest the physical development that occurs during puberty would be a factor in VOT production (Ma et al., 2017; see Yu et al., 2014). The third hypothesis was that gendered behavior would influence sex-specific speech. Specifically, male-typical behaviour will lead to shorter relative VOT times, while more feminine typical behaviour should lead to longer relative VOT. Sex specific production has been shown to emerge before the onset of puberty, suggesting a social component to VOT production (Ma et al., 2017; see Yu et al., 2014; Whiteside & Irving, 1997; Whiteside & Irving, 1998).

The following two sections discuss the implications of the findings from the Mandarin and English analysis individually. The third section compares and contrasts the findings from the two languages. The fourth section outlines how this research contributes generally to our knowledge of sex-specific stop-consonant production. Section 5.5 compares the finding from this thesis to other sex-specific sounds, specifically vowels and /s/. The purpose of this section is to highlight what this research contributes to our knowledge of gendered speech in general. The final section outlines the potential shortcomings of this research, with suggestions for future research in this area.

5.1 FINDINGS FROM MANDARIN

Consistent with previous studies of the Mandarin language (Li, 2013; Ma et al., 2017), there was an overall significant sex difference in VOT production. /d/ production was significant across all three age groups, while /t/ production was only found to be significant in the 6-8 and 10-12 age groups. The initial tests of the factors found that males and females differed significantly in speech rate, height, and their responses to the gender typicality questionnaire and each were considered a potential factor that could contribute to the sex difference in speech production and was added to the AMM.

For /d/, physical size and speech rate were not able to reduce the overall significance. It was only when gender typicality was added to the model that the significance was reduced. The results suggest that gendered behaviour mediates VOT output. However, this is most likely, not the case. During the AMM, the total effect size was only reduced by about 1 millisecond when the statistical significance was reduced. Given the range of variation both across sexes and within sexes, this small difference in VOT suggests that while there may be a statistical difference, there may not be a practical difference for communicating personal sex, with a significant amount of overlap for each sex.

For /t/, the speech rate reduced the significance after age, and physical size were accounted for. For both models, it appears that physical size does not seem to affect the overall VOT of speakers. This would suggest that differences in the larynx and vocal cord size do not affect VOT output. However, once the speech rate was account for, the sex difference was reduced. Speech rate was found to be significantly different for each sex, with male speakers speaking at a faster rate than female speakers. Given that VOT and speech rate are both temporal measures, it is possible that the significant difference in the speech rate between male and female speakers could account for this difference.

One interesting observation from the Mandarin data was the reduction in the sex difference between male and female speakers in the 14-16 age group. This result is similar to what Ma et al. (2017) reported for ages 12-14. However, Ma et al. found that male speakers' mean VOT increased during this period, while the current data shows that female speakers reduced mean VOT lengths. Adult speakers of Mandarin have been found to produce sex-specific VOT, and late adolescent speakers move towards this adult-like pattern of speech (Li, 2013; Ma et al., 2017). It appears that some factor seems to affect Mandarin speakers during this age period, but whether that factor is a common pattern of development across languages, or language specific is unclear.

5.2 FINDINGS FROM ENGLISH

The results of the English analysis suggest that there is only a minimal VOT sex difference in production. While the overall results suggested a significant difference when all consonants were considered together, when the voiced and voiceless consonant sounds were analyzed separately there was no difference. Previous research on sex differences in English has been controversial (see Morris et al., 2008; Robb et al., 2005; Yu et al., 2014). These studies have cited the need to control for stimuli type, speech rate, and in some cases, recording context, which have all been shown to affect VOT. Each of these factors were controlled for in this study. These results are consistent with the few studies that have controlled these factors. For example, Morris et al. found that after controlling for these factors that there was no sex difference in VOT, but females did tend to produce longer, but not significant, VOT.

Similar to the Mandarin corpus most of the differences observed in the English corpus were observed before the onset of puberty, when the vocal tracts of both sexes are mostly the same (Fitch & Giedd, 1999). One of the novel observations in the English group was the presence of prevoicing in the male speakers of the 6-8 age group. Prevoicing is the activation of vocal cords before the stop consonant release and occurs in voiced stop-consonants (i.e./d, g, b/; see figure 5.1), which results in negative VOT. The overall 6-8 age VOT values for /b/ and /d/ were negative. The mean VOT value of /g/ was 6.07 seconds, however there were several cases of prevoiced /g/. While the mean values of female VOT were positive, the range of stop-consonants included both voiced and prevoiced stop-consonant production. Prevoicing is typically not thought of as a feature of Canadian English, but it is commonly seen in other languages, such as French (Netelenbos & Li, 2016).The standard deviation for the voiced stop consonants in the youngest group show there is a considerable amount of variation in production for both sexes. Further research will be needed to understand why male prevoicing was so pronounced in the youngest age group.

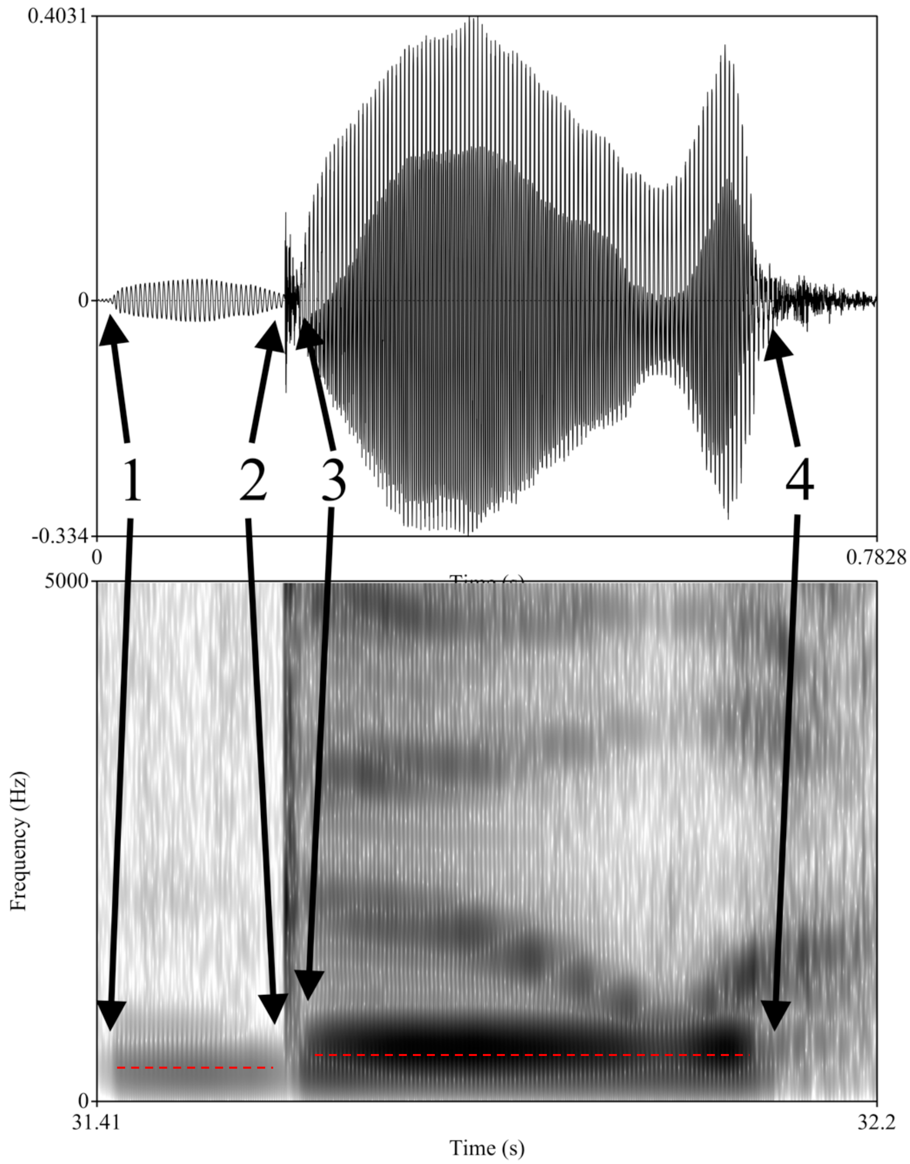

Figure 5.1a: The waveform and spectrogram for the prevoiced word 'goo.' The four measurement points for VOT are marked on both the waveform and the spectrogram. Points of interest recorded: voicing onset (1), burst (2), vowel onset (3), and vowel end (4). The voicing bar can be seen in the spectrogram both before and after the burst onset (dotted red line). Prevoicing can also be observed in the waveform between points 1 and 2 as a regularly repeating pattern.

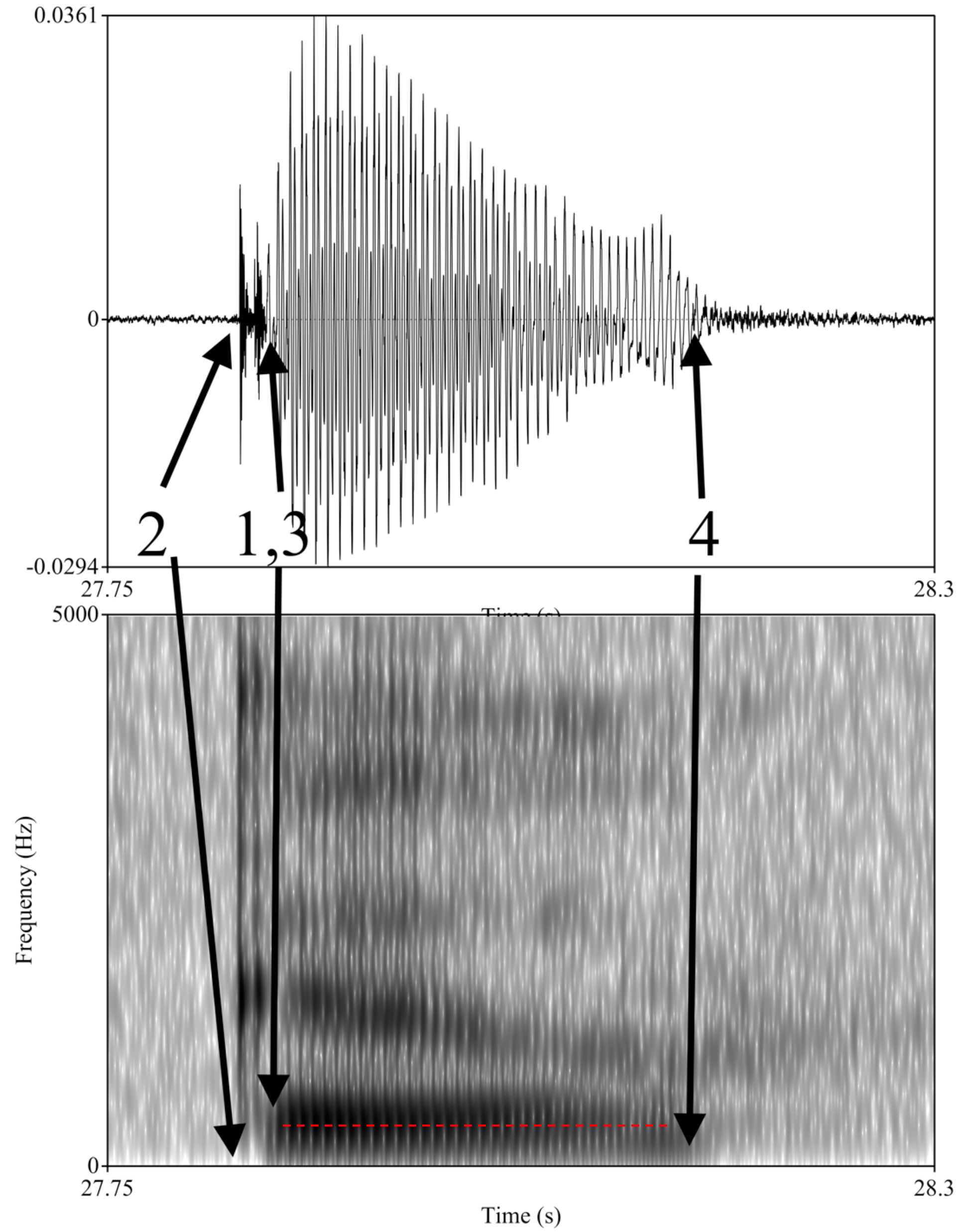

Figure 5.1b: The waveform and spectrogram for the word 'goo' with no prevoicing. The four measurement points for VOT are marked on both the waveform and the spectrogram. Points of interest recorded: voicing onset (1), burst (2), vowel onset (3), and vowel end (4). The voicing bar can be seen in the spectrogram both before and after the burst onset (dotted red line).

From the factors that could have played a role in gendered production, the measures of physical size significantly differed for each sex. However, physical development did not have a strong effect on VOT output. There was an increase in VOT length during the 10-12 age group for males, which is associated with the onset of puberty. Yu et al. (2014) observed a similar VOT increase in length during puberty. The increased VOT length during the 10-12 age group would suggest that a possible link between puberty and VOT development. However, these changes did not seem to persist after puberty, which suggests that vocal tract length is likely not the factor that contributes to this change and must be the result of some other mechanism.

In addition to physical size, the responses to the gendered behaviour were also found to be significantly different between sexes. The fact that there was very little difference in VOT production between sexes, suggests that stop-consonant production in English does not seem to be affected by gendered behaviour.

Unlike the Mandarin corpus, the speech rate was not found to differ for each sex significantly. This was true for both the vowel duration and word duration. Given the temporal nature of both VOT and speech rate and the significance found in the Mandarin results, speech rate may have a mediating effect on VOT output.

5.3 CROSS LANGUAGE COMPARISON

The current study systematically compares English and Mandarin Chinese for the effect of sex on VOT in children's and adolescent's stop productions. Such a cross-language study could avoid confounds due to methodological inconsistencies and thus help reveal cross-language similarities or differences that could shed light on the origin of sex-specific VOT pattern (see Morris et al., 2008; Robb et al., 2005; Yu et al., 2014). This was done by controlling for stimulus type, speech rate, and analysis procedures.

The most notable difference between the two languages is that a sex difference was much more robust for Mandarin than it was for English. This result is consistent with previous research of the two languages, with a more robust sex effect being observed in Mandarin than English (Li, 2013; Ma et al., 2017; Morris et al., 2008; Yu et al., 2014). For voiced stops, Mandarin speakers' VOT production was relatively stable throughout all three age groups. This was also the case for English speakers, except for the youngest group for male speakers, which produced prevoicing in voiced stops. However beyond this difference, there was very little difference between males and females in both languages. Most of the sex specific differences were seen in the voiceless stop-consonant category. Due to this small difference, the remainder of the comparison will focus on voiceless stops, unless otherwise stated.

For voiceless aspirated stops, both groups had a significant sex difference in production in the youngest age group. However, only Mandarin was significantly different for the 10-12 age group, and neither was significant in the 14-16 age group. Males during the 10-12 age group in both languages showed an increase in VOT length. The results of the analysis indicate that how stop consonants develop is language specific. This is particularly interesting, since the stop-consonant production of adult speakers of the two languages is comparable (Lisker & Abramson, 1964). The reason for this difference is likely due to language-specific differences in language development, however further data would need to be collected to confirm this speculation.

Another difference that was observed between the two corpora was sex-specific speech rate. Speech rate was sex-specific in Mandarin and not in English. Previous research has found that speech rate affects VOT (Kessinger & Blumstein 1997; Morris et al., 2008).

The AMM models for the Mandarin analysis also revealed that the speech rate was highly correlated with VOT. Faster speech rates lead to a reduced VOT (Kessinger & Blumstein 1997). One of the sources for sex specific VOT may be the result of the speech rate differences between males and females. The AMM analysis found that controlling for speech rate reduced the overall significance of Mandarin voiceless stops as a whole. However, this does not explain the reduction of the sex gap for the 14-16 Mandarin group, where speech rate was significantly different, but VOT was not. Further research will be needed to investigate the relationship between speech rate and VOT more fully.

5.4 DEVELOPMENTAL AND LANGUAGE SPECIFIC FACTORS AFFECTING VOT

Both languages offer valuable insight into the production of sex-specific stop-consonant production. There appears to be common developmental patterns, and language specific patterns that affect the production of VOT. The following section outlines how these factors could potentially affect stop consonant production, starting first with biological factors.

This research provides evidence that biological factors are in part responsible for stop-consonant development. As noted above, both languages showed an increased mean VOT for male speakers' production of voiceless stop consonants during the 10-12 age group, which corresponds with the onset of puberty. Similar developmental studies have also observed this pattern (Ma et al., 2017; Yu et al., 2014). This cross-language similarity leads us to propose that the origin of such difference may be, in part, biological-based. However, it is not entirely certain what the specific mechanism of that change is. As this pattern does not seem to persist into the 14-16 age group, this pattern is unlikely the result of anatomical development, and warrants further investigation.

Another potential explanation for this change could be the changing hormone levels of speakers. Previous research has shown that differences in hormone levels during different periods of the menstrual cycle affect speaker VOT (Wadnerkar, Cowell, & Whiteside, 2006; Whiteside, Hanson, & Cowell, 2004). High levels of estrogen and progesterone have been shown to increase mean VOT during menstruation. The role of hormones plays a profound effect on individuals throughout the lifetime and affect male and females differently (Erhart & Meyer-Bahlburg, 1981; Van Goozen et al., 1995). This has been shown to affect a variety of behaviours (See Van Goozen et al., 1995). However, the connection between hormone levels and speech output is not well studied, and so far, only accounts for female speakers during their menstrual cycles. There is currently very little research on the effects of hormones on male speakers, or how differences in hormone levels create differences between male and female speakers. It does, however, provide evidence that there may be underlying biological mechanisms that affect gendered speech that are not accounted for by anatomy alone.

Further research will be needed to determine whether underlying there exist other biological mechanisms that could affect stop-consonant output. However, the data collected for this thesis seems to suggest that language-specific factors play a more significant role than biological mechanisms. This is most clearly seen by the differences in production across the two languages. Even though Mandarin and English have a similar set of stop-consonant sounds, how this sex difference was expressed differed. If the differences in stop-consonant production were based mainly on biological factors, we would expect to see consistent patterns of production and development across languages.

Further investigation of other languages makes it clear that sex-specific production of stop-consonants is not expressed the same in languages beyond Mandarin and English. For example, in the study of sex-specific Korean stops, Oh (2011) found that male speakers generally produce longer VOT than females. In a study of speakers of Parisian French, Pépiot (2015) found that female speakers produced longer voiceless VOT than males, but also produced relatively shorter voiced stop-consonants than males. The findings from this thesis combined with these examples strongly suggest that the sex-specific production of stop-consonants is more likely due to differences in language-specific influence rather than biological differences.

One possible factor that may help explain differences in VOT time could be speech rate. Both speech rate and VOT are temporal measures of speech. It is also the case that they share an inverse relationship; the faster an individual speaks, the shorter relative VOT they will produce (Kessinger & Blumstein, 1997). While this research specifically controlled for speech rate in order to make a proper comparison between the two languages, there is evidence to suggest that the sex difference in production for both languages may be more due to a difference in speech rate, and not a difference in stop-consonant production. For the Mandarin corpus, there was a significant difference in speech rate and VOT recordings. However, during the AMM, the sex-difference was closed for voiceless stops when speech rate was added. Similarly, in the English, no sex difference was found for speech rate, where there was no difference sex difference for voiceless and voiced stop-consonants separately. This would suggest that there may not be a robust sex difference in stop-consonant production in the Mandarin and English languages when the speech rate is controlled. While a complete answer to this question is beyond the scope of this thesis, this question will need to be addressed in further detail.

5.5 IMPLICATIONS FOR SEX-SPECIFIC PHONETIC PRODUCTION

One of the primary purposes of this thesis was to explore sex-specific phonetic production of speech generally. While there have been several studies that have investigated differences in sex differences in phonetic production, there are relatively few studies that have attempted to investigate the origins of these sex differences and their pattern of development (Ma et al., 2017; see Yu et al., 2014; Whiteside & Irving, 1997; Whiteside & Irving, 1998). Additionally, most of these studies have focussed on English, without a crosslinguistic approach. As a result, further research is still needed to understand sex-specific phonetic production and how it affects speech production. The research in this thesis can only offer a small glimpse into the general speech patterns of female and male speakers. This makes this research valuable not only in understanding stop-consonant production but also in the study of sex-specific speech production in general.

The findings from this thesis are perhaps most useful when put into the context of what is known about other sex-specific phonetic production, specifically vowel and sibilant production. While these three types of sounds are very different, a comparison between the three offers a more complete picture of sex-specific production. Any comparison between stop consonants, vowels, and sibilants cannot be perfect since each is has a different measure. Stop consonants are measured using VOT, vowel quality is measured using pitch and formant frequency, and sibilants use spectral frequencies to measure quality. Despite these differences, several comparisons can be made between the three sounds.

Perhaps the most notable difference between vowel sounds and stop consonants is that vowel production is much more affected by vocal tract anatomy than stop consonants. While there are language specific factors that affect the production of vowels, the length and shape of vocal tract anatomy is a crucial factor in the sex-specific production of vowels (Fitch & Giedd, 1999). In contrast, there is no clear link between vocal tract length and stop consonant production. Similarly, sibilants, such as /s/, do not seem to have a clear connection between vocal anatomy for sex-specific production (Li et al., 2016). One reason for this difference is that sex-specific vowel production relies on vocal cord vibration, while stop-consonants and sibilants do not. This suggests that sex-specific speech production is not restricted to only one feature.

Another feature that contrasts between vowels and stop-consonants is that stop consonants do not seem to have a typical general pattern of sex-specific production. The patterns for sex-specific stop-consonants varied between Mandarin and English, with no consistent pattern between the two languages. In contrast, male speakers produce lower frequency vowels than females across languages (Hillenbrand et al., 1995; Takafuta, 1972; Boë et al., 1975; Rose 1991). While there are differences in the way vowels are produced across languages, this pattern is constant. In comparison, stop-consonants sex-specific production seems to be language specific. This is also similar to sibilant production. While there is a robust sex-specific production for /s/ in English (Kinsmen & Li, 2013; Li et al., 2016) a sex difference is not universally seen across languages (Li, 2017; Simpson & Weirich, 2015). Similarities across languages suggest a common pattern of language development factor affecting speech, while differences would suggest language specific development. For stop-consonants, this adds to evidence that production seems to be more language specific.

There are also several notable differences between stop consonants and sibilant production. Notably, the sex-specific differences for /s/ is far more robust than stop-consonants (Simpson, 2009). The findings in this thesis did not find a strong sex-difference in either Mandarin or English. In contrast, the sex-differences in /s/ production have been well documented (Fox & Nissen, 2005; Li et al., 2016; Simpson, 2009; Stuart-Smith, 2007). Additionally, sex-specific /s/ production is mediated by gendered behaviour. This difference not only explains differences between sexes, but also variation within sex (Li et al., 2016). For stop-consonants, it was not the case that gendered behaviour had a significant effect on sex-specific production. While there needs to be further research, the sex difference in stop-consonants does not seem to be near as strong as sibilants.

In the context of other sex-specific phonetic production, the sex effect in stop-consonants does not seem as strong as other sounds. However, given the limited amount of research in this area, there is still more than needs to be learned before a proper conclusion can be drawn. This research is unique in its scope and depth, as it controls for several speech measures, investigates the developmental trajectory of speech, and compares this development across languages. While there are several limitations, which are outlined below, research of this kind gives unique insight into gendered speech production.

5.6 LIMITATIONS AND FUTURE RESEARCH DESIGN

There were several limitations to the research that was conducted for this thesis. The first major limitation is regarding the stimuli data that was available for this analysis. While several similarities can be drawn between the two groups, this comparison can only be partial. One of the significant differences between the two corpora was the stop-consonant stimuli. While the English corpus was able to provide six different stop consonant sounds (/p, t, k, b, d, g/), the Mandarin corpus only provided two (/t, d/). While previous developmental research found that the Mandarin stop-consonant sounds not found in this research share a similar developmental pattern as /t/ and /d/ (see Ma et al., 2017), future research would benefit from having a wider variety of Mandarin sounds.

While the number of stop-consonant sounds available from the English corpus offered a greater variety of sounds, the number of English-speaking participants was relatively small when compared to the Mandarin corpus. This is particularly true for the 10-12 and 14-16 age groups, who had 35 and 28 participants, respectively. This is particularly problematic when comparing speech across sexes, which further divides the already limited number of participants. For example, the number of 14-16-year-old speaking males in Mandarin was 33, compared to only 12 English speaking males. Fewer participants makes the data more susceptible to individual variation, which could potentially change the results.

Similarly, an additional limitation is the lack of adult participants. Previous developmental research of this kind often uses adult speakers as a standard for comparison (Yu et al., 2014; Ma et al., 2017). This would have been particularly beneficial for this research since the 14-16 voiceless stop consonants showed no sex difference in either language. Typically, adult speakers in both languages produce sex specific VOT. While this thesis has looked at the developmental trajectory of VOT before, during the onset, and following puberty, there may be further development between late adolescence and adulthood. Given that the data used for this analysis came from two pre-existing corpora, to make these changes would be difficult. However, these considerations should be made for future research.

Finally, there are limitations to the use of the AMM approach that was used for this analysis. The sequential addition of factors in the AMM model can only reveal which factors are correlated with one another. This can be problematic for factors that are already known to be highly correlated, such as the use of sex and gendered behaviour in this analysis. Rather than reducing the statistical significance, it is possible that significance is simply being split between the two similar factors. It is also the case that the order that factors are added to the model could affect the overall outcome of the model. For these reasons, consideration needs to be taken into how the models are put together.

A possible alternative to the AMM approach would be to use a model selection approach. In model selection, all possible factors could be added to determine which factors contribute to variation in VOT production. However, model selection would not be able to address the sex differences in VOT specifically directly. The AMM approach directly addresses this problem, which makes it preferable for this analysis. Due to the mild sex effect in VOT, the AMM model is sufficient for these data. However, future use of this kind of AMM should be used cautiously.

In addition to the suggestions mentioned above for future research, future research would benefit from the inclusion of perceptual experiments. These kinds of experiments, which test for listener perception of speaker sex, have been conducted for both vowel sounds (Hillenbrand, 2009) and sibilants (Munson, 2007; Munson et al., 2015). Vowel pitch and formant frequency are reliable cues for an individual's sex, both before and after the onset of puberty (Cartei, Cowles, Banerjee, & Reby, 2014; Lee, Potamianos, & Narayanan, 1999; Perry, Ohde, & Ashmead, 2001; Li et al, 2016). Differences in /s/ production have been shown to signal gender in English (Fox & Nissen, 2005; Fuchs & Toda, 2010; Heffernan, 2004). However, at this time, this kind of perceptual research has not been conducted for stop consonants. A perceptual study would offer a broader context into the sex-specific production of stop-consonants and how they are used to communicate.

Specifically, it would be useful to know whether individuals are able to communicate sex or gender through stop-consonant use.

5.7 CONCLUSION

The present research contributes to the growing field that investigates sex-specific phonetic production. Specifically, this thesis investigated how the differences in male and female stop consonant production is developed in both Mandarin and English. The findings from this research have several implications that can be used in future research.

The first major finding is that the sex difference in stop-consonant product is language specific. While Mandarin and English stop-consonants share a number of similarities, how these sounds are used by each sex is unique to the language. The findings from this study, and a retroactive investigation into previous developmental studies (see Ma et al., 2017; Yu et al., 2014) seems to suggest that the sex difference is much more robust in Mandarin than in English. These findings suggest that each language may have a unique system for communicating sex or gender to speakers, with the specific use of individual sounds for each language.

The second major finding is that sex-specific VOT may be mediated by speech rate differences. While previous studies have found that speech rate and VOT are related, this study found that much of the differences between male and female speech may be accounted for with differences in speaking rate. This sheds new light on sex-specific speech, as speaking rate could potentially affect a number of speech sounds. While speech rate differences between males and females have been observed, there is currently little research into the origin or development of this difference. Studies similar to the present study could prove useful in studying this difference in speech rate.

The third major finding from this research is that stop-consonants may not be a reliable sex-specific sound. Other sounds, such as vowels and /s/ have been shown to have a robust sex effect, and with studies suggesting that information regarding speaker sex is easily communicated to the listener. Comparably, the sex difference in stop-consonants is at best modest. This suggests that certain sounds may be more suited for gendered speech than others. However, the qualities that make some sounds more suitable for gendered speech are currently not well studied.

While there are still many puzzles about sex-specific speech that will need to be researched, this present study gives some insight into how sex-specific speech develops. This line of research is useful in understanding the differences in how males and females use language to communicate, even from an early age. It has potential for application in both speech therapy and voice recognition programs. While stop consonants do not seem to strongly signal speaker sex, this study serves as a potential model for future research in gendered speech. Finally, this research underscores the nuanced ways in which sex interacts with speech throughout one's lifetime.

REFERENCES

Bartlett, N. B., & Vasey, P. L. (1996). A retrospective study of childhood gender-atypical behavior in Samoan fa'afafine. *Archives of Sexual Behavior*, 35, 559–566.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

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

Boë, L-J., M. Contini, and H. Rakotofiringa. 1975. Étude statistique de la fréquence laryngienne. *Phonetica* 32.1–23.

Boersma, P. & Weenink, D. (2017). PRAAT. Institute of Phonetic Sciences University of Amsterdam, The Netherlands. Free software retrieved in 2017 from http://www.fon.hum.uva.nl/praat/.

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.

Chao, K. Y., & Chen, L. (2008). A cross-linguistic study of voice onset time in stop consonant productions. *Computational Linguistics and Chinese Language Processing*, 13(2), 215-232.

Cho, T., & Ladefoged, P. (1999). Variations and universals in VOT: Evidence from 17 endangered languages. *Journal of Phonetics*, 27, 207-229.

Cohen-Kettenis, P. T., Wallien, M., Johnson, L. L., Owen-Anderson, A. F. H., Bradley, S. J., & Zucker, K. J. (2006). A parent-report gender identity questionnaire for children: a cross-national, cross-clinic comparative analysis. *Clinical Child Psychology and Psychiatry*, 11(3), 397–405.

Edwards, J., & Beckman, M. E. (2008). Methodological questions in studying consonant acquisition. *Clinical linguistics & phonetics*, 22(12), 937-956

Eguchi, S. & Hirsh I.J. (1969) Development of speech sounds in children. *Acta Otolaryngologica Supplementum*. 257:307–356.

Ehrhardt, A. A., & Meyer-Bahlburg, H. F. (1981). Effects of prenatal sex hormones on gender-related behavior. *Science*, 211(4488), 1312-1318.

Ford, C. E., & Mori, J. (1994). Causal markers in Japanese and English conversations: A cross-linguistic study of interactional grammar. *Pragmatics*, 4(1), 31-61.

Fox, R., & Nissen, S. (2005). Sex-Related Acoustic Changes in Voiceless English Fricatives. *Journal Of Speech Language And Hearing Research*, 48(4), 753

Feldstein, S., Dohm, F. A., & Crown, C. L. (1993). Gender as a mediator in the perception of speech rate. *Bulletin of the Psychonomic Society*, 31(6), 521-524.

Fuchs, S., & Toda, M. (2010). Do differences in male versus female /s/ reflect biological factors or sociophonetic ones? *In S. Fuchs, Toda, M. & Zygis, M. (Ed.), An interdisciplinary guide to turbulent sounds*. Berlin: Mouton de Gruyter.

Goodwin, R. D., & Gotlib, I. H. (2004). Gender differences in depression: the role of personality factors. *Psychiatry Research*, 126, 135–142.

Hasek, C. S., Singh, S., & Murry, T. (1980). Acoustic attributes of preadolescent voices. *The Journal of the Acoustical Society of America*, 68(5), 1262-1265.

Hazan, V., Romeo, R., & Pettinato, M. (2013). The impact of variation in phoneme category structure on consonant intelligibility. *Proceedings of Meetings on Acoustics*, 060103–060103.

Heffernan, K. (2004) Evidence from HNR that /s/ is a social marker of gender. *Toronto working papers in Linguistics* 23(2). 71–84.

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., 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.

Hitchcock, E. R., & Koenig, L. L. (2013). The Effects of Data Reduction in Determining the Schedule of Voicing Acquisition in Young Children. Journal of Speech, *Language, and Hearing Research*, 56(2), 441-457.

IBM Corp. Released 2016. IBM SPSS Statistics for MAC, Version 24.0. Armonk, NY: IBM Corp.

Jiang, H. (2011). Gender Difference in English Intonation. In ICPhS (pp. 974-977).

Johnson, K, (1991) Differential effects of speaker and vowel variability on fricative perception. *Language and Speech*, 34(3), 265–279.

Johnson, L. L., Bradley, S. J., Birkenfeld-Adams, A. S., Radzins Kuksis, M. A., 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.

Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: evidence from acoustic studies. *Journal of Speech and Hearing Research*, 19(3), 421-447.

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics*, 25(2), 143-168.

Kessinger, R. H., & Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, 26(2), 117-128.

Kewley-Port, D. & Preston, M. S. (1974) Early apical stop production: A voice onset time analysis. *Journal of Phonetics*, 2,195–210.

Kinsman, M. & Li, F. (2013) The relationship between gender-differentiated productions of /s/ and gender role behaviour in young children. Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech), Lyon, France, pp.1283-1286.

Koenig, L. L. (2000). Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds. *Journal of Speech, Language, and Hearing Research*, 43(5), 1211-1228.

Kong, E. J., Beckman, M. E., & Edwards, J. (2012). Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *Journal of Phonetics*, 40(6), 725-744.

Labov, W. (1972). Sociolinguistic Patterns. Philadelphia: University of Pennsylvania Press.

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(3), 1455-1468.

Li, F. (2008). The phonetic development of voiceless sibilant fricatives in English, Japanese and Mandarin Chinese (Doctoral dissertation, The Ohio State University).

Li, F. (2012). Language-specific developmental differences in speech production: a cross-language acoustic study. *Child Development*, 83(4), 1303–1315.

Li, F. (2013). The effect of speakers' sex on voice onset time in Mandarin stops. *The Journal of the Acoustical Society of America*, 133(2), EL142–EL147.

Li, F., Rendall, D., Vasey, P., 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.

Li, F. (2017). The development of gender-specific patterns in the production of voiceless sibilant fricatives in Mandarin Chinese. *Linguistics*, 55(5), 1021-1044.

Linville, S. (1998) Acoustic correlates of perceived versus actual sexual orientation in men's speech. *Pholia Phoniatrica et Logopaedica* 50, 35–48.

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

LoMauro, A., & Aliverti, A. (2018). Sex differences in respiratory function. *Breathe*, 14(2), 131-140.

Lowenstein, J. H., & Nittrouer, S. (2008). Patterns of acquisition of native voice onset time in English-learning children. *The Journal of the Acoustical Society of America*, 124(2), 1180–1191.

Ma, J., Chen, X., Wu, Y., & Zhang, L. (2017). Effects of age and sex on voice onset time: Evidence from Mandarin voiceless stops. *Logopedics Phoniatrics Vocology*, 19, 1-7.

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

Markova, D., Richer, L., Pangelinan, M., Schwartz, D. H., Leonard, G., Perron, M., ... & Paus, T. (2016). Age-and sex-related variations in vocal-tract morphology and voice acoustics during adolescence. *Hormones and behavior*, 81, 84-96.

Morris, R. J., McCrea, C. R., & Herring, K. D. (2008). Voice onset time differences between adult males and females: Isolated syllables. *Journal of Phonetics*, 36(2), 308-317.

Mulac, A., Wiemann, J. M., Widenmann, S. J., & Gibson, T. W. (1988). Male/female language differences and effects in same-sex and mixed-sex dyads: The gender-linked language effect. *Communications Monographs*, 55(4), 315-335.

Munson, B. (2007). The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. *Language and speech*, 50(1), 125-142.

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.

Munson, B., Crocker, L., Pierrehumbert, J., Owen-Anderson, A., & Zucker, K. (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(4), 1995-2003

Murry, T., & Singh, S. (1980). Multidimensional analysis of male and female voices. *The journal of the Acoustical society of America*, 68(5), 1294-1300.

Netelenbos, N., Li, F., & Rosen, N. (2016). Stop consonant production of French immersion students in Western Canada: A study of voice onset time. *International Journal of Bilingualism*, 20(3), 346-357.

Nittrouer, S. (1993). The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. *Journal of Speech, Language, and Hearing Research*, 36(5), 959-972.

Oh, E. (2011). Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics*, 39(1), 59-67.

Pépiot, E. (2015). Voice, speech and gender:. male-female acoustic differences and cross-language variation in english and french speakers. Corela. *Cognition, représentation, langage*, (HS-16).

Perry, T., Ohde, R., & Ashmead, D. (2001). The acoustic bases for gender identification from children's voices. *The Journal Of The Acoustical Society Of America*, 109(6), 2988-2998.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24(2), 175-185.

Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J., & Feinberg, D. R. (2014). Return to Oz: Voice pitch facilitates assessments of men's body size. *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), 1316.

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology*, 18, 10-13.

R Development Core Team, (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from 〈http://www.R-project.org/〉 .

Roger, D. B., & Schumacher, A. (1983). Effects of individual differences on dyadic conversational strategies. *Journal of Personality and Social Psychology*, 45(3), 700.

Roger, D., & Nesshoever, W. (1987). Individual differences in dyadic conversational strategies: A further study. *British Journal of Social Psychology*, 26(3), 247-255.

Rogers, R. G., Everett, B. G., Onge, J. M. S., & Krueger, P. M. (2010). Social, behavioral, and biological factors, and sex differences in mortality. *Demography*, 47, 555–578.

Robb, M., Gilbert, H., & Lerman, J. (2005). Influence of gender and environmental setting on voice onset time. *Folia Phoniatrica Et Logopaedica*, 57(3), 125-133.

Rose, P. 1991. How effective are long term mean and standard deviation as normalisation parameters for tonal fundamental frequency? *Speech Communication* 10.229–247.

Rosselli, M., Ardila, A., Matute, E., & Vélez-Uribe, I. (2014). Language development across the life span: A neuropsychological/neuroimaging perspective. *Neuroscience journal*, 2014.

Ryalls, J. Zipprer, A. & Baldauff, P. A. (1997) Preliminary investigation of the effects of gender and race on voice onset time. *Journal of Speech, Language, and Hearing Research*, 40, 642-645.

Sachs, J., Lieberman, P., & Erickson, D. (1973). Anatomical and cultural determinants of male and female speech. *Language Attitudes: Current Trends And Prospects*, 74-84.

Scharf, G.& Masure, H. (2002) Voice onset time in normal speakers of a German dialect: Effects of age, gender and verbal material," in Investigations in Clinical Phonetics and Linguistics, edited by F.Windsor, M. L. Kelly, and N. Hewlett (Erlbaum Associates, Mahwah, NJ), pp. 327–339.

Shi, Y. (2014). The development of gender-differentiated/ɕ/production in Mandarin speaking children.

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

Skoog Waller, S., Eriksson, M., & Sörqvist, P. (2015). Can you hear my age? Influences of speech rate and speech spontaneity on estimation of speaker age. *Frontiers in psychology*, 6, 978.

Sweeting, P. M., & Baken, R. J. (1982). Voice onset time in a normal-aged population. *Journal of Speech, Language, and Hearing Research*, 25(1), 129-134.

Takefuta, Y., E. G. Jancosek, and M. Brunt. 1972. A statistical analysis of melody curves in the intonation of American English. In Proc. VIIth ICPhS, Montreal 1971, 1035–1039.

Simpson, A. P., & Weirich, M. (2015). Gender-specific differences in sibilant contrast realizations in English and German. In ICPhS.

Stuart-Smith, J. (2007). Empirical evidence for gendered speech production:/s/in Glaswegian.

Swartz, B. L. (1992) Gender difference in voice onset time. *Perceptual and Motor Skills*, 75, 983-992.

Trudgill, P. (1983). On dialect. *Social and geographical perspectives*. Oxford: Blackwell.

Turner, J., Netelenbos, N., Rosen, N., & Li, F. (2014). Stop consonant production by French-English bilingual children in Southern Alberta. *Canadian Acoustics*, 42(3).

Vander Laan, D. P., Petterson, L. J., & Vasey, P. L. (2015). Elevated childhood separation anxiety: an early developmental expression of heightened concern for kin in homosexual men?. *Personality and Individual Differences*, 81, 188–194.

Van Goozen, S. H., Cohen-Kettenis, P. T., Gooren, L. J., Frijda, N. H., & Van De Poll, N. E. (1995). Gender differences in behaviour: Activating effects of cross-sex hormones. *Psychoneuroendocrinology*, 20(4), 343-363.

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of phonetics*, 71, 147-161.

Wadnerkar, M. B., Cowell, P. E., & Whiteside, S. P. (2006). Speech across the menstrual cycle: A replication and extension study. *Neuroscience Letters*, 408(1), 21-24.

Whiteside, S. P., Hanson, A., & Cowell, P. E. (2004). Hormones and temporal components of speech: sex differences and effects of menstrual cyclicity on speech. *Neuroscience Letters*, 367(1), 44-47.

Whiteside, S. P., & Irving, C. J. (1997). Speakers' sex differences in voice onset time: Some preliminary findings. *Perceptual and motor skills*, 85(2), 459-463E.

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., Henry, L., & Dobbin, R. (2004). Sex differences in voice onset time: A developmental study of phonetic context effects in British English. *The Journal of the Acoustical Society of America*, 116(2), 1179-1183.

Xia, X. (2013). Gender differences in using language. *Theory and Practice in Language Studies*, 3(8), 1485.

Yu, V., De Nil, L., & Pang, E. (2014). Effects of Age, Sex and Syllable Number on Voice Onset Time: Evidence from Children's Voiceless Aspirated Stops. *Language And Speech*, 58(2), 152-167.

Zlatin, M. A., & Koenigsknecht, R. A. (1976). Development of the voicing contrast: A comparison of voice onset time in stop perception and production. *Journal of Speech, Language, and Hearing Research*, 19(1), 93-111.

APPENDIX 1: Mandarin Word Stimuli

This appendix contains a complete list of word stimuli used for Mandarin word Production. The words are denoted in Chinese characters, the Chinese Phonetic Alphabet (Pinyins), the International Phonetic Alphabet (IPA), and an English translation from top to bottom respectively. The following table is organized by leading consonant and initial vowel sound.

Consonant	Vowel /a/			/i/			/u/		
/t/	塔 tǎ [tʰǎː] “ tower ”	毯子 tǎn zi [tʰæ̌ntsz] “ blanket	糖 táng [tʰáŋ] “ sugar”	踢球 tī qiú [tʰʲīː tɕʰə́u] “ kick the ball”	梯子 tī zi [tʰʲīːtsz] “ ladder”	体育课 tǐ yù kè [tʰʲǐːyːkʰɤ̀ː] “ physical education class”	兔子 tùzǐ [tʰʷùːtsz] “ rabbit”	土豆 tǔ dòu [tʰʷǔːtə̀u] “ Potato ”	图画 tú huà [tʰʷúːxʷàː] “ picture ”
/d/	大海 dà hǎi [tàːxǎi] “ sea”	袋鼠 dài shǔ [taìʂʷǔː] “ Kangaroo”	大象 dà xiàng [tàːɕàŋ] “ Elephant”	弟弟 dì dì [tʲìːtʲì] “ little brother”	钉子 dīng zi [tʲə̄ŋtsz] “ nail”	笛子 dí zi [tʲíːtsz] “ flute”	独木桥 dú mù qiáo [tʷúːmʷùːtɕʰáu] “ wooden bridge”	肚子 dù zi [tʷùːtsz] “ belly”	读书 dú shū [tʷúː ʂʷūː] “ reading ”

APPENDIX 2: Mandarin Gendered Behaviour Questionnaire

The following appendix contains the Mandarin-language gendered behaviour form that was used to collect data for the Mandarin Corpus. The appendix contains the original Mandarin form followed by an English translation.

Original Mandarin Questionnaire

提示: 请就您孩子现在或小时候的表现，在题后括号中填入相应数字：

若" 从未如此"填"1"
若"很少如此"填"2"
若"有时如此"填"3"
若"经常如此"填"4"
若"总是如此 "填"5"
若"不适用 （从未接触此活动，游戏，或玩具）"填"

1. 喜欢玩娃娃, 例如芭比娃娃( )
2. 喜欢踢足球或打篮球 ( )
3. 喜欢过家家游戏( )
4. 喜欢玩玩具枪( )
5. 喜欢玩毛绒玩具( )
6. 喜欢玩机械类玩具，例如机器人,赛车。( )
7. 喜欢像白雪公主那样的童话故事( )
8. 喜欢玩扮演太空人或士兵的游戏( )
9. 喜欢玩扮演护士或老师类的游戏( )
10. 喜欢在游戏中搭建堡垒( )
11. 喜欢玩"丢手绢〞之类的游戏( )
12. 生活中喜欢穿女孩子气的衣服,例如裙子( )
13. 生活中喜欢男孩子气的衣服,例如短褲( )
14. 在类似过家家的扮演游戏中，喜欢穿女性的服装( )
15. 在类似过家家的扮演游戏中，喜欢穿男性的服装( )
16. 喜欢打扮和化妆( )
17. 喜欢追捕打斗的游戏( )
18. 在学校裡，喜欢和女孩一起玩( )
19. 喜欢使用工具，例如锤子，螺丝刀( )
20. 在学校里，喜欢和男孩子一起玩( )
21. 喜欢模仿电影电视里的女性角色( )

22. 喜欢汽车(　)
23. 喜欢把毛巾之类的东西围在腰上当裙子(　)
24. 喜欢模仿电影电视里的男性角色(　)
25. 喜欢编织和缝纫(　)
26. 在男孩子中很受欢迎(　)
27. 在女孩子中很受欢迎(　)
28. 喜欢和女性长辈呆在一起或做事(　)
29. 在扮演游戏中，喜欢演女孩或女人的角色。(　)
30. 擅长模仿女性(　)
31. 擅长模仿男性(　)
32. 对体育比赛感兴趣(　)
33. 喜欢打弹球(　)
34. 喜欢跳舞(　)
35. 喜欢唱歌(　)
36. 喜欢摔跤或武术(　)
37. 喜欢玩做饭的游戏(　)
38. 喜欢跳绳(　)
39. 喜欢跳皮筋(　)
40. 喜欢玩买东西的游戏(　)
41. 喜欢跳格子的游戏(　)
42. 喜欢玩橡皮泥 (　)
43. 喜欢爬树 (　)
44. 喜欢下棋 (　)
45. 喜欢打弹弓 (　)
46. 喜欢看有关太空或恐龙的书(　)

Questionnaire translated into English

INSTRUCTIONS: Please rates the frequency of the occurrence of the following behaviors exhibited by your child in the past six months, and answer the questions in a way that best describes your child.

1 = Never 2 = Seldom 3 = Frequently 4 = Very often 5= Always 0=not applicable (never have been introduced to this games/been in the situations where this type of tendency could be shown)

1. He/she plays with Barbie or similar types of dolls ( )
2. He/she plays football or basketball ( )
3. He/she plays house ( )
4. He/she plays toy guns ( )
5. He/she plays with stuffed animals ( )
6. He/she plays with mechanical or machine-like toys, like robots ( )
7. He/she likes fairy tales like Snow White ( )
8. He/she plays as spaceman or soldier ( )
9. He/she plays teacher or nurse games ( )
10. He/she likes building fort in games ( )
11. He/she likes to play games like "Puts a handkerchief behind you" a ( )
12. He/she likes to dress in female clothing, such as skirts ( )
13. He/she likes to dress in male clothing, such as shorts ( )
14. In dress-up game , he/she likes to dress-up in women's clothing ( )
15. In dress-up game , he/she likes to dress-up in men's clothing ( )
16. He/she lies dressing up and make up ( )
17. He/she plays rough-and-tumble games ( )
18. At school, he/she plays with boys ( )
19. At school, he/she plays with girls ( )
20. He/she likes to use tools, such as hammers, screw drives, etc. ( )
21. At school, he/she plays with boys ( )
22. He/she imitates female characters seen on T.V. or in the movies ( )
23. He/she likes real automobile ( )
24. He/she wears things like towels around waist as a skirt ( )
25. He/she imitates male characters seen on T.V. or in the movies ( )
26. He/she likes knitting or sewing ( )
27. He/she is popular among boys ( )
28. He/she is popular among girls ( )
29. He/she prefers staying with female relatives ( )
30. In playing "mother/father", "house" or "school" games, he/she takes the role of a girl or woman ( )
31. He/she is good at imitating females ( )
32. He/she is good at imitating males ( )
33. He/she is interested in sports competition ( )
34. He/she likes singing ( )

35. He/she likes dancing ( )
36. He/she plays wrestling or Wushu ( )
37. He/she plays Cooking games ( )
38. He/she plays jump-rope ( )
39. He/she plays jump-rubber band ( )
40. He/she plays hopscotch ( )
41. Hi/he plays with plasticine ( )
42. He/she plays store ( )
43. He/she climbs trees ( )
44. He/she plays with slingshots () chess ( )
45. SLING SHOT
46. He/she reads books about dinosaurs and space ( )

APPENDIX 3: English Word Stimuli

This appendix contains a complete list of word stimuli used for English word Production. The following table is organized by leading consonant and initial vowel sound.

	Vowel		
Consonant	/ae/	/i/	/u/
b	battle, bat, back	beetle, beer, bee	boomerang, boot, boombox
d	daffodil, dam, dad	deal, deep, deer	duke, dude, doodle
g	gallop, galaxy, gas	geese, geek, gear	goo, goofy, goose
p	panda, paddle, Pacman	peacock, peel, peak	poodle, pooh, pool
t	tattoo, tadpole, tap	tea, teeth, tear	toothpaste, tools, two
k	can, cat, castle	key, kiwi, keyboard	cougar, cuckoo, cooler

APPENDIX 4: English Gendered Behaviour Questionnaire

The following appendix contains the English-language gendered behaviour form that was used to collect data for the English Corpus.

GENDER IDENTITY QUESTIONNAIRE

Instructions: Please answer the following behavioral statements as they currently characterize your child's **childhood** behaviour. For each question, circle the response that most accurately describes your child.

1. Favorite playmates are
a. always boys
b. usually boys
c. boys and girls equally
d. usually girls
e. always girls
f. does not play with other children

2. Plays with girl-type dolls, such as "Barbie"
a. as a favorite toy
b. frequently
c. once-in-a-while
d. rarely
e. never

3. Plays with boy-type dolls, such as "G.I. Joe" or "Ken"
a. as a favorite toy
b. frequently
c. once-in-a-while
d. rarely
e. never

4. Experiments with cosmetics (makeup) and jewelry
a. as a favorite activity
b. frequently
c. once-in-a-while
d. rarely
e. never

5. Imitates female characters seen on TV or in the movies
a. as a favorite activity
b. frequently
c. once-in-a-while
d. rarely
e. never

6. Imitates male characters seen on TV or in the movies
a. as a favorite activity
b. frequently
c. once-in-a-while
d. rarely
e. never

7. Plays sports with boys
a. as a favorite activity
b. frequently
c. once-in-a-while
d. rarely
e. never

8. Plays sports with girls
a. as a favorite activity
b. frequently
c. once-in-a-while
d. rarely
e. never

9. In playing "mother/father," "house," or "school" games, your child takes the role of
a. a girl or woman at all times
b. usually a girl or woman
c. half the time a girl or woman and half the time a boy or man
d. usually a boy or man
e. a boy or man at all times
f. does not play these games

10. Plays "girl-type" games such as "princess"
a. as a favorite activity
b. frequently
c. once-in-a-while
d. rarely
e. never

11. Plays "boy-type" games such as "weapons"
a. as a favorite activity
b. frequently
c. once-in-a-while
d. rarely
e. never

12. In dress-up games, your child likes to dress up as
a. a girl or woman at all times
b. usually a girl or woman
c. half the time a girl or woman and half the time a boy or man
d. usually a boy or man
e. a boy or man at all times
f. does not play these games