Listeners' Ability to Identify the Gender of Preadolescent Children Across Multiple Linguistic Contexts

Sharalee Ann Blunck
Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/etd
Part of the Communication Sciences and Disorders Commons

BYU ScholarsArchive Citation
Blunck, Sharalee Ann, "Listeners' Ability to Identify the Gender of Preadolescent Children Across Multiple Linguistic Contexts" (2011). All Theses and Dissertations. 2613.
https://scholarsarchive.byu.edu/etd/2613

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Listeners’ Ability to Identify the Gender of Preadolescent Children across Multiple Linguistic Contexts

Sharalee A. Blunck

A thesis submitted to the faculty of Brigham Young University in partial fulfillment of the requirements for the degree of Master of Science

Shawn L. Nissen, Chair
Ron W. Channell
Christopher Dromey

Department of Communication Disorders Brigham Young University
June 2011

Copyright © 2011 Sharalee A. Blunck All Rights Reserved
ABSTRACT

Listeners’ Ability to Identify the Gender of Preadolescent Children across Multiple Linguistic Contexts

Sharalee A. Blunck
Department of Communication Disorders
Master of Science

The purpose of this study was to determine whether 20 listeners could identify the gender of 10 preadolescent children from speech samples. An additional aim was to evaluate whether listeners identified children more accurately when listening to speech samples when more linguistic context was available. The listeners were presented with a total of 190 speech samples in four different categories of linguistic context: segments, words, sentences, and discourse. The listeners were instructed to listen to each speech sample and decide whether the speaker was a male or female. In addition, the listeners were instructed to rate their level of confidence in their decision on a 1-10 scale. Results showed listeners identified the gender of the speakers with a high degree of accuracy, ranging from 86% to 95%. In addition, statistical analysis showed significant differences in the accuracy of listener judgments among the four levels of linguistic context, with segments having the lowest (83%) and discourse the highest accuracy (99%). At the segmental level, the listeners’ ability to identify the each speaker’s gender from a speech sample was greater for vowels than for fricatives, with both types of phoneme being identified at a rate well above chance. Significant differences in identification were found between the /s/ and /ʃ/ fricatives, but not between the four corner vowels. The perception of gender is likely multifactorial, with listeners possibly using phonetic, prosodic, or stylistic speech cues to determine a speaker’s gender.

Keywords: gender, dimorphism, speech perception, segmental, suprasegmental
ACKNOWLEDGEMENTS

I would like to thank my thesis chair, Dr. Nissen, for your patience, understanding, and optimism. Thank you for the time you spent reviewing the many drafts of my thesis, making suggestions and corrections, and providing feedback. I appreciate your encouragement throughout the entire process. I would also like to thank my committee members, Dr. Dromey and Dr. Channell, for your time, suggestions, and insights.

I would like to thank my entire cohort—Lacey, Maggie, Aersta, Chelsea, Keri, Jen, Brandon, and Matt—for your friendship and support. I will truly miss the friendship and bond we shared throughout graduate school.

My sincere gratitude goes to my parents who funded my seven years of education at BYU. Thank you for always being there, for encouraging me, and for cheering me on through my many interests and activities in college. I would also like to thank my wonderful siblings and in-laws for your love and support.

My deepest gratitude and love goes to my wonderful husband Stephen. Thank you for willingly and patiently listening to me talk about my thesis for an entire year. Thank you for encouraging me every step of the way along the difficult path of graduate school. You are my greatest supporter and very best friend.
Table of Contents

Table of Contents ........................................................................................................ iv
List of Tables .................................................................................................................. v
Introduction .................................................................................................................... 1
Method .............................................................................................................................. 8
Participants ..................................................................................................................... 8
Stimuli ............................................................................................................................. 9
Procedures ..................................................................................................................... 11
Statistical Analysis ....................................................................................................... 12
Results ............................................................................................................................ 12
Discussion ..................................................................................................................... 18
References .................................................................................................................... 24
Annotated Bibliography ............................................................................................... 27
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Gender Identification by Listener</td>
<td>13</td>
</tr>
<tr>
<td>2. Gender Identification by Level of Context</td>
<td>14</td>
</tr>
<tr>
<td>3. Gender Identification by Speaker</td>
<td>15</td>
</tr>
<tr>
<td>4. Gender Identification of Same and Opposite Speaker Gender</td>
<td>15</td>
</tr>
<tr>
<td>5. Gender Identification at the Segmental Level</td>
<td>16</td>
</tr>
<tr>
<td>6. Gender Identification at the Word Level</td>
<td>17</td>
</tr>
<tr>
<td>7. Gender Identification at the Sentence Level</td>
<td>17</td>
</tr>
<tr>
<td>8. Average F0 and % Accuracy for Each Speaker</td>
<td>18</td>
</tr>
</tbody>
</table>
Introduction

Listeners’ ability to distinguish between male and female speakers with a relatively high degree of accuracy has been well established. However, the factors that enable listeners to perceptually identify a speaker’s gender are not well understood, especially in younger speakers. Some researchers believe the mechanisms that facilitate speaker gender identification may be acoustic-phonetic differences present in relatively short segments of speech (Bennet & Weinberg, 1979; Perry, Ohde, & Ashmead, 2001; Whiteside & Marshall, 2001), while others suggest that male and female speakers differ in their use of prosodic aspects of speech that extend over longer periods of speech (Andrews & Schmidt, 1997; Fitzsimons, Sheahan, & Staunton, 2001). The aim of this study is to investigate the perceptual cues by which listeners identify the gender of young speakers by examining their ability to identify a speaker’s gender across multiple levels of linguistic context.

Evidence of perceptible gender-related speech differences has been found in studies where listeners identified the gender of the speaker from a short segment of speech. Lass et al. (1976) found that adult listeners were able to correctly identify the gender of adult speakers 96% of the time from an isolated vowel. In addition, Lee, Dutton, and Ram (2010) showed that adult listeners correctly identified the gender of adult females with 96% accuracy and adult males with 82% accuracy from a fricative-vowel sound combination.

Many of the perceptual differences between the speech of men and women can be explained by differences in vocal tract anatomy. Recent magnetic resonance imaging (MRI) studies have reported that in general men have a longer vocal tract (18 cm) than women (15 cm), resulting in relatively lower vowel formant frequency patterns for male speakers (Vorperian et al., 2009). Another difference between adult male and female vocal tract anatomy is vocal fold length. The average vocal fold length is approximately 16 mm in men and 10 mm in women.
Men also typically have greater vocal fold mass and thickness. The differences in vocal fold length and tissue mass generally result in a lower average fundamental frequency (F0) in male speakers (Seikel, King, & Drumright, 2005; Titze, 1989). Sexual dimorphism of vocal tract structures typically begins to appear around 12 years of age (Vorperian et al., 2009). In addition, studies have shown that variations between men and women in laryngeal shape may contribute to differences in speech acoustics (Titze, 1989).

Although perceptually salient gender-related differences in adult speech are largely due to anatomy, research has shown that suprasegmental characteristics (such as such as speaking rate, rhythm, intonation, and stress) also play a significant role in the speech of males and females. Linguistic prosody assists listeners in correctly interpreting the speaker’s intended message by specifying the sentence type, clarifying lexical ambiguities, and signaling the end of a topic of conversation. Affective prosody conveys emotions and attitudes (Lynes, 2005). For example, happy expressions are typically produced with a higher and more variable F0, and sad expressions are usually produced with a lower and less variable F0, as well as with a decreased rate of articulation (Banse & Scherer, 1996). Prosody also enables speakers to place emphasis on the most important part of their utterance in order to direct listeners towards a particular interpretation (Blasko & Hall, 1998).

Suprasegmental differences between men and women have been examined by several researchers. Fitzsimons et al. (2001) found significant differences in adult male and female speech in F0 range and F0 slope. Similarly, other authors found that male speakers use a smaller F0 range than female speakers (Whiteside, 1996; Wu & Childers, 1991). Whiteside (1996) also found females tended to pause more than their male counterparts. In the same study, the author found males tended to centralize vowels and reduce the duration of consonants, possibly
contributing to an overall quicker speaking rate in male speakers. Other researchers have also found that adult male speakers tended to produce significantly shorter utterances and spoke at a faster rate than adult female speakers (Fitzsimons et al., 2001).

Prosody has also been examined in studies of transgendered individuals. These studies indicate the important role of prosody in perceptually distinguishing between male and female voices. Andrews and Schmidt (1997) found that a biologically male participant who was perceived by listeners to produce the greatest perceptual difference between his male and female voice also had the lowest F0. The participant did, however, also demonstrate considerable variability in speech rate, use of pause, duration of vowels, and length of time to read a passage. The authors concluded that high pitch, breathiness, and animation were associated with a feminine voice. Animation included variability in loudness, rate, duration, and intonation contours. Wolfe et al. (1990) found that transgendered participants who were categorized by listeners as females exhibited greater intonation variability in addition to higher F0s than participants categorized as males. More specifically, participants categorized as females had a higher percentage of upward intonations (frequency changes without interruption of phonation) and downward frequency shifts (changes in frequency that took place between the terminal frequency of a given phonation and the initial frequency of a subsequent phonation). In addition, these participants had less extensive downward intonations and a smaller percentage of level intonations and level shifts (frequency changes or shifts less than 5 Hz).

Perceptual experiments have indicated that listeners can identify the gender of children with a relatively high degree of accuracy based on their speech. A study by Bennet and Weinberg (1979) found that adult listeners were able to correctly identify the gender of 6- and 7-year-old children 68% of the time. Additional studies have reported that adult listeners can
correctly identify the gender of speakers as young as 4 years of age with as high as 81% accuracy (Perry et al., 2001; Sachs, Lieberman, & Erickson, 1973).

Many researchers have concluded that speech production differences in younger speakers cannot be solely explained by dimorphism of the vocal tract structures (Fox & Nissen, 2005; Nissen & Fox, 2005; Sachs et al., 1973; Whiteside & Marshall, 2001). A study by Fitch and Giedd (1999) using MRI showed no significant sex-related differences in vocal tract morphology until children reached peripubertal or pubertal stages of development, which for many children occur around 12 years of age. Research by Lieberman et al. (2001) found very limited sexual dimorphism in pharyngeal shape and size before 13.75 years of age. In a study by Vorperian et al. (2005), MRI examinations showed no sexual dimorphism of the vocal tract in children between birth and age 6 years 9 months. A recently conducted large-scale MRI study by Vorperian et al. (2009) examined the oral and pharyngeal structures of the vocal tract in 605 children. Results of this study found that prior to puberty children have ongoing, and at times, accelerated periods of vocal tract growth, although children younger than 12 years of age generally tend to have no appreciable sexual dimorphism in the vocal tract structures.

Although most researchers claim sexual dimorphism of the vocal tract does not occur until the initial stages of puberty, some researchers have hypothesized slight anatomical differences in younger children. Hasek, Singh, and Murry (1980) suggested possible differences in laryngeal size and thyroid angle among male and female children as early as 3 years of age. Vorperian et al. (2005) also hypothesized that even though significant sexual dimorphism was yet to occur in prepubescent children, boys may have slightly larger vocal tracts at an early age. Minor sex-related variations in vocal tract anatomy may be present in younger children;
however, these slight and debatable differences do not appear to fully account for the perceptual distinctions found in the speech communication of preadolescent boys and girls.

In view of these anatomical data, it can be reasoned that gender-related perceptual differences in the speech of younger children are in part the result of learned sociophonetic factors. Sachs et al. (1973) suggested that perceptual differences in children’s speech may be due to formant frequency and intonation patterns that adhere to culturally determined male and female archetypes. Fitch and Giedd (1999) concluded that because they found no differences in vocal tract length among preadolescent males and females, acoustic differentiation in speech patterns is likely the result of behavioral rather than anatomical factors. In particular, young males may protrude their lips more when speaking, thus behaviorally creating a longer vocal tract, which results in lower formant frequencies. Sachs et al. (1973) also suggested that perceptual distinctions may be caused by variation in lip configuration during articulation, which results in allophonic differences in phonemes. Bennet (1980) suggested that male participants used greater lip rounding, smaller jaw openings, and/or a lower larynx position than female participants.

Some researchers have attributed differences in preadolescent children’s voices to a combination of both anatomical differences and behavioral characteristics (Bennet, 1980; Vorperian et al., 2005; Whiteside, 2001). Hasek et al. (1980) attributed perceptual differences to a combination of anatomy and cultural stereotypes that might cause young males to speak with a lower F0 than what would be expected considering their vocal tract anatomy.

From an acoustic viewpoint, researchers have attempted to characterize the voices of preadolescent males and females by examining segmental differences in their speech. Multiple studies have analyzed vowel production of children and found notable differences in formant
frequencies among preadolescent males and females. One study found that formant frequencies of male children were consistently lower than formant frequencies of female children, and the extent to which formant frequencies differed as a function of gender depended on the formant number and vowel category (Bennet, 1980). Other studies have found similar results (Busby & Plant, 1995; Sachs et al., 1973; Whiteside & Hodgson, 2000a, 2000b). Another study looked at voice onset time and discovered gender differences in the voice onset time patterns in the speech of preadolescent males and females. These differences were associated with age and development, the extent of which varied depending upon the type of plosive examined (Whiteside & Marshall, 2001). Fox and Nissen (2005) evaluated voiceless fricatives in children 6 to 14 years of age and found that preadolescent children exhibited gender-specific differences in fricative articulation, though the acoustic differences were reduced or absent in the 6- to 7-year-old children. In another study the researchers examined the acoustic patterns of voiceless stop consonants in children 3 to 5 years of age. Results showed significant gender differences for the spectral measures of slope, mean, and skewness in the 5-year-old children (Nissen & Fox, 2009).

Although the suprasegmental aspect of adult male and female speech has been examined by a number of studies, less emphasis has been placed on understanding possible gender-related differences in suprasegmental aspects of children’s speech. Ferrand and Bloom (1996) found that males between the ages of 7 and 8 began to restrict their intonational ranges. This decrease in rising and falling shifts was not found in females of the same age. The authors attributed these intonational differences between the genders to sociocultural factors. Sachs et al. (1973) suggested that boys and girls often use different vocabulary and speak with different intonation patterns.
Although there are many studies that have documented gender-related speech production differences in children (Fox & Nissen, 2005; Nissen & Fox, 2005; Sachs et al., 1973; Whiteside & Marshall, 2001), the perceptual speech cues used by listeners to identify the gender of younger speakers remain unclear, especially considering the limited research that has been conducted on how listeners might use prosody in such decisions (Sachs et al., 1973). One way to provide some insight as to the origin of the perceptual differences in the speech of children is to present various levels of linguistic context to listeners (i.e., sound segments, words, sentences, and discourse) and instruct listeners to identify the stimuli as male or female. It is hypothesized that if listeners are able to judge accurately a majority of the time at the segment level, it can be suggested that formant frequencies are the basis for such judgments. On the other hand, if listeners are able to judge accurately much more often at the sentence or discourse level, it can be suggested that suprasegmental aspects of speech, such as intonation, pause, stress, rate, or rhythm, may play an important role in speaker gender identification.

Historically, most studies have only examined speech differences across a single level of linguistic context. Multiple studies, both acoustic and perceptual, have analyzed isolated phonemes, such as vowels (Bennet, 1980; Busby & Plant, 1995; Childers & Wu, 1991; Hasek et al., 1980; Lass et al., 1976; Perry et al., 2001; Whiteside, 2001), fricatives (Fox & Nissen, 2005), or stops (Nissen & Fox, 2009; Whiteside & Marshall, 2001). Several studies have investigated speech production differences within fricative-vowel combinations (Lee et al., 2010) or short phrases and sentences (Klatt & Klatt, 1990; Whiteside & Hodgson, 2000). However, only a limited number of studies have evaluated gender-related differences in children’s speech across more than a single level of context (Lee, Potamianos, & Narayanan, 1999; Sachs et al., 1973). Thus, in this study children’s speech will be evaluated across four different levels of linguistic
context (sound segment, word, sentence, and discourse). Three research questions will be addressed in this study. First, can listeners accurately identify the gender of a speaker from a speech sample? Second, does the accuracy of the listener’s identifications become more accurate with greater amounts of linguistic context? Third, does the listener’s ability to identify the speaker’s gender from a single sound segment improve depending on the phoneme being spoken?

Method

Participants

Speakers. The perceptual stimuli used in this study were collected as part of a previous investigation into the prosodic production abilities of typically developing children (Dromey, 2010). The speech samples were collected from 20 children between 8:0 and 9:11 years of age (M = 9:2), with an equal number of male and female participants. A pilot experiment from the study by Dromey rank-ordered the child speakers according to how well a small number of listeners could identify their gender based on a conversational sample. The current study used recordings from the 10 child speakers that showed the highest degree of identification. All participants were monolingual speakers of American English and had minimal exposure to a second language (i.e., participants had not lived outside of the United States for more than six months and had parents or guardians who also spoke English as their native language). The parents or guardians of the children reported that the participants had no diagnosed history of speech, language, or hearing problems at the time the speech samples were collected. All the participants were required to pass a hearing screening, exhibiting pure-tone air-conduction thresholds of ≤ 25 dB HL at octave frequencies from 500 to 8000 Hz. If participants exhibited perceptual signs of poor vocal health (e.g., cold, laryngitis, vocal hoarseness, etc.), the recording
sessions were postponed until a later date. The participants were recruited from the Brigham Young University community and surrounding areas. Approval to conduct the study was obtained from the Brigham Young University Institutional Review Board (IRB) and consent of the participants’ legal guardians was obtained prior to the collection of any data.

**Listeners.** In the present study, a group of 20 adult listeners recruited from the Brigham Young University community were asked to identify the gender of the speaker from the speech stimuli presented to them. The listeners were native speakers of American English and reported no history of speech or language disorders. At the time of their participation all of the listeners exhibited pure-tone air-conduction thresholds of \( \leq 25 \text{ dB HL} \) at octave frequencies from 500 to 8000 Hz. A consent form was signed by each individual prior to participation in the study.

**Stimuli**

The listeners were presented with 190 different speech samples, consisting of 60 sound segments (10 speakers x 6 phonemes), 60 words (10 speakers x 6 words), 60 short sentences (10 speakers x 6 sentences), and 10 sections of discourse. The speech recordings were elicited by asking the speakers to read a list of words, by having the speakers participate in a picture description task, and by engaging the speakers in conversation. A more detailed description of the linguistic structure of the recordings and how the speech samples were elicited can be found in Dromey (2010).

**Sound segments.** A series of corner vowels (/i/, /æ/, /u/, /ɑ/) were extracted from citation words read by the speakers. Additionally, two sibilant voiceless fricatives (/s/ and /ʃ/) were extracted from sentences elicited through picture description. Vowel segments were selected as stimuli because previous studies have shown that preadolescent male children exhibit lower vowel formant frequencies than female children (Bennet, 1980; Busby and Plant, 1995;
Whiteside and Hodgson, 2000). Fricative sounds were selected as stimuli because studies have found significant gender-related differences in young children’s productions of these types of consonants in terms of spectral slope, mean, and skewness (Nissen & Fox, 2005, 2009). Using Adobe Audition (Adobe Systems Incorporated, 2003), the onset and offset of the vowel sounds were segmented by analyzing the periodic energy as seen through waveform and spectrograph display. The onset of the fricative consonants was characterized by a sharp increase in diffuse noise energy and the rapid increase in zero crossings. The end of the fricative consonants was defined by a sharp decrease in diffuse noise energy. All sound segments were extracted from the same linguistic context across all speakers. All the stimuli were cut at a zero-crossing and ramped over the initial and ending 10 ms of the waveform. In addition, all files were high-pass filtered to eliminate any extraneous noise below 65 Hz. To make the tokens relatively equivalent with regard to intensity, the average RMS of each token was measured and digitally adjusted to a standard level, taking care not to exceed peak recording levels. As a final step, all the stimuli were listened to by the experimenter and any sudden changes in loudness were attenuated and any extraneous noises were edited out of the stimuli.

**Words.** Using Adobe Audition, the following stimulus words were extracted from the speaker recordings: *horse, pool, fence, jumping, pizza,* and *swimming.* The first three words were monosyllabic and the second three words were bisyllabic. Similar to the sound segments described above, the stimulus words were segmented at a zero-crossing, ramped, and equalized with regard to relative intensity.

**Sentences.** The following sentences were selected from the speaker recordings and used as perceptual stimuli:

*The boy is swimming in the pool.*
The horse is jumping the fence.
The girl is mowing the lawn.
The boy is eating the pizza.
The lady is picking the flower.
The boy is baking the cookies.

**Discourse.** To present listeners with a greater amount of linguistic context than the sentence level provides, a 1-minute conversation sample was selected from the speaker recordings.

**Procedures**

The listeners were asked to listen to each stimulus item and decide whether the speaker was a boy or a girl. In addition, the listeners were instructed to rate their level of confidence on a labeled scoring sheet for each perceptual decision using a scale from 1 - 10 (not confident at all - completely confident). The stimulus items were presented to the listeners in groups according to the level of linguistic context (e.g., sound segments, words, etc.). The order of presentation for each stimulus group, as well as the presentation of individual items within each group was randomized.

The signal was routed from a computer hard drive to the listener via Sennheisser HD 650 headphones. The listener was seated in a single-walled sound booth meeting ANSI S3.1 standards with ears covered (American National Standards Institute, 1999). Each participant listened to a sample token before rating the experimental stimuli prior to data collection. The listeners self-selected the intensity level of the presented stimuli, with a starting level of approximately 60 dB HL. The testing took place during one forty-minute session. The
randomization, stimulus presentation, and subsequent recording of the listeners’ perceptual judgments were controlled by custom software developed by Dr. Richard Harris.

**Statistical Analysis**

Prior to statistical analysis an arc sine transform was used to convert the percentage data. Analyses of variance (ANOVA) were used to determine significant variation in the listeners’ ability to accurately perceive the speakers’ gender as a function of the amount of linguistic context and the segment type. Results of significant $F$-tests included a measure of effect size, in particular partial eta squared or $\eta^2$ (the value of $\eta^2$ can range from 0.0 to 1.0, and can be considered a measure of the proportion of variance explained by a dependent variable when controlling for other factors). Greenhouse-Geisser adjustments were used to adjust the $F$-test degrees of freedom when significant deviations from sphericity were found. Furthermore, pairwise comparisons for significant within-subject factors were conducted. In order to test intra-rater reliability, 20% of the samples were randomly selected and replayed to the listeners a second time. The overall intra-rater reliability was found to be 89.90%, with a correlation of $r = .80$, $p < .001$.

**Results**

Results showed that listeners can identify the gender of a speaker from a speech sample with a relatively high degree of accuracy. Across all levels of context, the overall accuracy with which listeners identified stimuli was found to be 90.66%. Individual listener percentages ranged from 86.32% to 94.74% accuracy, with an average confidence level of 7.26. Please refer to Table 1 for a detailed listing of the identification percentages and confidence ratings as a function of each listener.
A repeated measures ANOVA was conducted to evaluate the differences in accuracy among the four levels of linguistic context. The analysis, $F(3, 57) = 95.321, p < .001, \eta^2 = .83$, demonstrated statistically significant differences between the four levels of context. Pairwise

Table 1

*Gender Identification by Listener*

<table>
<thead>
<tr>
<th>Listener</th>
<th>Gender</th>
<th>Number Stimuli</th>
<th>% Accuracy</th>
<th>Mean Confidence</th>
<th>SD of Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male</td>
<td>190</td>
<td>93.68</td>
<td>7.64</td>
<td>2.58</td>
</tr>
<tr>
<td>2</td>
<td>Male</td>
<td>190</td>
<td>90.00</td>
<td>7.19</td>
<td>2.20</td>
</tr>
<tr>
<td>3</td>
<td>Female</td>
<td>190</td>
<td>87.89</td>
<td>7.98</td>
<td>2.84</td>
</tr>
<tr>
<td>4</td>
<td>Female</td>
<td>190</td>
<td>92.63</td>
<td>6.76</td>
<td>3.37</td>
</tr>
<tr>
<td>5</td>
<td>Female</td>
<td>190</td>
<td>89.47</td>
<td>6.56</td>
<td>3.02</td>
</tr>
<tr>
<td>7</td>
<td>Female</td>
<td>190</td>
<td>89.47</td>
<td>6.74</td>
<td>3.17</td>
</tr>
<tr>
<td>8</td>
<td>Male</td>
<td>190</td>
<td>91.05</td>
<td>7.79</td>
<td>1.60</td>
</tr>
<tr>
<td>9</td>
<td>Male</td>
<td>190</td>
<td>91.58</td>
<td>5.70</td>
<td>2.95</td>
</tr>
<tr>
<td>10</td>
<td>Female</td>
<td>190</td>
<td>92.63</td>
<td>8.39</td>
<td>2.30</td>
</tr>
<tr>
<td>11</td>
<td>Male</td>
<td>190</td>
<td>86.32</td>
<td>7.65</td>
<td>2.59</td>
</tr>
<tr>
<td>12</td>
<td>Male</td>
<td>190</td>
<td>89.47</td>
<td>7.37</td>
<td>1.53</td>
</tr>
<tr>
<td>13</td>
<td>Female</td>
<td>190</td>
<td>93.16</td>
<td>8.12</td>
<td>2.37</td>
</tr>
<tr>
<td>15</td>
<td>Male</td>
<td>190</td>
<td>89.47</td>
<td>6.29</td>
<td>2.72</td>
</tr>
<tr>
<td>16</td>
<td>Female</td>
<td>190</td>
<td>87.37</td>
<td>6.44</td>
<td>2.79</td>
</tr>
<tr>
<td>17</td>
<td>Female</td>
<td>190</td>
<td>94.21</td>
<td>6.71</td>
<td>2.39</td>
</tr>
<tr>
<td>18</td>
<td>Female</td>
<td>190</td>
<td>88.95</td>
<td>7.22</td>
<td>3.19</td>
</tr>
<tr>
<td>19</td>
<td>Female</td>
<td>190</td>
<td>88.42</td>
<td>6.68</td>
<td>2.54</td>
</tr>
<tr>
<td>20</td>
<td>Female</td>
<td>190</td>
<td>92.11</td>
<td>7.53</td>
<td>2.38</td>
</tr>
<tr>
<td>21</td>
<td>Male</td>
<td>190</td>
<td>94.74</td>
<td>8.40</td>
<td>2.15</td>
</tr>
<tr>
<td>22</td>
<td>Female</td>
<td>190</td>
<td>90.53</td>
<td>8.12</td>
<td>2.16</td>
</tr>
</tbody>
</table>
comparisons showed statistically significant differences between all four levels of linguistic context. The comparison between the sentence and discourse levels of context was significant at $p < .02$, while pairwise comparisons between all other levels of context were significant at $p < .001$. As the amount of linguistic context increased, the accuracy of the listeners’ identifications also increased. Percentages ranged from 83.42% at the segmental level to 99.00% at the discourse level. Refer to Table 2 for a detailed listing of the identification percentages and confidence ratings as a function of level of context.

Table 2

*Gender Identification by Level of Context*

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Number Stimuli</th>
<th>% Accuracy</th>
<th>Mean Confidence</th>
<th>SD of Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segments</td>
<td>1200</td>
<td>83.42</td>
<td>5.23</td>
<td>2.66</td>
</tr>
<tr>
<td>Words</td>
<td>1200</td>
<td>89.83</td>
<td>7.80</td>
<td>2.28</td>
</tr>
<tr>
<td>Sentences</td>
<td>1200</td>
<td>97.33</td>
<td>8.45</td>
<td>1.94</td>
</tr>
<tr>
<td>Discourse</td>
<td>200</td>
<td>99.00</td>
<td>9.15</td>
<td>1.48</td>
</tr>
</tbody>
</table>

The accuracy with which listeners identified the gender of individual speakers ranged from a low of 82.37% for speaker 7 to a high of 98.68% for speaker 6, with a SD of 5.49%. See Table 3 for a detailed listing of the identification percentages and confidence ratings as a function of each speaker.

Data also showed that male speakers were identified with a slightly higher, yet non-significant, degree of accuracy than female speakers, with average accuracies of 91.84% and 89.47%, respectively. Male speakers were also identified with a slightly higher degree of listener confidence, with average confidence levels at 7.42 for males and 6.91 for females.
Table 3

*Gender Identification by Speaker*

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Gender</th>
<th>Number Stimuli</th>
<th>% Accuracy</th>
<th>Mean Confidence</th>
<th>SD of Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Female</td>
<td>380</td>
<td>85.26</td>
<td>6.75</td>
<td>2.58</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
<td>380</td>
<td>95.79</td>
<td>6.39</td>
<td>2.60</td>
</tr>
<tr>
<td>3</td>
<td>Female</td>
<td>380</td>
<td>90.53</td>
<td>7.33</td>
<td>2.56</td>
</tr>
<tr>
<td>4</td>
<td>Female</td>
<td>380</td>
<td>88.42</td>
<td>7.27</td>
<td>2.80</td>
</tr>
<tr>
<td>5</td>
<td>Female</td>
<td>380</td>
<td>87.37</td>
<td>6.82</td>
<td>2.72</td>
</tr>
<tr>
<td>6</td>
<td>Male</td>
<td>380</td>
<td>98.68</td>
<td>8.32</td>
<td>2.36</td>
</tr>
<tr>
<td>7</td>
<td>Male</td>
<td>380</td>
<td>82.37</td>
<td>6.55</td>
<td>2.86</td>
</tr>
<tr>
<td>8</td>
<td>Male</td>
<td>380</td>
<td>91.58</td>
<td>7.34</td>
<td>2.55</td>
</tr>
<tr>
<td>9</td>
<td>Male</td>
<td>380</td>
<td>98.42</td>
<td>8.16</td>
<td>2.33</td>
</tr>
<tr>
<td>10</td>
<td>Male</td>
<td>380</td>
<td>88.16</td>
<td>6.71</td>
<td>2.83</td>
</tr>
</tbody>
</table>

The perceptual data were analyzed to determine whether listeners identified speakers of the same gender with a greater or lesser degree of accuracy. Results indicated that female listeners identified female speakers with a significantly lower degree of accuracy than male speakers at a rate of 88.60% and 92.54% respectively, $F(1, 118) = 10.899, p < .01, \eta^2 = .09$. Male listeners identified both genders with the same degree of accuracy (90.79%). Refer to Table 4 for a detailed listing of same and opposite gender identification percentages and confidence ratings as a function of listener gender.

Table 4

*Gender Identification of Same and Opposite Speaker Gender*

<table>
<thead>
<tr>
<th>Listener Gender</th>
<th>Mean % Accuracy Same Gender</th>
<th>Mean % Accuracy Opposite Gender</th>
<th>Mean Confidence Same Gender</th>
<th>Mean Confidence Opposite Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>88.60</td>
<td>92.54</td>
<td>7.05</td>
<td>7.49</td>
</tr>
<tr>
<td>Male</td>
<td>90.79</td>
<td>90.79</td>
<td>7.31</td>
<td>7.20</td>
</tr>
</tbody>
</table>
Within the segmental level of linguistic context, results showed a higher degree of accuracy for vowels (87.13%) than for fricatives (76.00%). Although listeners’ accuracy in identification of fricatives was well above chance, the listeners exhibited lower confidence ratings for fricatives (3.74) than for vowels (5.98), or any other level of linguistic context. A repeated measures ANOVA was conducted to evaluate differences in speaker identification across individual vowel and fricative sound segments. No significant differences between the four vowels were found. However, the ANOVA did indicate statistically significant differences between the listeners’ identification of speaker gender when presented with /s/ and /ʃ/ segments, $F(1, 19) = 6.057, p < .02, \eta^2 = .24$. Refer to Table 5 for a detailed listing of the identification percentages and confidence ratings within the segmental level of context.

Table 5

<table>
<thead>
<tr>
<th>Segment</th>
<th>Number</th>
<th>% Accuracy</th>
<th>Mean Confidence</th>
<th>SD of Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>200</td>
<td>89.50</td>
<td>5.87</td>
<td>2.51</td>
</tr>
<tr>
<td>/æ/</td>
<td>200</td>
<td>91.00</td>
<td>6.08</td>
<td>2.40</td>
</tr>
<tr>
<td>/u/</td>
<td>200</td>
<td>84.00</td>
<td>5.82</td>
<td>2.40</td>
</tr>
<tr>
<td>/ɑ/</td>
<td>200</td>
<td>84.00</td>
<td>6.16</td>
<td>2.46</td>
</tr>
<tr>
<td>/s/</td>
<td>200</td>
<td>80.00</td>
<td>4.32</td>
<td>2.59</td>
</tr>
<tr>
<td>/ʃ/</td>
<td>200</td>
<td>72.00</td>
<td>3.16</td>
<td>2.12</td>
</tr>
</tbody>
</table>

The word and sentence level results showed less than 10% variation in accuracy of identification within individual word items and less than 4% for the sentence stimuli. See Tables 6 and 7 for a detailed listing of the identification percentages and confidence ratings within the word and sentence levels of context.
Table 6

*Gender Identification at the Word Level*

<table>
<thead>
<tr>
<th>Word</th>
<th>Number Stimuli</th>
<th>% Accuracy</th>
<th>Mean Confidence</th>
<th>SD of Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fence</td>
<td>200</td>
<td>93.00</td>
<td>7.64</td>
<td>2.32</td>
</tr>
<tr>
<td>Horse</td>
<td>200</td>
<td>88.00</td>
<td>7.60</td>
<td>2.39</td>
</tr>
<tr>
<td>Pool</td>
<td>200</td>
<td>90.00</td>
<td>7.85</td>
<td>2.28</td>
</tr>
<tr>
<td>Pizza</td>
<td>200</td>
<td>93.50</td>
<td>8.34</td>
<td>2.11</td>
</tr>
<tr>
<td>Jumping</td>
<td>200</td>
<td>90.00</td>
<td>7.54</td>
<td>2.23</td>
</tr>
<tr>
<td>Swimming</td>
<td>200</td>
<td>84.50</td>
<td>7.83</td>
<td>2.29</td>
</tr>
</tbody>
</table>

Table 7

*Gender Identification at the Sentence Level*

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Number Stimuli</th>
<th>% Accuracy</th>
<th>Mean Confidence</th>
<th>SD of Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baking</td>
<td>200</td>
<td>98.00</td>
<td>8.46</td>
<td>1.86</td>
</tr>
<tr>
<td>Eating</td>
<td>200</td>
<td>95.50</td>
<td>8.43</td>
<td>1.91</td>
</tr>
<tr>
<td>Jumping</td>
<td>200</td>
<td>96.00</td>
<td>8.49</td>
<td>1.83</td>
</tr>
<tr>
<td>Mowing</td>
<td>200</td>
<td>99.00</td>
<td>8.31</td>
<td>2.01</td>
</tr>
<tr>
<td>Picking</td>
<td>200</td>
<td>98.00</td>
<td>8.31</td>
<td>2.23</td>
</tr>
<tr>
<td>Swimming</td>
<td>200</td>
<td>97.50</td>
<td>8.70</td>
<td>1.77</td>
</tr>
</tbody>
</table>

Finally, using Praat analysis software, a post-hoc acoustic analysis was conducted on the sentence stimuli from each speaker. The mean speaking F0 was found to be 241.89 Hz (*SD* = 20.87) for females and 231.30 Hz (*SD* = 10.14) for males. A statistical correlation between speaker F0 and listener accuracy in identifying the speaker gender was found to be non-significant, *r* = -.25. Refer to Table 8 for a detailed listing of the mean F0 values and rate of identification percentages.
Table 8

*Average F0 and % Accuracy for Each Speaker*

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Gender</th>
<th>Average F0</th>
<th>% Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Female</td>
<td>224.48</td>
<td>85.26</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
<td>251.62</td>
<td>95.79</td>
</tr>
<tr>
<td>3</td>
<td>Female</td>
<td>253.32</td>
<td>90.53</td>
</tr>
<tr>
<td>4</td>
<td>Female</td>
<td>215.45</td>
<td>88.42</td>
</tr>
<tr>
<td>5</td>
<td>Female</td>
<td>264.58</td>
<td>87.37</td>
</tr>
<tr>
<td>6</td>
<td>Male</td>
<td>226.04</td>
<td>98.68</td>
</tr>
<tr>
<td>7</td>
<td>Male</td>
<td>247.08</td>
<td>82.37</td>
</tr>
<tr>
<td>8</td>
<td>Male</td>
<td>229.08</td>
<td>91.58</td>
</tr>
<tr>
<td>9</td>
<td>Male</td>
<td>220.24</td>
<td>98.42</td>
</tr>
<tr>
<td>10</td>
<td>Male</td>
<td>234.05</td>
<td>88.16</td>
</tr>
</tbody>
</table>

**Discussion**

The first aim of this study was to determine whether listeners could identify the gender of a child from a speech sample. Results showed that listeners can identify the gender of a child speaker from a speech sample with relatively high accuracy. These results were similar to those of Bennet and Weinberg (1979), Perry et al. (2001), and Sachs et al. (1973). These studies found that adult listeners were able to correctly identify the gender of children a majority of the time when listening to recorded speech samples. These studies reported mean accuracy rates ranging from 68% to 81%, compared to an overall accuracy of 90.66% for the present study. Variation in accuracy rates among these studies may be due to differences in speaker age (4-16 years), as well as the fact that listeners were making identification decisions based on stimuli with less or more linguistic context. Perry et al. (2001) used CVC syllables, while Bennet and Weinberg (1979) presented listeners with vowels and sentences. The results of the present study were most
similar to those reported by Sachs et al. (1973), who presented listeners with vowels, sentences, and a short reading passage.

The second aim was to examine whether listeners were more accurate in identifying the gender of the speaker when listening to a sample with more linguistic context. The present study found that as the amount of linguistic context increased, the accuracy of the listeners’ identifications also increased, with percentages ranging from 83.42% when listening to sound segments to 99.00% when presented with a 1-minute speech sample from each speaker. Although segments were identified with the lowest percentage of accuracy among the four levels of linguistic context, they were still identified at a rate well above chance.

Listeners’ higher accuracy for words than segments may be attributed to the fact that words contain multiple instances of phoneme segments. Differences may also be the result of additional suprasegmental information that can be heard when listening to words as opposed to sound segments in isolation, such as stress patterns and coarticulation. Some authors have hypothesized that children learn culturally determined suprasegmental patterns that are appropriate or archetypal for each gender (Fitch & Giedd, 1999; Sachs et al., 1973). A study by Andrews and Schmidt (1997) also reported gender-related variation due to an increased fluctuation in intonation and breathiness in the female voice.

Those suprasegmental aspects of speech present in sentences; such as speaking rate, rhythm, and sentential stress patterns; may underlie the higher identification scores for sentences. These results are supported by data reported by Ferrand and Bloom (1996), who found that young male speakers between the ages of 7 and 8 years began to restrict their intonational ranges while females did not. Although not documented in depth with child speakers, gender-related
differences in speaking rate for adults have also been found (Fitzsimons et al., 2001; Whiteside, 1996).

While the difference was less pronounced, listeners identified the gender of the speakers with greater accuracy when presented with 1-minute sections of discourse than when hearing sentence length stimuli. Sachs et al. (1973) suggested that male and female children often use different vocabulary when speaking. Although the topics discussed in the conversational samples in the present study were meant to be gender-neutral, it is possible that when the segmental or suprasegmental cues were ambiguous with regard to speaker gender, listeners then used possible differences in lexicon, providing a more broadly-based perception of speaker gender.

The third aim of the current study was to evaluate the extent to which listeners’ abilities to identify a speaker’s gender from a single sound segment improved depending on the phoneme was examined. Listeners’ accuracy varied depending on the phoneme being spoken, with higher accuracy for vowels than for fricatives. Percentages ranged from 84% to 91% across the four vowels, and were 72% to 80% for /s/ and /ʃ/ fricatives. These results were similar to Lee et al. (2010), who also found vowels were identified more accurately than fricatives. The accuracy with which vowels were identified provides support for the perceptual salience of formant frequency and F0 patterns in determining the gender of a speaker. Bennet (1980), who analyzed six vowels spoken by children, found that formant frequencies of boys were consistently lower than those of girls. Other studies found similar results when examining vowels in children (Busby & Plant, 1995; Sachs et al., 1973; Whiteside & Hodgson, 2000a, 2000b). As discussed in the literature review, authors have attributed differences in formant frequencies between genders to anatomical factors, behavioral factors, or a combination of both.
While some authors have found that F0 does not differ as a function of gender until children reach puberty (Busby & Plant, 1995), other authors have found differences in F0 beginning around age 8 (Whiteside & Hodgson, 2000a). It is interesting to note in the present study that the female with the highest average F0 was not the most accurately identified female, although the two males with the lowest average F0s were the most accurately identified males. Such results indicate that while F0 may play a role in gender identification, it is not the only factor listeners rely on when determining gender. As mentioned in the results section, a statistical correlation between speaker F0 and listener accuracy in identifying the speaker gender was found to be non-significant, \( r = -.25 \). This lack of a significant correlation indicates that listeners were likely using cues other than pitch to determine the speaker gender. Although not analyzed in this study, formant frequency patterns may have a more significant role in their decisions (Bennet, 1980).

Of interest is the result that the fricatives /s/ and /ʃ/ were identified at a rate much greater than chance, even though the fricatives had extremely short durations and contained no voicing. A study by Schwartz (1968) found that adult females tended to have higher frequency spectra when producing /s/ and /ʃ/ than adult male speakers. Another study found gender-specific differences in the acoustic properties of fricatives when analyzing spectral slices and spectral moments of adults (Jongman, Wayland, & Wong, 2000). Evidence that supports differences in fricative production in male and female children is found in Fox and Nissen (2005). In their analysis, the authors found spectral peak means were similar for /t/ and /ð/ among male and female speakers, but were significantly different for /s/ and /ʃ/. The authors also found similar spectral means for /t/ and /ð/ in male and female speakers, but found females had significantly higher spectral means for /s/ and /ʃ/. In addition, the authors found gender-related differences in
spectral skewness and slope. In particular, male speakers were higher than female speakers in the production of /f/ and lower than female speakers in the production of /s/. The authors concluded that, because dimorphism of the vocal tract does not occur for most children until peripubertal and postpubertal stages of development, a portion of the gender-linked differences found in children may be attributed to learned or behavioral factors. The authors found these differences begin to emerge in children as early as 6-7 years of age. A follow-up study (Nissen & Fox, 2005) found evidence that gender differences in fricative production started at around 5 years of age.

As mentioned in the results section, statistically significant differences in accuracy of identification were not found between the four vowels. However, a statistically significant difference was found between the /s/ and /ʃ/ fricatives. Fox and Nissen (2005) found spectral means declined for /ʃ/ as age increased regardless of speaker gender; however, spectral means for /s/ declined only in males as age increased. In addition, the authors found that the spectral kurtosis values for /s/ were significantly higher for female speakers than for male speakers, though this difference was not found in /ʃ/. The authors also conducted a series of separate discriminant analyses, which showed that an adult male model of fricative classification successfully categorized /s/ in male children a higher percentage of the time than the adult female model across all age groups. Similarly, the adult female model successfully categorized /s/ in female children a higher percentage of the time than the adult male model in all but the two youngest age groups (6-7 years and 8-9 years). These results give some support to the theory that across a combination of acoustic parameters, children’s speech may follow gender-specific patterns of articulation which may be in part due to learned behavior.
Future research in this area might involve larger numbers of both speakers and listeners across a more diverse age range. In addition, it would be of interest to examine the impact that individual acoustic parameters (e.g., overall duration, spectral mean, intensity) of a phoneme has on the perception of speaker gender, possibly through the manipulation of synthetic speech stimuli. This study correlated the perceptual judgments of the listeners with the acoustic measure of F0; however, it would also be of interest to more fully examine other perception-production links by conducting a detailed spectral and acoustic analysis of the speaker productions.

The high accuracy with which listeners identified stimuli within each of the four levels of linguistic context, as well as the differences in accuracy of identification across the four levels, indicate that the perceptual cues which enable listeners to identify gender are multifactorial. A combination of acoustic distinctions, possibly resulting from both anatomical and behavioral differences, and suprasegmental or prosodic aspects of speech, likely provide perceptual cues which enable listeners to identify gender in children. Despite the limitations of the current study, these data may provide additional insight into the perceptual cues used by listeners to identify the gender of a speaker and the associated phonetic, prosodic, or stylistic speech factors.
References


Annotated Bibliography


Eleven biologically male speakers were recruited for this study. All speakers described themselves as heterosexual transgendered individuals. The speakers were recorded reading a standard passage in both their masculine voice and their feminine voice. The authors analyzed the recordings for F0, frequency range, overall passage duration, and duration of a sample of stressed vowels. A group of 88 listeners evaluated the 22 speech samples and rated them using 18 semantic differential scales and 57 adjectives. Results showed that the listeners heard significant differences between the masculine and feminine voice presentations. The authors found that high pitch, breathiness, and animation were associated with a feminine voice. In addition, variations in loudness, duration, rate, intonation contours, and frequency all contributed to the perception of a male or female voice.


Fourteen professional actors were recruited to participate in this study. The professionals were recorded portraying fourteen different emotions varying in intensity and valence. A total of 224 recordings were acoustically analyzed. Analysis showed differences in fundamental frequency, mean energy, energy distribution in the spectrum, and speech rate among the various emotions. The recordings were also presented to judges who were asked to identify the emotions of the speakers. Results showed that judges were able to accurately identify almost all of the emotions with an accuracy much better than chance.


Speech from 73 children was recorded using whispered and normally phonated vowels, normally spoken sentences, and sentences spoken in a monotonous tone of voice. The four utterance types were selected to manipulate properties of speech thought to influence a listener’s opinion of sexual identity. Adult females (N=116) listened to the recordings and indicated whether they thought each speaker was a boy or girl. Listeners were able to correctly identify the sex of the 6- and 7-year-old children 68% of the time. Results showed that vocal tract resonance properties of preadolescent children appeared to provide primary cues about sexual identity. The authors hypothesized two possible reasons for this result: first, preadolescent boys and girls may have different vocal tract sizes which enable listeners to make judgments as to sexual identity, and second, preadolescent children may articulate the vowel /æ/ differently depending on sex.

The author examined vowel formant frequencies in 7- and 8-year-old children. Forty-two children were recorded saying six vowels in a CVC context. Results showed that formant frequencies of boys were consistently lower than those of girls. The extent to which formant frequencies differed as a function of sex depended on the formant number and vowel category. The author suggested the differences across sex were a result of differences in pharynx size between males and females as well as sex-specific articulatory behaviors. Particularly, males may have used more lip rounding, smaller jaw openings, and/or a lower larynx position than females.


This study was comprised of three experiments. In the first, three novice talkers and one expert talker were recorded saying sentences. The sentences were then evaluated to determine whether reliable prosodic boundaries were produced. Frequency and duration cues were measured for consistency and strength. The purpose of the first experiment was to assess the stimuli to be used in the latter experiments. The authors found that speakers exhibited significant changes in prosody at intended phrase boundaries. In the second experiment, 16 participants listened to the stimuli and were asked to determine whether the initial phrase of a sentence was short or long. When presented with monotonic sentences, the mean accuracy was 57%, whereas when presented with the same sentences with normal prosody, the mean accuracy was 91%. The authors concluded that listeners are able to use prosodic information to make quick decisions about phrase boundaries. In the third experiment, 40 participants listened to sentences in which the prosodic information was inconsistent with later morphosyntactic information. Results showed that the inconsistent prosodic information resulted in less accurate judgments and slower comprehension. The authors concluded that prosody plays an important role in perception and comprehension of spoken language.


Systematic differences in the vowel formant frequency values produced by preadolescents of different ages and genders were examined in this study. The first three formant frequency values were measured for boys and girls ages 5, 6, 9, and 11 years old. Recordings were made while the participants produced 11 test words containing the 11 non-diphthongized vowels of Australian English. Results showed that F0 did not differ as a function of gender, though it did decrease with increase in age. Results showed that formant frequencies also decreased with increase in age, and the values for boys were lower than girls.

This study evaluated the ability of digital speech processing and pattern recognition techniques to identify the gender of a speaker from a speech segment. The authors did a detailed analysis of vowels, including F0, formant frequencies, bandwidths, and amplitudes. Results showed that both F0 and formant frequencies were lower for males than for females. In general, bandwidths of the formants were narrower for males than those produced by females. In addition, the amplitudes were typically higher for vowels produced by males than those produced by females. The authors also found that the second formant frequency was a slightly better gender identifier than the F0. This led the authors to conclude that formant characteristics contain adequate information for gender recognition without F0.


The purposes of this study were to determine whether children used adult-like patterns of contrastive stress, if boys and girls marked contrastive stress differently, and whether listeners could identify the gender of a child from a spoken sentence. Twenty children and a comparison group of 10 adults participated in this study. Stimuli were pictures of everyday events, and the pictures were designed to elicit four lexically identical sentences from each participant. The author did not find consistent increases in frequency, intensity, or duration for emphatically stressed words. In addition, the author did not find consistent differences between male and female participants in the use of prosody to mark contrastive stress. In the perceptual task, adults were able to identify the gender of child speakers with 87.1% accuracy.


Magnetic resonance imaging was used to examine the vocal tract morphology of 129 participants between the ages of 2 and 25. Results showed a significant positive correlation between body size and length of the vocal tract. The authors also found a significant difference in vocal tract morphology between males and females, although these sex differences did not appear until children reached puberty. The authors hypothesized that the discernable difference in voices of preadolescent boys and girls is primarily due to behavioral differences, not anatomical differences.


Ten male and 10 female participants were instructed to read 10 sentences first with a declarative intonation and then with an interrogative intonation. The authors found a dependence of pitch on duration that differed between male and female participants. They also found significant differences in adult male and female speech in pitch range as
well as in pitch slope. In addition, males produced significantly shorter utterance lengths and spoke at a significantly greater rate than females.


The authors examined intonational differences in the speech of preadolescent children. Participants were boys and girls in four age groups: 3-4, 5-6, 7-8, and 9-10 years of age. Conversational speech samples were recorded and analyzed for minimum and maximum F0, average F0, and standard deviation of F0. In addition, the authors examined the number and direction of F0 changes per utterance. The mean F0 for males began to decrease in the 7-8 age group. In addition, males of the same age group began to show a decrease in the percentage of rising and falling shifts. These findings were not observed in the female groups. The authors attributed the lowering of F0 and the decrease in intonational shifts to both anatomical and sociocultural factors.


Three points of interest were examined in this study: the pattern of normal development in the acoustic properties of fricatives, whether sex-specific patterns of fricative articulation are present in the speech of preadolescent children and when these differences emerge, and to what extent a segment classification model based on the acoustic characteristics of fricatives produced by a specific sex could successfully categorize the fricatives of younger speakers. For this study, 50 male and 50 female speakers, ages 6 to 52 years old, produced words containing voiceless fricatives in the syllable initial position. Results relative to the present study showed that preadolescent children exhibited sex-specific differences in fricative articulation, although the acoustic differences were reduced or absent in the youngest children. The authors concluded that because vocal tract dimorphism does not occur until puberty for most children, sex-related differences in children may be associated in part with learned or behavioral factors.


The purpose of this study was to identify differences in acoustic attributes of preadolescent children according to age and sex. Samples of the sustained vowel /a/ were elicited from 180 children ages 5, 6, 7, 8, 9, and 10. F0 began to differ between sexes at the age of 7 or 8, and F0 decreased significantly in boys between the ages of 5 and 10. The authors suggested that the lowering of F0 for preadolescent males may be a result of differences in thyroid size and laryngeal angle, or cultural stereotypes that may encourage young males to speak at a lower F0 than would be natural for their phonatory apparatus.

Twenty adult speakers (10 male and 10 female) participated in this study. The participants were recorded saying the eight English fricatives in CVC syllables within the phrase “Say — again.” The fricatives were always in initial position and followed by six different vowels. A total of 144 samples were collected from each participant. The fricatives were analyzed for spectral peak location, spectral moments, locus equations, root-mean-square amplitude, and relative amplitude. Results relative to the present study showed that spectral peak location was significantly higher for females than for males. In addition, females demonstrated significantly higher values than males for spectral mean, kurtosis, and variance. In contrast, females exhibited significantly lower values than males for spectral skewness. Results also showed that relative amplitude values were smaller for females than for males.


In order to examine breathiness in voice quality, 10 female and six male speakers were recorded imitating two sentences. The authors noted in their literature review that the use of a breathy voice quality, as well as a more dynamic intonation contour for females, appear to be learned behaviors. The authors also pointed out that some male and female differences appear long before puberty. From their study, the authors concluded that females tend to be more breathy than males, although there were large differences between participants within each gender.


This study evaluated the relative importance of F0 and formant frequencies in sex identification tasks. The authors examined 20 adult speakers, 10 male and 10 female. Six isolated vowels were whispered and produced normally by each speaker. Three tapes were constructed from the recordings: voiced, whispered, and filtered. Fifteen listeners identified the sex of each speaker and indicated a level of confidence in their decision. 96% of the judgments were correct for the voiced tape, 91% for the filtered tape, and 75% for the whispered tape. The authors concluded that F0 appears to be a more important acoustic cue in identifying speaker sex than the formant frequencies.

Fifteen males and 13 females were examined for changes in the size and shape of the pharynx, the vocal tract, and the spatial positions of the hyoid, mandible, larynx, and hard palate. A longitudinal series of lateral radiographs were taken of the participants between the ages of 1 month and 14 years. Results showed significant changes in certain aspects of vocal tract shape. For example, the authors found the ratio of the pharynx height to oral cavity length decreases significantly between birth and 6 to 8 years of age. On the other hand, results showed few changes in spatial relations between the vocal folds, the hyoid, the mandible, and the hard palate. The authors noted very little sexual dimorphism in pharyngeal shape and size before 13.75 years of age.


The purpose of the study was to examine whether accurate gender identification of a speaker underlies the ability to estimate relative F0 height without cues usually present for speaker normalization. The authors explain that judgment of relative height of a tone is made in reference to the speaker’s F0 range. For example, a tone considered high for a male speaker may be acoustically equivalent to a tone considered low for a female speaker. Sixteen adult male and 16 adult female speakers were recorded producing the Mandarin syllable sa in all four tones. Three stimuli were produced from the syllable sa: fricative vowel, fricative only, and vowel only. Sixty listeners, both Mandarin and English, listened to the stimuli and identified gender. Fricative vowel stimuli were identified with the greatest accuracy, then vowel only, then fricative only. The authors concluded that accurate gender identification is a result of an individual’s ability to estimate relative F0 height without the aid of cues typically considered necessary for speaker normalization.


Speech samples from 436 children and 56 adults were used to evaluate F0, formant frequencies, changes in magnitude and variability of duration, and spectral envelope of children’s speech. Each participant produced 10 monophthongal vowels, five diphthongal vowels, and five phonetically rich meaningful sentences. The authors found that both magnitude and variability of segmental durations decrease significantly between the ages of 9 and 12, and generally reach adult levels around age 12. Differentiation between male and female F0 and formant frequency patterns starts near age 11 and becomes fully established around 15 years of age.

The purpose of this study was to determine whether two different training programs were effective in improving students’ abilities to match assigned targets of vocal intensity and fundamental frequency. The seven students who participated in the study were from the broadcast journalism department at BYU. Five of the students were assigned to the primary group and received a numerically coded method of prosodic instruction. Two of the students were assigned to the secondary group and received a color-coded method of prosodic instruction. The students were recorded reading five news stories before instruction and after eight weeks of prosodic instruction. Results showed there were no significant improvements in students’ abilities to match target level of emphasis in either training group. Potential reasons for the ineffectiveness of the training programs are discussed.


Thirty children between 3 and 6 years of age and a group of ten comparison adults participated in this study. Participants were recorded saying four voiceless fricatives /f/, /θ/, /s/, and /ʃ/ five times each in an initial syllable position, for a total of 60 tokens per participant. Fricatives were analyzed for duration, amplitude, and the spectral measures of slope, mean, variance, skewness, and kurtosis. The authors found female speakers exhibited higher spectral slope values for all places of fricative articulation except /s/. Results also showed males had higher spectral means than females for /f/, /θ/, and /s/, and lower spectral means than females for /ʃ/. Analysis of spectral kurtosis showed males had significantly lower values than females for /s/, though significant gender differences were not found for the other fricatives.


The purpose of this study was to describe the acoustic patterns of the voiceless stop consonants /p t k/ produced by typically developing children and a group of adults. In addition, the authors examined the extent to which spectral characteristics and individual amplitude differed as a function of age, sex, place of articulation, and vowel context. The participants in the study were 30 children between the ages of 3 and 5 and a comparison group of 10 adults. Participants were recorded saying words containing one of the three voiceless stops in initial position followed by a monophthongal vowel. Results related to the current study showed significant sex and age group differences for the spectral measures of slope, mean, and skewness. Acoustic differences for adult speakers can be explained by sexual dimorphism that results in differences in vocal tract length and shape and as well as vocal fold size. However, because studies indicate that dimorphism of the vocal tract is not present in preadolescent children, the authors hypothesized that sex-related differences in acoustic properties in the speech of children may be associated with learned or behavioral factors.

The acoustic characteristics of children’s voices and speech that enable listeners to identify gender were examined in this study. In Experiment I, speech samples and gross physical measurements were taken from 4-, 8-, 12-, and 16-year-olds (with 10 boys and 10 girls in each age group). F0 and formant frequencies were obtained from speech samples of seven vowels of American English. In Experiment II, 20 adults used a 6-point rating scale to rate the syllables produced by the children in Experiment I. The results have two relevant indications: first, vowel formant frequencies differ as a function of gender in children as young as 4 years old, while formant frequencies and F0 differ as a function of gender after 12 years old; and second, listeners can identify the gender from the speech and voice of children as young as 4 years old, and in young children, it appears that listeners base their ratings on vowel formant frequencies.


The purposes of this study were to determine whether preadolescent males and females have acquired voice characteristics that enable their voices to be identified as to sex, and to explore whether differences in formant frequencies contribute to this differentiation. Twenty-six children between 4 and 14 years old were recorded repeating a short sentence, reading a passage from a book, and repeating three sustained vowels. Eighty-three adult judges listened to the recordings and identified each voice as a boy or girl. Results showed that the judges guessed correctly 81% of the time. Results also showed that the average F0 was higher for the boys than for the girls, though average formant frequency was lower for the boys than for the girls. The authors hypothesized that the differences observed could be a result of children learning culturally determined patterns that are viewed as appropriate for each sex.


This book explained and discussed the anatomy and physiology that make speech, language, and hearing possible. Chapters concerning the anatomy and physiology of phonation were applicable to the present study.


Nine female and nine male speakers participated in this study. The speakers were recorded producing four fricatives /f/, /θ/, /s/, and /ʃ/ in isolation for a total of 72 stimuli. Ten listeners listened to the stimuli and rated each stimulus as male or female. Results showed listeners judged correctly 93% of the time for /s/, 90% of the time for /ʃ/, 74% of
the time for /f/, and 69% of the time for /θ/. Spectrographic analysis of /s/ and /ʃ/ showed that female spectra tended to be higher in frequency than male spectra.


This article discussed differences in the vocal tract between adult males and females. The authors stated that the anterior-posterior dimension of the male thyroid cartilage is about 20% larger than that of the female. In addition, the vocal fold of an adult male is approximately 16 mm long whereas the vocal fold of an adult female measures only about 10 mm. Men also have thicker vocal folds than women. The authors stated that there are no major differences in the type of tissue that makes up the vocal folds. Differences in vocal tract shape are also discussed. The authors concluded that differences in vocal fold length between men and women account completely for differences in average F0, mean airflow, and aerodynamic power produced.


Imaging studies of 63 individuals between birth and age 6 years 9 months were evaluated, as well as a group of 12 adults. Results showed that the vocal tract length increases from birth to age 6 years 9 months, with the most rapid period of growth occurring between birth and 16 months of age. The authors found no sexual dimorphism between birth and age 6 years 9 months in vocal tract length, though it is possible that even though vocal tract length remains similar across genders, males may have larger vocal tracts. Thus the acoustic differences in preadolescent children may be attributed to anatomic or articulatory differences rather than vocal tract length.


The authors evaluated 605 imaging studies of individuals between birth and 19 years old. The goals of the study were to characterize the anatomic growth trend and growth rate of the vocal tract, as well as to quantify the sex-specific growth curves of the vocal tract length and its oral and pharyngeal portions as neural (brain and cranium) or somatic (hard and soft tissues of the face). The authors explain that the development of the vocal tract includes growth of hard and soft tissues which can be grouped into different schedules of growth and maturation. Results showed significant sex differences in eight of the nine variables examined, with the most significant differences appearing after the age of 12. The prepubertal sex differences were seen at specific ages and for select structures. The authors suggested future studies examine the anatomic basis for the acoustic differences in prepubertal children.

In this study, five sentences were read ten times each by three male and three female adult speakers, for a total of 300 sentences. Results showed that the mean sentence durations were longer for females than for males. These results were statistically significant. The authors suggested pausing, consonant cluster reduction, voiceless fricative and affricate realizations, vowel reduction, consonant devoicing, and vowel elision all contributed to the differences in rate between male and female speakers. The authors also found that female speakers demonstrated statistically significantly higher VOT values.


Frequency and F0 data that were presented by Lee et al. (1999) were re-evaluated by the authors of this study. This was done using a sex-specific developmental perspective and exploring some of the possible explanations for differences among the sexes. The authors analyzed the mean formant frequency values for the ten monophthongs recorded by Lee et al. (1999) in individuals seven years of age to adult. Results showed both age and sex-specific developmental patterns. Differences may be attributed to anatomical, physiological, sociophonetic, and cultural factors.


Acoustic differences in age, sex, and phonetic context for a phrase final vowel were evaluated. Three groups of children (ages 6, 8, and 10) and a group of nine adults were recorded saying phrases elicited through a picture naming task. F0 decreased with an increase in age, although the decrease was much more apparent for children than for adults. The age at which the sharpest decrease in F0 occurred differed for male and female children, with female children showing a gradual decline and male children showing a marked decline after age eight. Formant frequency decreased with an increase in age for F1, F2, and F3 for both genders, although female children demonstrated higher formant frequency values than male children. The authors suggested the differences in F0 between males and females in preadolescents could be evidence for sexual dimorphism of the vocal tract beginning around age 8. However, the authors also suggested the differences could be attributed to behavioral rather than anatomical factors.

Formant frequency values, coarticulation, and temporal patterns were examined in three groups of children (ages 6, 8, and 10) and in a group of nine adults. Speech samples were elicited through a picture naming task. Results showed a decrease in F0 with an increase in age, though the decrease was much more apparent for children than for adults. Results also showed a decrease in formant frequency with an increase in age for both genders, although female children demonstrated higher formant frequency values than male children. The authors also found differences in articulation rate between the sexes, with male children showing a faster articulation rate beginning around the age of 6.


The authors examined 30 children, ages 7, 9, and 11. There were 10 children in each age group and an equal number of boys and girls. Each child produced five repetitions of the plosives /p b t d/ in an intervocalic, syllable initial position within the phrases of ‘silver /pi/,’ ‘silver /bi/,’ ‘silver /ti/,’ and ‘silver /di/.’ Voice onset time data were examined for age, sex, and plosive differences. The results confirmed three hypotheses: there are sex differences in the voice onset time patterns of preadolescent children, sex differences are associated with age and development, and the extent of age and sex differences depends upon the plosive being examined. The authors suggested that the differences due to age and sex are a result of changes in anatomy and physiology as well as sociophonetic influences.


The purpose of the study was to compare the F0 and intonational patterns in male-to-female transsexuals. Twenty genetic males identified as transsexuals participated in the study. Speech samples were obtained by recording the responses of participants to questions about home and family. In addition, speech samples from 20 normal speakers from another study were also used. The recordings were presented to ten listeners who rated the samples as male or female. A second group of eight listeners rated the speakers from 1-7 (extremely feminine-sounding speech and voice to extremely masculine-sounding speech and voice). The speech samples obtained from the transsexuals were also acoustically evaluated for average F0, intonations, and shifts. Results showed that transgendered participants who were categorized by listeners as females exhibited greater intonation variability in addition to higher F0s than participants categorized as males. More specifically, participants categorized as females had a higher percentage of upward intonations (frequency changes without interruption of phonation) and downward frequency shifts (changes in frequency that took place between the terminal frequency of a given phonation and the initial frequency of a subsequent phonation). In addition, these participants had less extensive downward intonations and a smaller percentage of level intonations and level shifts (frequency changes or shifts less than 5 Hz).
The purpose of this study was to examine the ability of digital speech processing and pattern recognition techniques to identify the gender of a speaker from a speech segment of various vowels, unvoiced fricatives, and voiced fricatives. The authors note that while humans are able to gather information about a speaker’s gender, age, emotional state, health, etc. from an acoustic signal, current automatic speech and speaker recognition systems are much less capable. In their review of literature, the authors describe the differences between male and female adult voices in three types of parameters: acoustical, physiological, and perceptual. For example, female voices are known to have a higher fundamental frequency than males. Females are also often considered to be more melodic and more breathy.