2010-07-12

An Acoustic and Perceptual Investigation of Contrastive Stress in Children

Anita Susan Dromey
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
https://scholarsarchive.byu.edu/etd/2278

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.
An Acoustic and Perceptual Investigation of Contrastive Stress in Children

Anita S. Dromey

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
Christopher D. Dromey
Ron W. Channell

Department of Communication Disorders
Brigham Young University
August 2010

Copyright © 2010 Anita S. Dromey
All Rights Reserved
ABSTRACT

An Acoustic and Perceptual Investigation of Contrastive Stress in Children

Anita S. Dromey

Department of Communication Disorders

Master of Science

Key aspects of prosody have been studied in adults for a number of years; however, less attention has been paid to the acoustic patterns of prosody in children. Thus, the purpose of the present study was to evaluate how a group of 20 pre-adolescent children use prosody to mark contrastive stress compared to a control group of adult speakers. It was of interest to investigate whether the children’s use of prosody differed between boys and girls or the part of speech being emphasized. The prosodic patterns of contrastive stress were evaluated in terms of duration, fundamental frequency, and intensity change relative to a baseline production of the same sentence. In addition, a perceptual experiment was conducted to determine if listeners could reliably identify the gender of the child speakers when listening to sentence length utterances.

Statistical analysis indicated that there were some differences in the duration and fundamental frequency change as a function of speaker age and the part of speech being emphasized, with relatively minor differences between genders. However it remains unclear if the acoustic differences found in this study were substantial enough to cause a salient perceptual difference. Although previous studies have identified increases in frequency, intensity, and duration as cues of contrastive stress, the present findings revealed patterns that did not consistently conform to these expectations. Limitations in the task design, individual speaker characteristics, and also the type of acoustic measure used may have contributed to these results.

Keywords: prosody, children, gender, fundamental frequency, intensity, duration
ACKNOWLEDGEMENTS

My heartfelt thanks go to Dr. Shawn Nissen for your willingness and encouragement in pursuing a research question into unfamiliar territory; what an adventure! Also thanks to my committee members Dr. Dromey and Dr. Channell for improving the project and for your discerning insights.

My appreciation goes to the research participants, who completed their tasks with such good humor that the data gathering became a pleasure.

Warmest thanks to my children Bronwen, Michael, Jonathan, and Hannah, who cheered me on, and who banded together as teammates, helping each other through this busy time for our family. I’m so proud of each of you.

And my deepest love and gratitude to my cherished husband Christopher, for your wisdom, experience, and unwavering support. Thank you.
# Table of Contents

Table of Contents ........................................................................................................ iv

List of Tables ................................................................................................................ v

List of Figures ............................................................................................................. vi

List of Appendices ..................................................................................................... vii

Introduction .................................................................................................................. 1

Review of Literature ................................................................................................... 2

Functions of Prosody ................................................................................................. 2

Age and Gender-related Differences in Speech .......................................................... 8

Purpose of the Study .................................................................................................... 15

Method ......................................................................................................................... 16

Experiment 1 – Acoustic Evaluation ........................................................................... 16

Experiment 2 – Perceptual Evaluation ....................................................................... 19

Results ........................................................................................................................ 21

Experiment 1 ................................................................................................................. 21

Experiment 2 ................................................................................................................. 24

Discussion .................................................................................................................... 27

Experiment 1 ................................................................................................................. 27

Experiment 2 ................................................................................................................. 33

Conclusions ............................................................................................................... 33

References .................................................................................................................. 35
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Differences Between Men and Women in the Production of Prosody</td>
<td>14</td>
</tr>
<tr>
<td>2. Change in Semitones for the Stressed Word Relative to the Baseline Condition by Age Group, Gender, and Part of Speech</td>
<td>22</td>
</tr>
<tr>
<td>3. Intensity Change in dB for the Stressed Word Relative to the Baseline Condition by Age Group, Gender, and Part of Speech</td>
<td>22</td>
</tr>
<tr>
<td>4. Duration Change in Milliseconds for the Stressed Word Relative to the Baseline Condition by Age Group, Gender, and Part of Speech</td>
<td>22</td>
</tr>
<tr>
<td>5. Accuracy of Gender Identification Overall and by Part of Speech, Listener, and Speaker</td>
<td>28</td>
</tr>
</tbody>
</table>
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Fundamental frequency change for the stressed word relative to the baseline condition by part of speech for adults and children.</td>
<td>23</td>
</tr>
<tr>
<td>2. Fundamental frequency change for the stressed word relative to the baseline condition by gender for adults and children.</td>
<td>25</td>
</tr>
<tr>
<td>3. Duration change for the stressed word relative to the baseline condition by part of speech for adults and children.</td>
<td>26</td>
</tr>
</tbody>
</table>
List of Appendices

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Informed Consent Documents</td>
<td>43</td>
</tr>
<tr>
<td>B. Elicitation Stimuli</td>
<td>47</td>
</tr>
</tbody>
</table>
Introduction

The accurate assessment and effective treatment of communication disorders in children requires a thorough understanding of how speech and language typically develop; clinicians often use developmental data to help inform their decisions about the need for speech therapy services, the nature of short and long term goals, and how clients are progressing. Previous research has established general trends on how young children typically acquire speech; however, many gaps remain in our understanding of specific aspects of speech development. In particular, additional research is needed to more fully examine how communication may differ between boys and girls.

It is clear that there are gender-related communication differences, because listeners can identify whether an individual is male or female by listening to them speak. The speech of men and women can largely be distinguished on the basis of gender-specific acoustic features which result from anatomical differences in the size and shape of the vocal tract. However, gender-related speech differences cannot be fully explained by anatomical differences alone, especially in children where sexual dimorphism of the vocal tract has yet to occur. Thus, it is reasonable to hypothesize that in children, speech development may be influenced by learned or behavioral factors that follow cultural or social expectations for male and female speakers. A number of studies have found gender-related developmental differences in how children produce sound segments; however less emphasis has been placed on such differences in suprasegmental or prosodic elements of children’s speech. Thus the aim of this study was to more fully examine one aspect of children’s use of prosody, namely the production of contrastive stress.
Review of Literature

As individuals speak they express meaning primarily by combining sounds to make syllables and words. Even so, without the use of prosody a speaker's message would be incomplete, since a portion of the meaning is conveyed to the listener through intonation patterns, stress (syllable level), emphasis (word and sentence level), speaking rate, and rhythm (Wingfield, Lahar, & Stine, 1989). These linguistic elements of prosody are perceptible as changes in the relative pitch, loudness, tempo, and sound quality (Fry, 1955; Lieberman, 1960). Since these components often influence more than a single phoneme or phonetic segment, they are considered suprasegmental in nature (Kent, 1997). Some researchers have even proposed that the prosodic elements of speech are the structures which help organize sound into meaningful units (Cutler, Dahan, & van Donselaar, 1997).

Functions of Prosody

Prosody in speech serves to convey a speaker’s emotion or affect as well as to clarify potential linguistic ambiguities.

Affective Functions

Affective or emotional prosody helps the listener discern the speaker’s attitude and emotional state. From it, the listener may better recognize the importance the speaker places on specific elements of the expressed message. If the speaker’s message contains prominent prosodic features, the listener is more likely to comprehend and recall the message spoken (Stine and Wingfield, 1987).

Express emotion. Listeners often depend on prosodic patterns to perceive the emotion underlying a speaker’s expressions. For this reason, prosody has been called the “emotional component of speech and language” (Viscovich et al., 2003, p. 760). Using tests such as the
Tuebinger Affekt Batterie and the New York Emotion Battery, researchers have found that a speaker’s fundamental frequency (F0) often changes in a measurable manner depending on the particular emotion of the speaker during expression of an utterance (Raithel & Hielscher-Fastabend, 2004). For example, expressions of happiness are generally produced with a higher and more variable F0 (Banse & Scherer, 1996; McRoberts, Studdert-Kennedy, & Shankweiler, 1995; Viscovich et al., 2003). On the other hand, sadness is often characterized by decreased prosodic variability (Scherer, Banse, Wallbott, & Goldbeck, 1991). Expressions of despair and elation, which usually accompany a state of high emotional arousal, are typically produced with higher fundamental frequencies (Banse & Scherer, 1996). These emotional prosodic cues provide the listener with a communicative context in which to interpret a speaker’s message.

Convey speakers’ priorities. Effective communicators will often use prosody to direct the listeners’ attention or focus to key words or phrases which the speaker is particularly interested in communicating by adjusting their vocal pitch, loudness, or tempo (Bolinger, 1978). Used this way, prosody provides listeners with the information needed to determine what parts of a message are most important. For example, in the utterance, The boy is swimming in the pool, if a speaker intends to focus the listener on the concept that a boy rather than a girl is swimming, they will typically change the acoustic properties of the word boy relative to other words in the sentence. Similarly, if the more important part of the communication is that the boy is swimming, and not floating on a raft, doggy-paddling, or some other activity, the acoustic changes will typically occur on the word swimming. While syllables and words make up the underlying linguistic structure of an utterance, it is the way a speaker emphasizes or focuses an utterance which guides listeners toward a particular interpretation of the message (Blasko & Hall, 1998).
**Facilitate comprehension in the listener.** Prosodic information not only helps focus the listener to the important details of a speaker’s message, but studies have also shown that prosodic cues facilitate the acoustic clarity of individual speech sounds (Cutler et al., 1997). Researchers have found that in stressed syllables, the segmental components are often more intelligible than in non-stressed syllables. Vowels in stressed syllables, for instance, often have more formant pattern separation from neighboring vowels, maintain a less variable steady-state, and are often longer in duration.

Prosodic cues, such as stress, have been found to increase the accuracy and speed of a listener's perceptual processing of both words and individual sounds (Cutler & Foss, 1977; Shields, McHugh, & Martin, 1974). Blasko and Hall (1998) explored how prosodic information facilitates comprehension of spoken messages. The researchers designed a study in which the prosodic elements of an utterance suggested a particular syntactic interpretation that was either consistent or inconsistent with subsequent information. Their results showed that not only were comprehension times longer for sentences in which prosody was inconsistent with later clarifying information, but listener response times were also greater when the participants were faced with the decision about which way to interpret the presented information.

**Promote recall in the listener.** Prosodic cues also play a role in a listener’s ability to recall what they have heard. To assess the effect prosody might have on listeners' ability to recall information from speech, Stine and Wingfield (1987) presented spoken information with and without normal prosody to both young and elderly adults. The results of the study showed better immediate recall of sentences presented with normal prosody for both groups of participants, with elderly adults exhibiting a greater increase in recall accuracy from the use of normal prosody. Another study by Wingfield, Wayland, and Stine (1992) examined the influence of
prosody on communication recall in groups of young and elderly adult listeners. The researchers had the participants listen to matched and unmatched sentences; the matched sentences were read using typical prosodic patterns considering the lexical and semantic content of the utterance, whereas the unmatched sentences were read with atypical prosodic patterns. The researchers reported that when the prosodic pattern of the sentence matched the syntactic structure, recall was better for both the young and elderly participants. When the prosodic pattern was in conflict with the lexical and semantic properties of the sentence, the elderly participants’ recall of the sentences matched how they had interpreted the prosodic structure of the utterance. These findings are supported by additional studies which have reported similar results (Cohen & Faulkner, 1986; Wingfield et al., 1989).

**Linguistic Functions**

In addition to the affective aspects of communication, the prosodic features of speech also serve several linguistic functions. Indeed, prosody is often used by speakers to signal the boundaries of an utterance, clarify lexical ambiguities, and signify a specific sentence type.

*Signify boundaries.* Research has shown that native English speakers rely on several different mechanisms to create syntactic boundaries in their speech, such as varying their acoustic intensity and F0, slowing down or lengthening the phonetic segments that precede a boundary, or inserting a pause (Lehiste, Olive, & Streeter, 1976). However, whether a listener can detect a syntactic boundary based solely on these suprasegmental cues is still unclear. Some researchers have reported that prosody alone is insufficient to accurately determine sentential boundaries. Cutler et al. (1997) concluded that listeners’ designations of boundary locations are determined by the syntactic structure of the sentences instead of the prosodic pattern of the utterance. Yet other studies have found that listeners are able to accurately and consistently
locate major syntactic boundaries from prosodic information alone, with no lexical information provided (Swerts & Geluykens, 1993; T’hart, Collier, & Cohen, 1990). Although the degree to which prosody helps listeners detect linguistic boundaries remains unclear, it is generally accepted that a speaker’s prosodic patterns are actively evaluated by the listener when interpreting an utterance. The majority of studies have found that prosody plays at least a supporting role in helping a listener detect boundaries within and between sentences, as well as between topics (Blasko & Hall, 1998; Cutler et al., 1997; Wingfield & Butterworth, 1984).

*Clarify lexical ambiguities.* Listeners may also use the additional perceptual cues from prosody to resolve lexical ambiguities and correctly interpret a speaker’s communicative intent (Blasko & Hall, 1998; Cutler et al., 1997; Goldstein, 1980; Wingfield et al., 1992). Even when it is in conflict with subsequent morphosyntactic information, prosody is often used by listeners to clarify lexical information (Blasko & Hall, 1998). In a study designed by Beach (1991), listeners used prosodic information to predict eventual sentence structure during online sentence processing. Listeners were presented with sentence fragments that were extracted from two syntactically different types of sentences. The first sentence type was subject-verb-object; for example, *The city council argued the mayor’s position / forcefully.* The second type was subject-verb-complement; for example, *The city council argued the mayor’s position / was incorrect.* The listeners were never given the whole sentence, yet results showed that listeners consistently chose the sentence ending that syntactically matched the prosodic information that was intended to be prototypic in the presented sentence fragment.

Because these findings confirm the important role of prosody in effectively communicating a message, researchers further questioned whether speakers routinely produce sufficient prosodic cues to allow listeners to resolve lexical ambiguities (Lehiste et al., 1976;
Wales & Toner, 1979). In their study, Albritton, McKoon, and Ratcliff (1996) instructed listeners to disambiguate utterances produced by two different groups of speakers, who were either naive or trained. The naive speakers were not considered professional voice users, whereas the trained group of speakers had received professional voice training as either actors or media broadcasters, most of whom were university students majoring in either broadcasting or performing arts. The trained speakers were asked to produce the sentence stimuli under two different conditions. In the first speaking condition they were instructed to read the sentences in a natural manner. In the second condition the speakers were made aware of the sentences' ambiguities prior to reading them and instructed to produce each sentence in a manner that would increase the likelihood that the listener would be able to disambiguate the intended meaning. Unfortunately, the study did not involve the untrained speakers in this second task. The results indicated that when the sentences were spoken in a natural manner, listeners were unable to disambiguate the intended meaning of the sentences from either the naive or trained speakers. However, when the trained group of speakers was informed of the ambiguities in the sentences, they produced the sentences with prosody that allowed listeners to resolve the ambiguities. Thus, it seems reasonable to infer that speakers can substantially clarify their messages with enhanced prosodic cues, yet such cues are not necessarily strong enough in conversational speech.

Signal sentence types. Intonation, an aspect of prosody that extends across the length of an utterance, is used by listeners to identify the type of sentence being expressed. In English, speakers typically use pitch declination (a lowering of F0 at the end of an utterance) to signal a declarative sentence. When producing an interrogative sentence speakers typically raise their pitch at the end of the sentence. For example, the sentence *Mary went to the store* can be
expressed as either a statement or a question depending on the slope of the intonation across the utterance.

*Emphasize contrasts.* Some researchers have concluded that a particular form of prosody known as *contrastive stress* is used by speakers to mark specific elements in an utterance, such as new versus previously-provided information, the deaccenting of redundant information, or the emphasis of a word for syntactic purposes (Cruttenden, 1986; Couper-Kuhlen, 1986; Crystal, 1969; Ladd, 1980). Although there are differences of opinion on the relative role of each component, most researchers agree on the acoustic features which speakers use to mark contrastive stress, namely changes in relative F0, duration, and intensity (Lieberman, 1960; Lehiste, 1970). Some studies have concluded that speakers primarily use changes in F0 to signal contrastive stress (Atkinson, 1978, O’Shaughnessy, 1979). However, others have also suggested that speakers may *trade* or replace F0 cues with the duration and/or intensity of a word to convey stress. For example, a study by Cooper, Eady, and Mueller (1985) found that the contrastive element in a series of sentences was generally characterized by longer durations on the target word, followed by a sudden drop in the F0 on the following word. O’Halpin (1997) also found that speakers mark contrastive stress by using a higher or more dramatic fall in pitch on the syllable following the stressed word.

*Age and Gender-related Differences in Speech*

Research has consistently shown that speech is different for men and women, which may affect how a listener interprets a speaker’s message. The speech of children has been found to be acoustically and perceptibly different from that of adults. There are both anatomic and behavioral bases for these differences.
Anatomic Variation in the Vocal Tract

Age and gender-based differences in speech are in part the result of a smaller larynx and vocal tract in women and children, which results in higher F0 and formant frequencies. Variations in body height generally correlate with the scaling of anatomic structures in the vocal tract (Ladefoged & Broadbent, 1957; Lieberman, 1984; Nearey, 1978; Peterson & Barney, 1952). More recently, magnetic resonance imaging (MRI) technology has been used to compare vocal tract size and growth characteristics of men and women, and how these structures develop from childhood. These studies confirm that men tend to have longer vocal tracts and larger laryngeal structures than women (Baer, Gore, Gracco, & Nye, 1991; Story, Titze, & Hoffman, 1996; Sulter et al., 1992), and that these differences typically begin around 12 years of age (Vorperian et al., 2009). In adulthood, the average length of the vocal folds is 29 mm in men and 22 mm in women. Men also typically have greater vocal fold mass and thickness.

Adult differences in vocal tract anatomy produce a naturally lower rate of vibration in the vocal folds of men compared to women (Aronovitch, 1976; Elyan, 1978; Lass, Hughes, Bowyer, Waters, & Bourne, 1976). Consequently, adult male F0 values tend to average approximately 120 Hz, while adult female speakers exhibit F0 values of approximately 220 Hz. Similarly, studies have also shown that the dissimilarities in vowel formant frequencies between men and women are due principally to differences in vocal tract length, shape, and size (e.g., Fant, 1960). The relatively larger and longer resonating cavities in the vocal tract of most male speakers results in lower formant frequency values (Childers & Wu, 1991; Coleman, 1971).

In children, anatomy appears not to be responsible for gender-related speech differences. Research examining the vocal tract anatomy of children generally indicates that gender-related dimorphism, which leads to differences in vocal tract dimensions, occurs during or after puberty.
A study by Fitch and Giedd (1999) used MRI to examine the vocal tract morphology (length, shape, and proportions) of 129 children and young-adults. The authors reported no significant anatomical sex differences in the vocal tract of pre-pubescent speakers younger than 12 years of age. In addition, more recent large-scale MRI studies (Fant, 1960; Vorperian et al., 2005; Vorperian et al., 2009) have supported the conclusion that although the vocal tract anatomy of prepubescent children undergoes periods of growth acceleration, vocal tract dimorphism for most children occurs at the peri-pubertal and post-pubertal stages of development.

Behavioral Factors for Speech Differences

Several studies with adults have shown that differences in the speech produced by men and women are greater than one would expect on the basis of anatomical differences in vocal tract dimensions alone (Mattingly, 1966). An early finding by Mattingly, based on a reanalysis of the Peterson and Barney (1952) data on vowel formants, was that differences in formant frequencies between men and women are not entirely attributable to differences in vocal tract size. Some researchers have concluded that gender-specific differences in speech may be due in part to learned or behavioral factors. These researchers have suggested that as a process of acculturation “men and women modify their articulation of the same phonetic elements to produce acoustic signals that correspond to the male-female archetype” (Sachs et al., 1973, p. 75).

In children, where sexual dimorphism of the vocal tract has yet to occur, behavioral factors may be responsible for the acoustic differences that have been documented between boys and girls. A significant contribution to this area of research was a large scale study (N = 436) by Lee et al. (1999) which found that the formant values for vowels produced by children 5 to 17 years of age differed according to the child’s gender. A subsequent reanalysis of these data by
Whiteside (2001) indicated that the formant differences could not be attributed solely to anatomic differences in vocal tract structures. Additional studies involving prepubescent children have also reported that the formant frequencies of boys and girls differ significantly (Bennett, 1981; Busby and Plant, 1995; Perry et al., 2001; Whiteside and Hodgson, 2000). Research investigating patterns of voice onset time (VOT) in the stop productions of preadolescent children (Whiteside & Marshall, 2001) has also found evidence of gender-specific speech development. This study reported that young girls exhibited longer VOT values than boys of a similar chronological age. Gender-related differences have also been found in how children utilize their voice, with findings showing that young boys are more prone to vocal nodules than are girls, possibly due to vocal aggressiveness (Roy, Holt, Redmond, & Muntz, 2007).

Several studies investigating the spectral properties of young children's obstruent productions have also found gender-specific differences (Fox & Nissen, 2005; Nissen & Fox, 2005, 2009). This research indicated that in children as young as 5 years of age the spectral characteristics of stop and fricative productions (i.e., spectral mean and slope) were significantly different for boys compared to girls. In a study by Fox and Nissen (2005), a discriminant analysis found that a combination of spectral parameters was similar to patterns exhibited by gender-matched adults. In other words, the boys exhibited production patterns similar to the men and the girls to those of the women. The authors concluded that these differences may be the result of behavioral or learned factors. However, the perceptual relevance and physiologic mechanisms for these acoustic differences remains unclear.

**Age and Gender Differences Specific to Prosody**

The conclusion that gender-linked differences in speech development are associated with learned or behavioral factors is further supported by studies that have examined aspects of
speech that are less impacted by anatomic variations in vocal tract structures, such as prosody. Research with adult speakers supports the notion that the nature of prosody differs by gender. It has been recognized that in conversational speech men often use a narrower F0 range than women (Eylan, 1978; Whiteside, 1996; Wu & Childers, 1991). Research has also indicated that men may use changes in F0 to accomplish different linguistic functions than women. A study by Whiteside (1996) reported that men tend to shift their speaking F0 to mark syntactic and phrase-final boundaries, while women marked these types of linguistic boundaries by pausing or using phrase-final lengthening.

This may be one reason why researchers have found differences between men and women in utterance duration. The study by Whiteside (1996) analyzed the speech of three men and three women to examine what happens acoustically to cause male speech to be relatively faster. Whiteside found the men and women were exhibiting gender-specific patterns in their use of pause. When the women paused, the duration of both the word and the phonetic segments prior to the pause increased. Moreover, the females tended to pause more often than the male speakers. In addition to using fewer pauses, the male speakers tended to reduce both vowels and consonants, which may be why they were found to speak more quickly (Whiteside, 1996).

Fitzsimons, Sheahan, and Staunton (2001) found a similar result, whereby men produced declarative and interrogative sentences with shorter durations than women, who tended to have higher pause-to-speech ratios. The researchers suggested the observed higher rate of pause in women as a possible explanation for longer verbal production times. In addition, Hillenbrand, Getty, Clark, and Wheeler (1995) found that women showed significantly greater vowel durations than either men or children.
Research has also reported gender-related differences in how speakers change vocal intensity to modify the prosodic patterns of their speech. Most studies in this area support the conclusion that men generally rely more on intensity to create prosody in their speech, while women typically rely more on changes in F0 (Byrd, 1992; Fitzsimons et al., 2001; Klatt & Klatt, 1990; Whiteside, 1996). While little is known about the acoustic differences in the way men and women produce contrastive stress, Cooper, Eady, and Mueller (1985) found that men produced contrastive stress by extending the duration of the emphasized word relative to other words in the sentence. However, they did not significantly change the F0 of the focus words when compared to the non-focus content words. Unfortunately, the study contained only male participants, so it remains unclear if female participants would have used similar prosodic patterns. A summary of the key findings from the literature on prosodic differences between men and women is presented in Table 1.

The use of prosody has also been investigated in children. Hornby and Hass (1970) found that 20 children, with a mean age of 4 years, were able to effectively use contrastive stress. Baltaxe (1984) compared the use of this form of prosody between typical, aphasic, and autistic children aged two to twelve years. The author found that all three groups of children reached only a 60% effective response rate. The typically developing children had the highest percentage of correct responses, followed by the children with aphasia, then the children with autism. The author interpreted the low response rate overall to the task complexity. Interestingly, the children produced a higher rate of responses for words functioning as the subject of the sentence, followed by the object, and then the verb. Another study completed by O’Halpin (1997) examined the way deaf children compared to typically developing eight year old children created contrastive stress. While the typically developing children used the typical stress cues of F0,
<table>
<thead>
<tr>
<th>Author and Year</th>
<th>Finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elyan, 1978; Whiteside, 1996; Wu &amp; Childers, 1991</td>
<td>Men often used a narrower F0 range than women in conversational speech.</td>
</tr>
<tr>
<td>Whiteside, 1996</td>
<td>Men marked syntactic and phrase-final boundaries by a shift their speaking F0. Women marked these boundaries by pausing or using phrase-final lengthening.</td>
</tr>
<tr>
<td>Whiteside, 1996</td>
<td>Men spoke more quickly due to less use of pause. Women paused more and thus increased the duration of both the word and the phonetic segments prior to the pause more often than men.</td>
</tr>
<tr>
<td>Whiteside, 1996</td>
<td>Men tended to reduce both vowels and consonants, (which may be why they were found to speak more quickly).</td>
</tr>
<tr>
<td>Fitzsimons, Sheahan, &amp; Staunton, 2001</td>
<td>Men produced declarative and interrogative sentences with shorter durations than women. Women tended to have higher pause-to-speech ratios.</td>
</tr>
<tr>
<td>Hillenbrand et al., 1995</td>
<td>Women showed significantly greater vowel durations than either men or children.</td>
</tr>
<tr>
<td>Byrd, 1992; Fitzsimons et al., 2001; Klatt &amp; Klatt, 1990; Whiteside, 1996.</td>
<td>Men generally relied more on intensity to create prosody in their speech. Women typically relied more on changes in F0 to create prosody in their speech.</td>
</tr>
<tr>
<td>Cooper, Eady, and Mueller, 1985</td>
<td>Men produced contrastive stress by extending the duration of the emphasized word. Men did not significantly change the F0 of the focus words compared to the non-focus words.</td>
</tr>
</tbody>
</table>
longer duration, and intensity to mark stress, the deaf children were found to use none of these acoustic cues to mark the target word. The author concluded that deaf children probably often fail to mark contrastive stress in their speech. Further studies have compared children with dysarthria with typical developing children, and have found that children with dysarthria rely more heavily on duration than do typically developing children (Patel & Campellone, 2009). Patel and Brayton (2009) studied whether children of different ages created contrastive stress differently. The authors found that 4 year old children relied more heavily on durational cues than 7 or 11 year olds.

**Purpose of the Study**

Although there is a body of literature investigating prosody in adults, additional research is needed on the way children use prosody. Thus, the purpose of this study was to find out more about how children create and apply the features of contrastive stress in their prosody, with a particular aim of identifying and describing any gender-based differences. Specifically, two experiments were conducted to evaluate the production and subsequent perception of contrastive stress. In the first experiment the production of contrastive stress was evaluated through examination of three relative acoustic measures: (a) F0 change, (b) intensity change, and (c) word duration change. The second experiment perceptually measured whether adults could tell boys from girls during the contrastive stress task. It was anticipated that this study would improve our understanding of prosody in children by addressing the following research questions:

1. Do young children use adult-like patterns of contrastive stress?

2. Do children mark contrastive stress differently as a function of speaker gender and the part of speech being emphasized?
3. Can listeners reliably identify the gender of the child speakers when listening to sentence length utterances?

Method

Experiment 1 – Acoustic Evaluation

Participants

Twenty children, between 8:0 and 9:11 years of age ($M = 9:2$), and a comparison group of ten adults between 21 and 28 years of age ($M = 23:9$) participated in this study. Each group had an equal number of male and female participants. All speakers were monolingual speakers of American English and had minimal exposure to a second language (i.e., not having lived outside of the United States for more than 6 months and having parents/guardians who also speak English as their native language). The adult participants, and the parents of the child participants reporting for their children, reported no diagnosed history of hearing, speech, or language problems. All the participants were required to pass a hearing screening prior to the collection of data. Thus, at the time of their participation all of the listeners exhibited pure-tone air-conduction thresholds $\leq 25$ dB HL at octave frequencies from 500 to 8000 Hz. The participants were recruited from the Brigham Young University community and surrounding areas. In addition, if participants exhibited perceptual signs of poor vocal health (e.g., laryngitis, vocal hoarseness, cold, etc.), the recording sessions were postponed until a later date.

Stimuli

Stimulus materials consisted of ten line-drawing pictures of simple everyday events. Each picture was designed to elicit four lexically identical sentences from each participant. One sentence required contrastive stress on the subject, verb, or object (the head word of the relevant phrase), as well as a baseline sentence produced in a naturalistic manner but without emphatic
stress being placed on any particular word. In this manner, a total of 40 sentences were recorded from each participant. The sentences produced across the ten picture elicitations were designed to be syntactically similar and of a similar length and complexity. A detailed list of the sentences used in the study can be found in Appendix B. In addition to the sentences analyzed in this study, a number of other speech samples (e.g., words in citation form, spontaneous speech samples, and picture descriptions) to be used in subsequent studies were collected during the recording session. However, these data were not analyzed in the current study and are therefore not described in further detail. Each recording session was completed in approximately 30 minutes.

**Procedures**

The researcher elicited baseline productions of the ten target sentence described above. Subsequently, the researcher presented to the participants the series of 10 pictures, one at a time. As each picture was presented, the participants were asked three questions: one each about the subject, the verb, and the object of the sentence. For example, the participants were shown a picture of a child in a swimming pool. They were then presented with the first question, intended to elicit a response about the subject of the target sentence, *Is a dog in the swimming pool?* The participant was then prompted to respond, *No, a CHILD is in the swimming pool.* The second question was constructed to elicit a response about the verb, for example, *Is a child playing video games in the pool?* The participant would then be prompted to respond, *No, a child is SWIMMING in the pool.* The third question was constructed to elicit a response about the object, for example, *Is a child swimming in the lake?* The participant was then prompted to respond, *No, a child is swimming in the POOL.* The order of the picture presentation was randomized for each participant. Prior to recording the productions of contrastive stress, each participant was familiarized with the research task. When the participant demonstrated familiarity with the task
by producing the target sentence appropriately, the experimental stimuli were then presented to
the participants.

**Recording**

The participants’ speech was recorded directly to a PC computer while each participant
was seated in a quiet-room. A high-quality head-set microphone (Shure 4011) positioned
approximately 2.5 centimeters from the participants' mouth was used to record the speech
samples. The recordings were sampled at a rate of 44.1 kHz and a quantization of 16 bits with
Adobe Audition software. Subsequently, sound files were archived to a PC computer hard drive
for further analysis. All recorded sentences were high-pass filtered at 70 Hz. In cases of
inaccurate articulation, peak clipping, or an error in the recording, the participant was asked to
repeat the test item and the stimulus was re-recorded.

**Measurement of Acoustic Variables**

*F0 measurement.* Mean F0 values of each target word were measured by extracting an F0
track (plotted over time) using Praat acoustic analysis software (version 5.1.20; Boersma &
Weenink, 2009). The extraction algorithm relied on autocorrelation, as described in Boersma
(1993). The F0 change measures were calculated by comparing the mean F0 of the target word
relative to the mean F0 of the same word in the baseline sentence with the identical linguistic
structure produced without any target emphasis. Since each individual has a different speaking
F0, it was necessary to convert from Hz to a semitone scale to enable comparisons across
speakers.

*Intensity measurement.* A relative measure of intensity was also obtained. This measure
was calculated by comparing the average root-mean-square (RMS) amplitude of each target
word to the mean intensity of the baseline production. Similar to the method described above, the
RMS was computed with Praat acoustic analysis software by extracting an intensity track from which a mean intensity value was calculated.

*Duration measurement.* The duration of each target segment was computed to the nearest millisecond (ms) using the Praat analysis software.

*Reliability of the Measures*

To examine the reliability of the extracted acoustic measures, speech samples from 10% of the speaker productions were selected and reanalyzed again by another individual. These additional sets of duration, intensity, and F0 measurements were extracted, recorded, and checked in the same manner as the original measures. Comparisons of the duration measures produced correlations of 0.95, F0 measures produced correlations of 0.85, and intensity measures were correlated at 0.99.

*Statistical Analysis*

The data in this experiment were analyzed using a repeated measures analyses of variance (ANOVA) to examine any significant acoustic variation (F0, intensity, duration) in the speakers’ productions as a function of participant age, gender, and the part of speech being analyzed. Partial eta squared ($\eta^2$) measures of effect size were also computed for any significant ANOVA results. Any required post hoc analyses consisted of pairwise comparisons, with Bonferroni adjustments for multiple comparisons.

*Experiment 2 – Perceptual Evaluation*

*Participants*

A group of 5 adult listeners between 18 and 40 years of age were recruited to evaluate the sentences produced by the child speakers in the first experiment of this study. The participants were native speakers of American English and reported no history of speech or language
disorder. At the time of their participation all of the listeners exhibited pure-tone air-conduction thresholds ≤ 25 dB HL at octave frequencies from 500 to 8000 Hz. All listeners were recruited from the Brigham Young University community and surrounding areas.

Stimuli

The listeners were presented with a total of 440 sentences previously recorded in Experiment 1. The sentences were composed of the four linguistic conditions (baseline, subject emphasis, verb emphasis, and object emphasis) from 5 randomly selected picture elicitations for each of the twenty child speakers (N = 400). In order to test intra-rater reliability, 10% of the samples (40 sentences) were randomly selected and replayed to the listeners. The intra-rater reliability was found to be 88.3% accuracy by comparing the reliability stimuli to the listener’s initial ratings.

Procedures

The listeners were instructed to listen to each stimulus sentence and perceptually evaluate whether the speaker was a boy or a girl. The sequence of presentation was randomized across the 440 stimulus sentences. The randomization, presentation, and subsequent recording of the listener’s perceptual judgments were controlled by customized software. The signal was routed from a computer hard drive via Sennheisser HD 650 headphones to the participant, who was seated in a single-walled sound booth meeting American National Standards Institute S3.1 standards with ears covered (American National Standards Institute, 1999). Prior to data collection, each participant listened to two sample tokens before rating the experimental stimuli. The listeners self-selected the intensity level of the presented stimuli, with a starting level of approximately 60 dB HL. The testing took place in a single 30-minute session.
**Statistical Analysis**

The data in this experiment were analyzed using descriptive statistics of the listeners’ ratings of the recorded speech samples. The listeners’ ability to correctly identify the gender of the speaker was calculated as a percentage across all the sentences as a group and across the individual listeners, speakers, and the part of speech intended to be emphasized.

**Results**

The acoustic results from the production experiment will be presented first. These show how the dependent variables of F0, intensity, and duration changed for the stressed words relative to the baseline condition (see Tables 2-4). The results of the perception experiment are reported next, showing how accurately the listeners identified the gender of each child across the different parts of speech that were stressed.

**Experiment 1**

A three-way repeated measures ANOVA was used to analyze the acoustic differences across the speaking conditions, with between-subjects factors of gender and age.

**Fundamental frequency**

F0 changed significantly across the parts of speech, \( F(2, 50) = 6.942, p = .002, \eta^2 = .22. \) Post hoc tests indicated \( (p < .05) \) that speakers decreased their F0 when emphasizing the subject of the target sentences \( (M = -.55 \text{ semitones}) \), yet increased their F0 when emphasizing verb \( (M = +.05 \text{ semitones}) \) and object \( (M = +1.26 \text{ semitones}) \) parts of the sentence. There was also a significant interaction effect between the part of speech and the age of the speaker, \( F(2, 50) = 3.418, p = .041, \eta^2 = .12. \) This was due to a slight decrease for the verb targets by the child speakers, whereas the adults exhibited a decrease in F0 for the subject parts of speech. These differences are illustrated in Figure 1. There was also an interaction of age group by gender, \( F(1,
Table 2. Change in Semitones for the Stressed Word Relative to the Baseline Condition by Age Group, Gender, and Part of Speech

<table>
<thead>
<tr>
<th>Age</th>
<th>Part of Speech</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Child</td>
<td>Subject</td>
<td>-0.35</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>Verb</td>
<td>-0.81</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Object</td>
<td>0.01</td>
<td>0.96</td>
</tr>
<tr>
<td>Adult</td>
<td>Subject</td>
<td>-1.03</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td>Verb</td>
<td>1.18</td>
<td>1.65</td>
</tr>
<tr>
<td></td>
<td>Object</td>
<td>1.67</td>
<td>3.72</td>
</tr>
</tbody>
</table>

Table 3. Intensity Change in dB for the Stressed Word Relative to the Baseline Condition by Age Group, Gender, and Part of Speech

<table>
<thead>
<tr>
<th>Age</th>
<th>Part of Speech</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Child</td>
<td>Subject</td>
<td>-0.58</td>
<td>2.67</td>
</tr>
<tr>
<td></td>
<td>Verb</td>
<td>-2.51</td>
<td>2.66</td>
</tr>
<tr>
<td></td>
<td>Object</td>
<td>-2.84</td>
<td>2.81</td>
</tr>
<tr>
<td>Adult</td>
<td>Subject</td>
<td>-1.25</td>
<td>2.05</td>
</tr>
<tr>
<td></td>
<td>Verb</td>
<td>-1.59</td>
<td>2.27</td>
</tr>
<tr>
<td></td>
<td>Object</td>
<td>-1.92</td>
<td>1.95</td>
</tr>
</tbody>
</table>

Table 4. Duration Change in Milliseconds for the Stressed Word Relative to the Baseline Condition by Age Group, Gender, and Part of Speech

<table>
<thead>
<tr>
<th>Age</th>
<th>Part of Speech</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Child</td>
<td>Subject</td>
<td>-42.45</td>
<td>50.53</td>
</tr>
<tr>
<td></td>
<td>Verb</td>
<td>-35.82</td>
<td>54.71</td>
</tr>
<tr>
<td></td>
<td>Object</td>
<td>-44.18</td>
<td>49.54</td>
</tr>
<tr>
<td>Adult</td>
<td>Subject</td>
<td>32.20</td>
<td>78.95</td>
</tr>
<tr>
<td></td>
<td>Verb</td>
<td>53.18</td>
<td>105.75</td>
</tr>
<tr>
<td></td>
<td>Object</td>
<td>30.68</td>
<td>61.91</td>
</tr>
</tbody>
</table>
Figure 1. Fundamental frequency change for the stressed word relative to the baseline condition by part of speech for adults and children.
Among the adults the female speakers decreased their F<sub>0</sub> for the target word, while the male speakers increased the F<sub>0</sub> of the emphasized word. The child speakers, however, exhibited an opposite pattern. These differences are illustrated in Figure 2.

**Intensity**

There were no significant main effects for intensity for the between-subjects factors of age or gender, or the within-subjects factor of POS. In addition, no significant interactions were found.

**Duration**

Results from the ANOVA indicated a significant difference between the adults and the children, \( F(1, 25) = 4.778, p = .038, \eta^2 = .160 \). Overall the children shortened the duration of the words to be emphasized (\( M = -26 \) ms), whereas the adults lengthened the target words (\( M = +37 \) ms). There was also a significant interaction between the part of speech and the adult vs. the child group, \( F(2, 50) = 4.21, p = .02, \eta^2 = .14 \). This effect was caused in the adult group by an increase in duration between the baseline production of the verbs relative to the subject and object targets examined. These differences are illustrated in Figure 3. The analysis indicated no other significant differences for duration.

**Experiment 2**

The perception task required 5 adult listeners to identify the gender of the children based on the children’s sentence productions. Due to malfunctioning equipment, however, the data for one of the listeners was not analyzable; therefore, data from 4 adult listeners were evaluated. The overall accuracy in identifying the gender of the child speakers was 87.1%. Differences in listener accuracy were found across the part of speech the child was intending to emphasize. These differences were minimal, ranging from 85.8% to 88.8%. The individual listeners’ ability
Figure 2. Fundamental frequency change for the stressed word relative to the baseline condition by gender for adults and children.
Figure 3. Duration change for the stressed word relative to the baseline condition by part of speech for adults and children.
to recognize the children’s gender ranged from 83.8 to 90.3%. Table 5 reports the gender identification accuracy for each of the speakers. Of the 20 participants, the gender of 14 (7 boys and 7 girls) was accurately recognized over 90% of the time; notably, the boys were generally recognized more accurately than the girls. The lowest gender recognition score for the boys was 66.3%, while two of the girls had gender recognition scores at less than 50%.

Discussion

The aim of this study was to examine children’s use of prosody to mark contrastive stress. The initial experiment was designed to describe the acoustic correlates of contrastive stress and examine the participants’ prosodic patterns as a function of speaker age, gender, and the part of speech being emphasized. The second experiment investigated the accuracy with which the gender of the speakers could be identified by a group of adult listeners.

Experiment 1

Previous work has shown that speakers primarily use changes in F0 to signal contrastive stress (Atkinson, 1978; O’Shaughnessy, 1979), with male speakers typically using a more narrow F0 range than women (Elyan, 1978; Whiteside, 1996; Wu & Childers, 1991). However, the present study found that neither the adult nor child speakers followed this trend consistently. Rather, the girls and the men had a slight increase in their F0 overall, while the women and the boys were actually found to decrease their F0 for the target production. The use of F0 change to mark contrastive stress was found to differ based on the part of speech, with a decrease for the verb targets by the child speakers and a decrease by the adults for the subject parts of speech. Although these differences were statistically significant, the statistical power of the interaction effect was relatively low ($\eta^2 = .12$) and it is unclear if the differences in F0 change would have been perceptually salient ($< 2$ semitones).
### Table 5. Accuracy of Gender Identification Overall and by Part of Speech, Listener, and Speaker

<table>
<thead>
<tr>
<th></th>
<th>Accuracy %</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall</strong></td>
<td>87.1</td>
</tr>
<tr>
<td><strong>Part of Speech</strong></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>87.3</td>
</tr>
<tr>
<td>Subject</td>
<td>88.8</td>
</tr>
<tr>
<td>Verb</td>
<td>86.5</td>
</tr>
<tr>
<td>Object</td>
<td>85.8</td>
</tr>
<tr>
<td><strong>Listeners</strong></td>
<td></td>
</tr>
<tr>
<td>Listener 1</td>
<td>83.8</td>
</tr>
<tr>
<td>Listener 2</td>
<td>88.0</td>
</tr>
<tr>
<td>Listener 3</td>
<td>86.3</td>
</tr>
<tr>
<td>Listener 4</td>
<td>90.3</td>
</tr>
<tr>
<td><strong>Speakers</strong></td>
<td></td>
</tr>
<tr>
<td>Male 1</td>
<td>100.0</td>
</tr>
<tr>
<td>Male 2</td>
<td>98.8</td>
</tr>
<tr>
<td>Male 3</td>
<td>91.3</td>
</tr>
<tr>
<td>Male 4</td>
<td>93.8</td>
</tr>
<tr>
<td>Male 5</td>
<td>97.5</td>
</tr>
<tr>
<td>Male 6</td>
<td>95.0</td>
</tr>
<tr>
<td>Male 7</td>
<td>76.3</td>
</tr>
<tr>
<td>Male 8</td>
<td>66.3</td>
</tr>
<tr>
<td>Male 9</td>
<td>93.8</td>
</tr>
<tr>
<td>Male 10</td>
<td>88.8</td>
</tr>
<tr>
<td>Female 1</td>
<td>96.3</td>
</tr>
<tr>
<td>Female 2</td>
<td>36.3</td>
</tr>
<tr>
<td>Female 3</td>
<td>98.8</td>
</tr>
<tr>
<td>Female 4</td>
<td>98.8</td>
</tr>
<tr>
<td>Female 5</td>
<td>73.8</td>
</tr>
<tr>
<td>Female 6</td>
<td>95.0</td>
</tr>
<tr>
<td>Female 7</td>
<td>96.3</td>
</tr>
<tr>
<td>Female 8</td>
<td>98.8</td>
</tr>
<tr>
<td>Female 9</td>
<td>98.8</td>
</tr>
<tr>
<td>Female 10</td>
<td>47.5</td>
</tr>
</tbody>
</table>
Intensity values for the stressed target words were found to decrease for both the adult and child participants. Considering previous research that has shown that men, in particular, generally rely on the acoustic cue of intensity to create prosody in their speech (Byrd, 1992; Fitzsimons et al., 2001; Klatt & Klatt, 1990; Whiteside, 1996) these data were unexpected. As will be discussed in greater detail below, such differences may be due to the child speakers continuing linguistic development, idiosyncratic speaker characteristics, or more likely, external factors related to the elicitation method or measurement technique.

For the adult speakers, findings for the dependent measure of duration change did follow an expected pattern, in that the durations of the target words tended to lengthen when emphasized. However, the child speakers failed to follow the adult-like pattern of using word lengthening to mark emphatic stress. It may be that the child speakers were continuing to develop their linguistic ability to mark contrastive stress and the skill of lengthening certain word targets for emphatic purposes has not yet developed. The method for calculating word duration in this study is consistent with numerous studies in acoustics, thus it is unlikely that the result was influenced by the measurement technique. However, it may be that since the baseline comparison was collected prior to the target sentences, the children would have been less familiar with the sentences and therefore exhibited extended durations. It is possible that incomplete cognitive-linguistic or neuromuscular development caused this familiarity effect to have a greater impact on the child speakers compared to the adults.

Generally, this study did not find consistent male-female acoustic differences in the use of prosody to mark contrastive stress for either the adults or the child speakers. Although previous research has indicated that some aspects of prosody are marked differently by men and women (e.g., Byrd, 1992; Fitzsimons et al., 2001; Klatt & Klatt, 1990; Whiteside, 1996), the
production of contrastive stress specifically has received less attention. If gender differences are not found in adults, it is not surprising that they were absent in typically developing preadolescent children.

Although previous studies have identified increases in frequency, intensity, and duration for emphatically stressed words, the present findings revealed patterns that did not consistently conform to these expectations. Differences between previous research and the findings of this study may have been due to several factors.

*Potential Factors Influencing the Results*

*Complex nature of prosody.* As previously described, speakers use prosody to accomplish a variety of communicative functions, both affective and linguistic in nature. Prosodic elements are generally considered to be suprasegmental. However, depending on which aspect of prosody is being expressed, the linguistic unit of importance could be at the level of the syllable, word, or sentence. In addition, the mechanisms by which prosody is produced vary across a number of different acoustic cues, such as F0, intensity, and duration change, as well as differences in tempo, pause, and possibly vowel quality. The research literature investigating prosody has provided insights into a number of these aspects, but does not provide a comprehensive understanding of the nature of prosody. Considering this complexity, it may be that the acoustic patterns found in additional aspects of prosody cannot be directly generalized to the process of contrastive stress.

*Participant training.* The participants (or at least the children) may have been uncertain about how to perform the task prior to data collection. This uncertainty, exacerbated by the contrast between the picture and the spoken sentence, may have caused the participants to become tentative when approaching the target word. The F0 data appear to support this
speculation. The children and the adults lowered their F0 on the subject, increased it a little when saying the verb targets, and showed the greatest increases on the object targets. Additional training prior to the collection of the experimental data may have increased the participants’ ability to produce the sentences in a more predictable manner. However, the disadvantage of providing additional training would be the risk that the participants might begin to merely echo the researcher’s patterns of speech instead of talking more naturally.

Task and stimulus design. It is possible that the nature of the elicitation task and the manner in which the stimuli were designed impacted the speakers’ productions. It is possible that the task was too redundant, too facile, or too decontextualized to support prosody that is more typical of spontaneous speech. The measures of F0 and intensity in the adults showed the opposite of what was anticipated based on a review of the literature. In other words, perhaps because the task was too easy or predictable, the adults quickly learned the expectation of the elicitation task, which then became repetitive or boring. If the speakers lost interest in the task, they may have been more likely to respond in an automatic way, which may have caused them to respond with unnaturally flat intonation in terms of both F0 and intensity. In a natural setting it would be unlikely that a speaker would need to use contrastive stress in such a frequent and repetitive manner.

Every attempt was made to structure the stimuli in a uniform manner. The word structure and lexical demands of the sentences were designed to be very familiar to the participants, especially the children. However, it is possible that the sentences had too little linguistic variety, considering the study was seeking measures of natural prosody.

Individual speaker characteristics. Another reason for the reported findings may be that the study involved a limited number of speakers, especially in the adult control group (N=10).
may be necessary to gather data from a larger group of participants to accurately examine the prosodic patterns in speech. This conclusion is supported by the high degree of variability found within the measures analyzed in this study (see Tables 2-4). Considering the multiple acoustic mechanisms by which a speaker can mark emphasis and the ability to trade or substitute prosodic cues (Beach, 1991), it is possible that individual speakers exhibit a high degree of idiosyncratic behavior when producing prosodic elements.

*Sensitivity of acoustic measurements.* Another reason for the unexpected results may be the way in which the data were measured. The F0 and intensity of the target words were calculated relative to the mean values of a baseline production of the same word produced in the same linguistic context. A baseline comparison was used to accommodate for the variations in F0 and intensity that naturally occur in speech, even in an utterance with no intended emphasis or stress. However it can be argued that the baseline utterances may have been produced in a way that actually suppressed some of the emphatic acoustic elements in the target words. In future research it may be valuable to consider the target word relative to both a separate baseline production and to the neighboring words within the same utterance. For example, the intensity measure could be based on a ratio of the target word to the average intensity of words immediately before and after it.

In addition, rather than computing the acoustic measures based on mean values across the entire target word, it may be more accurate to measure the peak values. For example, it may be that a word production with a sharp burst in intensity has a similar or greater perceptual impact on a listener compared with a production that has a sustained increase in intensity. It also may be possible to calculate a relative value based on both the mean and peak values.
Experiment 2

The second experiment was designed to determine whether adults could reliably tell boys from girls in the speaking tasks. Overall, the adult listeners identified the gender of the child speakers with an accuracy of 87.1%. This result is similar to the findings of Sachs et al. (1973) which found that listeners could identify the gender of a child’s speech with 81% accuracy. In a study by Perry et al. (2001), listeners correctly identified the gender of speakers as young as age 4 at a rate significantly better than chance. Since the present study involved girls and boys starting at age eight, they were well within the age range for which accurate identification might be expected. Since previous research has shown that only after age 12 are there significant gender differences in anatomical growth of the vocal tract (Fant, 1960; Vorperian et al., 2005; Vorperian et al., 2009), it can be suggested that children’s speech may exhibit gender specific characteristics that are learned or behavioral in nature (Fox & Nissen, 2005; Nissen & Fox, 2005; 2009; Sachs et al, 1973; Whiteside, 2001). The results of this experiment indicate that such differences may be found in the suprasegmental or prosodic aspects of speech. However, additional perceptual studies should be conducted to address this question more directly. It would be of particular interest to investigate why the gender of several children was judged incorrectly (e.g., female speaker 2 at 36.3%). A cursory examination of the baseline acoustic data indicated that the confused judgments were likely not a result of an increase or decrease in speaking F0.

Conclusions

Despite these limitations, it is hoped that the findings found in this study will promote greater understanding of speech prosody, namely how adults and children use acoustic cues to signal contrastive stress. Findings of the perceptual experiment from this study indicate the need for further investigations into possible gender specific differences in children’s speech.
communication and whether prosody plays a role in any such distinctions. In addition, it is anticipated that the methodological insights discussed will facilitate more accurate and efficient studies in this area in the future.
References


Appendix A – Informed Consent

Child Assent to be a Research Subject

We want to tell you about a research study we are doing. A research study is a special way to find out about something. We are trying to find out more about speech patterns in children. You are being asked to join the study because you have never had a speech or hearing problem.

If you decide that you want to be in this study, this is what will happen. It takes less than half an hour.

1. We will check your hearing to see if it is okay.
2. You will look at pictures and tell us what you see.
3. You will play a game similar to "Go Fish".
4. We will record your speech with a microphone.

Can anything bad happen to me?
Nothing in this study will hurt you.

Can anything good happen to me?
Being in this study won’t help you, but we hope to learn more about how children speak.

Do I have other choices?
You can choose not to be in this study

Will anyone know I am in the study?
We won’t tell anyone you took part in this study. When we are done with the study, we will write a report about what we found out. We won’t use your name in the report.

You will receive $10 in the form of cash or a gift certificate for being in this study. Before you say yes to be in this study, be sure to ask the person helping with the study to tell you more about anything that you don’t understand.

What if I do not want to do this?
You don’t have to be in this study. It’s up to you. If you say yes now, but you change your mind later, that’s okay too. All you have to do is tell us.

If you want to be in this study, please sign or print your name.

☐ Yes, I will be in this research study. □ No, I don’t want to do this.

________________________________________  __________________________  ____________
Child’s name                              Signature of the child          Date

________________________________________  __________________________  ____________
Person obtaining Assent                  Signature                        Date
Parental Permission for a Child to Be a Research Subject

Introduction
The purpose of this research experiment is to examine differences in the way that words and sentences are spoken by people of different ages. Your child is being invited to participate in this study because he/she is a native speaker of English with no history of any speech, language, or hearing disorders. This experiment is being conducted under the supervision of Dr. Shawn Nissen, an associate professor in the Department of Communication Disorders at Brigham Young University.

Procedures
In this experiment your child will be asked to (1) participate in a standard hearing and speech screening, and (2) produce a series of everyday words and short sentences. These words and sentences will be collected by asking your child to describe a series of pictures depicting everyday events, such as a child swimming or playing baseball, and participate in a game similar to "Go Fish". Your child’s speech will be recorded with a microphone into a computer. The entire session will take approximately 45 minutes.

Risks/Discomforts
There are minimal risks for participation in this study.

Benefits
There are no direct benefits to participants. However, it is hoped that through your child’s participation researchers will learn more about developing speech patterns in children.

Confidentiality
All information provided will remain confidential and will only be reported as group data with no identifying information. All data, including digital recordings of your child’s responses will be kept on a password protected computer in a locked laboratory and only those directly involved with the research will have access to them.

Compensation
Your child will be paid $10.00 in the form of cash or a gift certificate for participation in this study.

Participation
Participation in this research study is voluntary. Your child has the right to refuse to participate and the right to withdraw later without any penalty.

Questions about the Research
If you have questions regarding this study, you may contact Dr. Shawn Nissen at (801) 422-5056 or at shawn_nissen@byu.edu.

Questions about your child’s Rights as a Research Participant
If you have questions regarding your child’s rights as a research participant, you may contact the BYU IRB Administrator, A-285 ASB, Brigham Young University, Provo, UT, 84602 or at (801) 422-1461.

I have read and fully understand the consent form. Any questions have been answered to my satisfaction. I give permission for my child to participate in this research.

Signed: ________________________________________ Date: _______________
(signature of participant’s parent or legal guardian)

Child’s Name: ___________________________________
Consent to be a Research Subject (Speaker)

Introduction
The purpose of this research experiment is to examine differences in the way that words and sentences are spoken by people of different ages. This experiment is being conducted under the supervision of Dr. Shawn Nissen, an associate professor in the Department of Communication Disorders at Brigham Young University. You are invited to participate because you are a native English speaker with no known history of a speech, language or hearing problem.

Procedures
In this experiment you will be asked to (1) participate in a standard hearing and speech screening, and (2) produce a series of everyday words and short sentences. You will be asked to describe a series of pictures depicting everyday events, such as a child swimming or playing baseball, and participate in a game similar to the child's game "Go Fish". You will be recorded with a microphone into a computer. The entire session will take approximately 45 minutes.

Risks/Discomforts
There are minimal risks for participation in this study.

Benefits
There are no direct benefits to participants. However, it is hoped that through your participation researchers will learn more about differences in speech patterns between children and adults.

Confidentiality
All information provided will remain confidential and will only be reported as group data with no identifying information. All data, including digital recordings of your responses will be kept on a password protected computer in a locked laboratory and only those directly involved with the research will have access to them.

Compensation
You will be paid $10 for your participation in this study.

Participation
Participation in this research study is voluntary. You have the right to withdraw at anytime or refuse to participate entirely without jeopardy.

Questions about the Research
If you have questions regarding this study, you may contact Dr. Shawn Nissen at (801) 422-5056 or shawn_nissen@byu.edu.

Questions about your Rights as Research Participants
If you have questions regarding your rights as a research participant, you may contact the BYU IRB Administrator, A-285 ASB, Brigham Young University, Provo, UT, 84602 or at (801) 422-1461.

I have read and fully understand the consent form. Any questions have been answered to my satisfaction. I give my consent to participate in this research.

Signature: _______________________________  Date: _________

Printed Name: _______________________________
Consent to be a Research Subject (listener)

Introduction
The purpose of this research experiment is to examine differences in the way that sentences are spoken by people of different ages. This experiment is being conducted under the supervision of Dr. Shawn Nissen, an associate professor in the Department of Communication Disorders at Brigham Young University. You are invited to participate because you are a native English speaker with no known history of a speech, language or hearing problem.

Procedures
Participation in this study will involve one visit of approximately 30 minutes, which will take place in a research laboratory in the John Taylor Building at BYU. You will be asked to listen to individual sounds, words, or sentences spoken by children and adults and respond regarding your perception of the gender of the speaker and the nature of the stress patterns you hear in their speech.

Risks/Discomforts
There are minimal risks for participation in this study.

Benefits
There are no direct benefits to participants. However, it is hoped that through your participation researchers will learn more about differences in speech patterns between children and adults.

Confidentiality
All information provided will remain confidential and will only be reported as group data with no identifying information. All data, including records of your listening responses, will be kept on password protected computers in a locked laboratory and only those directly involved with the research will have access to them.

Compensation
You will be paid $10 for your participation in this study.

Participation
Participation in this research study is voluntary. You have the right to withdraw at anytime or refuse to participate entirely without jeopardy.

Questions about the Research
If you have questions regarding this study, you may contact Dr. Shawn Nissen at (801) 422-5056 or shawn_nissen@byu.edu.

Questions about your Rights as Research Participants
If you have questions regarding your rights as a research participant, you may contact the BYU IRB Administrator, A-285 ASB, Brigham Young University, Provo, UT, 84602 or at (801) 422-1461.

I have read and fully understand the consent form. Any questions have been answered to my satisfaction. I give my consent to participate in this research.

Signature: ____________________________ Date: _______

Printed Name: ____________________________
Appendix B – Elicitation Stimuli


Target: The boy is swimming in the pool.
Subject stress elicitation: The dog is swimming in the pool?
Verb stress elicitation: The boy is reading in the pool?
Object stress elicitation: The boy is swimming in the ocean?
Target: The lady is picking the flower.
Subject stress elicitation: The man is picking the flower?
Verb stress elicitation: The lady is watering the flower?
Object stress elicitation: The lady is picking the apple?

Target: The horse is jumping the fence.
Subject stress elicitation: The turtle is jumping the fence?
Verb stress elicitation: The horse is kicking the fence?
Object stress elicitation: The horse is jumping the moon?
Target: The man is riding the horse.
Subject stress elicitation: The lady is riding the horse?
Verb stress elicitation: The man is feeding the horse?
Object stress elicitation: The man is riding the car?
Target: The boy is hitting the baseball.
Subject stress elicitation: The monkey is hitting the baseball?
Verb stress elicitation: The boy is kicking the baseball?
Object stress elicitation: The boy is hitting the balloon?

Target: The dog is following the girl.
Subject stress elicitation: The girl is following the girl?
Verb stress elicitation: The dog is leading the girl?
Object stress elicitation: The dog is following the tractor?
Target: The boy is eating the pizza.
Subject stress elicitation: The mother is eating the pizza?
Verb stress elicitation: The boy is throwing the pizza?
Object stress elicitation: The boy is eating the jello?

Target: The boy is carving the pumpkin.
Subject stress elicitation: The baby is carving the pumpkin?
Verb stress elicitation: The boy is eating the pumpkin?
Object stress elicitation: The boy is carving the table?
Target: The girl is mowing the lawn.
Subject stress elicitation: The dog is mowing the lawn?
Verb stress elicitation: The girl is watering/raking the lawn?
Object stress elicitation: The girl is mowing the driveway?

Target: The boy is baking the cookies.
Subject stress elicitation: The grandma is baking the cookies?
Verb stress elicitation: The boy is eating the cookies?
Object stress elicitation: The boy is baking the carrots?