August 2013

Perceptual Correlates of Acoustic Measures of Vocal Variability

Bree Ann Cumbers
University of Wisconsin-Milwaukee

PERCEPTUAL CORRELATES OF ACOUSTIC MEASURES OF VOCAL VARIABILITY

by

Bree A. Cumbers

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Science

in Communication Sciences and Disorders

at

The University of Wisconsin-Milwaukee

August 2013

ABSTRACT

PERCEPTUAL CORRELATES OF ACOUSTIC MEASURES OF VOCAL VARIABILITY

by

Bree A. Cumbers

The University of Wisconsin-Milwaukee, 2013
Under the Supervision of Marylou Pausewang Gelfer

This study investigated relationships between acoustic measures of vocal variability (pitch sigma, SFF range) and perceptual ratings of vocal variability during a reading task. Fifteen male (19-30 years of age) and nineteen female speakers (20-30 years of age) who were recorded reading the Grandfather Passage provided the stimuli for the listening task. From these samples, 30 were selected as representing a continuum of degrees of vocal variability. Male (N = 15) and female (N = 15) samples were presented to listeners separately. Thirty graduate students in Communication Sciences and Disorders who had a course background in voice supplied the perceptual judgments of these samples. The listeners provided perceptual judgments of vocal variability on a 7-point Likert scale (1 defined as "complete monotone" and 7 defined as "extreme variability"). Results indicated a strong positive correlation between acoustic measures of vocal variability and listener judgments of pitch variability, significant at the $p < .01$ level.

This study also investigated whether acoustic measures of vocal variability (pitch sigma, SFF range) in males differ significantly from these acoustic measures of vocal variability in females. Results showed no significant differences between male and female voices for either acoustic measure. Additional research is needed to determine whether there are differences between male and female voices in terms of perceptual measures of vocal variability.

This study also reported speaking fundamental frequency (SFF) characteristics of young adults during reading. Chosen measures included mean SFF, pitch sigma, and SFF range. Results showed that males averaged an SFF of 122.73 Hz, a pitch sigma of 2.18 STs, and an SFF range of 11.33 STs. Females averaged an SFF of 215.92 Hz, a pitch sigma of 2.27 STs, and an SFF range of 12.05 STs. Comparisons with earlier literature revealed differences, possibly relating to adjustment of analysis range.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to extend my heartfelt thanks to the following people:

My thesis advisor, Dr. Marylou Gelfer. Your interest and passion for research is truly inspiring. I could not have asked for a better advisor, editor, and teacher to guide me through this process. Many, many thanks, Dr. Gelfer.

My thesis committee, Dr. James Bashford and Dr. Heuer. Thank you for all of your honest feedback.

Communication Sciences and Disorders Graduate Program Class of 2013 and 2014. I am so grateful to those that took time away from their busy lives to participate in this study. Thank you!

My family and friends. Thank you for all of your support.

Brandon. Thank you for supporting me every step of the way and helping me to always see the bright side in life. You make me smile, and I love you.

Perceptual Correlates of Acoustic Measures of Vocal Variability

Introduction

Webster's dictionary defines monotone as "a series of sounds of uniform pitch" **(**Allee, 1986, p. 243**).** Most people do not speak at the same pitch, that is, in a monotone, due to linguistic and communicative demands. Speakers have to vary their pitch to enhance their message. Sentence type, word and sentence stress, and affective content are all denoted through variations in the voice (Baken and Orlikoff, 2000). Failure to vary one's pitch would significantly diminish the ability to communicate emotions and in some cases to express the message itself. For example, a speaker who says "I love this place" could use pitch variability to convey either the surface meaning of the sentence or, through sarcastic inflections, the opposite of the surface meaning.

Vocal variability is the amount of variation that can be perceived in a voice. Acoustically, these variations can be measured by examining the amount of change in the fundamental frequency of a speech or voice sample. *Fundamental frequency* (F_o) is the rate at which the vocal folds vibrate (i.e., open and close during production of a sound). The vibratory rate of the vocal folds is measured in terms of cycles per second, or hertz (Hz). In general, when the vocal folds vibrate at a faster rate, a higher pitched sound is perceived. According to Ferrand (2007), pitch, the perceptual correlate of F_o, "… is how we perceive the sensation of sound as being high or low on a musical scale" (p.33). So examining changes in F_o should correlate with the magnitude of perceived movement of a person's pitch from high to low or low to high and back again.

The primary factors that affect the rate of vocal fold vibration are related to the thickness, length, and elasticity of the vocal folds. A shorter, thicker, and somewhat lax vocal fold will vibrate at a slower rate (and thus produce a lower pitch) than a longer, thinner, more tense fold, which will produce a higher pitch (Boone and McFarlane, 2000). Contraction and relaxation of muscles in the larynx alter the length and thickness of the vocal folds. Change in the length and thickness of the vocal folds enables both men and women to produce a large range of frequencies. According to Boone and McFarlane (2000), the average length of the male vocal fold is approximately 17-20 mm relative to 12-17 mm, the average length of the female vocal fold. These differences in length can explain the speaking fundamental frequency differences between men and women.

Speaking Fundamental Frequency

When talking about conversational or connected speech, the concept of *speaking fundamental frequency* (SFF) is more useful than F_o. While F_o is a general term that can be applied to any vocal production (for example, a prolonged vowel), SFF is the average rate at which the vocal folds vibrate during *connected speech*. During speech, the vocal folds typically produce a *range* of frequencies and are continually changing their vibratory patterns. Therefore, identifying SFF requires that a measure of central tendency be utilized. Measuring one's SFF during connected speech will be indicative of the central tendency of the frequency that is commonly employed by the speaker. A lower SFF will be perceived as lower in pitch compared to a higher SFF.

Measures of central tendency used in determining SFF could include mean, median, or mode. *Mean* SFF is what most people consider the "average," or the sum of all measurements (frequencies, in this case) divided by the total number of measurements (waves, in this case; Baken and Orlikoff, 2000). Typically, mean SFF is the most commonly reported measure of central tendency. *Median* SFF is located at the middle of the fundamental frequency distribution of a speech sample. In other words, median SFF is the point at which half of the values in the sample are located above it and half are located below. The median is found by arranging the frequency values in the speech sample in increasing (or decreasing) order and locating the midpoint (Baken and Orlikoff, 2000). The *modal* SFF is the particular frequency value that occurs the most often in the speech sample (Baken and Orlikoff, 2000). Both median and modal SFF are not used as extensively as mean SFF in current research. Baken and Orlikoff (2000) summarized a variety of studies that investigated normative SFF values. These authors reported that the mean SFF during reading for males from 20-92 years of age was found to be between 107.1 – 146.3 Hz in the research cited, whereas the mean SFF for females from 20-94 years of age was typically between 188.6 – 224.3 Hz. These values were derived by Baken and Orlikoff (2000) from a summary of studies that were available at the time of their book's publication.

Maximum Phonational Frequency Range

There are many different ways to quantify frequency *variability* for a given speaker. The way to determine the greatest extent of variability that the human vocal mechanism is capable of is to measure the entire range of frequencies that a person is able to produce with maximal effort. This range of frequencies, from lowest to highest, is called *maximum phonational frequency range* (MPFR; Baken and Orlikoff, 2000). Various researchers have examined MPFR and found that this range was about three octaves for normal young adults (Baken and Orlikoff, 2000) or approximately 36 semitones (ST)[1]. However, the range of frequencies used in *conversational* speech is quite different than the MPFR.

Speaking Fundamental Frequency Range

The range of fundamental frequencies in conversational speech can be obtained by finding the difference between the highest and lowest frequency found in a connected speech sample. This range, called the SFF range, has been reported as a measurement of fundamental frequency variability for decades. Researchers have examined SFF range in both reading and spontaneous speech. In their review of normative studies, Baken and Orlikoff (2000) reported that the SFF range for males was typically found to be between 16.84-19.40 semitones (ST) during reading and between 16.79-19.78 ST during spontaneous speech. Similar to males, females demonstrated an SFF range from 17.70-19.10 ST during reading. These values come from a summary of studies that were reported in Baken and Orlikoff (2000).

Semitones and Vocal Variability

Semitones are typically used to describe variability measures. This is because hertz values do not represent pitch perception accurately, and present a distorted view of actual variability present in a voice. As mentioned previously, a semitone is based on the chromatic musical scale. This scale consists of 12 notes, where each note is separated from the next by a musical half-step, or a semitone (Behrman, 2007). One octave consists

[1] A semitone is based on "the Western cultural musical scale that consists of 12 tones. Each tone or pitch is separated from its neighbor by a semitone, the interval of one half-step" (Behrman, 2007, p. 45). Semitone relationships to musical notes, Hz values, and semitone numbers have been standardized by the Acoustical Society of America (1960).

of 12 half-steps, or semitones. This means that a semitone is 1/12th of an octave. In addition, an octave is perceived whenever frequency in hertz doubles. The latter is the source of disconnect between perceived pitch and frequency measured in hertz. For example, the difference between D_4 (D in the fourth octave), which is 293.7 Hz, and $D\#_4$, which is 311.1 Hz, is 17.4 Hz. However, the difference in the octave above, between D_5 to $D\#_5$, is 35 Hz (587.3 for D5 and 622.3 Hz for D#5). This difference between notes in hertz in the upper versus lower octaves reflects the ears' "…greater sensitivity to differences between lower pitches compared to higher pitches" (Behrman, 2007, p.46). Thus, the interval between D_4 and $D\#_4$ *sounds* the same to the human ear as the interval between D_5 and $D\#_5$, even though the latter includes a wider range of hertz.

An example of the difficulty caused by measuring SFF range in hertz is as follows: a male whose frequencies range from 100-200 Hz while speaking has an SFF range of 100 Hz. Since the hertz value doubled, this is 1 octave, or 12 semitones. However, a female whose speaking frequencies range from 200-400 Hz has an SFF range of 200 Hz. The female's hertz value also doubled so, in perceptual terms, this is also a range of 1 octave, or 12 semitones, and would be perceived as such by a listener. In terms of octaves and semitones, the ranges of the male and female speakers are perceptually equal. However, if SFF range is considered in hertz, the male appears to have the lesser range (or less variability). Individuals with lower SFFs will always appear to have smaller ranges than those with high SFFs. Semitones are absolutely necessary when reporting measures of variability because they allow such measures to be related to meaningful perceptual units. Hertz should always be avoided when reporting measures of variability due to the non-linearity (i.e., logarithmic nature) of pitch perception and the lack of equivalence between hertz and perceptual units.

Although using semitone measures is more representative of a speaker's perceptual variability during conversation than reporting variability in hertz, there are still several challenges associated with SFF range no matter how it is measured. The most obvious challenge is that the SFF range only reports the overall extreme values found in the speaking sample, and thus may give a distorted view of variability. For example, consider the fundamental frequencies found in the sample below:

110.7	109.7	110.1	125.3	111.3	109.4

The mean fundamental frequency of the sample is 112.7 and the range is 15.9 Hz (125.3-109.4). The range is clearly enlarged by the outlying value of 125.3 Hz, thus creating an inaccurate picture of fundamental frequency variability. Additionally, when measuring extreme values, it is difficult to determine what is an artifact and what is actual vocal fold vibration. At the extremes, researchers have to examine the waveform and attempt to discern what is legitimately produced by the human vocal mechanism and what sounds do not truly represent vocal fold vibration. This causes the measure of SFF range to be somewhat subjective in nature. Identification of the lower limit of SFF range is further complicated due to the need to exclude vocal fry from SFF range calculations.

Vocal fry, or pulse register, is at the lower end of the frequency scale where the vocal output is perceived as pulsatile in nature (Baken and Orlikoff, 2000). A vocal register refers to a "…specific range of fundamental frequencies characterized by a particular mode of vocal fold vibration resulting in a particular quality" (Ferrand, 2007, p.139). Typically, three registers are identified, which Hollien (1974) labeled as pulse, modal, and loft. Ferrand (2011) states that “pulse is the lowest register, modal refers to the middle range of frequencies most often used in conversational speech, and loft is the highest register” (p. 59). Each register consists of distinctive perceptual qualities and is produced by a differing mode of vocal fold vibration. It has been hypothesized that in pulse register, the vocal folds are held so tightly together that they only briefly “pop” apart, giving this register its characteristic rough and very low-pitched sound (Hollien, 1974). Since vocal fry does not represent “normal” vocal fold vibration, it can skew measures of fundamental frequency range toward the lower end (Ferrand, 2011). Unfortunately, many speakers occasionally go into pulse register during speech, and these phonations are difficult to separate out acoustically from modal register phonations. Because of the misleading effects of extreme frequency values and difficulty identifying the use of pulse register by speakers, SFF range may be somewhat problematic as a measure of fundamental frequency variability.

Pitch Sigma

Unlike SFF range, which presents only the extremes of an individual’s SFF variability, *pitch sigma* can be used as a more representative measure of a speaker’s dispersion of fundamental frequency. Baken (1987) states that pitch sigma is a “measure of the average distance of values from the mean” (p. 151). This measure is the *standard deviation* (SD) of the frequencies included in a speech sample, which is the “square root of the sum of the squares of the deviations from the mean” (Baken and Orlikoff, 2000, p. 171-172). This concept can be represented in the following algebraic formula:

$$SD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\bar{x} - x_i)^2}.$$ (Baken and Orlikoff, 2000, p. 171)

Since pitch sigma measures the *average* distance of values from the mean, the extreme values would not distort the measure of variability as significantly as they would for SFF range.

In research, pitch sigma is often reported as SFF SD, a description of an *individual's* average pitch variability, which can then be averaged for a group of speakers. However, fundamental frequency research also reports *group* SFF SD, which represents the dispersion of mean SFFs calculated within the group. Individual SFF SD and group SFF SD are not at all the same, although both are often represented as "SFF SD." Because of the use of this term in two ways, discerning which measures are truly pitch sigma and which are group SDs in older research is difficult. Individual SFF SD is usually reported in semitones for the reasons previously discussed.

Research on normative values of SFF range and pitch sigma in both males and females has been conducted for decades. When mentioned, researchers often report that these measures represent the amount of pitch variability (a perceptual quality) present in a speaker's voice. However, very few researchers have reported on whether SFF range or pitch sigma values actually correlate with listener perceptions of vocal variability. While measures of variability have been elicited in different contexts, including reading and spontaneous speech, research is needed to determine the relationship between acoustic measures of SFF range and pitch sigma and the perceptual phenomenon of vocal variability. This knowledge would assist Speech-Language Pathologists (SLPs) when they are conducting voice therapy. A strong correlation between acoustic and perceptual measures of vocal variability would indicate that perceived pitch variability can be reliably quantified and used as a pre- or post-treatment measure. In addition, this relationship can provide information on the perceptual salience of SFF range and pitch sigma values.

Review of the Literature on Pitch Sigma and Speaking Fundamental Frequency Range

One of the earliest researchers in SFF, Snidecor (1951), examined pitch characteristics in young adult female speakers during a reading task. Twenty-five speakers were chosen based on their superior general effectiveness in speech. Of these 25, eight subjects were selected based on their superior voice usage, as determined by expert listener judgments of an articulation examination, an oral reading performance, and an impromptu speech. The subjects were allowed ample time to practice the passage silently and aloud. The speakers then performed four readings of the passage with a five-minute interval between each reading. The subjects were instructed to read the passage as though they were reading to an audience of approximately 25 people. The best reading was chosen based on a ranking process that was completed by fourteen trained observers. The six highest ranked recordings were then selected for further study. The phonophotographic technique was used to analyze the speech samples, and the pitch sigma was found to be 1.52 tones[2] (3.04 semitones) for the group of women. In Snidecor (1951), the total pitch range (SFF range) was found to be 10.5 tones (21 ST). The median

---

[2] A tone is a musical *whole* step on the Western musical scale. Two semitones are the equivalent of one tone. Thus, doubling measures in tones should convert such measure to semitones and allow for comparisons among researchers.

90 percent, or functional pitch range (a measure that is not currently used), was 5 tones (10 semitones).

While Snidecor was one of the first researchers to examine pitch sigma in adult females, there are several limitations to generalizing from this work. The number of participants (n=6) is quite small, and the researchers only mention that the participants are “adult female” but do not specify an age range. The method used to determine pitch measurements (i.e., the phonophotographic technique) relies on hand counting low-pass filtered speech waves, and without reliability measures may be imprecise. Also, the researcher reported pitch sigma in tones, while semitones are currently used. Another issue was that the participants were instructed to read the Rainbow Passage as if they were speaking to an audience of 25 people. Speaking at a louder than normal level can cause changes in a speaker’s fundamental frequency characteristics, thus altering the measurements obtained. Finally, the researcher did not include any correlations between perceptual judgments and acoustic measures of vocal variability.

Pitch sigma and SFF range were again reported in Mysak (1959), who studied measures of “pitch flexibility” in regard to age in males. Participants were chosen if they qualified as a reader and were not impaired by any physical, auditory, or speech deficits. Three age groups arose from the selection process: an elder group I (65-75 years, with a mean age of 73.3 years), an elder group II (80-92 years, with a mean age of 85.0 years), and a middle-aged group (32-62 years, with a mean age of 47.9 years). These groups consisted of 12, 12, and 15 individuals for elder group I, elder group II, and the middle-aged group, respectively. The individuals comprising the middle-aged group were all sons of the men from the older groups. The procedures involved having the participants record both a reading and spontaneous speech sample. The reading sample consisted of the first paragraph of the Rainbow Passage, which had been previously practiced by the participants. The spontaneous speech sample involved the topic “What I Like To Do Most In The Summertime,” and was identical for all subjects. The recordings were analyzed using a custom-built Fundamental Frequency Recorder that had been modified to include a Comparator-Counter attachment.

Pitch sigma values for spontaneous speech were 2.9 semitones for the middle-aged group, 2.8 semitones for the elder group I, and 3.4 semitones for the elder group II. Pitch sigma values for oral reading were 2.9 semitones for the middle-aged group, 3.0 semitones for the elder group I, and 3.3 semitones for the elder group II. Speaking fundamental frequency (SFF) range values for spontaneous speaking were 16.6, 17.0, and 19.4 semitones for the middle-aged, elder group I and elder group II, respectively. Last, SFF range values for oral reading were 16.9 semitones for the middle-aged group, 17.7 semitones for the elder group I, and 19.6 semitones for elder group II. Overall, variability measures for these groups indicated greater variability with increasing age. Additionally, these measures showed greater variability during oral reading when compared to spontaneous speech. While a good addition to the new literature on SFF, with results supportive of Snidecor (1951), this study nevertheless had several weaknesses. Mysak was interested in examining familial relationships in measures of voice, so the participants from the middle-aged group were all sons from the elder groups. The results indicated that there were no SFF relationships found, but this limited the generalizability of results from the study as the familial link between the groups could be a confounding factor. In addition, while Mysak reported measures of variability, he did not compare these acoustic measurements to perceptual judgments.

McGlone and Hollien (1963) examined pitch characteristics of aged women. Subjects included 20 volunteers that lived at the Kansas Masonic Home in Wichita. These volunteers were placed into Group A and Group B; Group A included 10 women (ages 65-79 years, mean age of 72.6 years) and Group B included 10 women (ages 80-94 years, mean age of 85.0 years). Interestingly, the mean ages and age ranges were nearly identical to Mysak's (1958) elder groups I and II. Subjects were included in the study if they were healthy, free of voice disorders, ambulatory, and able to read. Subjects were also required to pass a hearing screening set at the level of 30 dB.

Subjects read the passage into a microphone system that was coupled with an Ampex Model 600 tape recorder. The taped material was then transferred to discs that were cut on a high quality Presto disc recorder. A phonellograph was used to convert recorded materials into a measurable pitch trace showing only the fundamental frequency. Measurements of frequency were made by hand. The unit was a modification of the photophoneloscope described by an earlier researcher. The speech sample that was analyzed consisted of the first paragraph of the Rainbow Passage. Prior to recordings being made, subjects were instructed to read the passage out loud at least three times, with the third reading usually recorded. The pitch sigma value for Group A was 1.48 tones (2.96 STs) and 1.35 tones (2.7 STs) for Group B. The functional pitch range (90% of SFF range) for Group A was 4.71 tones (9.42 STs) and 4.28 tones (8.56 STs) for Group B. The total SFF range for Group A was 9.56 tones (19.12 STs) and 8.87 tones

(17.74 STs) for Group B. From this study, the researchers concluded that pitch variability changes little with advancing age.

Hollien and Jackson (1973) examined SFF characteristics in young adult males. The subjects included 157 males from 17.9-25.8 years old, with a mean age of 20.3 years old. Inclusionary criteria for the study included enrollment at the University of Florida, speaker of a Southern dialect, and an absence of laryngeal pathology or voice training. Both spontaneous speech samples and reading samples were collected from participants. The reading sample consisted of a prose passage, approximately 3 minutes in length, by R.L. Stevenson called *Apology for Idlers.* The topic for the spontaneous speech sample was not identical for all participants; participants were allowed to chose from one of four topics and prepare a 3 minute talk based on their chosen topic.

A device called the Fundamental Frequency Indicator (FFI) was used to analyze the speech samples. This instrument consisted of both an analog and digital component. The analog component isolated the fundamental frequency by filtering out the harmonics in the speech wave. These harmonics were filtered by sending the speech sample through eight low-pass parallel filters that were assembled to cut off the high frequencies in half-octave steps (thus ensuring the harmonics would be excluded from the speech wave). What remained was the fundamental frequency, in the form of a sine wave, which would be sent to a pulse generator (Schmitt trigger) that produced and recorded a signal for all of the sine wave outputs. The Schmitt trigger, which converted the analog signal into an all-or-none pulse, was the initial digital (binary) conversion. These recorded Schmitt trigger pulses represented fundamental frequency cycles, and were fed into a computer (i.e., the digital component) that was responsible for computing mean SFF and standard deviations of SFF (Hanley and Peters, 1971). Pitch sigma was found to be 1.6 tones (3.2 semitones) for both the reading and spontaneous speech tasks.

In a study on pitch characteristics of female voices, Linke (1973) also reported measurements of pitch sigma and SFF range. The participants included 27 female subjects that were selected from speech classes at the State University of Iowa. The participants were chosen from a preliminary group of 60 subjects. Criteria for inclusion in the study included speaker of an American dialect, free of any articulatory or vocal quality impairments, and identification by instructors as representing a wide range of effectiveness of voice usage, from not very effective to extremely effective. The preliminary group was instructed to read a 55-word sample of expository prose as if speaking to an audience of 25 people. These samples were then judged by a panel of 30 graduate students and instructors in speech and speech pathology. The samples were rated on a 9-point scale of general vocal effectiveness. A rating of 1 indicated superior voice usage, whereas a rating of 9 indicated a very ineffective vocal performance. The median scale values of each voice sample, as determined by the judges, were used in the selection process. Selected samples were chosen in order to form a uniform distribution of vocal effectiveness that covered the entire range of scale values.

Following judgments, the tape recorded samples were transferred to lacquer discs for phonellographic analysis. Analysis was completed using the methodology described above. The total range (i.e., SFF range) was 11.66 tones (23.32 semitones) while the standard deviation of the fundamental frequency measures (i.e., pitch sigma) was 1.52 tones (3.04 semitones). The standard deviation of “reduced frequency” measures was found to be 1.21 tones (2.42 semitones). The reduced standard deviation was computed using only the frequencies that were above the lowest tone possible for each of the speakers to sustain, or the lower limit of the Maximum Phonational Frequency Range (MPFR). The researchers stated that phonations below the lower limit of the sustained tone range were omitted because the reduced range "…might be a better index to the perceived pitch variability of the voice and hence might be more closely related to judgments of general effectiveness…" (p. 182). Interestingly, it appeared that Linke (1973) was attempting to eliminate vocal fry phonations or other artifacts from his vocal frequency measures, although he did not specifically state that purpose.

Despite the contributions made to the literature, there were some weaknesses in this study. Similar to Snidecor (1951), the age range in Linke (1973) was identified as "young adult" with no specific age ranges listed. The procedure used in this study also replicated Snidecor's study in that participants were instructed to speak to a group of 25 people, which is not representative of normal interpersonal communication. The study also did not address the correlation between measures of variability and perceptual measures. However, on the positive side, Linke (1973) did examine how variability correlated with judgments of effective voice usage in females. He found that the correlation between the standard deviation of all frequency measures and effective voice usage was essentially unrelated (r = 0.06). However, the *reduced* frequency standard deviation and perceptions of effective voice usage yielded a correlation coefficient of r = 0.67.

In another study of pitch characteristics, Horii (1975) measured SFF range and pitch sigma, or standard deviation of fundamental frequency, in 65 adult males whose ages ranged from 26-79 years (mean of 54.1 years; Experiment II). Recordings were made during an oral reading of the first paragraph of the Rainbow Passage where subjects were instructed to familiarize themselves with the passage and then read aloud at a comfortable vocal intensity. The recordings of the speech sample were then digitized and stored on a computer magnetic tape. A computer program that utilized a peak-picking[3] method was used for obtaining fundamental frequency data. This was the first study to use a completely digital method of fundamental frequency extraction. The individual standard deviation was found to be 2.41 STs. The range of individual standard deviations was 1.46 – 3.54 STs.

Pitch sigma was again examined in Stoicheff (1981). This study looked at SFF characteristics as a function of age in women and aimed to make a statement about age-related changes in women's SFF. Participants included 111 healthy, nonsmoking adult females of various occupations from 20-82 years of age. The participants were grouped by age in 10-year increments, and data on pitch sigma were examined across the decades. The sample size per group ranged from 15-21 participants. Exclusionary criteria included: history of formal vocal training, failing a hearing screening test, speech or voice pathology, speaker of a foreign dialect, or evidence of a cold or sore throat at the time of testing. Subjects were instructed to read the first passage of the Rainbow Passage as though they were speaking to an audience of 25. This was consistent with the instructions used by Snidecor (1951) and Linke (1973). Speaking fundamental frequency measures were obtained using FFI. Results revealed that pitch sigma values, denoted as frequency distribution standard deviation (FDSD) in this particular study, were larger in the older groups than in the younger groups. In other words, as measured by the standard

---

[3] This method of measuring frequency involves counting the total number of peaks in the sine wave and dividing by time in seconds or milliseconds.

deviation of the fundamental frequencies, there was increased variability in older females compared to younger females. Exact distributions for the 20-, 30-, 40-, 50-, 60-, and 70-year-age groups were 3.78 STs, 3.92 STs, 4.00 STs, 4.33 STs, 4.25 STs, and 4.70 STs respectively. These findings were similar to the results obtained in Mysak (1959).

Stoicheff also examined post hoc the effects of menopause on speaking fundamental frequency values. Results for the middle-aged women might have been more valid if the groups had been separated initially based on menopausal status (i.e., pre-menopause, peri-menopause, or post-menopause) and not age. Regroupings were later made on the 40- and 50-year-old age groups to account for menopause. This data revealed a statistically significant difference ($p < .05$) in FDSD between the pre- and peri-menopause groups (3.92 STs and 3.97 STs respectively) compared to the post-menopause group (4.48 STs). The researcher interpreted this finding as evidence that a female’s variability around their mean fundamental frequency increases after menopause. However, a factor that could confound the data includes the procedures used to elicit the speech sample. As noted with previous studies, speaking to an audience of 25 people is not representative of typical speech and therefore limits the generalizability of the study. In addition, while the researchers reported measures of variability, they did not compare these acoustic measurements to perceptual judgments.

In 1982, Hudson and Holbrook examined fundamental frequency characteristics as a function of race. Specifically, the researchers examined SFF in 200 young adult African Americans (100 male, 100 female) ranging in age from 18-29 years. Subjects were volunteers from Florida A & M University who were taking introductory speech classes. Judgments of all subjects were made by ASHA-certified, experienced speech-language pathologists to ensure that subjects had normal voice and rhythm. Further exclusionary criteria included failing the hearing screening test or a history of formal voice or speech training.

Participants recorded both a reading and a spontaneous speech sample. The reading sample consisted of the Rainbow Passage and was recorded after participants read the material twice (once silently and once orally). The spontaneous speech sample consisted of the middle 40-second segment that had been extracted from a larger 120-second response to the following question: "Since you have been in school have you encountered any difficulties with registration or financial aid? If so, discuss these problems as well as possible solutions." Extracting the middle 40-second segment was done with the goal of minimizing anticipated pauses that are typical of the initiation of speech, and reducing the potentially confounding effects of decreased fundamental frequency that are typically present when speech is terminated. Recordings were made on an Ampex AG 440 B tape recorder that was coupled to an Electro Voice condenser boom microphone placed 10 centimeters below the subject's chin. The recordings were then analyzed using the FLORIDA I. This instrument was described as "a frequency-to-voltage converter, which automatically tracks the fundamental frequency by suppressing the harmonic partials in a complex wave form and registers the duration of the fundamental vocal frequency energy falling within a preset band-pass" (Hudson and Holbrook, 1982, p. 26).

The measure of central tendency that the researchers chose to examine was the modal SFF value. As previously mentioned, modal SFF (i.e., the mode) is the frequency value that occurs the most often in the speech sample. After analysis, the fundamental frequency that accounted for the greatest duration of time found in the sample was denoted as the modal fundamental frequency. For each subject, the range from the lower to the upper limit of the values was obtained and the standard deviation was then determined. The standard deviation of modal values for the male subjects during speaking was found to be 2.59 tones (5.18 STs) and during reading was 2.53 tones (5.06 STs). For the female subjects, the modal value during speaking was 1.64 tones (3.28 STs) and during reading was 1.65 tones (3.3 STs). Mean total vocal range (i.e., SFF range) was also reported. The SFF range for males during speaking was 6.27 tones (12.54 STs) and 5.77 tones (11.54 STs) during reading. Females were reported to have an SFF range of 6.18 tones (12.36 STs) during speaking and 5.61 tones (12.22 STs) during reading.

The greatest limitation found in Hudson and Holbrook (1982) was the use of modal frequency as the chosen measure for analysis. In the discussions section, the researchers compared their results of the standard deviation of modal values (in tones) to studies that presented the standard deviation of *mean* values. Additionally, the researchers report that the standard deviation values they found paralleled the results of Hollien and Jackson (1973) and Mysak (1959). However, the middle-aged male group (ages 32-62 years) in Mysak (1959) had pitch sigma values of 2.9 STs in both reading and spontaneous speech, which is 2.16 STs (in reading) and 2.28 STs (in speaking) below the values reported for Hudson and Holbrook's male group. Similarly, Hollien and Jackson (1973) reported pitch sigma values 3.2 STs in both reading and spontaneous speech. This is a difference of 1.86 STs in speaking and 1.98 STs in reading. It is evident that Hudson and Holbrook's pitch sigma values are at least 1.8 STs greater than previous studies and, therefore, do not parallel past results as the researchers reported they do.

Healey (1982) looked at speaking fundamental frequency characteristics of 10 male nonstutterers (ages 16-52 years, mean age of 27.4 years). Subjects presented with normal hearing and were free of vocal abnormalities or upper respiratory disorders during the time of testing. The pitch variability data were derived from recordings of declarative ("The fighting was tough. *The men dug in.* That won them the war,") and interrogative ("He asked her, *John ran away?* Yes, he did she replied,") utterances that were embedded in the middle of two short phrases. The participants rehearsed the sentences once before recordings were made. The second utterance was then used for the data analysis. Recordings were made using a Revox A77 MKIII tape recorder and an electret condensor microphone. Subjects were seated in a sound-treated room with a standardized equipment set-up and mouth-to-microphone distance. An oscilloscope was used to obtain a waveform of each sentence and then transferred it to an X-Y plotter (Hewlett-Packard 7010A) for a permanent record of the waveform to be used for later measurement.

Data were analyzed via hand measurements of waves in the speech sample. Period values were first measured in millimeters and then converted into hertz. The SFF SD for the declarative sentence was found to be 1.74 STs, while the interrogative sentence was 3.19 STs. The SFF range was found to be 7.28 STs and 8.80 STs for the declarative and interrogative utterances, respectively. These data are limited due to the small sample size and also the speech context used to derive pitch variability data. Sentences were used for this study whereas speech contexts in other studies normally involved a reading passage or spontaneous speech. It is difficult to compare these data to other studies due to this variation in speech context. In addition to the small sample size and speech context, the technology used to obtain pitch variability data included making hand measurements, which was outdated even at the time of the study. For these reasons, this study is limited in its use.

Brown, Morris, and Michel (1989) also examined pitch sigma values during a reading sample. Subjects included 25 young women (ages 20-32 years, mean age of 27.5 years) and 25 aged women (ages 70-90 years, mean age of 79.4 years). All subjects were ambulatory, Caucasian, and healthy. All of the subjects reported that they did not have a history of respiratory disease, neurological disease, and/or structural abnormalities to the speech mechanism. Normal hearing was also a requirement of the study. Subjects were seated in a sound-treated booth with a standardized mouth-to-microphone distance and equipment. The reading sample used for analysis was the first paragraph of the Rainbow Passage. Data were analyzed using the Fundamental Frequency Indicator (FFI). In the young group, SFF SD was reported to be 2.00 STs while the aged groups had an SFF SD of 2.44 STs. This increased variability found in aged speakers (irrespective of gender) is consistent with variability data reported in Stoicheff (1981) and Mysak (1959).

Perceptual Studies

A considerable amount of research is available on normative values of SFF range and pitch sigma, as evidenced by the previously-mentioned studies. However, of the studies reviewed, not one of them examined the correlation between acoustic measures and listener perception of vocal variability. In fact, very few studies are available that include any perceptual judgments of pitch variability at all.

In one study that did not include perceptual judgments, Benjamin (1981) examined frequency variability, both acoustically and perceptually, in the aged voice, although he did not correlate the two sets of measures. Participants were placed in a young group and older group; the young group included 10 males and 10 females. The male group ranged in age from 21-32 years (mean age of 29.8) and the female group ranged in age from 21-32 years (mean age of 29.0 years). The older group also included 10 males (aged 68-82 years, mean age of 74.5) and 10 females (aged 68-82 years, mean age of 73.6 years). All of the participants were nonsmokers, consumed less than 1 ounce of alcohol daily and did not have a history of speech therapy, laryngeal pathology, or an upper respiratory infection. Additionally, all participants had to pass a hearing screening. Recordings were made on an Ampex AG 500-2 tape recorder that was coupled to an 802 Ampex microphone positioned 15 centimeters below the subject's chin. Participants were instructed to read the first paragraph of the Rainbow Passage, and these recordings were then input to a Textronik 6087 Visipitch. Only the fourth and fifth sentences were used for fundamental frequency analysis. The acoustic measure of variability reported by the researchers included speaking fundamental frequency range (SFF range). The young male range was found to be 8.7 STs and the young female range was found to be 7.7 STs. The older male range and older female range was 11.7 STs and 9.7 STs, respectively. Pitch sigma was not reported.

Twelve fourth term graduate students in speech-language pathology were recruited as listeners. These listeners were trained to rate voices on specific characteristics using a modified Wilson Voice Profile. Judgments on pitch variability were made as “monotone,” “normal,” or “displaying excessive variability.” Benjamin (1981) concluded that older voices were not perceived as presenting with significantly more pitch variability than younger voices. She also concluded that aged speakers present with significant differences in SFF range relative to young adult speakers, although, *subjectively*, excess variation was not a perceived characteristic.

Benjamin (1981) was one of the few researchers that examined listener perceptions and acoustic measures of pitch variability, however several limitations are present. First, although she gathered both objective and perceptual data on pitch variability, the two measures were never correlated. This is unfortunate because knowing whether SFF range and perceptual ratings are mutually predictive, and to what degree, would provide valuable information on how this acoustic measure relates to listener perceptions. In addition, the acoustic measure that the researchers chose to use was speaking fundamental frequency range. While SFF range is normally reported as a measure of pitch variability, data on pitch sigma is also of interest, and may be a more stable measure of variability than SFF range.

While Benjamin (1981) did not examine correlations, Nadig and Shaw (2011) did examine this concept and attempted to answer the question: Does listener perception of pitch variability have direct acoustic correlates? The participants in this study were 13 typically developing children that were aged 8-14 years, from monolingual English households, who had normal or better than normal language abilities. Spontaneous speech samples were obtained during a face-to-face conversation where participants were asked questions about pets, interests, hobbies, or siblings. Audio recordings were obtained through a microphone that was positioned on the ceiling, approximately 5 feet above the participants. Each child's longest uninterrupted speech segment was extracted for analysis. Due to the natural give-and-take of conversation, the speech samples examined for the experiment were very brief and ranged from 10-13 seconds. The speech samples were analyzed using PRAAT software which automatically extracted maximum and minimum pitch for SFF range calculations. The researchers reported that the median SFF range was 124 Hz.

Thirty-two students in the School of Communication Sciences and Disorders from McGill University were recruited to rate the conversational speech samples. Listeners rated the voices on a seven point scale, with 1 being flat (monotone) and 7 being too variable (sing song). A 4 indicated normal pitch variation. The audio stimuli were presented to the listeners in a classroom via a PowerPoint presentation. The presentation began with a tutorial with examples of flat versus variable changes in pitch. The raters were then presented with a PowerPoint slide displaying the child's age and gender while the child's conversational speech sample was played. All raters were instructed to use their first impression to rate pitch variability, relative to the child's age and gender. Two practice trials of children that were not included in the study were presented in order to familiarize the listeners with the experimental protocol. The raters were presented with each conversational sample once and then given 35 seconds to complete the perceptual ratings for that child. The median rating for pitch variation was 3.81, where four was indicative of normal variation in pitch based on the child's age and gender. Correlations between pitch variability and perceptual ratings were also determined. Acoustically, "pitch variability" was considered to be SFF range. Pearson product-moment correlation analysis was completed and found that these two variables were not significantly correlated ($r = .25$, $p = .24$). While SFF range and perceptual ratings were not *significantly* correlated, researchers did find that, in general, as perceptual ratings of pitch variability increased, so did the acoustic measure of SFF range.

The Nadig and Shaw (2011) study is one of the few that examined correlations between acoustic and perceptual measures of pitch variability but, as in other research on this topic, there are aspects of the study that limit its usefulness. The main limitation was that the researchers reported SFF range in hertz, not semitones. Additionally, the speech samples used for analysis and perceptual ratings were very brief (ranging from 10-13 seconds) and thus perhaps did not provide a long enough sample for listeners to accurately judge for vocal variability.

Unlike Nadig and Shaw (2011), Philhour (1948) examined the correlations between the dispersion metric of pitch sigma (measured in semitones) and perceptual ratings of pitch variability. The primary purpose of Philhour's study was to investigate the correlations between various acoustic measures of the voice variability and listener judgments of speech. The five different types of judgments made by each listener included pitch level, pitch range, pitch variability, effectiveness of pitch usage, and general effectiveness. The acoustic measures of voice variability that Philhour correlated with perceived pitch variability were: standard deviations of the distributions of frequency measurements for each subject (i.e., pitch sigma), mean extent of frequency changes, mean extent of inflections, mean extent of shifts, mean rate of frequency changes per second, mean rate of frequency changes during inflection, and mean number of changes in direction of frequency modulation per second[4]. Interestingly, Philhour did not correlate SFF range with perceived pitch variability; instead, he correlated it with perceived pitch range.

[4] Refer to Philhour (1948) for a complete description of these acoustic measures of pitch variability.

Speakers in the study were 50 males who were thought to use different levels of pitch variability (ranging from monotonous pitch to extreme variability). The speakers varied in amount and type of speech training background and their general effectiveness of speaking. Some speakers were enrolled in public speaking classes, some were skilled speakers who were prominent participants in radio announcing, acting, or debate, and still others were enrolled in dramatic interpretation courses. Phonograph recordings of all the speakers were made and then projected into a measurable form by a phoneloscope. The technique employed was similar to earlier research. All speakers were asked to read the Rainbow Passage as though they were speaking to an audience of 25 people. Of the fifty speech samples, 24 were used in the experiment. Samples were omitted if the speaker read the passage inaccurately, deviated from General American dialect, had voice qualities that were not considered normal, or had poor inter-rater agreement among listeners. Prior to audio recordings being made, speakers were allowed to practice the passage silently several times and, if they chose, practice the passage aloud. Recordings were not made until speakers felt competent in their ability to read the selection aloud. Pitch sigma measures ranged from 1.7 – 5.9 STs and the mean was found to be 3.36 STs.

Thirty listeners supplied the perceptual ratings. Twenty-eight of these listeners were instructors in Communication Skills or Speech, or were graduate students in speech. The remaining two were trained musicians. Listeners rated pitch variability on a 7-point scale, where 1 was defined as "extreme monotony," a 4 was considered "average," and 7 was defined as "extreme variability." The definition of pitch variability that the researcher gave to the listeners was: "We want you to judge the amount of tonal movement - - the amount of pitch change apparent in each reading." (p. 72). The listeners were given practice speech samples in order to establish criteria for their judgment. The stimuli used during the training protocol included "…recordings rejected as stimuli and some of the twenty-four to be judged later [which were played] while the observers just listened" (p.16). Listener instructions were as follows: "You are to establish your standards of extreme variability of pitch and extreme monotony of pitch from the records you will hear in the practice session. That is, your judgments should be relative, based upon these samples of speech rather than some preconceived ideas of pitch monotony and pitch variability" (p.72).

The listeners then had a short practice session to familiarize themselves with the rating scales. The stimuli during this training task included seven recordings that did not include any of the 24 that were later judged. Three of the seven recordings were repeated (for a total of ten practice ratings) and the listeners were told to check their ratings on these repeated records. The researcher stated that differences for the same speech sample were generally small, and this procedure served to increase the confidence of the listeners in their ability to make accurate judgments. Out of the seven physical measures that were correlated with perceptual ratings, pitch sigma was found to correlate most highly with perceived pitch variability ($r = .88$; significance level not provided). The second most highly correlated physical measure with pitch variability was mean extent of inflections ($r = .82$).

As mentioned previously, Philhour did not correlate SFF range with perceived pitch variability. The researcher did, however, have the listeners rate all of the speech samples for perceived pitch range[5]. Pitch range was described to the listeners as the "extent of the difference or the tonal distance between the highest and the lowest note used" (p.15, 1948). The acoustic measures that were correlated with perceived pitch range were total frequency range and median 90% range. When correlated with perceptual judgments of SFF range, SFF range was found to have an r of .64; and median 90% range had an r of .83. From these data, Philhour (1948) concluded that total frequency or SFF range was not a satisfactory acoustic variable as a description of a speaker's perceived pitch range, as the correlation coefficient he obtained was not in the range considered "strong."

Philhour was clearly ahead of his time in his desire and ability to establish the relationship between perceptual measures and the nascent acoustic measures available at the time. He provided very valuable information in this regard, but limitations in the procedures must be considered. For example, the phoneloscope used to derive the acoustic measures of SFF and its variability required wave-by-wave hand measurements that required considerable subjective judgment. Reliability for these acoustic measures, either intra- or inter-rater, were not presented. In addition, the method for calculating intra-judge reliability for the perceptual protocol was not sufficient. The researcher only repeated one sample during the rating procedure in order to check intra-rater reliability of the judges. More samples should have been repeated during the listening task in order to obtain a more accurate measure of intra-judge reliability. Moreover, the training protocol used experimental subject recordings that were later judged by the listeners. Three voices were repeated during the training protocol and the listeners were encouraged to compare

[5] Listeners rated pitch range on a 7-point scale, where 1 was defined as "narrow range," a 4 was considered "medium range," and 7 was defined as "wide range."

their own ratings during the different presentations of the same voice. This made the listeners aware of a possibility of repeated voices and the need for consistency.

While this study provided valuable information, additional studies are also needed that address the weaknesses that were present in Philhour's (1948) study. The technology and methodological choices are outdated, and the instrumentation available today would presumably provide more accurate measures of speaking fundamental frequency and standard deviations. Additionally, the researcher did not correlate SFF range with perceived pitch variability and thus, this relationship remains unknown. Due to these weaknesses, the relationship between perceptual judgments of pitch variability and acoustic measures remains unconfirmed.

Critique and Rationale

Normative pitch sigma and SFF range values have had a substantial history of research. Over the years, technology and the procedures used to obtain these measures have become progressively more refined, and have increased the accuracy of pitch sigma.

The data from Tables 1-1 and 1-2 shows that pitch sigma has been examined in both males and females in the context of reading and spontaneous speech. A variety of ages has been measured as well, and considerable attention has been given to pitch sigma values as a function of age (Mysak, 1958; McGlone and Hollien, 1963; Stoicheff, 1981; Brown et al., 1989). With the exception of McGlone and Hollien (1963), these studies have reported that pitch sigma increases with age. Unfortunately, comparisons among studies are difficult, due to lack of specificity with which speaker subjects were described. Some studies used an ambiguous age range which they defined as "young adult" (Snidecor, 1951; Linke, 1973) while others used a large age range extending several decades (Healey, 1952; Horii, 1975).

Table 1-1: Summary of Previous Pitch Sigma and SFF Range Research in Male Speakers

Researchers	Number of Subjects	Age Range	Pitch sigma (in STs) Reading	Pitch Sigma (in STs) Oral Speaking	SFF Range (in STs) Reading	SFF Range (in STs) Oral Speaking
Hollien and Jackson (1973)	157	17;9-25.8	3.2	3.2		
Hudson and Holbrook (1982)	100	18-29	5.06 (modal value)	5.18 (modal value)	11.54*	12.54*
Healey (1982)	10	16-52		1.74 (declarative sentence) 3.19 (interrogative sentence)		7.28 (declarative sentence) 8.80 (interrogative sentence)
Horrii (1975)	65	26-79	2.41			
Mysak (1959)	15 12 12	32-62 65-75 80-92	2.9 3.0 3.3	2.9 2.8 3.4	16.9 17.7 19.6	16.6 17.0 19.4

*Reported by the researchers in the original article in tones. For this table, tones were converted to semitones.

Table 1-2: Summary of Previous Pitch Sigma and SFF Range Research in Female Speakers

Researchers	Number of Subjects	Age Range	Pitch Sigma (in STs) Reading	Pitch Sigma (in STs) Oral Speaking	SFF range (in STs) Reading	SFF Range (in STs) Oral Speaking
Snidecor (1951)	25	Young Adult	3.04		21*	
Linke (1973)	27	Young Adult	3.04 (“reduced frequency” pitch sigma = 2.42)		23.32*	
Hudson and Holbrook (1982)	100	18-29	3.3 (modal value)	3.28 (modal value)	12.22*	12.36*
Brown et al. (1989)	25 25	20-32 70-90	2.00 2.44			
Stoicheff (1981)	21 18 21 17 15 19	20-29 30-39 40-49 50-59 60-69 70 & up	3.78 3.92 4.00 4.33 4.25 4.70			
McGlone and Hollien (1963)	10 10	65-79 80-94	2.96 2.7		19.12* 17.74*	

*Reported by the researchers in the original article in tones. For this table, tones were converted to semitones.

Of all of the studies noted in Table 1-1 and Table 1-2, none reported *perceptual* measures of vocal variability. There is a paucity of research available in this area, with even fewer studies available that examine actual correlations between acoustic measures (specifically, pitch sigma or SFF range in ST) and perceptual measures of vocal variability. However, of the three perceptual studies reviewed in the current research, some valuable information has been obtained relating to the area of perception of variability.

Benjamin (1981) looked at SFF range and compared these values to perceptual ratings of vocal variability. The research suggested that older voices present with significantly greater SFF ranges relative to young adult speakers, but the perceptual ratings showed that excess variation was not a perceived characteristic in the older age group. This study tells us that although a person may present with a large SFF range, his or her voice will not necessarily be perceived as containing a notable amount of variability. This study would have been enhanced if the researcher had looked at pitch sigma in addition to SFF range, to see if older voices would be characterized as having higher pitch sigma values, and if individuals with greater pitch sigma values were perceived as having more variability. Correlations between acoustic and perceptual measures would also have enhanced the findings in the study.

More recently Nadig and Shaw (2011) also looked at SFF range and perceptual ratings of vocal variability. Unlike Benjamin (1981), these researchers reported correlations between the two measures. However, the limitations of the study detract from the usefulness of the information. Most importantly, pitch sigma was not used as an acoustic measure of variability. The researchers used only SFF range and further limited the study by failing to convert SFF range values to semitones. As noted previously, semitones are absolutely necessary when reporting measures of variability. The researchers found that SFF range and perceptual ratings of vocal variability were not significantly correlated (r = .25), although in general, as ratings of pitch variability increased, so did the acoustic measure of SFF range.

The most useful study in the area of perceptual ratings of vocal variability was Philhour (1948). Philhour examined many acoustic measures of vocal variability (i.e., pitch sigma, mean extent of frequency changes, mean extent of inflections, mean extent of shifts, mean rate of frequency changes per second, mean rate of frequency changes during inflection, and mean number of changes in direction of frequency modulation per second) and correlated all of them with perceptual ratings of perceived pitch variability. Pitch sigma was found to be the acoustic measure most highly correlated with perceptual ratings of perceived pitch variability (r = .88). However, the limitations in procedures and technology detract from the usefulness of the study. The researcher used the phonophotographic technique for fundamental frequency analysis, a technique that relied on hand counting low-pass filtered speech waves, which is subjective and potentially imprecise. He also instructed speakers to read as though they were speaking to an audience of 25 people, which is an outdated technique of determining speaking frequency values. The method of determining intra-judge reliability was not adequate for the study due to the researcher obtaining consistency of judgments on only one recorded sample. Furthermore, the researcher used some of the 24 experimental voices during the training protocol and also made the listeners aware of the possibility of repeated voices and the need for consistency (since listeners were encouraged to compare their own ratings on two different presentations of the same voice). Due to the problematic procedures, methodological choices, and technology employed in Philhour (1948), the relationship between perceptual judgments of pitch variability and acoustic measures is still uncertain.

In addition to the limitations described above, no previous research on pitch sigma and vocal variability has compared male and female speakers. If the data in similar types of studies in Tables 1 and 2 are averaged, results suggest potential gender differences. Excluding Hudson and Holbrook (1982) which used median values, and Healey (1982) who looked separately at declarative and interrogative sentences, male pitch sigma was found to average 2.96 STs for reading, while female pitch sigma averaged 3.43 STs for reading. These results are consistent with a general belief that women use more vocal variability than men (Gelfer and Mordaunt, 2012; Andrews, 2006), however a direct comparison between male and female vocal variability has never been made.

There is research evidence providing some indirect support for differences in male and female use of vocal variability. For example, in a study examining intonation in male-to-female transgender individuals perceived as female versus those perceived as male, Wolfe, Ratusnik, Smith, and Northrop (1990) found evidence of more vocal variability in the group that was perceived as female. In a similar study by Gelfer and Schofield (2000), some differences in vocal variability between the perceived female transgendered individuals versus the perceived male group were found although they did not reach the level of statistical significance. Andrews and Schmidt (1997) found that biological males who identified as heterosexual cross-dressers were perceived as more “animated” (as opposed to “monotone”) when using their feminine voice versus their masculine voice. Acoustic findings from this study suggested that variation in frequency may not be the most salient factor in being perceived as feminine, but that combined with variation in loudness, intonation contours, rate, and duration, overall “variability” can provide cues for gender identification. Therefore, this is an area that needs further study.

Purpose

The purpose of this study was to investigate relationships between acoustic measures of the variability of fundamental frequency (pitch sigma and SFF range) and perceptual ratings of vocal variability during a reading task. A second purpose was to determine whether vocal variability in males differs significantly from vocal variability in females. The study aimed to answer the following research questions:

1. Do perceptual judgments of pitch variability correlate with measures of pitch sigma or SFF range for both male and female speakers?
2. Do perceptual judgments of pitch variability correlate with measures of pitch sigma or SFF range more strongly for either male or female speakers?
3. Does one gender display significantly more pitch variability than the other as measured by pitch sigma or SFF range?

No studies to date have reported correlations of perceived pitch variability on the basis of gender. Knowledge of this correlation would be useful in research and voice therapy for transgender individuals. For example, if females are perceived to have greater variability, male-to-female transgender individuals will have to learn to vary their voice in order to be perceived as feminine sounding. Additionally, the first research question would assist SLPs in their understanding of the degree to which differences in pitch sigma and SFF range are perceptually significant. A strong correlation would indicate that perceived pitch variability can be reliably quantified using a relatively simple calculation and used as a pre- or post-treatment measure. Quantifying pitch variability could have possible uses in normative and/or outcomes research as well.

Method

Speaker Participants

Two groups of speakers were used to answer the research questions in the current study; an expanded group and an experimental group. Speakers in the expanded group (19 adult females and 15 adult males) supplied acoustic data for an investigation of gender differences in measured SFF variability during *normal* speech. For subjects in the expanded female group, the mean age was 22 (years); 2 (months) with a range of 19;2 – 30;11. For subjects in the expanded male group, the mean age was 22 (years); 10 (months), with a range of 19;5 – 30;6. All of these subjects provided reading samples for determination of the acoustic measures of speaking fundamental frequency (SFF), pitch sigma, and SFF range.

In order to determine the relationship between perceptual judgments of pitch variability and acoustic correlates over a wide range of variability types, an experimental group of speakers was formed. This group consisted of 13 adult females and 13 adult males that were chosen from the pool of speakers in the expanded group due to their use of a wide range of vocal variability. For subjects in the experimental female group, the mean age was 21;1, with a range of 19;2 – 28;2. For subjects in the experimental male group, the mean age was 23;2, with a range of 19;5 – 29;0. Experimental samples included readings produced by speakers in their normal voice, in a monotone (as instructed), and in an expressive voice (as instructed).

All speaker participants were native English speakers, who reported a medical history free of respiratory dysfunction, and voice and neurological disorders. Additional inclusionary criteria were voice, articulation, and hearing abilities within normal limits, as determined from screening procedures. In order to pass the hearing screening, participants had to have hearing thresholds of 25 dB or better at three different frequencies in the speech range (American Speech-Language Hearing Association [ASHA], 1997).

Speaker subjects were recruited through a variety of methods, including personal contacts and presentations to large psychology and health sciences classes (see Appendix A). Following the presentations, interested individuals provided contact information on sign-up sheets. These individuals were contacted and screened (either during a phone interview or an in-person screening) prior to participation in the study. Initial screening consisted of questions related to the individual's age, respiratory status, history of hearing loss, and history of speech, voice or neurological disorders (see Appendix B). Individuals who indicated a positive history for respiratory disorders, hearing loss, or speech, voice or neurological disorders were thanked for their interest but were not tested further.

Speaker Selection Procedures

Potential participants who passed the initial screening were invited to continue with eligibility testing and possible experimental participation at the UWM Speech and Language Clinic. Upon arrival, they were educated about the details of participation in this research, and all potential risks and benefits. The risks for participating in the study were minimal, but perhaps some speakers could have experienced slight embarrassment reading the passage aloud in front of the researchers. If after learning about the study and its risks they agreed to participate, potential participants were asked to sign an informed consent document approved by the institution's Institutional Review Board (IRB; see Appendix C).

To continue determining eligibility, a speech/voice screening and hearing screening were administered to potential participants. The screening tool used to determine appropriateness of articulation and voice was a modified version of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V; Kempster, Gerratt, Abbott, Barkmeier-Kraemer, and Hillman, 2009). This protocol was used to assess vocal quality based on perceptual impressions of *roughness*, *breathiness*, and *strain* during sentence production tasks, spontaneous speech, and sustained vowels. The appropriateness of the speaker's *pitch*, *loudness*, and *resonance* was also assessed. In addition to assessing voice, the speech samples elicited during this protocol were used to determine the presence of any speech sound errors (see Appendix D).

A hearing screening was then administered on a Beltone Model 119 audiometer that had been calibrated within a year of the study (see Appendix E). The hearing screening given for the purposes of this study had three components: Participants were first asked questions related to hearing loss and current hearing status. An otoscopic inspection was then completed, followed by a pure-tone screening in both ears at 1000, 2000, and 4000 Hz at a loudness level of 25 dB. These hearing screening procedures were in accordance with the hearing screening procedures outlined in ASHA (1997). Individuals who passed all hearing screening components, and were judged by the authors to be within normal limits for overall voice and articulation, were included in the study. Approximately three subjects were excluded from the study due to failure to pass the screening tests.

Procedures for Eliciting and Recording Speech Stimuli

For the experimental session, each speaker was seated individually in a sound-treated booth (Acoustic Systems M0366558). Speakers were given the Grandfather Passage (see Appendix F) and allowed to familiarize themselves with its content. When the speakers felt comfortable with reading the passage, they were instructed to begin reading it aloud, starting with the title, and using a loudness level typical of a one-to-one speaking situation with a listener positioned a few feet away. The experimenter then recorded their production. Samples were collected using a Shure Model SM58 microphone with a constant 12 inch mouth-to-microphone distance. The microphone fed into an AudioBuddy Dual Mic Preamp connected to a Dell Optiplex 980 desktop computer.

Samples were digitally recorded using the Real Time Pitch subprogram of the KAYPentax MultiSpeech analysis software system (Model 5121; version 3.2.0). If interruptions or errors occurred during the reading, the subjects were instructed to redo the reading sample. Upon completion, reading samples were saved to the computer.

In order to obtain a range of vocal variability, all speakers were asked to read the passage two more times, first in a “monotone” voice and then in an “expressive voice.” In an effort to limit any confusion about the nature of the task, speakers were given two exemplars of a monotone voice (one male and one female speaker, both reading the Rainbow Passage). After hearing the monotone exemplars, all speakers were instructed to read the passage in a monotonous voice. Again, the experimenter recorded this reading and the sample was saved to the computer. Next, two exemplars of an expressive voice (again, one male and one female speaker who both read the Rainbow Passage) were given. After hearing the expressive exemplars, a final sample was obtained with the speakers instructed to read the passage in a very animated or expressive voice. For all three readings, the speakers were instructed to begin with the title of the passage.

Acoustic Analysis

Before acoustic measures on the data were obtained, the speech samples were trimmed so that each one encompassed the first word of the paragraph all the way to the word “trifle.” To block off the appropriate signal, the cursors on either side of the screen were used, and the signal outside of the selection was removed. The frequency analysis range was set at a minimum of 100 Hz and maximum of 600 Hz for females. The analysis range for males was set at a minimum of 50 Hz and maximum of 300 Hz. Display and analysis durations were set at 60 seconds (i.e., roughly the amount of time it takes to read the entire passage) and the pitch smoothing level was set to *medium*. An acoustic analysis of all reading samples was then calculated through use of the Real Time Pitch program’s statistical application by the primary investigator. A preliminary analysis was completed on the trimmed samples, and data on the speaker’s mean speaking fundamental frequency (SFF), lower and upper SFF limits, pitch sigma, and SFF range were recorded. The pitch tracing was then observed by the investigator to determine whether there were outlier data points or noise present in the signal as displayed on the screen that should be excluded from the acoustic analysis. An isolated point suggested that the frequency value was not speech-related and should therefore be eliminated from the data analysis, while a point that was connected to a frequency contour was preliminarily treated as valid data. The isolated points were then evaluated by listening to the sample to determine whether or not they could be supported by perceptual impressions. Points that did not correlate with perceptions of low- or high-pitched voicing during that time interval were then eliminated from the analysis through manipulation of the analysis range. When the upper and lower limits of the analysis range were set appropriately to eliminate as many outliers as possible, a final statistical analysis was completed (again, using the statistics application) and the speaker's mean SFF, lower and upper SFF limits, pitch sigma, and SFF range were again recorded (see Appendix G).

All of the speech samples were then re-analyzed by a second examiner using the same procedures. The criteria used to ensure adequate inter-rater reliability of the acoustic measures were: mean SFF and lower and upper SFF limits within 5 Hz; SFF range within 1 semitone (ST); and pitch sigma within .03 units between raters. If the results obtained by the second examiner met these criteria when compared to the results of the primary examiner, differences in acoustic findings were averaged. If the results obtained by the second examiner did not meet the criteria when compared to the results of the primary examiner, then the two examiners analyzed the sample together to resolve their differences.

Creation of a Stimulus CD

Two CDs were made, one containing 15 speech samples from the 13 female speakers and another with 15 speech samples from the 13 male speakers in the smaller group. The samples on each CD were selected to represent a continuum of variability from extremely monotone to very expressive. For each speaker, the researcher and advisor selected from among the normal, monotone, and expressive readings to find the best exemplars of degrees of variability. Two male and two female speakers had exceptional monotone and expressive samples, and these speakers were repeated twice in the stimulus set. This meant that the female and male stimulus CDs each contained a total of 15 different samples from 13 different speakers. The speakers that were repeated twice had such considerable differences in pitch variability that the researchers felt it would be unlikely for listeners to detect that the same speaker provided two different samples. Eight randomly-selected samples from each stimulus set were repeated for intra-rater reliability purposes (approximately 50%). The samples on the stimulus CD were presented in a quasi-random order so that two identical samples never occurred together.

Normative Data

Acoustic data from the expanded male (N = 15) and female (N = 19) groups were employed to determine acoustic characteristics of the normal voices of male and female speakers, and to examine the differences in acoustic measures of pitch variability between genders. All speakers provided three speech samples (i.e., a normal sample, monotone sample, and expressive sample) as part of the recording procedure (as described above in *Procedures for Eliciting and Recording Speech Stimuli).* Acoustic data obtained from all of the speakers' normal samples were employed to report normative data.

Listener Participants

The listeners used in the study were 30 graduate students (mean age of 25 [years]; 10 [months], and a range of 22;7 – 40;5) in Communication Sciences and Disorders (CSD), who had completed at least one professional-level CSD course in voice disorders.

This decision was based on the results of Gelfer (1988), who found that speech-language pathologists (SLPs) exhibited higher levels of concordance (W=.50) than untrained listeners (W=.36) during perceptual ratings of vocal variability. All listeners were native English speakers and passed a hearing screening to ensure normal hearing. In addition, all listeners had to meet criteria for intra-rater reliability (see *Analysis of Perceptual Results*). Listener participants were excluded from the study if they did not meet any of these criteria.

Listener subjects were recruited through a variety of methods, including personal contacts, announcements in classes, and sign-up sheets for Communication Sciences and Disorders graduate students. Individuals who signed up or otherwise indicated their interest were contacted by the researcher and were screened either in person or on the phone. This screening consisted of a series of questions related to age, history of hearing loss, native language, and their graduate level class history (see Appendix H). Individuals who did not meet all of the inclusionary criteria were thanked for their interest but were not tested further.

Listener Selection Procedures

Individuals who passed the initial screening were invited to come to the UWM Speech and Language Clinic to participate in the study. Subjects were educated about the details of participation in the study, and all potential risks and benefits. Before participation, all persons were asked to sign an informed consent document approved by the institution's IRB (see Appendix C). At that point, a hearing screening was administered on a Beltone Model 119 audiometer, calibrated within a year of the time of the study (see Appendix E). The hearing screening protocol given to the listeners consisted of the same three components outlined in ASHA (1997), previously described in the section *Speaker Selection Procedures.* All persons who passed the screening protocols were included in the study.

Listening Protocols

Small groups of listeners were seated in a quiet classroom. Ambient noise levels were measured at 50 dB (re: Weighting Network C on a Radio Shack digital sound level meter in the sound field) or less. To determine which gender the listener would rate first, a quasi-random selection procedure was used, so that 50% of all listener subjects heard the female speakers first and 50% heard the male speakers first. In order to familiarize the listeners with the range of stimuli, a 15 second sample of all of the voices of the selected gender were played prior to the start of the rating procedure, while the participants were instructed to just listen. Subjects were then informed that all of the voices would be replayed and that they were to rate the voices in terms of vocal variability as compared to the other voices in the stimulus set. Listeners were given response sheets that they used to rate their perception of each speaker's pitch variability. Pitch variability was rated on a 7-point Likert scale, where 1 was defined as "extreme monotony," a 4 was considered "average," and 7 was defined as "extreme variability" (see Appendix I for the instructions that were given to the subjects). Three seconds were given after each recording to allow sufficient time for listeners to rate the voice. Eight of the voices were repeated to obtain information for intra-rater reliability, for a total of 23 perceptual ratings per stimulus set. Five speakers were rated per response sheet (see Appendix J). These procedures were then repeated for the second stimulus set containing the voice samples of the opposite gender.

Analysis of Perceptual Results

The perceptual results were first examined to ensure that all listeners met reliability criteria. To do so, listeners had to average absolute value of less than 1.5 units of difference between the first and second samples of the sixteen repeated speakers, and obtain a correlation of $r > .5$ when the first judgment was correlated with the second. The next step in the analysis was to determine whether the perceptual ratings met the requirements necessary to use parametric statistics (i.e., equal variances and normal distribution). Finally, an average over all listener ratings for each individual speaker was obtained. This average for each speaker was used during the various correlational analyses with the acoustic variables.

Statistics

The procedures that were described above were used to obtain acoustic measures of mean speaking fundamental frequency (SFF), upper and lower limit of SFF, SFF range, and pitch sigma along with perceptual ratings of vocal variability for each speaker during connected speech. Pitch sigma and SFF range measures were calculated in semitones. Prior to any statistical analyses, the perceptual judgments were analyzed to determine if all criteria for using parametric statistics were met. After this determination, the appropriate correlation analyses were employed to examine the relationship between the acoustic and perceptual variables.

Six correlations were obtained to answer the first and second research questions. Two correlations calculated the relationship between acoustic measures (pitch sigma, SFF range) and perceptual measures of vocal variability for all 30 samples used during the listening task. An additional two correlations examined the relationship between the acoustic measures and perceptual measures of vocal variability for the 15 male samples, and the final two correlations assessed the relationship between the acoustic measures and perceptual measures of vocal variability for the 15 female samples. Last, an appropriate test for differences between means for independent measures was used to determine whether statistically significant differences existed between the normal samples of males and females in terms of acoustic measures of variability.

Results

Listener Reliability Results

In order to be included in the data analysis, listeners had to meet specific intra-rater reliability criteria for the eight randomly-selected repeated speech samples. These criteria included an average absolute value of less than 1.5 units of difference between ratings of the first occurrence and the second occurrence of each of the sixteen repeated samples, and a correlation of $r > .5$ when the first judgment was correlated with the second. All 30 listeners met these intra-rater reliability criteria. The average number of units of difference for all listeners was |.43|, with a range of |.19| – |.75|. The average correlation was $r = .919$, with a range of .822 – .975. All correlations were significant at the $p < .01$ level. These data indicated a high degree of internal consistency in each listener's perceptual judgments of vocal variability. Thus, the data from all 30 listeners were utilized to determine the average perceptual rating for each speech sample.

Justification for Use of Parametric Statistics

In deciding whether to use parametric or non-parametric statistics, it was necessary to determine whether our perceptual data met the criteria needed to use parametric statistics (Schiavetti and Metz, 2006). First, we considered whether our measurements were interval-level data. Since our data were taken from a Likert scale we assumed that these data were at equally spaced intervals. We then considered if we had a large enough subject number; there were 30 total listeners that provided perceptual ratings so it appeared there were enough subjects to meet this criterion as well. Next, we looked at whether the data were normally distributed; we found that when all of the perceptual ratings were combined, they were indeed normally distributed. In addition, perceptual data for male speakers and female speakers independently were also examined, and found to be normally distributed. The last criterion to be met was equal variances between groups to be compared. To determine this, we looked at whether there were equal variances between males and females, and found that these groups did have equal variances. Since the perceptual data met all of the criteria, we were able to use parametric statistics.

Research Results

The first research question asked whether perceptual judgments of pitch variability correlated with acoustic measures for *both* male and female speakers. Acoustic measures for SFF, pitch sigma, and SFF range (in ST) can be seen for the experimental samples in Table 3-1. In answering this first research question, correlational analyses between pitch sigma and perceptual ratings were performed. Results can be seen in Table 3-2. The strong correlation results suggested listener judgments of pitch variability were well represented by both pitch sigma and SFF range.

The second research question asked whether perceptual judgments of pitch variability correlated with acoustic measures more strongly for either male or female speakers. In order to answer this question, we grouped the data by gender. Summary statistics for males and females can be seen in Table 3-1. Correlations are presented in Table 3-2. Results showed that when pitch sigma was correlated with mean perceptual ratings, male correlations and female correlations both indicated strong positive relationships. A z-score was then obtained to determine whether these correlation coefficients were significantly different from one another (Lowry, 2013). A two-tailed test indicated that these correlations were *not* significantly different from one another ($z = -.89$, $p = .37$), thus correlations between perceptual judgment and pitch sigma were similarly strong for both genders.

In terms of correlations obtained between SFF range and mean perceptual ratings, the male group and the female group both had correlation coefficients that indicated strong positive relationships, as seen in Table 3-2. A two-tailed z-score indicated that these correlation coefficients were also *not* significantly different from each other ($z = .25$, $p = .80$; Lowry, 2013). Results obtained from these z-scores indicated that relationships between listener judgments and both acoustic measures of pitch variability were similar for male and female voices.

Additional correlations were obtained to examine the relationship between pitch sigma and SFF range. For the male and female combined group, when pitch sigma and SFF range measures are correlated, we found an r of .979 ($p < .01$). Males and females were also examined independently. Male correlations indicated an r of .990 ($p < .01$) and female correlations indicated an r of .964 ($p < .01$). Overall, these data indicated that as pitch sigma increased, SFF range increased as well.

The third and final research question asked whether one gender displayed significantly more pitch variability than the other as measured by pitch sigma or SFF range during normal reading. The expanded subject group was used to answer this question. Acoustic measures for SFF, pitch sigma, and SFF range (in ST) for normal samples in the expanded group can be seen in Table 3-3. An independent samples t-test was conducted to compare pitch sigma in males and females. There was no significant difference in pitch sigma for males (M = 2.18; SD = .99) and females (M = 2.27; [SD = .55]; t [32] = -.357; p = .723). Another independent samples t-test was conducted to compare mean differences in SFF range between males and females. There was no significant difference in SFF range for males (M = 11.33; SD = 4.30) and females (M = 12.05; [SD = 2.37]; t [32] = -.621; p = .539). These results suggested that there were no significant differences between male and female speakers in terms of mean pitch sigma or mean SFF range values, and that males and females have similar pitch variability in reading when measured acoustically.

Table 3-1: Speaking Fundamental Frequency (SFF) and Acoustic Measures of Pitch Variability during Reading for the Experimental Samples

Measure	Both Genders (N = 30)	Male (N = 15)	Female (N = 15)
Mean SFF in HZ (Minimum-Maximum)	____	125.01 (98.49 – 173.86)	213.05 (159.05 – 270.11)
Pitch Sigma (in ST) (Minimum-Maximum)	2.25 (0.57 – 4.42)	2.12 (0.57 – 4.42)	2.38 (0.78 – 4.23)
SFF Range (in ST) (Minimum-Maximum)	11.8 (4 – 22)	11.33 (4 – 22)	12.27 (5 – 21)

*Significant at the $p < .01$ level

Table 3-2: Pearson Product-Moment Correlation Coefficients Relating Mean Perceptual Judgments of Pitch Variability with Acoustic Measures of Pitch Variability for the Experimental Samples

Correlation	Both Genders (N = 30)	Male (N = 15)	Female (N = 15)
Pitch Sigma and Perceptual Judgments	.927*	.911*	.956*
SFF Range and Perceptual Judgments	.890*	.902*	.881*

*Significant at the $p < .01$ level

Table 3-3: Speaking Fundamental Frequency (SFF) and Acoustic Measures of Pitch Variability for Reading in Males and Females (Expanded Group)

Measure	Male (N=15)	Female (N=19)
Mean SFF in Hz (Minimum-Maximum)	122.73 (100.39 – 173.86)	215.92 (185.88 – 247.16)
Pitch Sigma in ST (Minimum-Maximum)	2.18 (0.77 – 4.30)	2.27 (1.07 – 3.71)
SFF Range in ST (Minimum-Maximum)	11.33 (4.00 – 21.00)	12.05 (6.00 – 18.00)

Discussion

The goal of this study was to evaluate relationships between acoustic and perceptual measures of pitch variability during a reading task. It also sought to answer the following questions:

1. Do perceptual judgments of pitch variability correlate with measures of pitch sigma or SFF range for both male and female speakers?
2. Do perceptual judgments of pitch variability correlate with measures of pitch sigma or SFF range more strongly for either male or female speakers?
3. Does one gender display significantly more pitch variability than the other as measured by pitch sigma or SFF range?

Based on the results of this study, it can be concluded that perceptual judgments of pitch variability and the acoustic measure pitch sigma have a strong positive correlation ($r = .927$) for male and female speakers. In addition, perceptual judgments of pitch variability and the acoustic measure SFF range also have a strong positive relationship ($r = .890$). While both have a strong correlation at $p < .01$, a somewhat stronger correlation was achieved when pitch sigma measures were used.

When correlations between pitch sigma and perceptual judgments or SFF range and perceptual judgments were examined for male and female speakers independently, significant differences in the strength of the correlations between genders were not obtained. This finding indicates that pitch sigma and SFF range are appropriate acoustic measures of vocal variability for both males and females.

This study also examined whether males and females display significant differences in pitch variability during a normal reading task when measured acoustically. Based on the results of this study, males and females did not show significant differences in measured pitch variability for either SFF range or pitch sigma. While there has always been a general belief that females use more vocal variability than males when speaking (Gelfer and Mordaunt, 2012; Andrews, 2006), this study showed that in terms of reading, men and women displayed quite similar vocal variability when measured acoustically.

Limitations of the Present Study

One potential limitation of the methodology of this study may have been the two male and two female speakers who provided exceptional monotone and expressive samples and were repeated twice in the stimulus set. While the researchers felt that these samples had such considerable differences in pitch variability that it would be unlikely for the listeners to detect that the same speaker provided two different samples, it is possible some of the listeners were able to detect this. This realization could have cued the listener to the speaker’s task of providing a variety of speaking samples that ranged in vocal variability, and thus impacted their perceptual judgments of the repeated speakers. It is important to note that following participation in the study, none of the listener participants mentioned that the same speaker had provided two different speech samples. However, many listeners did comment on the fact that they had noticed samples were repeated for intra-rater reliability purposes.

Comparisons to Previous Literature

A primary reason this study was conducted was because the relationship between perceptual judgments of pitch variability and their acoustic correlates was unconfirmed. For example, the only recent study to examine the relationship between SFF range and perceptual judgments of pitch variability was Nadig and Shaw (2011); however, they reported SFF range in hertz (Hz), not semitones. Nadig and Shaw (2011) found a correlation of $r = .25$ ($p = .24$) between perceptual measures of pitch variability and SFF range. The present study found a much stronger correlation between SFF range and perceptual judgments ($r = .890$, $p < .01$) when SFF range was converted to semitones. The stronger relationship found in the present study between SFF range and perceptual judgments may have been due to Nadig and Shaw's use of Hz to measure range. Hertz are geometrically related to octaves in the musical scale, thus yielding very different acoustic measures of range in Hz compared to listener perceptions (see *Semitones and Vocal Variability).* Measuring SFF range in semitones (ST), a frequency unit based on perception, may have substantially improved the relationship between perceptual judgments and acoustic measures.

While Philhour (1948) did not examine the relationship between SFF range and listener judgments of pitch *variability*, he did correlate SFF range with "perceived pitch range." When correlated with perceptual judgments of pitch range, total frequency range was found to have an r of .64. Philhour then calculated the "90% range," obtained by eliminating the lowest 5% and the highest 5% of the frequency measures in each speaker's data set. When this measure was used, Philhour (1948) then found a correlation of $r = .83$ between the 90% range and listener judgments of perceived pitch range.

Although we cannot directly compare these correlations to those found in the current study because of differences in measurement procedure and terminology, it is interesting to note that the relationship between perception of pitch range and range measures was improved when the upper and lower 5% of the frequency points were removed from the data set, similar to the current researchers' use of analysis range adjustments to eliminate frequency inaccuracies.

There was also only one previous study (Philhour, 1948) that examined the relationship between pitch sigma and perceptual judgments of pitch variability. Despite the older technology that Philhour used, there was agreement between the correlations found in his study and those found in the current study. Philhour obtained a correlation of $r = .88$ (significance levels not stated) between measures of pitch sigma and listener judgments of pitch variability. The present study found a comparable correlation of $r = .927$ ($p < .01$). While these correlations are similar, the updated technology and data evaluation procedures used in this study may have contributed to the somewhat stronger correlation. In addition, it is possible that the stronger correlation for pitch sigma and perceptual judgments found in this study was due to the training procedure that the current study used to familiarize listeners with the continuum of vocal variability prior to any listener judgments taking place.

The current study also reported normative data for pitch sigma and SFF range for young adult males and females during reading. Previous studies examining acoustic SFF variability data for young adult males and females presented with somewhat dissimilar data from the current investigation (see Tables 1-1 and 1-2 in the section *Critique and Rationale)*. Hudson and Holbrook (1982) found that for reading, men 18-29 years (N =

100) had a pitch sigma of 5.06 STs, compared to the present study's findings of a pitch sigma at 2.18 STs. The difference in pitch sigma values may be because Hudson and Holbrook used modal values when reporting pitch sigma instead of mean values. Hollien and Jackson (1973) also examined pitch sigma in men (17.9 – 25.8 years; N = 157) during reading based on mean SFF values, and reported pitch sigma values of 3.2 STs, which is one semitone higher than the current finding of 2.18 STs. But despite the differences in pitch sigma between the present study and previous studies, measures of SFF range agreed well. Hudson and Holbrook found an SFF range of 11.54 STs, while the present study found 11.33 STs.

Previous studies of vocal variability measures in young adult females can also be compared to the present study. Hudson and Holbrook (1982) reported a pitch sigma of 3.3 STs for 100 female subjects, ages 18-29. As with their male data, Hudson and Holbrook's data for females showed a larger pitch sigma than the current study's result of 2.27 STs. For SFF range, Hudson and Holbrook's (1982) findings of 12.22 STs agreed well with the current study's findings of 12.05 STs. More similar to the current study's pitch sigma data were Brown, Morris, and Michel's (1989) findings of 2.00 STs during reading for 25 young women (ages 20-32 years). Unlike Hudson and Holbrook, Brown et al. used mean values when reporting pitch sigma.

In an older study of female variability data, Snidecor (1951) found that for reading, "young adult" females (N = 25) had a pitch sigma of 3.04 STs, somewhat above the current findings of 2.27 ST. In contrast, Snidecor's SFF range value of 21 STs was nine semitones above the current finding of 12.05 STs. The lower values reported by the current study might be attributed to the attempt to eliminate noise-related data points from the acoustic analysis, resulting in a reduced SFF range. In fact, Snidecor (1951) also included data for a “90% range,” as did Philhour (1948), from which he eliminated the upper 5% and lower 5% of frequency data in his samples. His “90% range” SFF range was 10 STs, similar to the 12.05 STs SFF range found in the present study. This similarity, coupled with the results of Philhour (1948) for his 90% range, suggests that the SFF range findings of earlier studies may have included frequency errors and inaccuracies that artificially increased measured SFF range.

Theoretical Implications

In the introduction, it was hypothesized that SFF range may not be the best indicator of vocal variability due to its reliance on the extreme values of a frequency data set. However, the present study indicated that perception of vocal variability correlated very strongly with SFF range ($r = .890$). The correlation between the perception of vocal variability and pitch sigma also yielded a strong relationship ($r = .927$). These data indicated that *both* SFF range and pitch sigma are good indicators of perceptual judgments of vocal variability (i.e., as SFF range and pitch sigma increase, perception of vocal variability will increase and vice versa).

The introduction also stated that there is a suggestion in previous literature that females exhibit more vocal variability than males. This study examined this belief with acoustic measures of vocal variability and did not find significant differences between males and females in terms of mean pitch sigma or mean SFF range values. This finding suggests that males and females have similar pitch variability when measured acoustically, at least in reading samples.

The normative data for pitch variability in the present study were taken from the expanded female and male speaker groups using all normal samples. Due to time constraints, the researchers were unable to examine the hypothesis of similar variability across genders in terms of perceptual judgments of vocal variability. Perhaps listeners might detect differences in vocal variability between genders that were not measured. Further, it should be noted that all vocal variability measures in the current study were based on reading. It is possible that in spontaneous speech samples, different patterns of variability may be seen for male and female speakers.

Clinical Implications

One of the most interesting findings of the present study was the difference between frequency measures *prior to* adjusting the analysis range and *after* adjustment. For all speech samples, a preliminary analysis was initially completed, and data on the speaker's mean speaking fundamental frequency (SFF), lower and upper SFF limits, pitch sigma, and SFF range were recorded. For a vast majority (93%) of the speech samples the presence of outlier data points necessitated adjustment of the analysis range, and a final analysis was then completed to ensure that only valid data points were included. The data obtained from this final analysis were used for answering all of the research questions for this study.

Figure 4-1: Two Examples of the Procedure Used for Adjusting the Acoustic Analysis Range during Preliminary and Final Acoustic Analyses

Figure 4-1a: Preliminary acoustic analysis for sample 1. Outlier points are noted by the arrow.

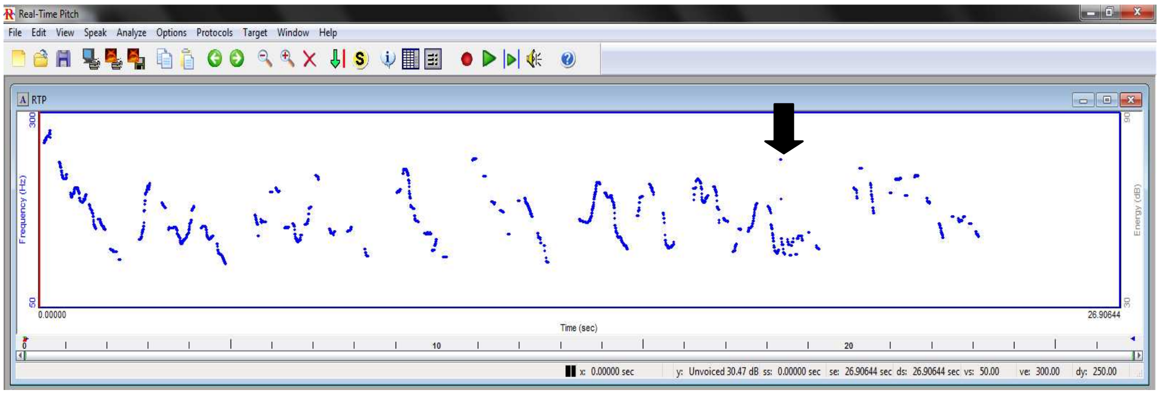


Figure 4-1b: Final acoustic analysis for sample 1. The outlier points could not be removed with analysis range adjustments; thus they were retained.

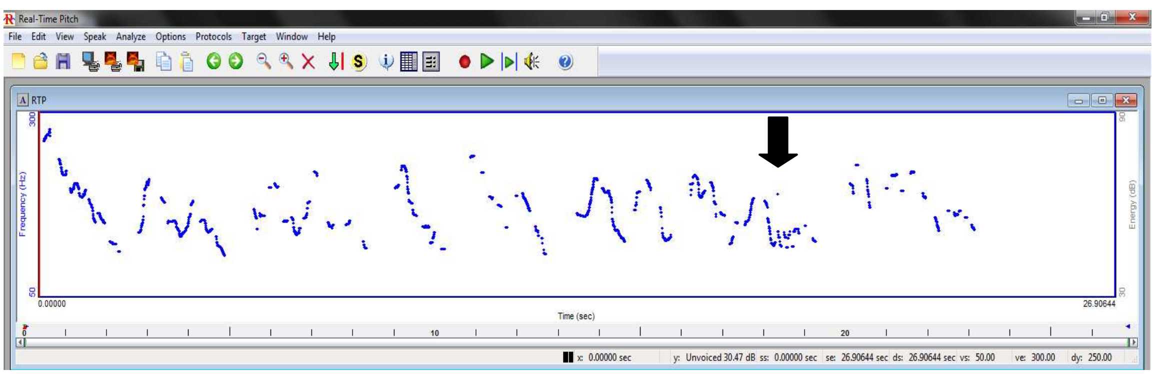


Figure 4-1c: Preliminary acoustic analysis for sample 2. Outlier points are noted by the arrow.

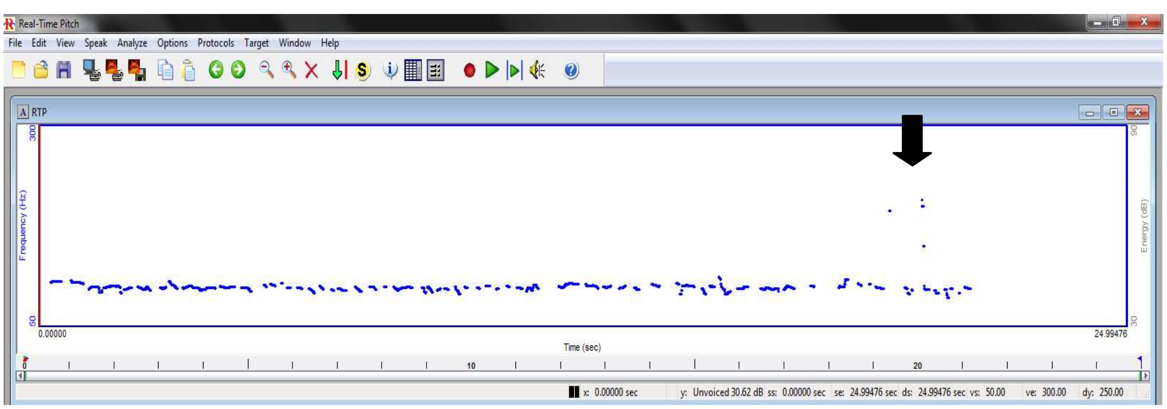
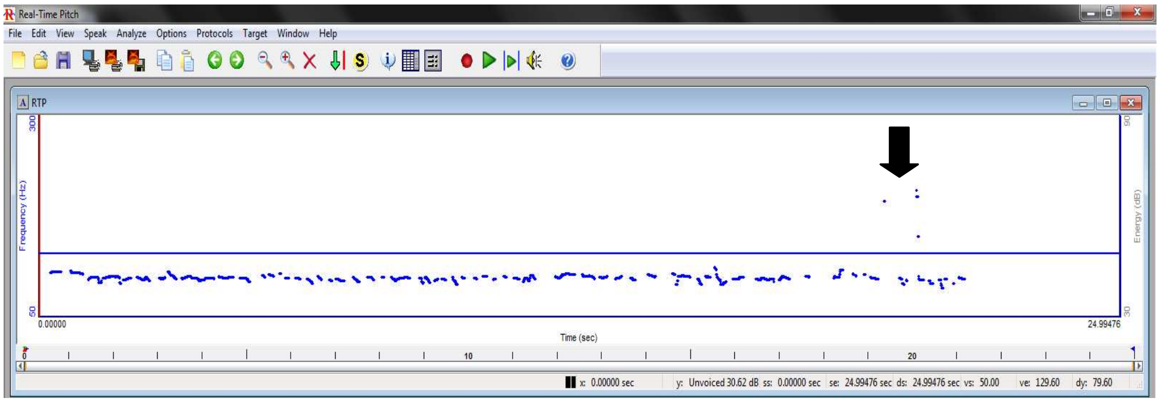


Figure 4-1d: Final acoustic analysis for sample 2. Analysis range adjustment of the upper limit (horizontal line) was used to remove the outlier points.



An example of our procedure for adjusting analysis range can be seen in Figure 4-1. The first pair of frequency tracings (Figure 4-1a and 4-1b) represents a speech sample with few outliers. The main outliers, seen around 18 seconds, could not be removed with analysis range adjustment without also removing legitimate data points such as those seen at 1 second. Thus, no adjustment of analysis range was done, and Figure 4-1b shows the data points that were included in the analysis. In contrast, in Figure 4-1c, outliers can be seen at 19-20 seconds. Including these outliers in the acoustic analysis would give an exaggerated outcome for measures of pitch variability. Consequently, analysis range was adjusted (as seen with the horizontal line) to remove these data points from the sample (see Figure 4-1d). Prior to removing the outliers, the investigator listened carefully to the sample to ensure that the outliers did not reflect a sudden upward pitch inflection. Figure 4-1d shows the data points that were included in the analysis (below the horizontal line).

In order to ensure a high degree of inter-rater reliability for setting the analysis range, all of the speech samples were analyzed separately by two examiners and then compared to determine similarity. Acoustic findings that met strict inter-rater reliability criteria (see *Acoustic Analysis* section for specific criteria) were averaged. When results did not meet the criteria, the two examiners analyzed the samples together to resolve their differences on where to set the analysis range. The researchers obtained the same acoustic findings on 50% of samples, averaged 4% of samples, and analyzed the remaining 46% together in order to come to an agreement on where to set the analysis range.

The relatively low agreement achieved among researchers (i.e., 50%) illustrates the difficulty present in adjusting the analysis range using the procedures outlined in the current study. While it was clear that outliers were present in most speech samples, it was less clear where exactly the upper and lower limits of the analysis range should be placed. In almost half the cases (46%), it was necessary for *both* investigators to examine the samples together, discuss which points would be considered outliers, and determine upper and lower analysis range limits. However, despite the complexity of the process, it proved to be crucial for obtaining acoustic data that correlated with listener perceptions of pitch variability. This necessity was discovered by the researchers after all of the data were analyzed, when we decided to investigate what happened to the correlations when the analysis range was not adjusted (i.e., when data from the *preliminary* analysis was used rather than data from the *final* analysis).

The two correlations we obtained based on data from the preliminary analysis examined whether perceptual judgments of pitch variability correlate with unadjusted acoustic measures for both male and female speakers. Table 4-1 shows the preliminary analyses for the experimental male and female speakers (mean SFF, pitch sigma, and SFF range). To explore the effect of adjusting analysis range, correlational analyses between preliminary pitch sigma data and perceptual ratings, and preliminary SFF range data and perceptual ratings, were compared to the correlation between these measures when data from the final analysis were used (see Table 4-2). Correlations based on the preliminary data were significantly weaker. In fact, virtually no relationship between preliminary SFF range and perceptual ratings of vocal variability was found**.** Thus, when using software programs that detect fundamental frequency in connected speech through auto-correlation and peak-picking (which may lead to aliasing) in clinical situations, it is essential to set the analysis range appropriately for each speech sample that is analyzed. This involves removing as many isolated points and artifacts from the analysis range as possible. While this process is time consuming and subjective, it appears to result in acoustic vocal variability data that correlate strongly with perceptual impressions.

The manual for Real-Time Pitch (RTP; KayPENTAX, 2009) and a technical article on the pitch extraction method used for RTP (Snell, 1995) support the notion that the individual doing the pitch analysis needs to exercise judgment in removing artifacts from the speech signal. Snell (1995) states that automatic pitch extraction techniques are not and probably will never be completely accurate, and that "the best pitch extraction system for processing a speech waveform file combines an accurate time-domain pitch extractor with an experienced human operator who listens to the signal and examines the graph of the waveform to correct the errors produced by the machine-based system" (p. 11). The manual for Real-Time Pitch (KayPENTAX, 2009) similarly notes that RTP "rarely misses glottal events, but it reports more spurious points than MDVP (sic)" (p. 18). Pitch-tracking errors can be reduced by adjusting the pitch analysis range, according to KayPENTAX (2009).

In summary, there is no substitute for good clinical judgment in obtaining reliable and valid acoustic data from the KayPENTAX software. In order for quality data to be obtained from automatic pitch extraction software, it is essential for the pitch analysis range to be set by a human listener (preferably experienced). In addition, specific procedures need to be developed to aid clinicians in configuring the software appropriately. Clinicians will always need to use these procedures as guidelines and combine them with knowledge of normal and expected frequency values in order for the data to be both reliable and valid.

Table 4-1: Speaking Fundamental Frequency (SFF) and Acoustic Measures of Pitch Variability for Reading in Males and Females using Preliminary Data (Unadjusted Analysis Range)

	Preliminary Analysis		Final Analysis with Adjusted Analysis Range	
Measure	Male (N = 15)	Female (N = 19)	Male (N = 15)	Female (N = 19)
Mean SFF in Hz (Minimum-Maximum)	125.24 (104.16 – 173.86)	218.62 (192.83 – 246.89)	122.73 (100.39 – 173.86)	215.92 (185.88 – 247.16)
Pitch Sigma in ST (Minimum-Maximum)	2.85 (1.8 – 4.3)	2.74 (1.76 – 3.89)	2.18 (0.77 – 4.30)	2.27 (1.07 – 3.71)
SFF Range in ST (Minimum-Maximum)	18.2 (15 – 22)	24.7 (19 – 30)	11.33 (4.00 – 21.00)	12.05 (6.00 – 18.00)

Table 4-2: Pearson Product-Moment Correlation Coefficients Relating Mean Perceptual Judgments of Pitch Variability with Acoustic Measures of Pitch Variability for the Experimental Samples using Preliminary Acoustic Data

Correlation	Preliminary Analysis	Final Analysis with Adjusted Analysis Range
Pitch Sigma and Perceptual Judgments	.652*	.927*
SFF Range and Perceptual Judgments	.068	.890*

*significant at $p < .01$

Implications for Future Research

Future research on correlations between perceptual judgments and acoustic measures of vocal variability should be conducted with racially diverse populations. This research may determine whether or not separate norms are needed based on a speaker's culture. By providing these correlations for diverse populations, data could be applied to speakers of different racial, ethnic, and linguistic backgrounds with more certainty.

The speaker recording procedures were designed to elicit a significant amount of vocal variability. That is, speakers were required to provide three different speech samples (normal, monotone, and expressive). From this, the researchers chose speech samples for the CDs that represented a continuum of vocal variability from very monotone to very expressive, in order to examine correlations at the extremes. In the future, obtaining perceptual measures of vocal variability, and analyzing correlations between perceptual judgments and acoustic measures of pitch variability during *normal* reading conditions would enhance the generalizability of the results to clinical situations. Additionally, this knowledge would help to clarify whether males and females have differences in vocal variability when measured perceptually, since this study shows that males and females do not display significant acoustic differences.

A considerable amount of time was spent on developing procedures for the acoustic analysis process and completing acoustic analysis on all speech samples to ensure that outlier data points were excluded from that data set. Correlations obtained between SFF range and perceptual judgments of pitch variability, and pitch sigma and perceptual judgments, when using *preliminary* acoustic data revealed the necessity of adjusting the analysis range for all speech samples. Nevertheless, using the procedures outlined in the current study (see *Acoustic Analysis)*, resulted in only about 50% agreement regarding where the analysis range should be placed. For this reason, it is important that future research identifies specific procedures to eliminate artifacts from the acoustic analysis so that the resulting frequency data set for each subject is both reliable and valid. It is also essential that these procedures emphasize the importance of using the examiner's knowledge, experience, and clinical judgment when obtaining acoustic data.

Conclusion

Through the methods defined in this study, it was determined that both pitch sigma and SFF range yield strong positive correlations with perceptual judgments of pitch variability. When male and female voices were separated, there were no gender differences in the strength of these correlations for either pitch sigma or SFF range. In addition, there were no significant differences between males and females in terms of mean pitch sigma or mean SFF range values.

This study also demonstrated that it is essential for clinicians to develop reliable and valid procedures for frequency analysis of speech samples. It is important for clinicians to adjust the analysis range in order to include only legitimate data points, thus excluding data points that are artifacts. This procedure allows for a better representation of an individual's acoustic vocal variability measures, and also has a strong positive correlation with listener perceptions of pitch variability.

REFERENCES

Acoustical Society of America (1960). *American Standard Acoustical Terminology.* New York, NY: Acoustical Society of America.

Allee, J. G. (1986). *Webster's dictionary*. Baltimore, MD: Ottenheimer Publishers, Inc.

American Speech-Language Hearing Association (1997). *Guidelines for audiologic screening*. Available from www/asha.org/policy.

Andrews, M.L. (2006). *Manual of Voice Treatment: Pediatrics through Geriatrics* (3rd ed.). Clifton Park, NY: Thomson/Delmar Learning.

Andrews, M.L. and Schmidt, C.P. (1997). Gender presentation: Perceptual and acoustical analyses of voice. *Journal of Voice,* 11(3), 307-313.

Baken, R.J. (1987). *Clinical measurement of speech and voice.* Boston, MA: College-Hill Press.

Baken, R.J., and Orlikoff, R. (2000). *Clinical measurement of speech and voice* (2nd edition). San Diego, CA: Singular Publishing Group.

Behrman, A. (2007). *Speech and voice science.* San Diego, CA: Plural Publishing, Inc.

Benjamin, B.J. (1981). Frequency variability in the aged voice. *Journal of Gerontology*, 36(6), 722-726.

Boone, D.R. and McFarlane, S.C. (2000). *The voice and voice therapy* (6th ed.). Needham Heights, MA: Allyn and Bacon.

Brown, W.S., Morris, R.J., and Michel, J.F. (1989). Vocal jitter in young adult and aged female voices. *Journal of Voice,* 3, 113-119.

Ferrand, C.T. (2007). *Speech science: An integrated approach to theory and clinical practice* (2nd ed.). Boston, MA: Allyn and Bacon.

Ferrand, C.T. (2011). *Voice disorders: Scope of theory and practice.* Boston, MA: Allyn and Bacon.

Gelfer, M.P. (1988). Perceptual attributes of voice: Development and use of rating scales. *Journal of Voice*, 2(4), 320-326.

Gelfer, M.P. and Mordaunt (2012). Pitch and intonation. In: R.K. Adler, S. Hirsch, and M. Mordaunt (Eds.), *Voice and Communication Therapy for the Transgender/Transsexual Client: A Comprehensive Clinical Guide* (2nd ed., pp. 187-223). San Diego, CA: Plural Publishing, Inc.

Gelfer, M.P. and Schofield, K.J. (2000). Comparison of acoustic and perceptual measures of voice in male-to-female transsexuals perceived as female versus those perceived as male. *Journal of Voice*, 14(1), 22-33.

Hanley, T. D., & Peters, R. (1971). The speech and hearing laboratory. In L. Travis (Ed.), *Handbook of Speech Pathology and Audiology* (pp. 106-108). Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Healey, H.E. (1982). Speaking fundamental frequency characteristics of stutterers and nonstutterers. *Journal of Communication Disorders,* 15, 21-29.

Hollien, H. (1974). On vocal registers. *Journal of Phonetics,* 2, 125-143.

Hollien, A. and Jackson, B. (1973). Normative data on the speaking fundamental characteristics of young adult males. *Journal of Phonetics,* 1, 117-120.

Horri, Y. (1975). Some statistical characteristics of voice fundamental frequency. *Journal of Speech and Hearing Research,* 26, 467-471.

Hudson, A., and Holbrook, A. (1982). Fundamental frequency characteristics of young black adults: spontaneous speech and oral reading. *Journal of Speech and Hearing Research*, 25, 25-28.

KayPENTAX (2009). Software instruction manual: Real-time pitch, model 5121 (Issue E). Lincoln Park, NJ: KayPENTAX.

Kempster, G.B., Gerratt, B.R., Abbott, K.V., Barkmeier-Kraemer, J., and Hillman, R.E. (2009) Consensus auditory-perceptual evaluation of voice: Development of a standardized protocol. *American Journal of Speech-Language Pathology*, 18, 124-132.

Linke, C.E. (1973). A study of pitch characteristics of female voices and their relationship to vocal effectiveness. *Folia Phoniatrica*, 25, 173-185.

Lowry, R. (2013). Significance of the difference between two correlation coefficients. Retrieved from: http://vassarstats.net/rdiff.html.

McGlone, R.E., and Hollien H. (1963) Vocal pitch characteristics of aged women. *Journal of Speech and Hearing Research*, 6(2), 164-170.

Mysak, E.D. (1959). Pitch and duration characteristics of older males. *Journal of Speech and Hearing Research*, 2, 46-54.

Nadig, A. and Shaw, H. (2012). Acoustic and perceptual measurement of expressive prosody in high-functioning autism: Increased pitch range and what it means to listeners. *Journal of Autism and Developmental Disorders*, 42, 499-511.

Philhour, C. W. (1948). *An experimental study of the relationships between perception of vocal pitch in connected speech and certain measures of vocal frequency*. (Doctoral dissertation, State University of Iowa).

Schiavetti, N. and Metz, D.E. (2006). *Evaluating Research in Communicative Disorders* (5th ed.). Boston, MA: Allyn and Bacon.

Snell, R. (1995). An overview of pitch extraction techniques. Victoria, Canada: Speech Technology Research Limited.

Snidecor, J.C. (1951). The pitch and duration characteristics of superior female speakers during oral reading. *Journal of Speech and Hearing Disorders*, 16, 44-52.

Stoicheff, M.L. (1981). Speaking fundamental frequency characteristics of nonsmoking female adults. *Journal of Speech and Hearing Research,* 24, 437-441.

Wolfe, V.I., Ratusnik, D.L, Smith, F.H., Northrop, G. (1990). Intonation and fundamental frequency in male-to-female transsexuals. *Journal of Speech and Hearing Disorders*, 55, 43-50.

Appendix A

Script for Speaker Recruitment

Hi everyone my name is Bree and I am a graduate student in the Communication Sciences and Disorders program. I am in the process of writing a thesis and I am here today to recruit potential participants. My thesis is designed to look at the perceptual impressions of voice and right now I need individuals who are native English speakers and who do not have a history of asthma. The total time commitment for your participation would be just 30 minutes or less.

I now want to take a minute to discuss what your commitment would all involve. The first step would be a short phone conversation with me asking you a few questions to make sure that you fit all of the criteria needed for this study. If you pass this screening, we would plan a day and time that fits into your schedule for you to come to the UW-M Speech and Language clinic which is located on the 8th floor of Enderis Hall. After the study is described in more detail, we would take a couple minutes to screen your speech and hearing. If you pass these screening measures, you would then qualify as a speaker for the study. At this point we would ask you to familiarize yourself with a one paragraph reading sample and, when you are comfortable, you would be asked to read this paragraph while being recorded. We would obtain a few recordings with you using different speaking styles. After this time, your participation in the study would be completed. Again, your total time commitment would only be 30 minutes, if not less, and the reading task is very simple. I am going to pass around a sign-up sheet so if you are a native English speaker, do not have asthma, and this sounds like something you would be interested in doing I encourage you to sign up. If you could just list your name, email, phone number, and when a good time is for me to reach you I would really appreciate it. Thank you for letting me take up a minute of your time and I hope to see some of you soon.

Appendix A (continued)

Participant Sign-Up Sheet

Name	**E-mail**	**Phone**	**What time of day is best to contact you?**
1.			
2.			
3.			
4.			
5.			
6.			
7.			
8.			
9.			
10.			
11.			
12.			
13.			
14.			
15.			
16.			
17.			
18.			
19.			
20.			

Appendix B

Speaker Questionnaire - Phone Screening

Participant Initials: __________ Date: __________

Are you between the ages of 18 and 35 years? ❑ Yes ❑ No

Is your native language English? ❑ Yes ❑ No

No: What is your native language?

Have you had any history of respiratory problems, such as asthma? ❑ Yes ❑ No

Yes: When and what type

Have you ever received speech or voice therapy? ❑ Yes ❑ No

Yes: Explain

Have you ever been diagnosed with a voice disorder? (e.g., vocal nodules) ❑ Yes ❑ No

Have you ever been diagnosed with a neurological disorder? ❑ Yes ❑ No

Do you have any history of hearing loss? ❑ Yes ❑ No

Yes: Describe

PASS: ❑ Yes ❑ No

Have you ever worked in any of the following fields:

Drama, broadcast journalism, vocal performance, or speech communication? ❑ Yes ❑ No

Do you have a history of voice/singing training?	❑ Yes	❑ No
Yes: Describe		
Are you a student majoring in drama, broadcast journalism, vocal performance, communication sciences and disorders, or speech communication?	❑ Yes	❑ No
Yes: What is your major?		

Appendix C

Notice of Institutional Review Board (IRB) Approval

UNIVERSITY of WISCONSIN
UWMILWAUKEE
Department of University Safety & Assurances

Jessica Rice
IRB Administrator
Institutional Review Board
Engelmann 270
P. O. Box 413
Milwaukee, WI 53201-0413
(414) 229-3182 *phone*
(414) 229-6729 *fax*

http://www.irb.uwm.edu
ricej@uwm.edu

New Study - Notice of IRB Expedited Approval

Date: October 25, 2012

To: Marylou Gelfer, PhD
Dept: Communication Sciences and Disorders

Cc: ---

IRB#: 13.142
Title: Perceptual Correlates of Acoustic Measures of Vocal Variability

After review of your research protocol by the University of Wisconsin – Milwaukee Institutional Review Board, your protocol has been approved as minimal risk Expedited under Categories 6 and 7 as governed by 45 CFR 46.110.

This protocol has been approved on **October 25, 2012** for one year. IRB approval will expire on **October 24, 2013**. If you plan to continue any research related activities (e.g., enrollment of subjects, study interventions, data analysis, etc.) past the date of IRB expiration, a continuation for IRB approval must be filed by the submission deadline. If the study is closed or completed before the IRB expiration date, please notify the IRB by completing and submitting the Continuing Review form found on the IRB website.

Unless specifically where the change is necessary to eliminate apparent immediate hazards to the subjects, any proposed changes to the protocol must be reviewed by the IRB before implementation. It is the principal investigator's responsibility to adhere to the policies and guidelines set forth by the UWM IRB and maintain proper documentation of its records and promptly report to the IRB any adverse events which require reporting.

It is the principal investigator's responsibility to adhere to UWM and UW System Policies, and any applicable state and federal laws governing activities the principal investigator may seek to employ (e.g., FERPA, Radiation Safety, UWM Data Security, UW System policy on Prizes, Awards and Gifts, state gambling laws, etc.) which are independent of IRB review/approval.

Contact the IRB office if you have any further questions. Thank you for your cooperation and best wishes for a successful project

Respectfully,

Jessica P. Rice
IRB Manager

Appendix D

Modified Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)

Subject Number: __________ Date: __________

Complete the following tasks:

1. Hold the vowels /a/ and /i/ for 3-5 seconds duration each.

2. Say the following sentences:

   a. The blue spot is on the key again.
   b. How hard did he hit him?
   c. We were away a year ago.
   d. We eat eggs every Easter.
   e. My mama makes lemon muffins.
   f. Peter will keep at the peak.

3. Give a brief (~15 second) response to: "Tell me about your voice."

Check all that apply:

_____ appropriate pitch

_____ appropriate loudness

_____ appropriate resonance

_____ roughness

_____ breathiness

_____ strain

_____ hypernasality

_____ hyponasality

Articulation Screening

List any phonemes that the speaker frequently or consistently misarticulated:

PASS: ❑ Yes ❑ No

Appendix D (continued)

Say the following sentences:

a. The blue spot is on the key again.

b. How hard did he hit him?

c. We were away a year ago.

d. We eat eggs every Easter.

e. My mama makes lemon muffins.

f. Peter will keep at the peak.

Appendix E

Hearing Screening

Subject Number: ___________ Date: __________

Birthdate (month and year): ___________ Age (in years and months): ___________

Do you think you have a hearing loss?	Yes	No
Having hearing aid(s) ever been recommended for you?	Yes	No
Is your hearing better in one ear?	Yes	No
If yes, which is the better ear? Right Left		
Have you ever had sudden or rapid progression of hearing loss?	Yes	No
If yes, which ear? Right Left		
Do you have ringing or noises in your ears?	Yes	No
If yes, which ear? Right Left		
Do you consider dizziness to be a problem for you?	Yes	No
Have you had recent drainage from your ear(s)?	Yes	No
If yes, which ear? Right Left		
Do you have pain or discomfort in your ear(s)?	Yes	No
If yes, which ear? Right Left		
Have you received medical consultation for any of the above conditions?	Yes	No

PASS **REFER**

Visual/Otoscopic Inspection

Referral for cerumen management _______Referral for medical evaluation ___________

PASS **REFER**

Pure-Tone Screen (25 dB HL) (R = Response, NR = No Response)

Frequency	1000 Hz	2000 Hz	4000 Hz
Right Ear			
Left Ear			

PASS **REFER**

Appendix F

The Grandfather Passage

You wished to know all about my grandfather. Well, he is nearly ninety-three years old. He dresses himself in an ancient black frock coat, usually minus several buttons; yet he still thinks as swiftly as ever. A long, flowing beard clings to his chin, giving those who observe him a pronounced feeling of the utmost respect. When he speaks his voice is just a bit cracked and quivers a trifle. Twice each day he plays skillfully and with zest upon our small organ. Except in the winter when the ooze or snow or ice prevents, he slowly takes a short walk in the open air each day. We have often urged him to walk more and smoke less, but he always answers, “Banana Oil!” Grandfather likes to be modern in his language.

Appendix G

Data Collection Form

Subject Number: ___________ Date: ___________

Normal: Preliminary Analysis

Mean SFF:
Lower SFF Limit:
Upper SFF Limit:
SFF Range:
Pitch Sigma

Normal: Final Analysis

Mean SFF:
Lower SFF Limit:
Upper SFF Limit:
SFF Range:
Pitch Sigma

Expressive: Preliminary Analysis

Mean SFF:
Lower SFF Limit:
Upper SFF Limit:
SFF Range:
Pitch Sigma

Expressive: Final Analysis

Mean SFF:
Lower SFF Limit:
Upper SFF Limit:
SFF Range:
Pitch Sigma

Monotone: Preliminary Analysis

Mean SFF:
Lower SFF Limit:
Upper SFF Limit:
SFF Range:
Pitch Sigma

Monotone: Final Analysis

Mean SFF:
Lower SFF Limit:
Upper SFF Limit:
SFF Range:
Pitch Sigma

Appendix H

Listener Questionnaire (Phone Screening)

Participant Initials: __________ Date: __________

Are you between the ages of 20 and 30 years? ❑ Yes ❑ No

Is your native language English? ❑ Yes ❑ No

Are you currently enrolled in a Communication Sciences and Disorders graduate program? ❑ Yes ❑ No

Do you have any history of hearing loss? ❑ Yes ❑ No

Yes: Describe

Have you taken a graduate level course in voice disorders? ❑ Yes ❑ No

PASS: ❑ Yes ❑ No

Appendix I

Listener Instructions

For this study, we want you to listen to the following voice recordings, and judge the amount of pitch variability displayed by each speaker. We want you to judge the amount to tonal movement – the amount of pitch change apparent in each reading. The voices have clear differences in quality, duration, and pitch level (or SFF), and you may like some of the voices more than others. However, we want you to try to disregard these factors, restricting your judgment to pitch *variability* only.

Before making any judgments, you will hear a brief sampling of the experimental voices that we will later ask you to rate. As you listen to the voices, we would like you to establish your standards of pitch variability from this stimulus set. That is, your judgments should be relative based upon these samples of speech rather than some pre-conceived ideas of pitch monotony and pitch variability. The voice or voices with the greatest amount of variability in this sample should be rated as a 7 while those with the least amount of pitch variability in this sample should be rated as a 1.

Once we have listened to the sampling of experimental voices, the rating procedure will begin. Please use the packets that have been given to you. I will present each Speaker Number to you visually. Write the Speaker Number in the blank preceding each rating scale. Listen to all of each sample of speech before making your judgment of it. (You will first hear [gender] voices. After you have completed the ratings of the [gender] voices, we will follow the same procedure for the [opposite gender] voices.)

Appendix J

Listener Response Sheet

Listener Number: ____________ Date: ____________
Gender: ____________

Speaker Number: ____________

1	2	3	4	5	6	7
Complete Monotone			Average			Extreme Variability

Speaker Number: ____________

1	2	3	4	5	6	7
Complete Monotone			Average			Extreme Variability

Speaker Number: ____________

1	2	3	4	5	6	7
Complete Monotone			Average			Extreme Variability

Speaker Number: ____________

1	2	3	4	5	6	7
Complete Monotone			Average			Extreme Variability

Speaker Number: ___________

1	2	3	4	5	6	7
Complete Monotone			Average			Extreme Variability