2016

Fundamental Frequency Characteristics of Modal and Vocal Fry Registers

Heidi Brubaker
heidib@bgsu.edu

Jason Albertson Whitfield Ph.D.
Bowling Green State University, jawhitf@bgsu.edu

Jane Schoonmaker Rodgers D.M.A.
Bowling Green State University, janesro@bgsu.edu

FUNDAMENTAL FREQUENCY CHARACTERISTICS OF MODAL AND VOCAL FRY REGISTERS

HEIDI BRUBAKER

HONORS PROJECT

Submitted to the Honors College

at Bowling Green State University in partial

fulfillment of the requirements for graduation with

UNIVERSITY HONORS

DECEMBER 12th, 2016

Advised by:

Jason Whitfield, Ph.D., Department of Communication Sciences and Disorders

Jane Schoonmaker Rodgers, D.M.A., College of Musical Arts

Abstract

This study examined the distribution of fundamental frequencies in the connected speech of 14 healthy young adults. Acoustic analysis of fundamental frequency was performed on previously collected speech samples of a phonetically balanced reading passage. For the first three sentences of the reading passage, fundamental frequency contours were extracted using PRAAT, a speech analysis software package. The accuracy of these contours were visually verified and manually corrected when needed. The distribution of the fundamental frequency histories for each sample were then analyzed using Gaussian Mixture Model analyses in MATLAB. For most speakers, four statistical modes were identified in the data based on model optimization. The lowest statistical mode was located in a frequency region that was consistent with the vocal fry register. This lowest statistical mode made up only around 5 percent of all glottal cycles, on average across both male and female participants. The results are discussed in relation to normal voice production, voice disorders, and vocal performance.

INTRODUCTION

Physiology and Acoustics of Voice

Speech can be defined in simple terminology as the production of sound from the movements of the vocal tract. There are two types of speech sounds, voiceless and voiced. Voiceless sounds are created by an aperiodic noise source generated by airflow through a contraction in the vocal tract, without vocal fold vibration. Voiced speech sounds are characterized by vocal fold vibration, or the quasi-periodic opening and closing of the vocal folds, which create a vibratory sound source (Slifka, 2006). This cyclic opening and closing of the vocal folds generates acoustic compressions and rarefactions that are the primary source of sound for speech. The rate at which the vocal folds open and close is referred to as vocal fundamental frequency (f_0). Perceptually, f_0 is perceived as pitch. Because the mass of the vocal folds determines f_0 (Whiteside, 2001), men have a lower f_0 than women and women have a lower f_0 than children. The f_0 of male speakers is around 100Hz on average, while female speakers have an f_0 that is usually around 200Hz and Children have an f_0 usually around 300Hz (Peaterson & Barney, 1952).

Phonation

Because the vocal folds are housed within the larynx, phonation is a strictly a laryngeal phenomenon. The vocal folds are paired membranous folds of tissue that are situated at the superior end of the trachea. Each vocal fold is attached posteriorly to the arytenoid cartilages, and they converge anteriorly to attach the posterior aspect of the thyroid cartilage. Prior to the initiation of phonation, they are brought toward midline by contraction of arytenoid muscles. The air space between the vocal folds is the glottis. The area below the glottis is called the subglottal space and the space above is the supraglottal space. In order for phonation to occur, there must be a pressure differential between the subglottal and supraglottal spaces, with the subglottal space having a more positive pressure than the supraglottal space. This pressure differential will cause the vocal folds be set into vibration. When the folds are opening they move laterally, being pushed by the positive subglottal pressure. The lower margin of the vocal folds open prior to the superior margin. During opening, therefore, the glottal configuration is convergent. The vocal folds are maximally open for only a fraction of a second. During this optimum period of vocal fold separation, the pressure gradient is at equilibrium. However, the increased elasticity of the vocal fold tissue draw the vocal folds back together. Again, the lower margin of the vocal folds lead the upper margin, creating a divergent glottal configuration during closing. This shape of the vocal folds creates a more negative pressure in the glottis, pulling the vocal folds towards midline. After complete closure, the cycle of phonation repeats.

Vocal Registers

There are at least three categories of vocal register discussed in the literature, including falsetto, modal, and fry (Hollien, 1968). Modal is what you would think of as normal or typical speech. This is the most frequently used register. The falsetto register is associated with a very thin and stiff vibration portion of the vocal fold, and is, therefore, associated with a higher f_0 than in modal speech that has a lighter and breathier sound quality. Vocal fry is the lowest register and, therefore, is associated with a low f_0 and sounds creaky, harsh and rough (Wolk, 2012). Physiologically, vocal fry is the result of very close approximation of the vocal folds and irregular vibration characterized multiple air pulses per vibratory cycle (Abdelli-Beruh, 2014).

This study will examine f_0 characteristics in connected speech that are consistent with both modal and vocal fry registers.

Relevance of Vocal Fry in Disordered Speech

Vocal fry can be harmful to vocal fold health over time (Wolk, 2012). As the use of vocal fry is increased proportionally in speech, the chances of disordered speech developing increase in conjunction with this change. Contact granulomas can form as a result of persistent fry (Wolk, 2012). Contact granulomas are essentially an ulcer formed on the vocal fold(s). Professional vocalists, teachers, lawyers, public speakers, and anyone else that uses their voice frequently are susceptible to the development of these voice disorders (Nix, 2005; Schmidt, 1998). When a contact granuloma is formed, individuals may experience symptoms such as hoarse voice, coughing, pain when speaking or coughing, and the sensation of a foreign object in the throat. Treatment ordered for contact granulomas is most frequently vocal rest for up to six weeks. Coaching and informing the patient about vocal health is advisable. As a last resort, surgical removal may be performed, though not typically prescribed.

Classroom teachers who exhibiting high degrees of vocal fry may have difficulties in teaching their students effectively. According to Feldman, it was found that teachers' speaking skills were among the three most important factors in students' evaluations of teachers. Voice use seems to be tied directly to teaching effectiveness (Feldman, 1986; Schmidt, 1998). A teacher with persistent fry use may be a distraction to their students and may correlate with the lower teacher evaluations.

Purpose and Hypotheses

The purpose of this study was to examine the distribution of f_0 values in connected speech of neurologically healthy younger adult speakers determine what the average proportion of f_0 values were in the frequency range that was consistent with the fry register. This study used a purely acoustic methodology as acoustic analyses can provide a relatively easy, inexpensive, and objective measure of vocal function. It was expected that the majority of f_0 values would be located in the higher frequency range associated with modal voice, while only a small portion of f_0 values would be in the lower frequency associated with the vocal fry register. Additionally, it is expected that the overall f_0 of the female speakers will be higher than males, regardless of register.

METHODS

Habitual reading samples of “The Rainbow Passage” (Fairbanks, 1960) that were collected as part of an earlier data collection were analyzed for the current study. Reading samples from a total of 14 participants were 8 of whom were male speakers, Mean Age = 25.5, Range = 24-30, and 6 of whom were female, Mean Age = 20.83, Range = 19-26, were included in the study. Participants were asked to produce the reading passage “at a comfortable rate and loudness, as if they were having a conversation with someone seated across the table.” Speech samples were recorded onto a portable digital audio recorder (Marantz PMD661; sampling rate = 44.1 kHz) using a table-top microphone (Shure SM-58) in a sound-treated booth.

For this study acoustic analyses were completed using PRAAT (Boersma & Weenink, 2015). First, spectrographic and waveform display with overlaid f_0 contour was examined to determine the appropriate frequency range of f_0 for each participant. Using this frequency range, pitch extraction was then completed for each cycle of vocal fold vibration in the first three sentences of the passage. The fundamental frequency of each glottal cycle was extracted using the cross-correlation method in PRAAT (Boersma & Weenink, 2015). Following the automatic extraction, the f_0 traces were inspected for accuracy and manually corrected. For this process, the f_0 trace was displayed simultaneously with the waveform display. f_0 values that were grossly inaccurate were removed from the f_0 history. Glottal cycles for which the f_0 value was inaccurate or not predicted by the automatic analysis were manually calculated by first measuring the period from the left-most negative trough in a cycle to the subsequent most negative trough corresponding to the beginning of the next cycle. The inverse of the period was then calculated and the f_0 was manually added to the history at that time point.

These f_0 histories were then imported into MATLAB to visualize the distribution. Following visual examination of the f_0 distributions, the data were analyzed using a custom MATLAB script to examine the distributional characteristics of the f_0 data for each participant. Using a Gaussian mixture model (GMM) analysis was used to examine the f_0 distribution data for each participant. This function uses an Expectation-Maximization algorithm to estimate parameters of a Gaussian distribution with an expected number of components (referred to as statistical modes). This technique can be useful in identifying the descriptive characteristics of each mode in a data set that has a bimodal or multimodal distribution.

The GMM model fit was optimized for f_0 data of each participant. The data were modeled as having one, two, three, and four predicted statistical modes, both a shared and unshared covariance structure, and diagonal and full variance structures. This yielded a total of sixteen models per participant. The model with the lowest AIC was selected as the best fit. For the best fit model, the mean, standard deviation, and relative proportion were calculated for each identified statistical mode.

Analyses were completed to insure the final model was a good fit. First, the f_0 distributions for each participant (observed values) were visually compared to the model (predicted values). Second, the percent root-mean-square (RMS) error between the observed and predicted values was measured to be 3.8 percent on average (*range*: 2.5 to 5.6 percent). Finally, a Kolmogorov-Smirnov (KS) tests were completed that compared the observed and data predicted by the model to determine if they were statistically different. These comparisons showed there was no significant differences between the observed and predicted values, p>0.05 for all comparisons. Based on these analyses, the optimized model for each participant fit the data well.

RESULTS

Modes identified by the analysis were labeled successive from lowest to highest frequency. Therefore the lowest statistical mode (SM1) represents the lowest collection f_0 values in the distribution, with each higher mode labeled in succession. For 13 of the 14 participants four modes were predicted. The remaining participant the analysis identified 3 statistical modes. For this participant, (M3), visual inspection of the f_0 distribution suggested no data were clustered in the lowest statistical mode, SM1. Therefore, the first identified mode corresponded to SM2. Figures 1 through 14 show the statistical modes that were identified for each participant. Figure one shows the frequency count of f_0 values across the range of fundamental frequency. The red vertical lines represent each statistical mode from lowest to highest (SM1-4) and the surrounding shaded region represents the two standard deviations of the mean of that statistical mode. Each statistical mode is labeled on Figure 1 for clarity.

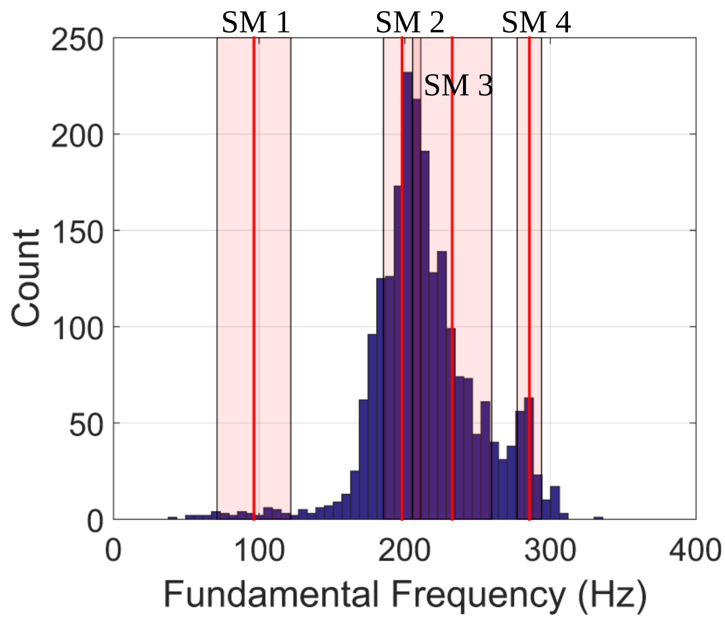

Figure 1. Fundamental frequency distribution for participant F1. Note: The red vertical lines represent each statistical mode from lowest to highest (SM1-4) and the surrounding shaded region represents the two standard deviations of the mean of that statistical mode.

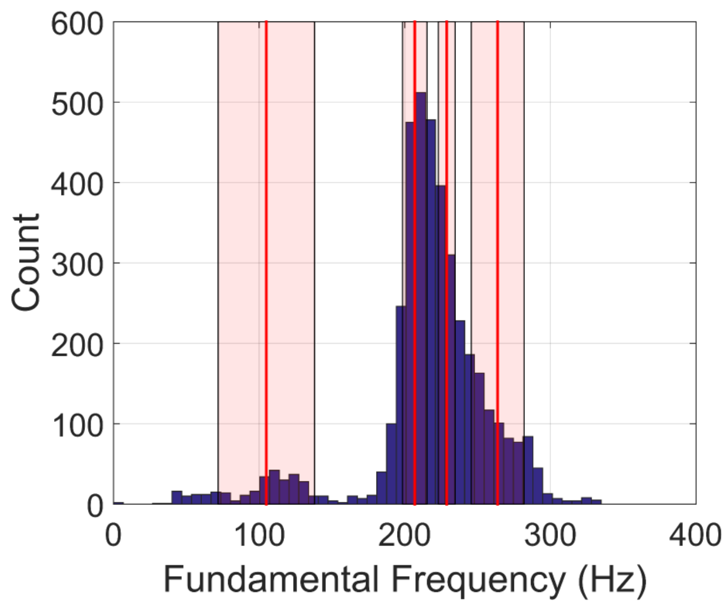

Figure 2. Fundamental frequency distribution for participant F2. Note: The red vertical lines represent each statistical mode from lowest to highest (SM1-4) and the surrounding shaded region represents the two standard deviations of the mean of that statistical mode.

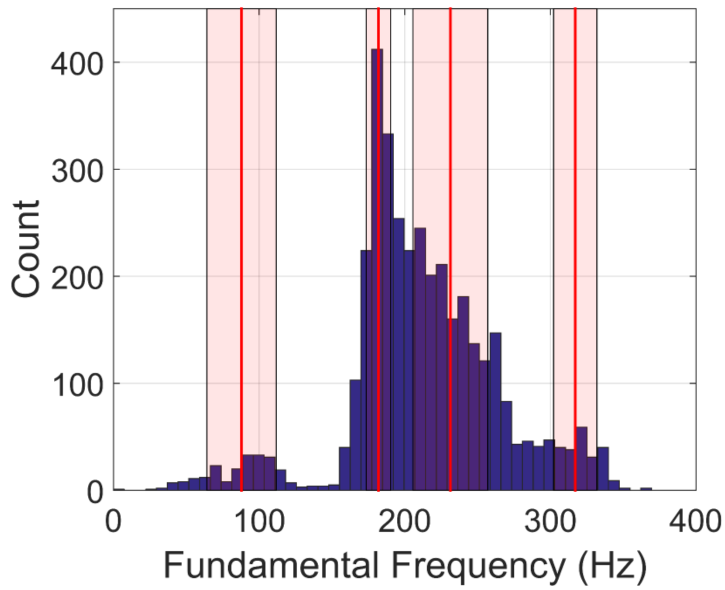

Figure 3. Fundamental frequency distribution for participant F3. Note: The red vertical lines represent each statistical mode from lowest to highest (SM1-4) and the surrounding shaded region represents the two standard deviations of the mean of that statistical mode.

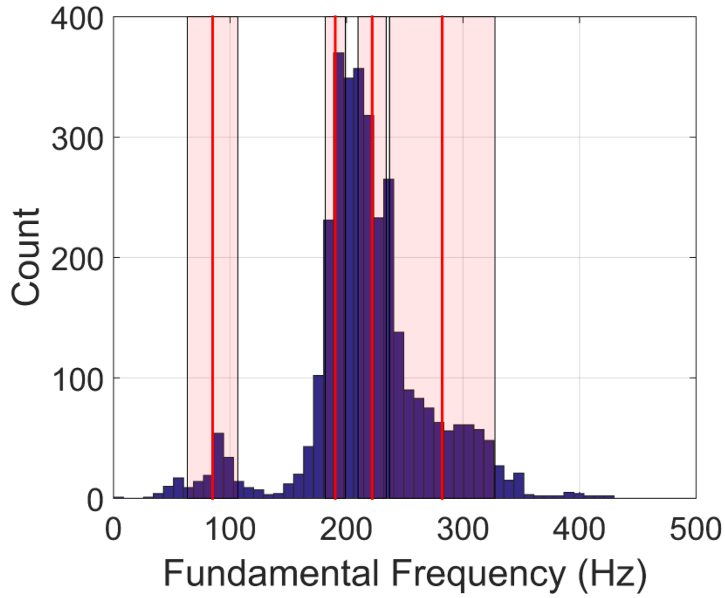

Figure 4. Fundamental frequency distribution for participant F4. Note: The red vertical lines represent each statistical mode from lowest to highest (SM1-4) and the surrounding shaded region represents the two standard deviations of the mean of that statistical mode.

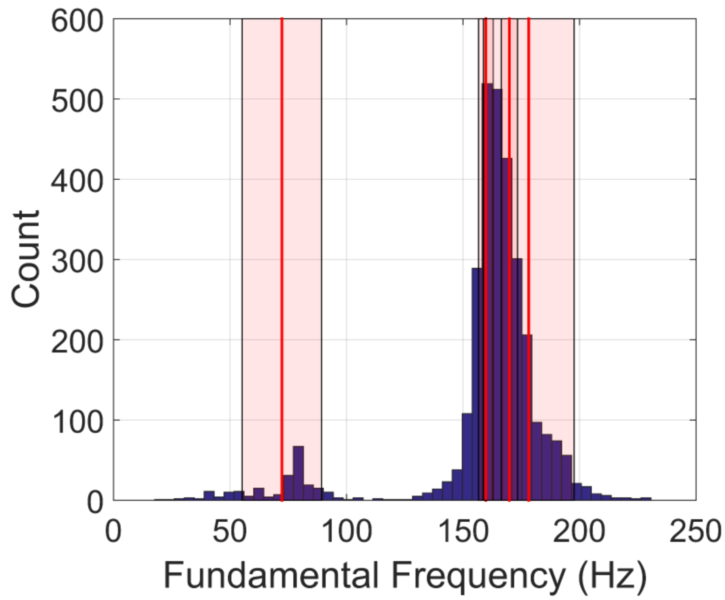

Figure 5. Fundamental frequency distribution for participant F5. Note: The red vertical lines represent each statistical mode from lowest to highest (SM1-4) and the surrounding shaded region represents the two standard deviations of the mean of that statistical mode.

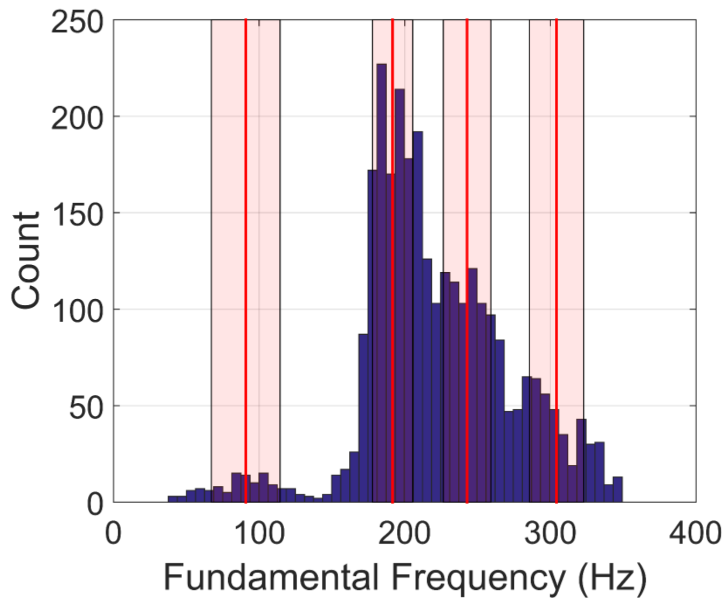

Figure 6. Fundamental frequency distribution for participant F6. Note: The red vertical lines represent each statistical mode from lowest to highest (SM1-4) and the surrounding shaded region represents the two standard deviations of the mean of that statistical mode.

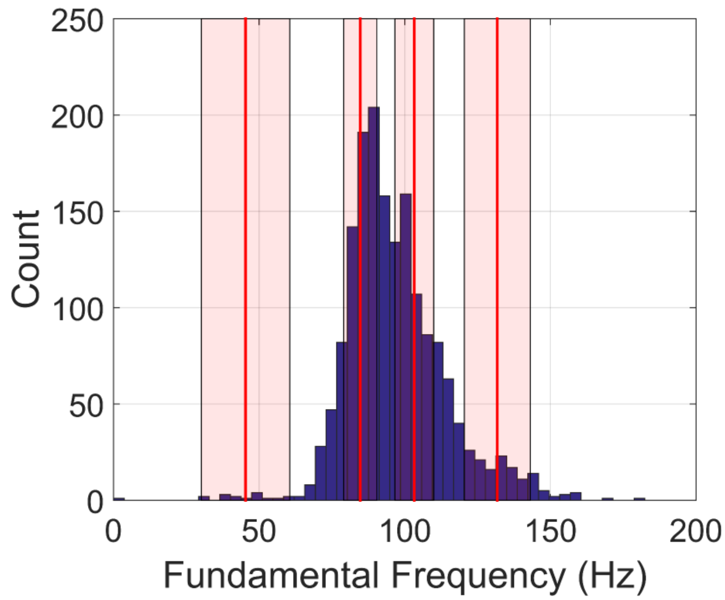

Figure 7. Fundamental frequency distribution for participant M1. Note: The red vertical lines represent each statistical mode from lowest to highest (SM1-4) and the surrounding shaded region represents the two standard deviations of the mean of that statistical mode.

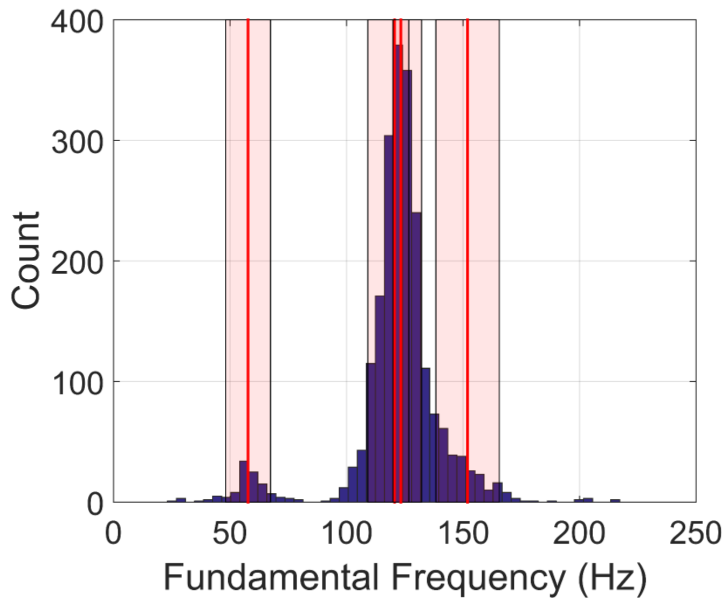

Figure 8. Fundamental frequency distribution for participant M2. Note: The red vertical lines represent each statistical mode from lowest to highest (SM1-4) and the surrounding shaded region represents the two standard deviations of the mean of that statistical mode.

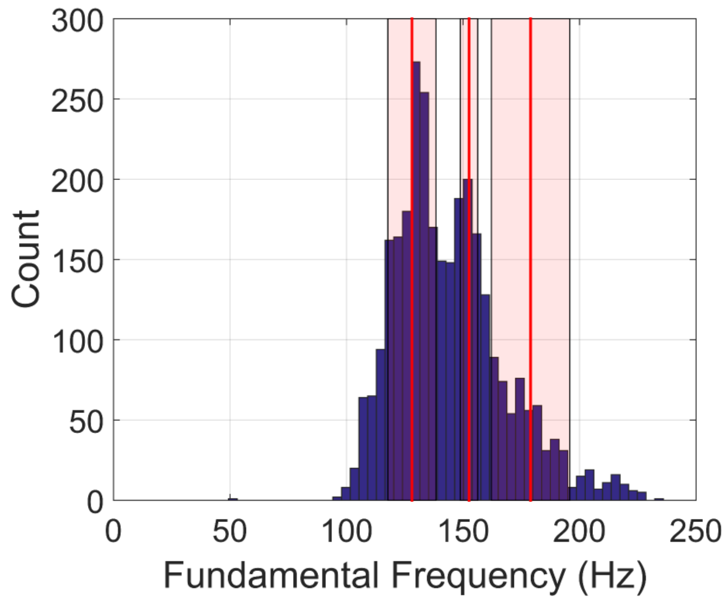

Figure 9. Fundamental frequency distribution for participant M3. Note: The red vertical lines represent each statistical mode from lowest to highest (SM1-4) and the surrounding shaded region represents the two standard deviations of the mean of that statistical mode.

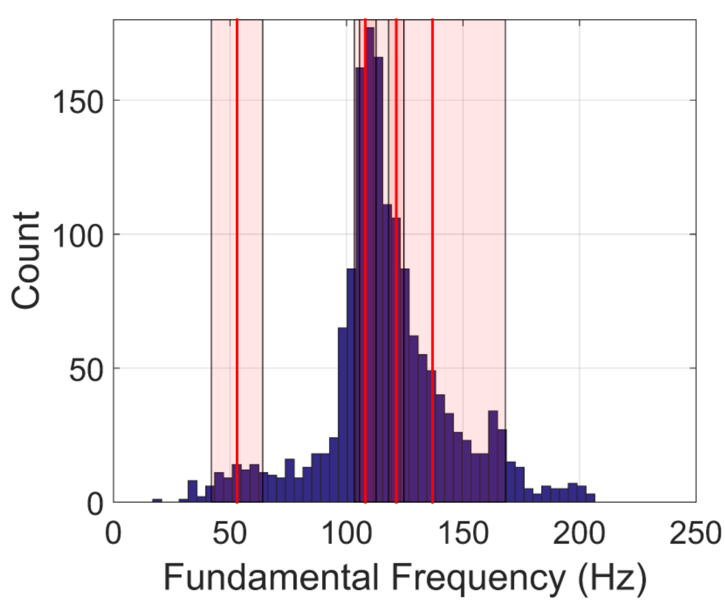

Figure 10. Fundamental frequency distribution for participant M4. Note: The red vertical lines represent each statistical mode from lowest to highest (SM1-4) and the surrounding shaded region represents the two standard deviations of the mean of that statistical mode.

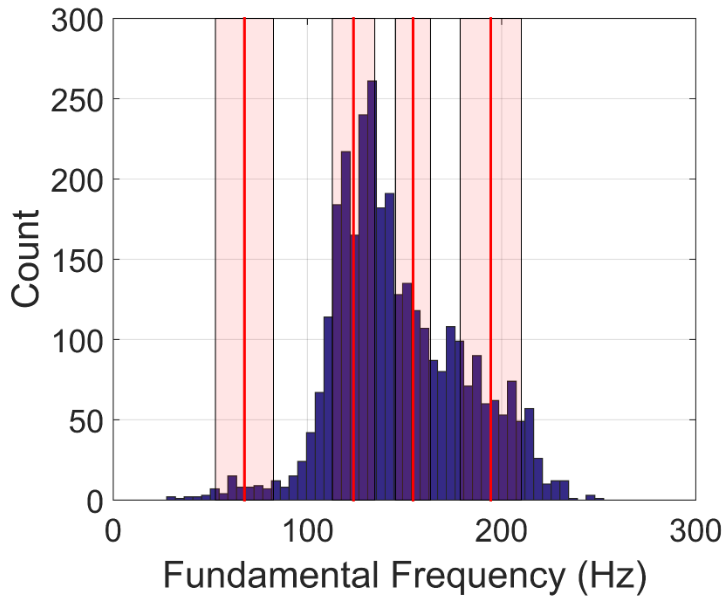

Figure 11. Fundamental frequency distribution for participant M5. Note: The red vertical lines represent each statistical mode from lowest to highest (SM1-4) and the surrounding shaded region represents the two standard deviations of the mean of that statistical mode.

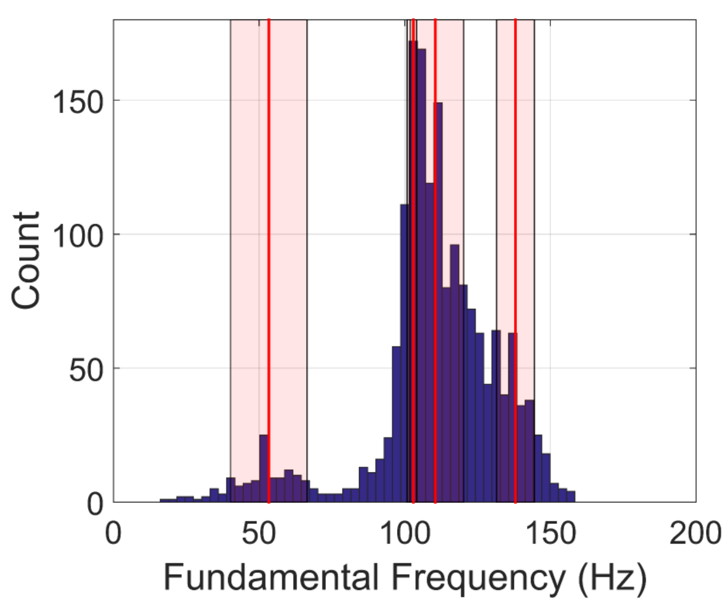

Figure 12. Fundamental frequency distribution for participant M6. Note: The red vertical lines represent each statistical mode from lowest to highest (SM1-4) and the surrounding shaded region represents the two standard deviations of the mean of that statistical mode.

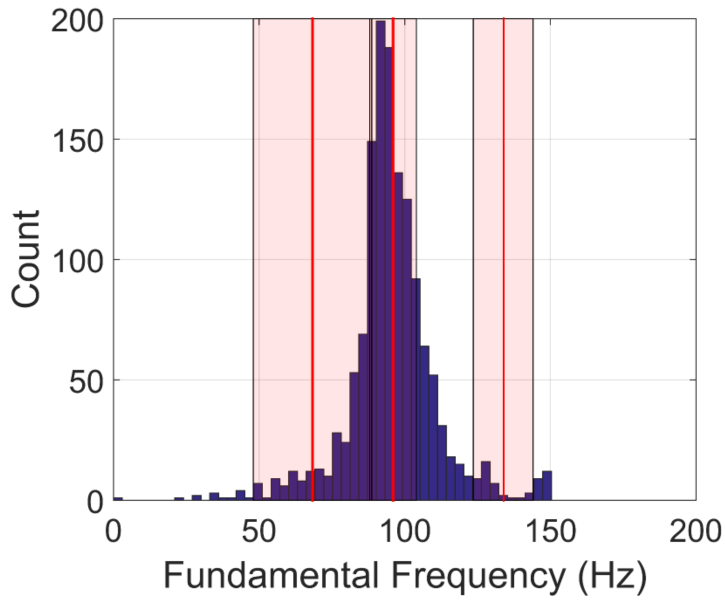

Figure 13. Fundamental frequency distribution for participant M7. Note: The red vertical lines represent each statistical mode from lowest to highest (SM1-4) and the surrounding shaded region represents the two standard deviations of the mean of that statistical mode.

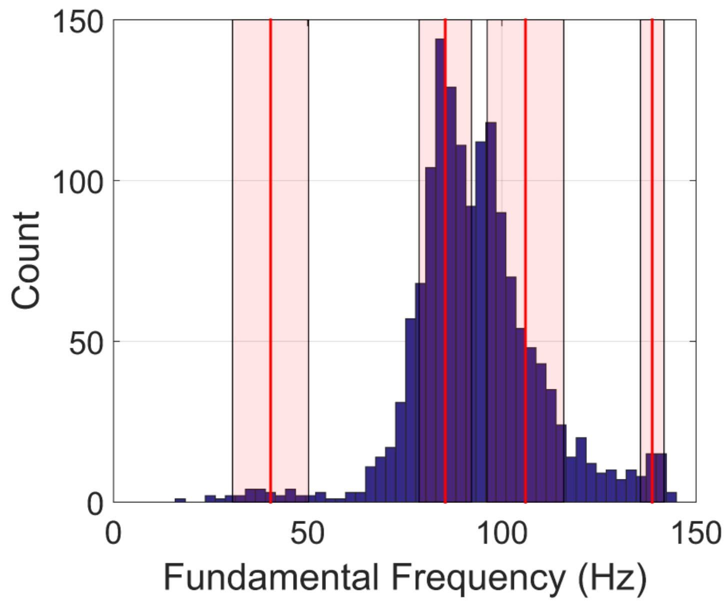

Figure 14. Fundamental frequency distribution for participant M8. Note: The red vertical lines represent each statistical mode from lowest to highest (SM1-4) and the surrounding shaded region represents the two standard deviations of the mean of that statistical mode.

Table 1. Table 1 provides summary statistics for the mean, standard deviation, and proportion of each statistical mode for males and females.

	Female			Male		
	Mean	SD	Proportion	Mean	SD	Proportion
SM 1	74.31	19.30	0.05	55.12	13.43	0.04
	(35.9)	(11.0)	(0.03)	(10.4)	(3.8)	(0.03)
SM 2	156.88	9.12	0.34	106.21	7.38	0.37
	(74.8)	(74.81)	(0.18)	(16.9)	(3.6)	(0.21)
SM 3	181.65	10.98	0.32	125.66	6.99	0.32
	(89.6)	(9.1)	(0.17)	(19.8)	(3.1)	(0.19)
SM 4	224.27	21.32	0.15	152.97	14.05	0.18
	(111.0)	(17.5)	(0.09)	(24.3)	(9.1)	(0.10)

Descriptive summary statistics were calculated for the mean, standard deviation, and relative proportion of f_0 values for the male and female speakers. These data are shown in Table 1. Females had a SM 2 that had a mean of 156.88 and a standard deviation of 9.12. Approximately 34 percent of all glottal cycles belonged to SM 2. The SM 3 had a mean of 181.65 and a standard deviation of 10.98. 32 percent of the time SM 3 was the occurring statistical mode in the speech selection. The SM 4 had a mean of 224.27 and a standard deviation of 21.32. The proportion of occurrence was 15 percent. The overall mean of females' f_0 was calculated to be 216.84 with a median of 212.68, suggesting that the overall mean does fell between the mean of SM 3 and SM 4 for the female speakers.

For female speakers, SM1 was 65 percent lower than the mean f_0, SM2 was 28 percent lower than the mean, SM3 was 16 percent lower than the mean, and SM4 was three percent higher than the mean. For the male speakers, SM1 was 52 percent lower than the mean f_0, SM2 was 8 percent lower than the mean, SM3 was 9 percent higher than the mean, and SM4 was 32 percent higher than the mean.

In male speakers, SM 2 had a mean of 106.21 and a standard deviation of 7.38. The relative occurrence of glottal cycles in SM 2 was 37 percent, very similar to the female values. SM 3 had a mean of 125.667 and a standard deviation of 6.99. 32 percent of the glottal cycles were assigned to SM 3 for the male speakers, a value that was again similar to the female data. The SM 4 had a mean of 152.97 and a standard deviation of 14.05. The relative proportion of occurrence was 18 percent. The male participants' overall mean f_0 was calculated to be 115.75 with a median of 113.59. Unlike the female participants, this mean f_0 falls between SM 2 and SM 3.

Thus, for both sexes, the mean frequency of SM1 was about half of the f_0 average for each sex. Additionally, the overall spread of the data was larger for males than females, with the male participants exhibiting less difference in the overall mean of the data and the mean of SM2 and SM3.

CONCLUSION

The current study examined the distribution of f_0 values in connected speech of neurologically healthy younger adult speakers to determine what the average proportion of f_0 values located in the frequency range consistent with the vocal fry register. As expected, the f_0 values consistent with the modal register (SM 2, SM 3, SM 4) were higher and of much greater proportion than the f_0 values that would be consistent with of vocal fry (SM 1). The overall f_0 of the female speakers was also higher than males, regardless of register.

On average, the f_0 values for the male speakers are lower than the female speakers, which is in line with sex differences that are reported in numerous studies in the literature (e.g., Childers and Wu, 1991; Deterding, 1997; Peterson and Barney, 1952; Wu and Childers 1991). Additionally, there was much less separation between in frequency between SM2, SM3, and SM4, suggesting that the f_0 values of these statistical modes likely overlap, while SM1 is well-separated from the upper modes.

A relatively low proportion of f_0 values were identified as belong to SM1. As displayed in Table 1, instances of SM1 occurred in 5 percent of women's collected speech data and 4 percent of men's recorded speech samples. This suggest these lowest frequencies were relatively infrequent in the connected speech of healthy younger adult speakers. If indeed these glottal cycles were produced using glottal fry, they likely occurred at the end of phrases and therefore represent a small minority of the total number of glottal cycles in the reading passage. To qualify this argument, visual inspection of the data suggest that there are individual differences in SM1 in both frequency count and shape of mode. For example, participants F2, F5, M2, and M4 exhibited a larger number of glottal cycles belonging to SM1, while participants M5, M7, and M8 very few if any cycles in this statistical mode. It would be interesting to determine if the participants with the larger proportion of glottal cycles assigned to SM1 would be perceived to use more glottal fry.

This research is only preliminary and has many possibilities for further research. The participants in this study were all healthy young adults. Future work should should apply this analyses to spontaneous speech samples to determine if more natural context result in a larger proportion of cycles produced in vocal fry or assigned to SM1. Additionally, it would be valuable to apply this type of analysis to speakers that were identified by clinicians to use pervasive vocal fry or exhibit a hyperfunctional voice disorder. It would also be of interest to exhibit the effect of various speaking styles, acting or other stage productions on the distributional characteristic of the statistical modes. Relative to singing, it would be interesting to determine if such an approach may help to classify singer's fach or type.

From a prosodic perspective, further research could also be devoted to understanding the relationship between these statistical modes and other acoustic structures such as phonetics, linguistics, and prosody. It is possible that the statistical modes are closely related to word stress or the placement in a phrase (intonation). A final area of interest is cross examining this research with respect to vocal intensity. Vocal fry is thought to be comprised of low intensity and low frequency sounds. Analyzing the statistical modes along with a distribution of intensity may yield a more evident bimodal distribution in the in two dimensional space than this study, which only considered the frequency domain.

Relative to limitations, this study used a purely acoustic methodology and portions of the speech signal that were produced using vocal fry were not perceptually identified. Additionally, this study examined speech production in the context of a reading passage. Results may have differed if based on the nature of the speaking task or speaking style. Finally, the study was limited to 14 participants due to the time intensive nature of the data analysis and manual examination processes required.

Beyond the scope of this project, there are additional considerations and controversies surrounding vocal fry in the music world. For example, some vocalists and singing teachers report using glottal fry as a singing exercise thought to have a variety of applications including warm up. A reason that is typically given for this practice is that it creates a glottal configuration with a lax vocal fold structure and full adduction. Additionally, others have suggested that glottal fry may be used as a means to treat vocal fold injuries, such as vocal cyst plus a reactive mass on the opposite side, vocal fold stiffness secondary to a vocal fold hemorrhage, or postoperative singers (Nix, 2005). However, there is much concern around this view in the clinical realm of Speech-Language Pathology, as glottal fry is potentially associated with a high degree of medial compression that may be associated with vocal hyperfunction. However, the low subglottal pressures associated with glottal fry likely blunt the acceleration and collision forces encountered during phonation. The opposing view point in the realm of voice pedagogy and singing is more consistent with the views of Speech-Language Pathology and the potentially injurious effects of vocal fry on vocal health. Singing teachers and vocalist that hold these beliefs do not support the use of fry as a vocal exercise in singing, as they profess it to be harmful to the voice and unhelpful in ailments cited above.

Vocalists have been historically thought to be proficient users of good vocal hygiene habits. Although vocal training may be a significant factor in speakers' responses to a single reading task, professional voice users are not immune to problems with the speaking voice. For example, Koufman and Blalock (1989) reported on a group of 67 adult professional voice users who demonstrated dysphonia resulting from apparent laryngeal hypertension. Clinical features included muscle tension in the neck, poor breath control, and abnormally, low-pitched speaking voice (Linville, 1995). Thus, it is imperative to continue education of vocalists in order to avoid vocal strain and damage.

REFERENCES

Abdelli-Beruh, N., Wolk, L., & Slavin, D. (2014). Prevalence of Vocal Fry in Young Adult Male American English Speakers. *Journal of Voice, 28*(2), 185-190.

Boersma, Paul & Weenink, David (2015). Praat: doing phonetics by computer (Version 6.0) [Computer program]. Retrieved December 4, 2016, from http://www.praat.org/

Childers, D. G., and Wu, K. (1991). "Gender recognition from speech. II. Fine analysis," *Journal of the acoustical society of America.* 90, 1841-1856.

Deterding, D. (1997). "Development of speech sounds in children," *Acta Oto-Laryngol.,* 257, 1 51

Fairbanks, G. (1960), *Voice and Articulation Drillbook*, New York: Harper & Row, pp. 124-139

Feldman KA. (1986).The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: a review and synthesis. *Res Higher Ed* 1;24:129-213.

Hollien, H., & Michel, J. (1968). Vocal Fry as a Phonational Register. *Journal of Speech Language and Hearing Research, 11*, 600-600. Retrieved December 11, 2015.

Koufman, J. A., & Blalock, D. P. (1989). Is voice rest never indicated? *Journal of Voice, 3*(1), 87-91.

Linville, S. (1995). Changes in glottal configuration in women after loud talking. *Journal of Voice*, *9*(1), 57-65.

Nix, J., Emerich, K., & Titze, I. R. (2005). Application of Vocal Fry to the Training of Singers. *Journal Of Singing*, *62*(1), 53-59.

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

Schmidt, C., Andrews, M., & Mccutcheon, J. (1998). An acoustical and perceptual analysis of the vocal behavior of classroom teachers. *Journal of Voice, 12*(4), 434-443. Retrieved December 8, 2015.

Slifka, J. (2006). Some Physiological Correlates to Regular and Irregular Phonation at the End of an Utterance. *Journal of Voice, 20*(2), 171-186.

Whiteside, Sandra P., Sex-specific fundamental and formant frequency patterns in a cross sectional study. *The Journal of the Acoustical Society of America*, 110, 464-478 (2001), DOI:http://dx.doi.org/10.1121/1.1379087

Wolk, L., Abdelli-Beruh, N., & Slavin, D. (2012). Habitual Use of Vocal Fry in Young Adult Female Speakers. *Journal of Voice, 26*(3), E111-E116. Retrieved December 14, 2015.

Wu, K., and Childers, D. G. (1991). “Gender recognition from speech. I. Coarse analysis,” *The Journal of the Acoustical Society of America*. 90. 1828-1840.