4-26-2016

Acoustic Correlates of Speech Naturalness in Post-Treatment Adults who Stutter: Role of Speaking Rate

Caylee Jessen
Western Michigan University, cayjes93@gmail.com

Acoustic Correlates of Speech Naturalness in Post-Treatment Adults Who Stutter:

Role of Speaking Rate

Caylee Jessen

Western Michigan University

Abstract

Purpose: While many stuttering therapy programs can result in a decrease in stuttering behaviors among participants, some clients leave therapy with speech sounding unnatural due in part to the therapy techniques themselves. The aim of the current study is to examine associations between selected acoustic measures of speaking rate and listener ratings of speech naturalness in a group of adults who stutter following participation in an intensive fluency shaping treatment program.

Methods: Twenty-nine adults who stutter, drawn from an existing database of speech samples, served as participants. All participants had recently completed a four-week group-based intensive fluency shaping treatment program. Pre- and post-treatment measures of stuttering severity using the Stuttering Severity Instrument (SSI) were determined by a team of expert listeners. A group of listeners used a 9-point interval scale to rate the speech naturalness of post-treatment video excerpts drawn from each participant. Acoustic measures taken from the audio channel of the video excerpts included overall speaking rate, articulation rate, and average duration of inter-phrase pausing within the speech sample.

Results: Speech naturalness ratings were not significantly correlated with post-treatment stuttering severity suggesting that speech naturalness was unrelated to frequency of stuttering events and more likely due to treatment variables. Significant, negative correlations were found between speech rate, articulation rate, and speech naturalness ratings. A non-significant positive correlation was found between average pause duration and speech naturalness ratings. Both speech and articulation rate were positively

correlated with SSI scores while median pause duration was negatively correlated with

SSI scores.

Conclusions: Speech naturalness ratings are correlated with measures of articulation rate

was found to be most predictive of naturalness ratings among the participants while pause

duration was the least predictive. Results from the current study may provide future

speech-language pathologists with information that could contribute to improved speech

naturalness as an outcome of intervention in stuttering treatment programs.

Background

Stuttering is commonly described as an interruption in the forward flow of speech (Van Riper, 1982). These interruptions, or disfluencies, can include repetitions, prolongations, tense pauses, interjections and hesitations. However, stuttering is much more than a disruption in speech and can also affect a person's cognitive, social, and emotional well-being. Stuttering typically develops before the age of three when speech and language development are on the rise (Yairi & Ambrose, 2005). The prevalence for the general population is approximately 1% with more males being affected (Bloodstein & Ratner, 2008).

Many adults who stutter want to decrease their stuttering and increase their speech fluency. Many stuttering therapy programs use techniques such as prolongation, the elongation of words or phrases, and easy onset, expanding and exaggerating the initial sounds of words to reduce or eliminate stuttering moments. While these approaches can lead to stutter-free speech production, the resulting speech pattern can be perceived as unnatural (Martin, Haroldson, & Triden, 1984). This can be discouraging to both persons who stutter and the clinicians that direct stuttering treatment programs. If the speech parameters that are associated with the unnatural speech patterns noted after stuttering therapy could be identified, treatment programs could focus on ways to enhance the naturalness of speech thereby improving treatment outcomes.

Measuring Speech Naturalness

In order to determine the naturalness of speech, stuttered or otherwise, researchers have used a variety of measurement scales. Some scales used were the Likert 9-point scale, direct magnitude estimation scale, and the three-point ordinal scale. Martin,

Haroldson, and Triden (1984) created the 9-point scale in an attempt to find the most

reliable and consistent way to judge naturalness in persons who stutter. A total of 60

raters (30 sophisticated and 30 unsophisticated) were presented with a horizontal line

marked with evenly spaces vertical lines numbered 1 through 9, each corresponding to a

degree of naturalness. Marking a 1 on the scale indicated the person's speech was

perceived as "highly natural" while marking a 9 meant the speech was "highly

unnatural". The listener group made judgments on the naturalness of speech recordings

made from 10 people who stutter, 10 people who do not stutter and 10 people who stutter,

but using the delayed auditory feedback, a fluency enhancing condition. To make sure the

ratings were reliable, the students were asked to come back one to three weeks later to

rerate the participants. Overall, Martin el al. found the samples of stuttered speech to be

perceived as more unnatural than the speech of adults who do not stutter. Consistency

among the listener ratings illustrated the scale created was beneficial for future studies.

     The interval scale (Martin et al., 1984) was compared with direct magnitude

estimation in a study by Metz, Schiavetti, and Sacco (1990). Direct magnitude

estimations were found by first presenting each listener with a speech sample of a person

who does not stutter as a stimulus. The students rated the remaining participants relative

to the original recording. Results from this study indicated that both direct magnitude

estimation and interval scaling were equally suitable for measuring speech naturalness.

     Both direct magnitude estimation and interval scaling are problematic because

they can be long and exhausting for the listeners. Therefore, Schaeffer and colleagues

(Schaeffer and Eichorn, 2001; Schaeffer, 2006) proposed an alternative scaling approach

in a study examining the role of sound and sentence duration on ratings of speech

naturalness. Their approach involved the use of a three-point ordinal scale where listeners chose to "Strongly Agree", "Agree", or "Disagree" that the speech sounded natural in hopes of simplifying the listening task. A clear drawback of this approach is that there are only a limited number of possible responses available to the listener.

In summary, a number of studies have concluded that naturalness of speech can be reliably measured. Of the different scaling methods used, the 9-point equal appearing interval scale has been the most consistently used in studies focused directly or indirectly on stuttered speech naturalness.

Speech Naturalness and Stuttering Treatment

In addition to examining different approaches for scaling speech naturalness, it is now important to find a possible cause behind each rating. Acoustic analysis of speech has been used to identify correlates of listener naturalness ratings in a group of recently treated adults who stutter and a non-stuttering control group (Metz et al., 1990). The study included 20 adults who stutter who had completed a 5-week residential stuttering treatment program and 20 non-stuttering adults with no signs of hearing or speech disabilities. Since this study focused on fluency shaping, a principal feature of the treatment program was controlling easy voice onset. Speaking samples obtained included an assigned reading passage or a story about a picture shown. Thirty undergraduate students in a speech-language pathology program were asked to rate the naturalness of the speech of treated adults who stutter and the adults who speak fluently. Half of the raters used interval scaling and the other half used direct magnitude estimation to rate the participants' naturalness. A number of acoustic variables were measured including voice onset time and sentence duration. The authors found that voice onset time, a

measurement of time between the noise burst of a plosive consonant and the voicing of the proceeding vowel, and sentence duration were two acoustic measures that were correlated with listener ratings of speech naturalness. Specifically, voice onset time was significantly correlated with naturalness for oral reading (r=. 68), while sentence duration was significantly correlated with naturalness for the picture description task (r=. 64). The data from this study indicated that durational variables such voice onset time and sentence duration are correlated with speech naturalness. This is relevant because speech prolongation (i.e. stretching or elongating vowels or words) is one of the most common treatment methods for improving fluency (Schaeffer, 2006).

Onslow, Hayes, Hutchins, and Newman (1992) investigated relations between this therapeutic teaching and speech naturalness. The study included a total of 14 male subjects, half adults who stuttered and half adults who did not. Prior to the study, the adults who stutter were enrolled in a prolonged speech residential treatment program. Undergraduate students served as listeners, rating speech naturalness of the participants on a 9-point scale. Of the two key findings that came out of this study, naturalness ratings for monologue speech were very similar to those of conversational speech, providing evidence that the ratings of one task can predict the ratings of the other. Second, the naturalness ratings of the speech treated adults who stutter was consistently worse than those who did not stutter, a result the authors attributed to a therapy technique used in therapy that focused on prolonging the sounds in each word. However, it should be noted that the authors did not attempt to quantify this conjecture.

The relationship between sound/word prolongation and speech naturalness was directly tested by manipulating the duration of recorded words using a resynthesis

technique (Schaeffer & Eichorn, 2001). The goal of the study was to determine the

association between vowel duration and listener ratings of naturalness. Twelve, one-

syllable words were spoken by a female non-stuttering speaker. The vowel within each

word was then resynthesized at the original duration, and prolonged to 150%, 200%, and

250% of the original duration. A large group of college students (N= 52) served as

listeners. None had previous knowledge with stuttering and were told what they will need

to do but not the purpose of the study. The subjects were asked to rate the speech samples

using the following 3-point scale: Strongly agree the speech is natural, agree the speech is

natural, or disagree the speech is natural. Samples were presented in a randomized order.

Only 22% of the subjects judged the samples to be natural for the samples prolonged to

be 250% of the original duration. At the 200 % of original duration, 52% of listeners

rated the speech as natural. At 150% of original duration, 83% of listener rated the speech

as natural. The authors concluded that words containing prolonged vowels can be

perceived as natural, but that there is a point where the level of prolongation is too great

to sound natural for the listener.

Schaeffer and Eichorn (2001) focused their study exclusively on naturalness

ratings of single words. It is unclear how these finding might generalize to connected

speech. In a follow-up study, Schaeffer (2006) studied the effects of phrase/sentence

prolongation on listener ratings of speech naturalness. A total of 60 phrases spoken by a

fluent speaker were recorded and resynthesized at a number of durations longer than the

original recording. Based on the results of the initial study, the author decided to lower

the levels in 20% increments to 150%, 170%, 190%, and 210% of the original duration.

Thirty participants with no prior understanding of stuttering or the study being conducted

served as listeners, rating the samples using the same scale as Schaeffer and Eichhorn (2001). For the 150% prolongation condition, results were very similar to the previous study with more than 80% of the listeners judging the recordings to be natural. When prolongation levels were raised to 210% in the current study, only 10% of listeners agreed that the speech sounded natural. This revealed that there is a distinct difference in how listeners perceive prolonged vowels and words versus prolonged phrases. Because prolonged vowels are surrounded by sounds that are at normal speech rate, it may not be as easy to distinguish as prolonged phrases (Schaeffer, 2006). Researchers also found that prolonged phrases that contained certain words had a higher naturalness rating than those that did not, implying that context plays a role in the perception of naturalness.

In summary, previous studies suggest that measures of speech timing may be correlated with listener ratings of speech naturalness in both adults who stuttering following treatment and in non-stuttered speech that has been artificially altered. The goal of the present study is to further examine the role that acoustic parameters of speaking rate have on the naturalness ratings in a group of adults who stutter just following participation in a stuttering treatment program.

Methods

Participants: Participants were drawn from the Walter Reed-Western Michigan
University Stuttering Database. This database includes clinical-behavioral and
physiological data obtained on 43 adults who stutter and 43 normally fluent adult
speakers. Participants were largely reserve and active duty members of the United States
Armed Services. All participants who stutter reported that they had stuttered since
childhood and were seeking treatment through the Walter Reed Stuttering Treatment
Program. Of the 43 adults who stutter, post-treatment data was available for 33
participants. Three participants had incomplete datasets and were excluded. One
participant exhibited a strong foreign accent that was judged to interfere with the speech
naturalness ratings. As a result, total of 29 participants were included in the present study.
Of the participants used in this study, the mean age was 25 years, 10 months with a range
of 24. Twenty-seven of the participants were male and two were female. Four
participants reported English was their second language and two participants had
previously attended the treatment program within the last two years.

Treatment Program: The treatment program was one-month long, intensive, group-based
fluency shaping program. The program was split into three parts. The first part
concentrated on the nature of speech, characteristics associated with stuttering, as well as
the feelings related to stuttering. Part two, which took the largest amount of treatment
time, focused on learning a set of fluency enhancing speech targets including abdominal
breathing, increased breath support, easy phonatory onset, continuous phonation, and
how to use phrasing. Many of these were established through the use of the computer-
aided fluency enhancing training (CAFET) program (Gobel, 1988). The computer

program provided biofeedback to participants on both respiratory and voice-acoustic

characteristics. To ensure reliable and consistent use of the CAFET program, a speech-

language pathologist monitored all treatment sessions.

Data Collection: All behavioral analyses and listener ratings were based on studio quality

video recordings of monologue, reading passages, and telephone calls made by each

participant.

Stuttering Severity: The Stuttering Severity Instrument (SSI-3) was used to assess

stuttering severity before and after treatment participation. SSI-3 items were scored by

two certified speech-language pathologists, highly familiar with stuttering. A consensus

judgment approach was used and required that judges agree on instances of stuttering or

associated behaviors. To ensure reliability, judges reviewed the videotapes and original

score sheets 6 months following the study, discussing and addressing any disagreements

arisen.

Speech Naturalness Ratings: Speech naturalness was assessed using a 9-point interval

scale ranging from highly natural (1) to highly unnatural speech (9) (Martin et al., 1984).

Listeners consisted of three separate groups of students enrolled in a university speech

and hearing program. Each group was comprised of approximately 20 listeners and each

group rated a different set of video samples. One-minute video segments from each

participant's monologue, reading, and telephone call recordings were presented in a

randomized order. Individual video segments were rated by multiple listeners allowing

the determination of inter-rater reliability. The intraclass correlation coefficients for each

of the listener groups were 0.98, .97 and 0.95, suggesting good reliability. This study

used the mean naturalness ratings from across the entire listener pool.

Acoustical Analysis: For this current study, the acoustic analyses were performed on the reading samples. Reading samples were selected because they yield a consistent task across all participants. A transcript of the reading passage is in Appendix A. The audio portion of the audio-video recording was extracted using *Audacity*, an audio-editing software suite, and saved as an uncompressed audio file.

TF32, an acoustic analysis software package was used to display audio samples. A synchronous waveform and a wide-band spectrogram display were used to examine the samples and determine the onset and offset of spoken phrases. Time values were identified at the beginning and cessation of the acoustical energy of each phrase. Pauses were found by looking for gaps in speech-related acoustical energy greater than 250 milliseconds in the sample. Once all speech and pause time was marked with TF32, a syllable count was taken of the reading passage for that particular sample. The following measures were then derived (See also Appendix B):

Speech rate: total number of syllables/(articulation time + pause time)

Articulation rate: total number of syllables/(articulation time)

Average pause duration: total pause time/total number of pauses

Statistical Analysis: Simple correlation was used to determine the association between the key behavioral and acoustic measures.

Reliability: Inter-rater reliability was found by having a student retest 3 samples of the acoustic analyses using the steps located in Appendix B. The mean speech rate found by the student was $\bar{x}=3.05$ while the mean speech rate found by the researcher was $\bar{x}=2.93$. The student's articulation rate was 4.56 syllables/second and the researcher's was found to be 4.35 syllables/second. The student's mean pause duration was 961 msec and the

researcher's was 952 msec. For the second sample, the student's mean speech rate, as

well as the researchers, was 2.50 syllables/second. The student's articulation rate was

3.63 syllables/second and the researcher's was 3.56 syllables/second. The student's mean

pause duration was 1618 msec and the researcher's was 1544 msec. Lastly, in sample

three, the student's mean speech rate was 3.42 syllables/second and the researcher's was

3.41 syllables/second. The student's articulation rate was 4.58 syllables/second and the

researcher's was 4.38 syllables/second. The student's mean pause duration was 620 msec

and the researcher's was 540 msec. Results show a significant similarity between the two

data sets.
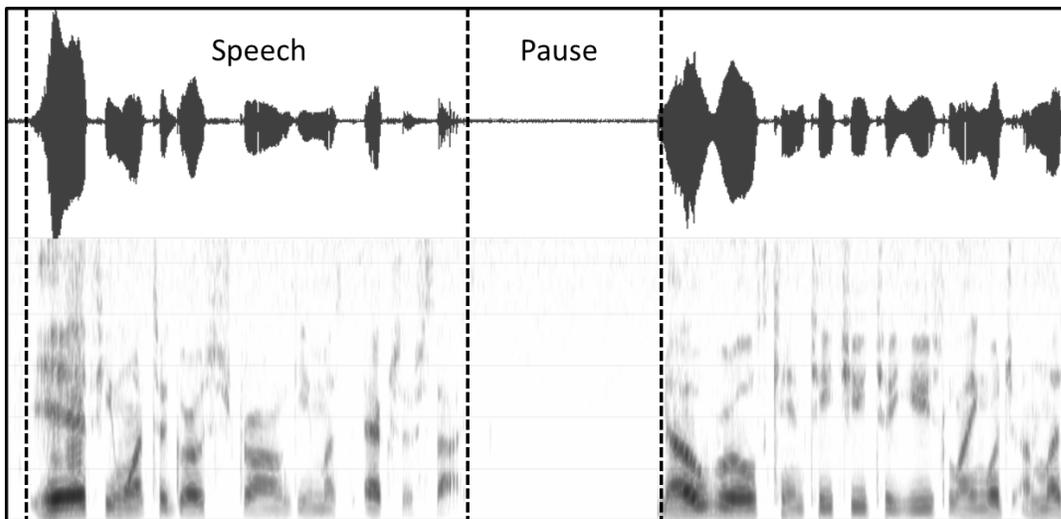
Figure 1. A broad-band spectrogram slice displaying pause boundaries and the onset and offset of speech.

Results

Descriptive Statistics:

Table 1 summarizes the group means, medians, standard deviations, minimum, maximum

and the range for the key behavioral and acoustic measures. Stuttering severity was

estimated using the SSI-3 score. All participants in the study exhibited a drop in SSI-3

scores following treatment. SSI-3 scores pre-treatment ($\bar{x}$=25.4, range 18) were higher

than post-treatment ($\bar{x}$=9.20, range 20) consistent with reduced frequency of disfluency.

Post-treatment naturalness ratings had an mean score of 3.2 (range 1.65 to 6.14).

The mean speaking rate among participants was 3.14 syllables/second (range 1.93 to 4.40

syllables/second). Articulation rate had a mean of 4.43 syllables/sec (range 3.56 to 5.55

syllables/sec). The last speech parameter found was pause duration with an mean of 966

msec (range 499 to 1754 msec.

Table 1. Mean data of both behavioral and acoustic measures including Stuttering Severity Index (SSI-3) pre- and post-treatment, naturalness ratings post-therapy, speech rate post-treatment, articulation rate post-treatment, an pause duration post-treatment.

Measure	Mean	Median	Standard Deviation	Min	Max	Range
SSI-3 Pre-Treatment	25.4	25	5.28	17	35	18
SSI-3 Post-Treatment	9.2	9	4.77	0	20	20
Naturalness Post-Treatment	3.2	2.8	1.28	1.65	6.14	4.49
Speech rate (Syllables/sec)	3.14	2.9	0.62	1.93	4.40	2.46
Articulation Rate (Syllables/sec)	4.43	4.38	0.53	3.56	5.55	1.92
Pause Duration (msec)	966	952	333	498	1754	1255

Table 2 provides the correlation coefficients between post-treatment naturalness scores and three acoustic variables (articulation rate, speech rate, and pause duration). Figure 2 shows the relationship between speech rate and speech naturalness ratings. As speech rate increases, naturalness ratings decrease (speech is rated as more natural). A statistically significant negative, correlation was found between naturalness ratings and speech rate ($r=.46$, $p=.01$). Figure 3 shows the relationship between speech naturalness ratings and articulation rate. As articulation rate increases, naturalness scores decrease (more natural). A statistically significant negative correlation was found between speech naturalness and articulation rate ($r=.53$, $p=.003$). Figure 4 shows the relation between

speech naturalness ratings and mean pause time. A non-significant correlation was found

between these two variables (r=.30).

Table 3 summarizes the correlation coefficients between post-treatment SSI-3 scores and

acoustic measures of speech rate, articulation rate, and pause duration. Figure 5 shows the

relationship between post-treatment SSI scores and speech rate. As the rate of speech

increases, post-treatment SSI scores increase. A significant positive correlation was found

between speech rate and post-treatment SSI scores (r=.49, p=.008). Figure 6 shows a

relationship among post-treatment SSI scores and articulation rate. As articulation rate

increases, post-treatment SSI scores increase as well. A significant, positive correlation

was found between the two variables (r=.42, p=.02). Lastly, Figure 7 displays the

relationship between post-treatment SSI scores and pause duration. As pause duration

decreases, the post-treatment SSI scores increase. A significant, negative correlation was

found between pause duration and post-treatment SSI scores (r=-.44, p=.02).

Table 2. Correlations between post-treatment naturalness ratings and measures of speech
rate, articulation rate, and pause duration.

	r	R^2	p-value
Speech Rate	-0.46	0.22	0.01
Articulation Rate	-0.53	0.29	0.003
Pause Duration	0.30	0.09	*NS*

Table 3. Correlations between post-treatment SSI-3 scores and measures of speech rate, articulation rate, and pause duration.

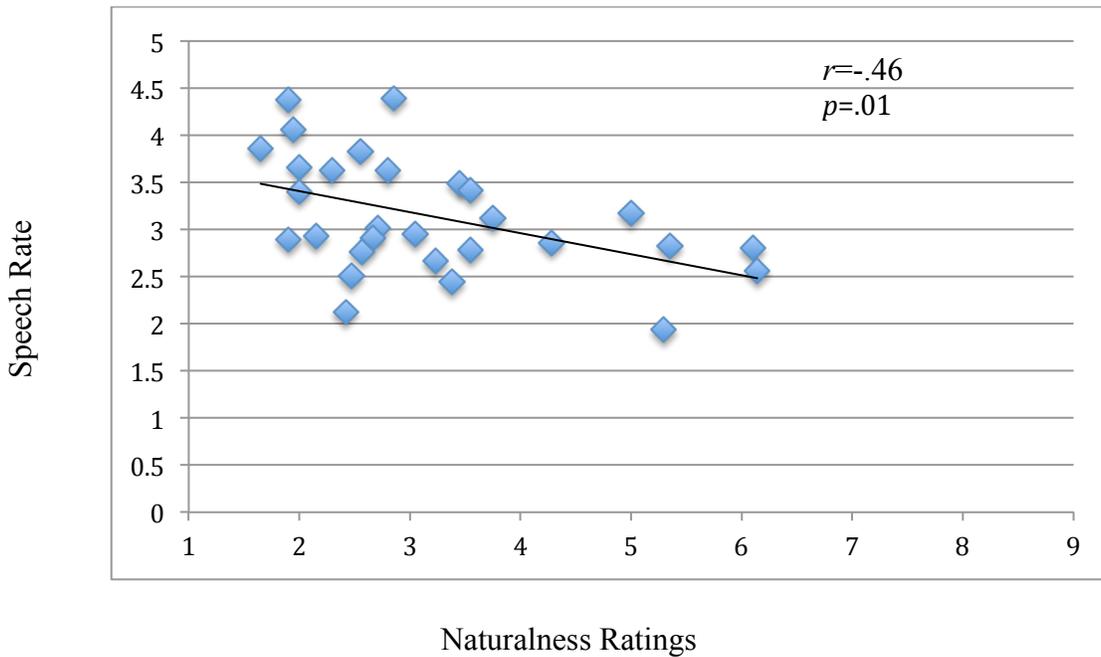|	r	R^2	*p*-value
Speech Rate	0.49	0.23	0.008
Articulation Rate	0.42	0.18	0.02
Pause Duration	-.44	0.20	0.02

Figure 2. Speech Rate is plotted against post-treatment naturalness ratings. *r*=-.46, *p*=.01.
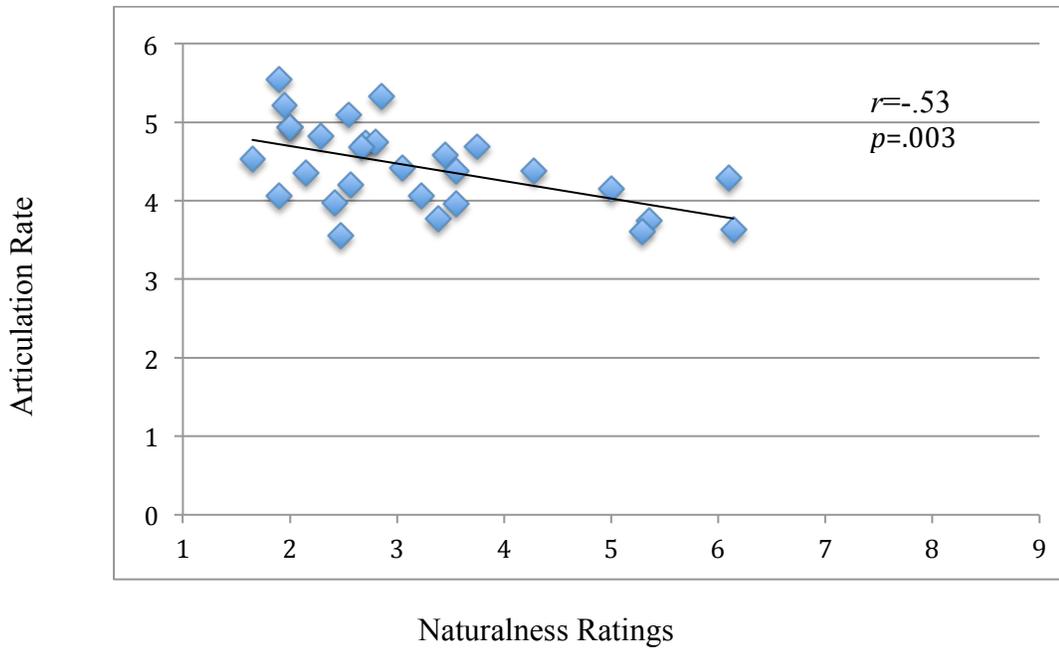
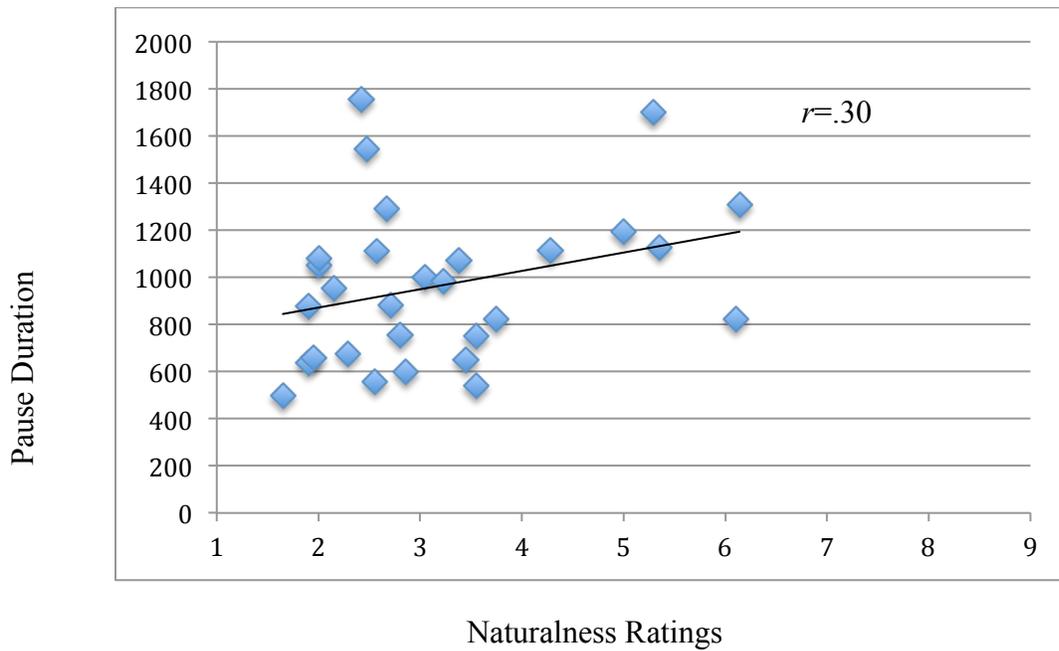Figure 3: Articulation rate plotted against post-treatment naturalness ratings. *r*=-.53, *p*=.003.

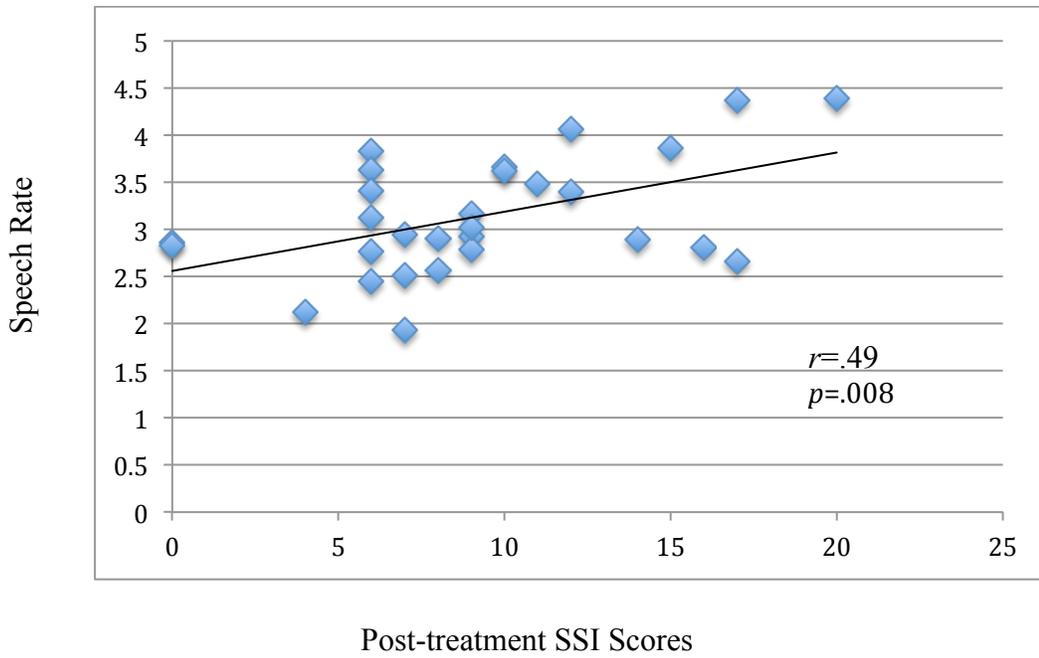Figure 4: Mean pause duration plotted against post-treatment naturalness ratings. *r*=.30, *NS*.

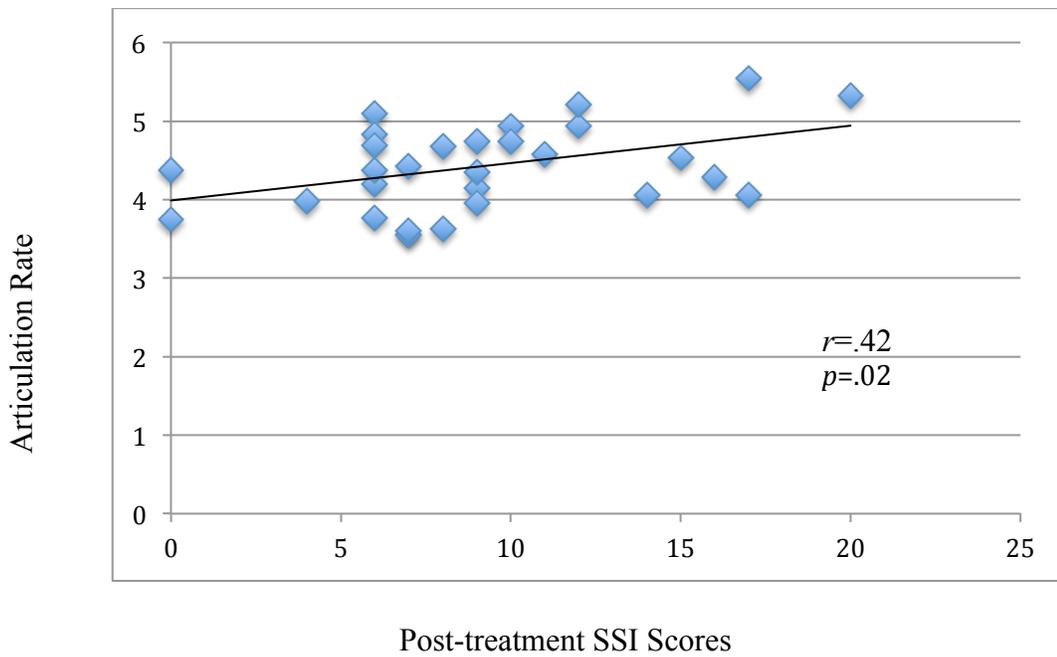Figure 5. Speech Rate plotted against post-treatment SSI-3 scores. *r*=.49, *p*=.008.

Figure 6. Articulation rate plotted against post-treatment SSI-3 scores. *r*=.42, *p*=.02.
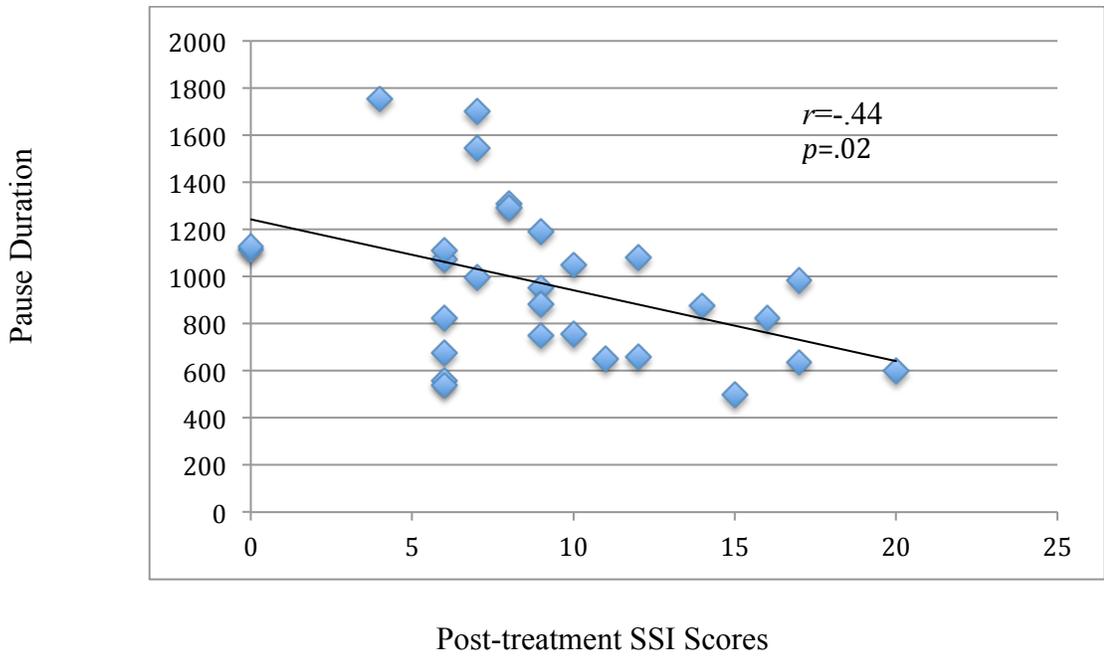
Figure 7. Mean pause duration plotted against post-treatment SSI-3 scores. *r*=-.44, *p*=.02.

Discussion

As previously mentioned, prolongation is a common technique used during stuttering therapy that decreases speech rate. The purpose of this study was to determine if acoustic measures of speaking rate such as articulation rate and pause time, were associated with post-treatment naturalness ratings of adults who stutter. When comparing speech rate to naturalness ratings, it was found that as the speaking rate decreased, the listeners' ratings of naturalness got worse. Listeners may be associating slow speaking rate with unnatural sounding speech. This supports the findings of Metz et al. (1990), Schaeffer (2006) and Onslow et al. (1992) who concluded that prolongation leads to unnatural sounding speech.

Another factor that decreases rate of speech is pause duration. The longer the pause, the slower the speaking rate. While there was a trend for longer pauses to be associated with more unnatural sounding speech, there was not a significant correlation between pause duration and naturalness ratings. Duration of pauses seems have less influence on listener perception of naturalness than the overall rate of speech. Another noteworthy finding was between naturalness ratings and articulation rate. Articulation rate is simply the rate of speech, excluding the pauses. Even though the correlations of speech rate and articulation were very similar, articulation rate was slightly stronger. Because articulation rate is the only variable that omits pauses, listener judgements could be relying more heavily on changes in the manner in which the speaker control timing of articulation rather than timing of pauses..

When comparing speech rate to post-treatment SSI scores, as speech rate increased, post-treatment SSI scores increased as well. Since many therapy programs

focus on techniques that decrease speech and articulation rate, it would make sense that an increase in rate would lead to worse SSI scores. However, increased speech rate has been associated with improved naturalness scores among participants. This could mean that listeners pay more attention to the rate of the speech than to the stuttering associated with it. A similar comparison was found between post-treatment SSI scores and articulation rate. As articulation rate increased, so did the severity of the stuttering. As previously discussed, naturalness scores improved when articulation rate was increased. Once again, listeners are more likely to ignore the severity of the stuttering as long as the articulation rate increased. Lastly, a negative correlation was found between post-treatment SSI scores and pause duration. The longer the duration of the pause, the better the post-treatment SSI scores. However, the longer the pause, the greater chance they will be perceived as sounding unnatural. This could mean that pauses alone are not the strategy to adopt.

One limitation of the study was that acoustic measures were made on reading passaged but naturalness ratings were made on monologue samples. Another limitation was the small sample size used and the gender imbalance. There were far more male participants than female.

Onslow et al. (1990) and Schaeffer (2006) identified prolongation levels that sounded unnatural, it might be of interest to find how long a pause can be before it is perceived as unnatural. Also, another variable that should be considered in future research is the quantity and location of the pauses in each sample. A sample with an abundance of short pauses or one with a few long pauses could be a predictor of speech naturalness. In addition, there are appropriate times for pauses during speech. Placing

pauses mid-word or phrase may be perceived as less natural than a pause associated with a phrase or clause ending. Therefore, during therapy, it is important to not only focus on the fluency of the speech itself, but also the fluency of the speech as a whole, including pause time in and between each utterance. Since there is notable evidence between articulation rate, speech rate and naturalness ratings and a majority of the strategies used in speech therapy slow down the rate of articulation movement, it is possible that after repeated practice, adults who stutter will eventually be able to increase their articulation rate and achieve more natural sounding speech.

References

Bloodstein, O., Ratner, N. B. (2008). *A handbook on stuttering (6th ed.).* Clifton Park,

NY: Delmar Learning.

Gobel, M. (1988). *CAFET: A computer-aided fluency establishment trainer.* Annandale,

VA: Annandale Fluency Clinic.

Martin, R., Haroldson, S., & Triden, K. (1984). Stuttering and speech

naturalness. *Journal of Speech and Hearing Disorders, 49*, 53-58.

Metz, D., Schiavetti, N., & Sacco, P. (1990). Acoustic and psychophysical dimensions of

the perceived speech naturalness of nonstutterers and posttreatment

stutterers. *Journal of Speech and Hearing Disorders, 55*, 516-525.

Onslow, M., Hayes, B., Hutchins, L., Newman, D. (1992). Speech naturalness and

prolonged-speech treatments for stuttering: Further variables and data. *Journal of

Speech and Hearing Research, 35*, 274-282.

Schaeffer, N., & Eichorn, N. (2001). The effects of differential vowel prolongations on

perceptions of speech naturalness. *Journal of Fluency Disorders, 26*, 335-348.

Schaeffer, N. (2006). Effects of differential prolongations of phrases on the perceptions

of speech naturalness. *Contemporary Issues In Communication Science and

Disorders, 33*, 152-159.

Yairi, E. & Ambrose, N. (2005). *Early childhood stuttering.* Austin, TX: Pro-Ed.

Van Riper, C. (1982) *The nature of stuttering* (2nd ed.). Englewood Cliffs, N.J.: Prentice-

Hall.

Appendix A
Reading Passage

If you want to grow strong you must exercise. No rules can teach you to use language

unless you practice using it. You must develop habits. You can no more learn to read by

memorizing rules than you can become a full back by memorizing the rules of football. In

both cases, rules will help you to practice, but it is the practice that develops your skill.

The exercises in this class are planned to give you the practice. They will help you to

form good habits and acquire skills of communication but it will require work on your

part. You must practice over and over the skills you want to learn. But even on the easiest

ones you must be sure to practice correctly. The exercise can strengthen only what it

practices. That stands to reason if you practice a mistake the mistake will grow if you

practice a skill the skill will grow.

Appendix B
Steps for Measuring Pauses in Speech Sample

1. Open selected sample in the tf-32 software and darken the gray scale floor of the spectrogram to -80 dB.
2. Adjust time scale of image.
3. For initial labeling of start:
4. Place the right cursor at the estimated start of acoustical energy for the first utterance. (Do not include preparatory throat clearing or paper shuffling or a single phoneme repetition if it occurs 250 ms or more before the speech continues).
5. Then place the left cursor anywhere within 300ms on the left of the right cursor.
6. Zoom into the section by hitting the page down key. Here, attempt to make a more accurate placement of the right cursor for the start of the utterance (acoustical energy)
7. Once you have achieved a satisfactory boundary; go to view>label>okay>view>mark label> type in start to the text box> okay.
8. You have marked the start and are ready to begin labeling to pauses that follow. We are looking for pauses or gaps in acoustical energy greater than 250ms (time measurements are noted in the upper right hand corner of the screen).
9. If you believe you see a pause greater than 250ms, place the left and right cursors around the estimated gap, **then zoom into the section by hitting the page down key.** Attempt to align the left cursor with the end of the previous utterance and the right cursor with the start of the next utterance. Remember, these endings are determined by presence of energy on the acoustic signal and the spectrogram. Attempt to align the cursor with the wave signal and activity on the spectrogram (darkening).
10. SIDE NOTE: When adjusting the cursors, **keep them on the display,** shifting them out of the screen will cause them to disappear.
11. Once satisfactory parameters are found, label the pause. Go to view>label>okay>view>mark label> type in p01 to text box> okay. We will begin with p01 and proceed with p02, p03, p04, etc. for all following pauses.
12. When the end of the sample has been reached, we will repeat the "start labeling" step in reverse.
13. Place the left cursor at the estimated end of the last sound and the right cursor 300ms to the right of the left cursor.
14. Zoom into the section. Adjust the left cursor to fit the ending of energy on both the time domain signal and the spectrogram.
15. Mark this label as "end."