4-25-2019

Effects of Stuttering Frequency, Speaking Rate and Treatment on Speech Naturalness in Adults Who Stutter

Emily Hausman
Western Michigan University, ehausman96@gmail.com

Effects of Stuttering Frequency, Speaking Rate and Treatment on Speech Naturalness in Adults Who Stutter

Emily Hausman

Western Michigan University

Abstract

Purpose: Individuals who stutter can often exhibit speech patterns that sound unnatural compared to individuals who do not stutter. Unnaturalness can be due to several underlying factors including the presence of stuttering itself and the unintended consequences of treatment approaches that alter speech motor patterns. Understanding factors that impact speech naturalness within the stuttering population can help improve the clinical management of the disorder. The current study investigates the relationship between pause duration, speech rate and stuttering frequency, and listener ratings of speech naturalness in a group of adults who stutter (AWS) before and after participation in a stuttering treatment program.

Methods: Participants include 34 individuals who stutter, who were enrolled in a group-based, intensive, four-week fluency-shaping stuttering treatment program. One-minute audio video recordings were obtained from monologue speaking activities recorded prior to and following stuttering treatment. The audio track from each video sample was extracted and submitted to an acoustic analysis software to extract two measures. Pause duration was determined by identifying the temporal extent of all pauses greater than 250 ms. Fluent syllable rate was determined by measuring the overall duration and syllable count drawn from three perceptually fluent runs of speech in each video sample. These measures were combined with speech naturalness ratings and stuttering frequency data reported in previous studies conducted on this participant pool. Paired *t*-tests were used to assess pre vs. post-treatment differences in mean pause time, fluent syllable rate, mean naturalness ratings and stuttering frequency. Linear regression was used to identify associations between pause time, syllable rate, stuttering frequency, and speech naturalness.

Results: Results revealed that following participation in stuttering treatment, the speech samples were rated as more natural, (t: 1.92; p=0.03), exhibited a dramatically reduced stuttering frequency (t: 10.18; p<0.00005), had significantly longer pause durations (t: -4.1145; p=.0001), and showed no change in fluent syllable rate. Linear regression results indicated that prior to treatment, stuttering frequency was the strongest predictor of speech naturalness, accounting for over 60 % of the variance in naturalness ratings. Pause duration and syllable rate provided little predictive value for pre-treatment speech naturalness ratings. Following treatment, pause duration was the strongest predictor of speech naturalness, accounting for more than 50 % of the variance in naturalness ratings. Syllable rate and stuttering frequency were not significantly associated with speech naturalness ratings.

Conclusion: While there is a great deal of individual variation, speech samples of persons who stutter are commonly rated as unnatural sounding to listeners. This is true even after participating in some treatment programs designed to reduce or eliminate stuttering. Prior to treatment, stuttering frequency is the best predictor of speech naturalness. However, following treatment, stuttering frequency ceases to have predictive value. Instead, it was found that the duration of pausing between successive runs of speech is a much better predictor of speech naturalness following treatment. These two findings suggest that, within the same group of speakers, speech naturalness ratings are influenced by different kinds of factors. One interpretation is that following treatment, participants used longer speech pauses to plan the integration of the fluency enhancing speech targets. The result was relatively fluent, but still unnatural sounding speech. Results from the current study may provide future speech-language pathologists with information contributing to improved speech naturalness as an outcome of intervention in stuttering treatment programs.

Chapter 1: Introduction

What is Stuttering?

Stuttering is a speech disorder that affects about one percent of the population. It is primarily characterized by intermittent disruptions of speech including sound, syllable and word repetitions, sound prolongations and hesitations or pauses in speech. These disruptions can have associated muscle tension and extraneous movements of the body. Stuttering typically begins in childhood between the ages of two and four years when speech and language skills are rapidly developing. While as many as 80 percent of children who begin to stutter will recover unassisted, for those who persist, stuttering will typically be a chronic, lifelong condition (Yairi and Ambrose, 2005). For those who chronically stutter, it is not uncommon to have additional characteristics including concomitant physical behaviors such as increased muscle tension, eye blinking, facial grimacing, avoidance of difficult words and speaking situations, the development of negative thoughts and emotions about speaking and other social situations. For many, social, emotional and cognitive factors may have a larger impact than primary stuttering behaviors. Emotional reactions seen in persons who stutter (PWS) can include fear, anxiety, feeling of being trapped, panic, embarrassment, shame, humiliation, anger and resentment (Yairi & Seery, 2015). Heightened emotional reactions are common in PWS and as a consequence, PWS commonly avoid engaging in typical activities of daily life and social interactions.

Living with stuttering can also negatively influence thoughts and attitudes about speaking and communication as a whole. Attitudinal surveys in children and adults who stutter reveal significant negative attitudes about communication (Yairi & Seery, 2015). Because of the impact social, cognitive and emotional variables can have on PWS, it is important to take those variables into consideration during the treatment process of stuttering therapy.

<u>Approaches to Stuttering Treatment</u>

While there is a long history of research interest in developing pharmaceutical and technological treatments for stuttering, the most common approach to stuttering treatment involves a range of behavioral interventions. Most mainstream stuttering interventions focus on treating one or more of the following issues: reducing the severity of stuttering events, increasing speech fluency (i.e. reduce the frequency of stuttering), helping the client develop more positive thoughts and feelings about speaking and stuttering, and assisting the client in removing barriers to social interaction and ultimately improving the client's quality of life. The first two of these issues address the primary stuttering behaviors. The first approach, commonly termed stuttering modification, involves teaching clients to systematically alter stuttering patterns to make them less disruptive to communication. The second term, commonly termed fluency-shaping, focuses on changing the whole manner of speaking rather than only the stuttering event. The main objective of fluency-shaping is to employ methods known to facilitate fluency and decrease stuttering. These methods can include stretching and slowing speech, altering speech rhythm, changing the manner of voice and consonant production, and increasing breath support for speech production. These new speaking patterns are practiced until fluency is obtained. Speech patterns are then adjusted slowly while still maintaining fluency to make the speech changes less noticeable to the speaker and client. These changes may be achieved simply through instruction and practice or can be facilitated using delayed auditory feedback (DAF), and forms of biofeedback (Yairi & Seery, 2015).

Fluency-shaping treatment approaches, which is the focus of the current study, has a large research body supporting its use when the aim is to reducing stuttering frequency (Bothe et al. 2006). For example, the rhythmic speech approach which utilizes methods like finger tapping and using a metronome to keep rhythm when speaking, have been successful in significantly reducing stuttering frequency. A study done by Trajkovski et al (2009) revealed results that provided experimental evidence for the potential efficacy of an early intervention for stuttering using rhythmic treatment termed syllable-timed speech (STS). All three participants used in the study by Trajkovski et al (2009), maintained a consistent reduction in their “beyond clinic” stuttering, which was associated with the introduction of the treatment used in the study. Such therapies are often individualized to meet specific needs of the client and such approaches can often have the unintended consequence of making the fluent speech produced by the PWS sound unnatural to both the person who stutters and listeners.

Stuttering and Speech Naturalness

Both the core behaviors of stuttering and, in some cases, the treatment strategies taught to clients in therapy can create speech patterns that sound unnatural to listeners. A primary goal of any therapeutic procedure for a speech disorder is to maximize how natural the speech patterns sound relative to a referent population. The importance of this distinction between measures such as stuttering behavior and speech naturalness is reflected in a significant body of research focused on speech naturalness as a psychophysical dimension in persons who stutter. Recognizing this distinction, the most recent edition of the Stuttering Severity Instrument (SSI-4), arguably one of the most common clinical assessment tools used to assess stuttering behavior, added a measure of speech naturalness (Riley, 1994).

Martin Haroldson and Triden (1984) was the first study to investigate the idea of scaling naturalness of persons who stutter (PWS) and comparing results to persons who do not stutter (PWNS). Listeners rated audio-recordings of speech samples with a nine-point, equal appearing rating scale. The scale was rated from one-nine where one was considered “highly natural” and nine was considered “highly unnatural.” The study had raters use the scales to evaluate naturalness in PWS, PWNS, and PWS during the use of delayed auditory feedback (DAF). DAF was utilized because of its known ability reduce the frequency of stuttering. The results of this study indicated that speech samples of PWS were considered significantly more unnatural than the speech samples of PWNS. While DAF reduced the frequency of stuttering, the raters still found that these samples sounded more unnatural than the group of individuals that did not stutter. Measures of rater reliability and agreement were found to be generally good within and across raters. Since this initial study, there have been many studies that have examined various factors related to speech naturalness in PWS (Metz et al. 1990; Martin & Haroldson, 1992; Onslow et al. 1992; Tasko, McClean & Runyan, 2007; Teshima et al. 2010). These studies have found that speech naturalness ratings in PWS can be influenced by a variety of factors such as whether the sample is audio-only or audiovisual (Martin & Haroldson, 1992), the past experience of the listener (Onslow et al. 1992; Teshima et al. 2010), the type of material and length of the speech sample (Onslow et al. 1992), and participation in stuttering treatment (Metz et al., 1990; Onslow et al. 1992); Tasko et al. 2007). This last issue of stuttering treatment is of particular relevance to the present study.

It was noted earlier that PWS can achieve a high degree of fluency as a result of stuttering treatment. However, many of these treatment approaches require the participant to markedly change the manner in which they speak. The rationale is that altering speech patterns in a particular way (e.g. slowing speech, using gentle phonatory onsets), will result in reduced stuttering. Once speech fluency is achieved, the participant will then gradually adjust speech patterns so that the new speech behaviors are less obvious to listeners. As a result, such treatment approaches can result in fluent, but unnatural sounding speech. This can therefore result in a situation where a client who stutters might be considered “successfully” treated because he/she is relatively fluent, but speech patterns are still considered unusual to both the client and the listener. Therefore, it is important to examine what factors influence speech naturalness ratings in PWS after they participate in treatment. In one of the first studies to address this issue, Metz, Schiavetti and Sacco, (1990), performed acoustic analysis of speech to identify correlates of listener naturalness ratings in a group of recently treated adults who stutter and adults who do not stutter. This study consisted of 20 adults who stutter (AWS) who had completed a five-week residential fluency-shaping treatment program, and 20 adults who do not stutter (AWNS), containing no hearing or speech disabilities. A key feature of the treatment program was controlling easy onset used by the speakers. The speaking samples used included a reading passage or a story about a picture shown to them. Thirty undergraduate students in a speech-language pathology program participated as raters, rating speech naturalness of both AWS and AWNS. During this study, half of the raters used an interval scaling method and the other half used a direct magnitude estimation approach. A wide range of acoustic measures were assessed. The two measures that were most strongly correlated with speech naturalness ratings were voice onset time and sentence duration. Voice onset time had a high correlation coefficient with the naturalness ratings of the oral reading (r=.68). Sentence duration on the other hand was highly correlated with the naturalness ratings for the picture description task (r=.64) (Metz, et al., 1990). This data indicated that durational factors such as voice onset and sentence duration have a strong relationship with speech naturalness ratings.

These results of Metz et al. (1990) are consistent with some subsequent studies that have been completed in our laboratory (Tasko et al., 2007; Jessen, 2016; Novelli, 2018). These studies indicate that post-treatment naturalness ratings can be correlated with measures such as syllable rate (Tasko et al. 2007; Jessen 2016) and relative rate of lung volume change during the speech breath (Tasko et al. 2007).

The current study is a follow up of a set of previous studies recently completed in our laboratory (Jessen, 2016; Novelli, 2018). Novelli (2018) conducted a large auditory perceptual analysis of speech naturalness in a group of AWS before and after treatment and across different speaking conditions. The author explored the validity of using a visual analog scale to measure speech naturalness rather than the traditional nine-point interval scale. Results indicated that the visual analog scale ratings were reliable and correlated highly with the interval scale ratings ($r=.62$). The ratings from Novelli's study combined with preliminary work from Jessen (2016) suggest that speech naturalness ratings may be correlated with behavioral measures of stuttering as well as acoustic measures related to speech rate.

Overall Research Aim and Study Hypotheses

The general aim of the current study is to examine the relationship between stuttering frequency, speech rate (hereafter termed syllable rate) and speech pause duration and speech naturalness in AWS before and after participation in an intensive fluency-shaping treatment program. Specifically, the following hypotheses were tested:

1. Pre-treatment speech samples will be rated as sounding less natural compared to post-treatment speech samples.
2. Speech naturalness ratings will be associated with stuttering frequency such that speech samples containing more stuttering will be rated as sounding less natural to listeners.
3. Speech naturalness ratings will be associated with syllable rate such that speech samples spoken at slower syllable rates will be rated as sounding less natural to listeners.
4. Speech naturalness ratings will be associated with mean pause duration such that speech samples containing greater pause durations will be rated as sounding less natural to listeners.

Chapter 2: Methods

This study (Protocol 17-09-28) was approved by Western Michigan University Human Subjects Institutional Review Board (WMU HSIRB).

Participants

Participants used in this study were drawn from the Walter Reed-Western Michigan University Stuttering Database. The database includes a wide range of clinical-behavioral and peripheral physiological data from 43 adults who stutter and 43 typically fluent age-matched controls. All participants were reserve and active duty members of the United States Armed Services. All participants who stutter were recruited from the client pool referred to the Army Audiology and Speech Center for possible participation in the Walter Reed Stuttering Treatment Program. Each participant reported to have stuttered since childhood and a diagnosis of stuttering was established by one of the treatment program's clinical speech language pathologists, all of whom had extensive clinical experience with stuttering. Of the 43 participants who stutter, only 39 participants completed the program and therefore were deemed eligible for this study. However, due to technical problems with data recording, a complete set of usable pre-treatment and post-treatment video samples were only available for 34 participants. This group of 34 became the participant pool for the current study. The vast majority of the participant group were male (33 males, 1 female) and the mean age of the group was 26 years with participant ages ranging from 19 to 43 years of age. Further details regarding data collection and processing of the Walter Reed-Western Michigan University Stuttering Database can be found in Tasko, McClean & Runyan (2007).

Description of Walter Reed Stuttering Treatment Program

The Walter Reed Stuttering Treatment Program is an intensive, four-week, residential program. Participants are engaged in therapy approximately, five to six hours a day, five days a week. The therapy is group-based with 3-6 members per treatment group. The program consists of three parts that overlap in time. The first part is largely instructional and focuses on teaching the participants the basics of speech production, characteristics of stuttering behaviors as well as the attitudinal, affective and social dimensions of stuttering. Group discussion of the attitudinal, affective and social dimensions of stuttering continued regularly through the full four weeks of the program. The second part occupies the largest proportion of therapeutic effort and involves teaching and practice of behavioral targets aimed at reducing the probability of stuttering (i.e. fluency-shaping). Speech behaviors targeted by the program include abdominal breathing, increased breath support and volume, continuous airflow, easy phonatory and articulatory onset, and continuous phonation. The behavioral targets mentioned were established with the computer assisted fluency enhancement training (CAFET) program (Goebel, 1998), which provides computerized feedback for respiratory and voice parameters. he third part focuses on providing opportunities to transfer fluency-enhancing skills to a variety of challenging speaking situations and discussion of importance of continued activities for fluency maintenance.

Procedure for Collection of Video Samples

During original data collection, all participants who stutter underwent a series of video recordings. The original video recordings took place in a dedicated recording studio using trained technicians at Walter Reed Army Medical Center. An initial set of recordings were made on all participants. This initial recording was completed within 24 hours of an in-depth instrumental recording of speech motor behavior. For participants who enrolled in the Walter Reed Stuttering

Treatment Program, the initial video recording also occurred within 48 hours of commencing treatment. Participants who completed the stuttering treatment program were recorded a second time, within 48 hours of completion of the final week of the program. During both recording sessions, three different speech tasks were recorded. One task required that participants engaged in about four minutes of extended monologue on the topic of their choice. Another task required participants to read aloud a passage. The third task required that participants engage in a series of telephone calls to local hotels regarding room rates. The purpose of making these video recordings was to allow for detailed behavioral analysis of different speech tasks prior to and following stuttering treatment. The current study focuses only on the monologue task of those with complete pre-treatment and post-treatment recordings.

Initial recordings were made on videotape. Following the collection of the video samples, the complete recordings were digitized into an Audio Video Interleaved (AVI) format. As part of a previous study (Novelli, 2018), one-minute samples were extracted from the original video recordings of the monologue speaking task using Microsoft Moviemaker (Microsoft, 1995). These samples were saved in a Windows Media Video (WMV) file format. As noted earlier, due to technical issues, not all video samples could be used for perceptual and acoustics analysis resulting in a total of 68 video samples (34 participants, each with a pre-treatment and post-treatment sample).

<u>Behavioral Analysis of Stuttering</u>

Estimates of stuttering frequency were drawn from a previous study that included an in-depth behavioral analysis of stuttering. Analysis details can be found in Tasko et al. (2007). Briefly, two certified speech-language pathologists highly experienced with stuttering independently orthographically transcribed the speech samples from all video recordings and identified and labeled all disfluent events using standard techniques. Following this analysis by the individual speech pathologists, the pair then reviewed the results together to compare results. In the instance of disagreements regarding the transcription, the judges would repeatedly watch the videos until a consensus was reached. This final consensus transcription was used to derive disfluency counts organized by type (i.e. blocks, prolongations, repetitions and interjections) and word/syllable counts. These data were used to derive estimates of stuttering frequency as well as stuttering severity using the third edition of the Stuttering Severity Instrument (Riley, 1994).

Speech Naturalness Ratings

Speech naturalness ratings were drawn from a previous study (Novelli, 2018). Novelli (2018) recruited 50 adults to serve as judges. All passed a standard hearing screening (25 dB_{HL} at 1, 2, and 4 kHz). The pool was largely young (Mean Age: 22 years; Range: 20-29 years) and female (Gender: 49 female, one male). All judges were college students enrolled in a speech, language and hearing sciences program. The judges were in the final semester of a bachelor’s degree course, the first year of a graduate degree course, or the final year of a graduate degree course. Therefore, all had some exposure to observing the speech of persons who stutter.

Each judge was randomly assigned 38 video samples drawn from a total of 152 video samples from the stuttering database. Samples included oral reading and monologue samples drawn from pre and post-treatment video recordings. Judges were instructed to view the video and rate the perceived speech naturalness using a digital “slider” control that provided a visual analog scale rating. Moving the slider to the left (toward a score of zero) indicated a more natural-sounding speech pattern. Moving the slider to the right (toward a score of 1000) indicated a less natural sounding speech pattern. Therefore, the final location of the slider yielded an integer between 0 and 1000 where the higher the number the less natural the speech sampled sounded to the judge. Ratings were averaged across a panel of 12-13 listeners and results indicated that there was good to excellent inter-rater agreement on naturalness ratings and that the visual analog scaling method provides similar information to the more traditional one-nine numerical scale routinely reported in the literature. Further methodological details can be found in Novelli (2018). For this study, only the 68 pre and post-treatment monologue samples were used.

Acoustic Analysis of Speech Samples

Audio File Extraction and General Acoustic Analysis Procedures

Prior to performing acoustic analysis of the speech samples, the audio waveform was extracted from the WMV video files using version 3.0.4 of the Vetinari VideoLan (VLC) Media Player program (Courmont et al., 1989). The audio waveform was saved in a Waveform Audio file (WAV) format with a 22050 Hz sampling frequency and 16-bit quantization. These audio files were then amplitude normalized and all subsequent analysis was completed using TF32 (Milenkovic, 2000), an acoustic analysis software package commonly used in speech analysis. All measures were performed using a synchronous waveform and broadband (BW: 300 Hz) spectrogram display. Time stamps associated with speech and pause time onsets and offsets were identified using the software's built-in time indexing feature.

Pause Duration Measurement

Measures of speech and pause duration were made using methods similar to Jessen (2016). Each audio sample was separated into alternating periods of speech runs and pause events. To determine the onset and offset of speech runs from the samples used in this study, a synchronous waveform and broadband spectrogram was used in the TF32 program (Milenkovic, 2000). Any gaps in acoustical energy (i.e. pauses) in speech larger than 250 milliseconds,

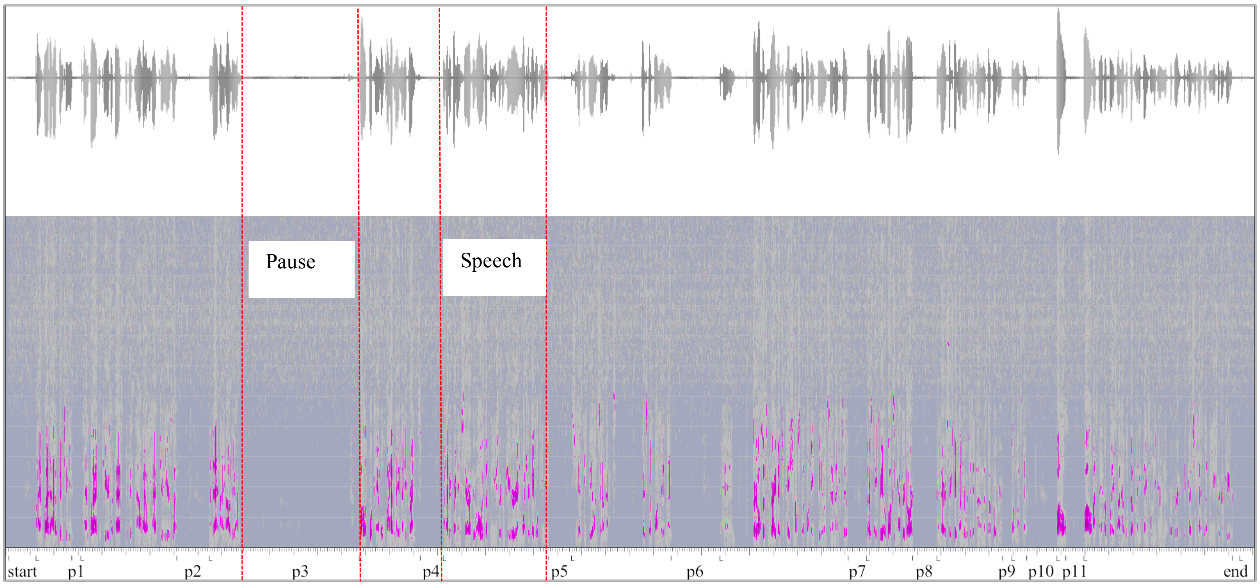

Figure 1. A broadband spectrogram displaying boundaries of the onset and offset of speech in a post-treatment sample.

unrelated to disfluencies such as silent blocks were identified for each monologue sample. Time values were measured at the beginning and end of each sample using start and end labels in the TF32 program. These time indices were used to determine the overall speech duration of each individual sample. Once all speech runs and pause events were determined, the time values were imported into MATLAB (Mathworks, Natick, MA) and the mean pause duration was calculated for each audio sample. Figure 1 includes a sample of the TF32 display. Each marked interval (e.g. p1, p2, p3, etc) represents events that met the operational definition of a speech pause. Intervals between pauses were, by default, deemed to be speech runs.

Fluent Speech Syllable Rate

The speech run intervals identified in the pause duration measurement procedure along with orthographic transcriptions of the samples generated for an earlier study (Tasko et al. 2007), were used to estimate the syllable rate of the fluent speech for each monologue sample. To establish an accurate syllable rate for each sample, fluent, stutter-free speech, was identified using the TF32 computer program. Three fluent runs of speech in each sample were labeled in both pre-treatment and post-treatment recordings. After identifying the three fluent speech runs, the transcripts available for each sample were used to develop a syllable count. The total syllable count across the three fluent speech intervals along with the total duration of the three intervals, was used to estimate an overall syllable rate. Figure 2 includes a TF32 display of the identified intervals of fluent speech. The image below was drawn from a post-treatment monologue sample. Each number underneath the three separate boundaries represents the number of syllables found in each fluent utterance.

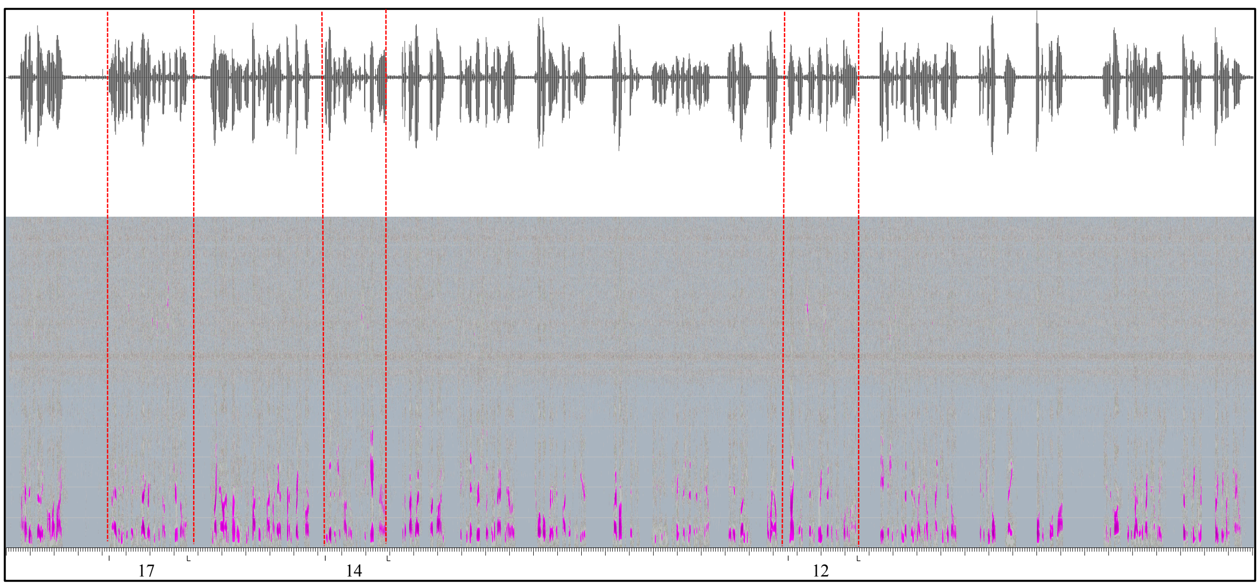

Figure 2. A broadband spectrogram of a post-treatment sample displaying boundaries around the fluent sections of speech used to obtain a syllable rate count. Numbers underneath each boundary correspond with the number of syllables for each fluent utterance.

Statistical Analysis

Speech naturalness ratings and stuttering frequency measures were submitted to a rationalized arcsine transform or a RAU transform prior to statistical analysis (Studebaker, 1985). This data transformation is recommended for data such as proportions/percentages and rating scale data, which have clear minimum and maximum boundaries. Variance estimates for data near these boundaries can be problematic and influence subsequent parametric statistical analysis.

All data was analyzed using a STATA statistical analysis software (StataCorp, College Station, TX). Measurement agreement for was assessed using a two-way random effects intraclass correlation coefficient (ICC). A mixed-effects model was used to compare the measures of (1) speech naturalness ratings, (2) stuttering frequency; (3) mean pause time; and (4) syllable rate for fluent runs of speech before and following treatment participation. Finally, linear regression analysis was used to determine statistical relationships between speech naturalness ratings, stuttering frequency, mean pause time and fluent syllable rate.

Chapter 3: Results

Rater Agreement

All acoustic measures of pause duration were completed by the investigator. Seven speech samples were randomly drawn from the pool of 68 speech samples (34 participants recorded before and following treatment) and measured a second time by the investigator to estimate within-rater agreement for the mean pause duration. This constituted approximately ten percent of the samples. A two-way random effects intraclass correlation coefficient (ICC) was used to assess within-rater reliability. The ICC is a value that ranges between 0 (no agreement) and 1 (perfect agreement). The ICC for a two-way random effects model was .97, with a 95% confidence interval between .82 and .99. Using conventional methods, this indicates that the agreement between the first and second sets of pause duration measures is considered excellent.

Descriptive Statistics

Table 1 contains a statistical summary of mean speech naturalness ratings, percent stuttered words, fluent syllable rate and mean pause duration, organized by treatment condition. Note that the speech naturalness ratings and stuttering frequency data are drawn from previous studies (Novelli, 2018; Tasko et al., 2007). It can be observed that the mean naturalness rating for the post-treatment samples is lower than the pre-treatment samples. Statistical analysis using a paired *t*-test revealed that the post-treatment speech naturalness ratings were significantly lower than the pre-treatment ratings (*t*-statistic: 1.92; df=33; p=0.03). It should be noted that naturalness ratings that are larger in magnitude reflect less natural sounding speech. herefore, as a group, the participants exhibited more natural sounding speech following treatment.

It can also be seen in Table 1 that the stuttering frequency (expressed as percent stuttered words) prior to treatment is much higher than stuttering frequency following treatment.

Statistical analysis using a paired *t*-test supports this observation (*t*-statistic: 10.18; *df*=33; *p*<0.00005). Prior to treatment, the group mean stuttering frequency was 9.9 percent stuttered words. Stuttering frequency dramatically decreased to 1.5 percent stuttered words following treatment. A key aim of the stuttering treatment program was to reduce stuttering frequency. This seems to have been achieved, at least for the group as a whole.

For the variable, syllable rate, there is little observable difference between the pre and post-treatment syllable rate. Prior to treatment, the group mean syllable rate was 4.6 syllables/sec and following treatment mean syllable rate was 4.5 syllables/sec. Not surprisingly, a paired *t*-test revealed no statistically significant difference was observed for syllable rate (*t*-statistic: .4139; *df*=33; *NS*). It appears that, for this participant pool, treatment participation had a negligible effect on syllable rate.

For the variable mean pause duration, there is a marked increase in pause duration following treatment participation. The group mean for the variable pause duration rose from 811 ms prior to treatment to 1073 ms following treatment. Statistical analysis using a paired *t*-test supports this finding (*t*-statistic: -4.1145; *df*=33; *p*=.0001). Therefore, as a group, participants exhibited significantly longer pause durations after treatment as compared to before treatment.

Measure	Treatment	Min	25^{th}	Median	75th	Max	Mean	SD
Naturalness	Pre	82	271	402	548	801	418	196
	Post	81	185	294	364	807	340	215
Percent Stuttered Words	Pre	1.0	4.6	7.3	16.2	23.5	9.9	6.6
	Post	0.0	0.3	1.2	2.5	5.8	1.5	1.5
Syllable Rate (syll/sec)	Pre	2.6	3.9	4.6	5.2	7.6	4.6	1.2
	Post	2.8	3.9	4.5	4.8	7.5	4.5	0.9
Pause Duration (ms)	Pre	532	675	790	941	1209	811	200
	Post	437	845	1039	1191	2162	1073	335

Table 1. Distributional summary, along with median, mean and standard deviation (SD) of speech naturalness ratings, percent stuttered words (stuttering frequency), syllable rate, and pause duration for both pre-treatment and post-treatment samples. Min: minimum value, 25^{th}: 25^{th} percentile, 75^{th}: 75^{th} percentile, Max: maximum value.

Figure 3 explores the relationship between pre-treatment and post-treatment speech naturalness ratings for individual participants using a scatter plot. Post-treatment mean naturalness ratings are plotted on the vertical axis and pre-treatment mean naturalness ratings are plotted on the horizontal axis. The diagonal dashed line represents a line of equality. Data points that fall on or near the line of equality indicate participants who showed minimal speech naturalness differences before and after treatment. Data points that fall below the line of equality indicate participants' post-treatment speech naturalness ratings are lower than pre-treatment ratings. Finally, data points above the diagonal represent post-treatment ratings that are higher (i.e. less natural sounding) than the pre-treatment ratings. Note that the majority of the data points are located below the line of equality indicating that a majority of the participants were rated as more unnatural prior to treatment. However, some data points are located on or very near to the line of equality, indicating that there was little change in speech naturalness from pre-treatment to post-treatment. The few data points located above the line of equality, reveal that some samples were rated more unnatural after treatment opposed to before treatment. In summary, because of the significant amount of data points below the line of equality, it can be concluded that speech naturalness is perceived as being more unnatural before treatment compared to after treatment.This finding supports the hypothesis that, as a group pre-treatment speech samples will be rated as sounding less natural compared to post-treatment speech samples. However, there are notable exceptions where participants exhibit either negligible reductions in speech naturalness ratings or large increases in speech naturalness ratings.

Figure 4 is identical in structure to Figure 3 but for stuttering frequency, expressed as percent stuttered words. Note that all data samples used in this study are below the line of equality. All data points are located below the line of equality, indicating that all participants showed some reduction in stuttering frequency during post-treatment samples. In other words, more stuttering was present during pre-treatment samples compared to post-treatment samples. For many participants, stuttering frequency falls to zero following treatment.

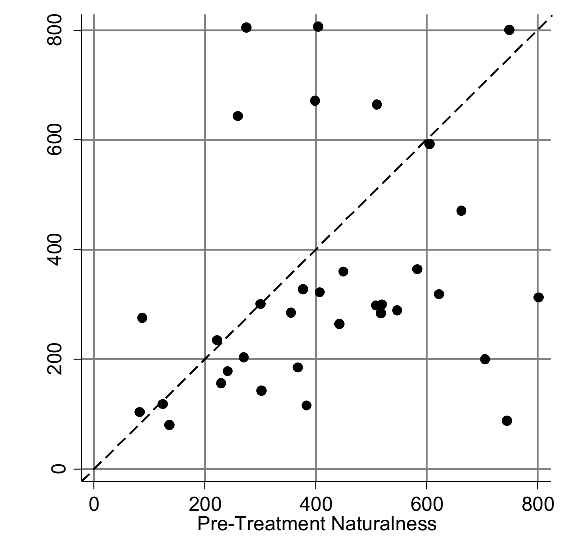

Figure 3. *A scatter plot with post-treatment speech naturalness on the Y-axis and pre-treatment speech naturalness on the X-axis. The diagonal dashed line represents a line of equality. Data points on or very near to the line of equality indicate little change in speech naturalness ratings across the two samples. Data points that fall below the line of equality indicate that post-treatment ratings are lower than pre-treatment ratings. Data points that fall above the line of equality indicate that post-treatment ratings are higher than pre-treatment ratings. Note that the majority of the data points fall below the line of equality. However, there are examples where the data points lie on or near the line of equality and a few data points where the post-treatment naturalness ratings are much higher than the pre-treatment ratings.*

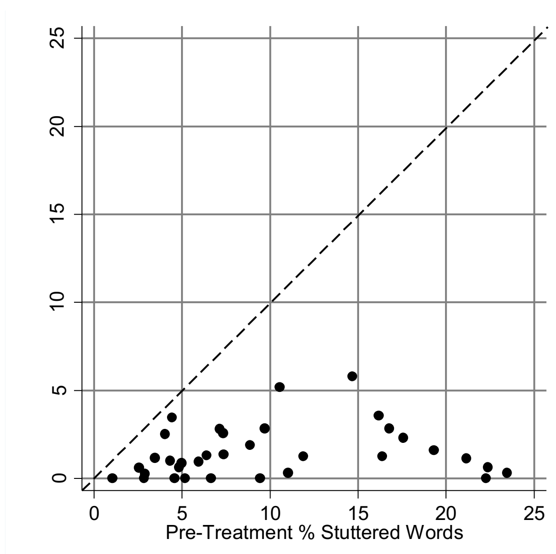

Figure 4. *A scatter plot with post-treatment percent stuttered words on the Y-axis and pre-treatment percent stuttered words on the X-axis. The diagonal dashed line represents a line of equality. Data points that fall below the line of equality indicate that there is a higher frequency of stuttered words during the pre-treatment samples. Note that all the data points fall below the line of equality. Therefore, all samples used in the study contain a higher frequency of stuttered words during pre-treatment samples.*

Relations between speech naturalness and percent stuttered words

Figures 5 and 6 explore the relationship between listener ratings of speech naturalness and percent stuttered words as measured by expert judges. Results for the pre-treatment samples are shown in Figure 5 and results for post-treatment samples are shown in Figure 6. Note that speech naturalness ratings and percent stuttered words are expressed using a rationale arcsine transformation to better meet the assumptions of statistical testing. For the pre-treatment data, there is a strong, positive, linear relationship between percent stuttered syllables and speech naturalness ratings. In other words, the AWS who exhibit a greater frequency of stuttering behaviors are more likely to be rated as sounding more unnatural than those who exhibit list stuttering behavior. Regression results (R^2: 0.62, $F(1, 32)$: 51, $p<0.00005$) provides statistical support for a strong positive association between percent stuttered words and speech naturalness. Approximately 62 percent of the variance in pre-treatment speech naturalness ratings is accounted for by stuttering frequency. Clearly, prior to treatment, speech naturalness ratings are strongly influenced by the stuttering frequency.

The relationship observed for the pre-treatment data is not replicated for the post-treatment condition (Figure 6). For the post-treatment data, there is little evidence for a linear relationship between percent stutters syllables and speech naturalness ratings (R^2: 0.002, $F(1, 32)$: 0.79, *NS*). One of the reasons for this disparity from pre-treatment findings is that as a group, the post-treatment stuttering frequency is very low (zero for many participants) and thus variability was quite restricted. These results indicate that following treatment, speech naturalness is not influenced by stuttering frequency. In summary, while stuttering frequency is a good predictor of speech naturalness prior to treatment it does not predict naturalness following treatment.

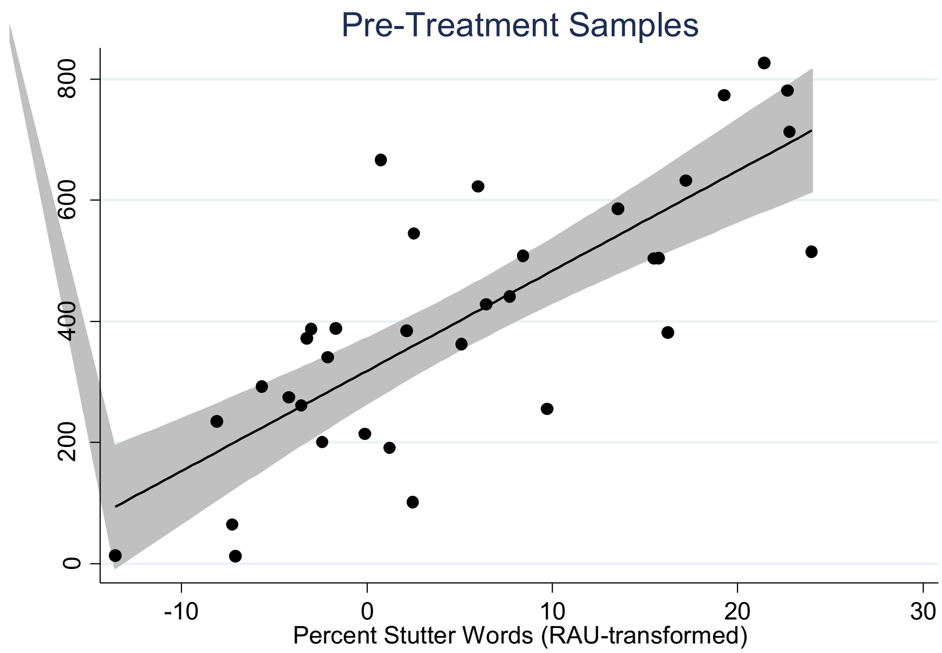

Figure 5. *A scatter plot with (RAU-transformed) mean naturalness ratings on the Y-axis and (RAU-transformed) percent stuttered words on the X-axis. The solid black line represents the line of best fit and the shaded grey area shows the 95% confidence interval about the line. A strong, positive linear relationship is observed between these two variables, a finding supported by linear regression analysis (R^2: 0.62, F(1, 32): 51, p<0.00005). Note this plot includes data for the pre-treatment condition only.*

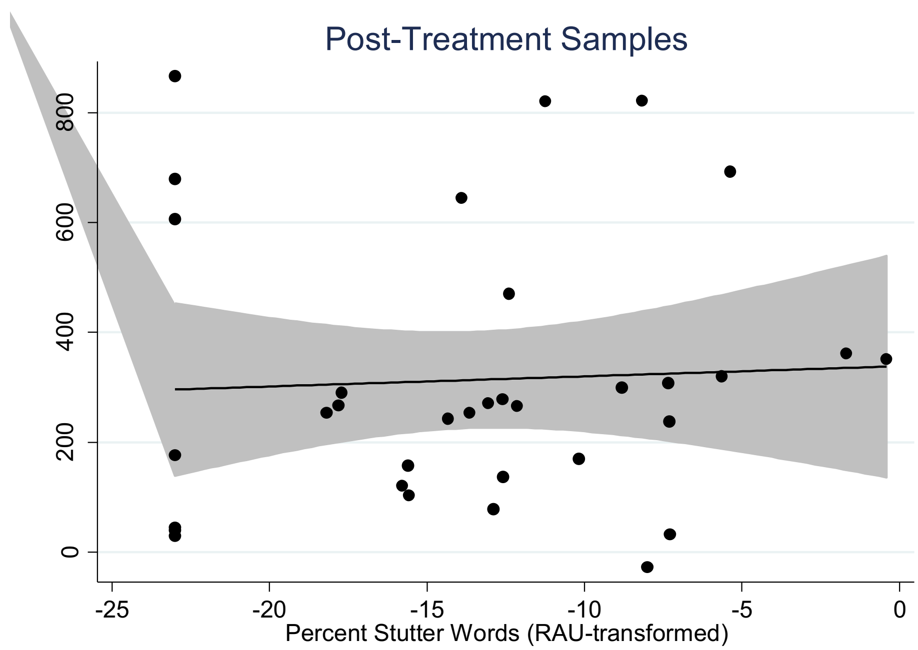

Figure 6. *A scatter plot with (RAU-transformed) mean naturalness ratings on the Y-axis and (RAU-transformed) percent stuttered words on the X-axis. The solid black line represents the line of best fit and the shaded grey area shows the 95% confidence interval about the line. There is no clear association between percent stuttered words and mean naturalness ratings, which is supported by linear regression analysis (R^2: 0.002, F(1, 32): 0.79, NS). Note this plot only includes data for the post-treatment condition.*

Relations between speech naturalness and syllable rate (syllables per second)

One of the possible correlates of speech naturalness in AWS is syllable rate. The relationship between speech naturalness and syllable rate for pre-treatment and post-treatment recordings is illustrated in the scatterplots in Figures 7 and 8 respectively. Prior to treatment participation, there was no evidence of a relationship between the speakers' syllable rate and perceived speech naturalness as evidenced by the small non-significant correlation coefficient observed in Figure 7. Regression results displayed in Figure 7: (R^2: 0.02, $F(1, 32)$: 0.66, *NS*) demonstrate neither a strong positive nor a strong negative linear association between the two variables. A similar finding was observed for the post-treatment condition (Figure 8). With the pre-treatment sample, the correlation coefficient between syllable rate and perceived speech naturalness was small and non-significant. Regression results displayed in Figure 8: (R^2: 0.03, $F(1, 32)$: 0.86, *NS*) reinforce a small, non-significant relationship between syllable rate and speech naturalness both before and after treatment. Overall, the syllable rate of the monologue samples appears unrelated to listener ratings of speech naturalness.

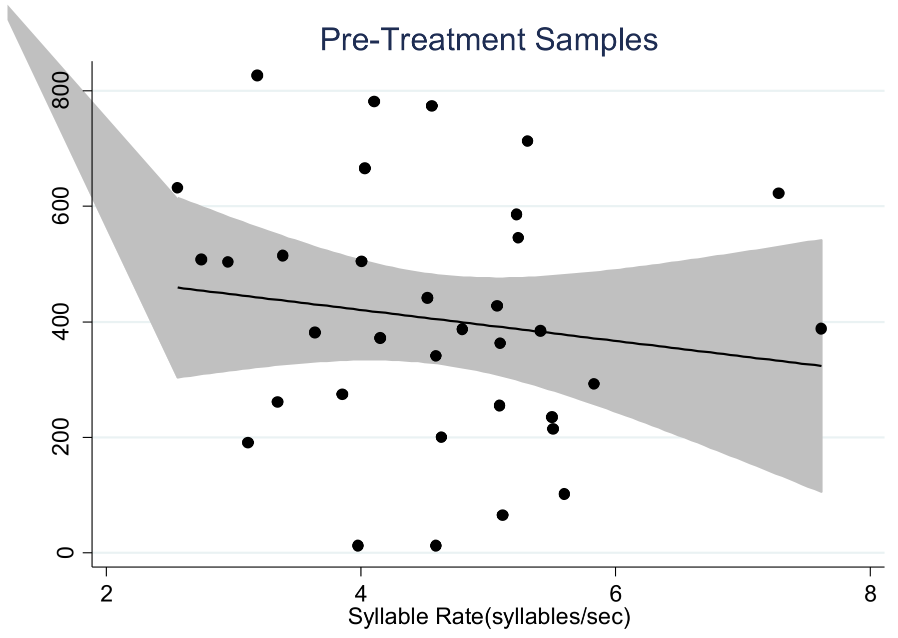

Figure 7. A scatter plot with (RAU-transformed) mean naturalness ratings on the Y-axis and syllable rate on the X-axis. The solid black line represents the line of best fit and the shaded grey area shows the 95% confidence interval about the line. Note this plot only includes data for the pre-treatment condition. There is no clear association between the two variables, which is supported by linear regression analysis (R^2: 0.02, F(1, 32): 0.66, NS).

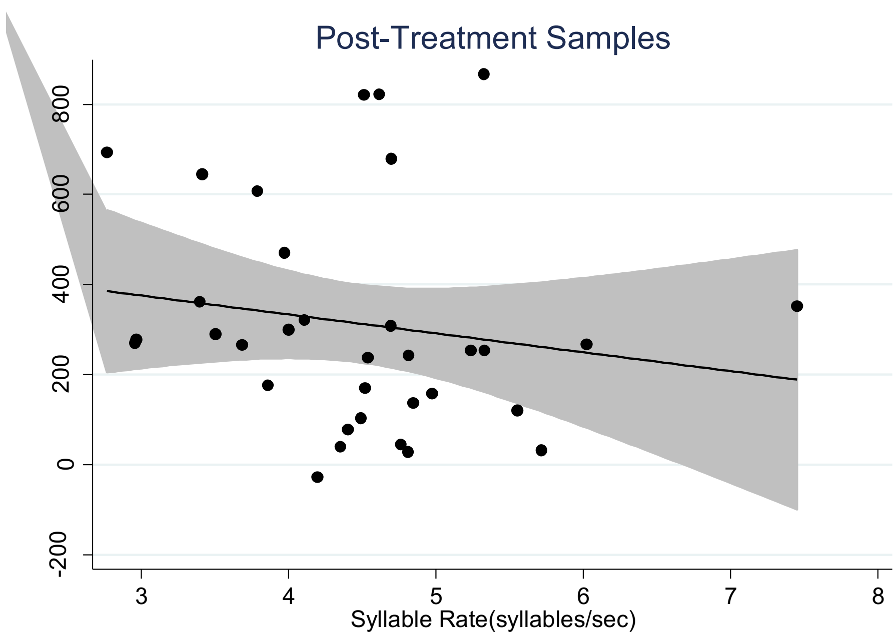

Figure 8. A scatter plot with (RAU-transformed) mean naturalness ratings on the Y-axis and syllable rate on the X-axis. The solid black line represents the line of best fit and the shaded grey area shows the 95% confidence interval about the line. This plot only includes post-treatment data. There is no clear association between the speech naturalness ratings and syllable rate for the post-treatment condition. This observation is supported by results from linear regression analysis (R^2: 0.03, F(1, 32): 0.86, NS).

Relations between speech naturalness and mean pause duration

Another possible correlate of speech naturalness in AWS is pause duration. Figures 9 and 10 explore the relationship between listener ratings of speech naturalness and mean pause duration. For the pre-treatment data found in Figure 9, there is no linear relationship between speech naturalness and pause duration. Regression results displayed in Figure 9: (R^2: 0.07, F(1, 32): 2.56, *NS*) reveal that there is more variability between the two correlates compared to other correlation coefficients, but there is no clear linear relationship. Therefore, during the pre-treatment samples, AWS did not exhibit significant pauses in speech that impacted speech naturalness ratings. In summary, pause time had little effect on speech naturalness during the pre-treatment data set.

The relationship observed for the pre-treatment data is not replicated for the post-treatment condition (Figure 10). For the post-treatment data, there is a strong, linear relationship between pause duration and perceived speech naturalness. Regression results displayed in Figure 10: (R^2: 0.55, F(1, 32): 39.44, *NS*) reinforce the strong correlation between pause duration and speech naturalness ratings in the post-treatment samples. Therefore, after treatment, subjects exhibited significant pauses in speech. As seen in Figure 10, the longer the pause, the more unnatural the sample was rated. In summary, pause duration has a significant impact on speech naturalness ratings during post-treatment only.

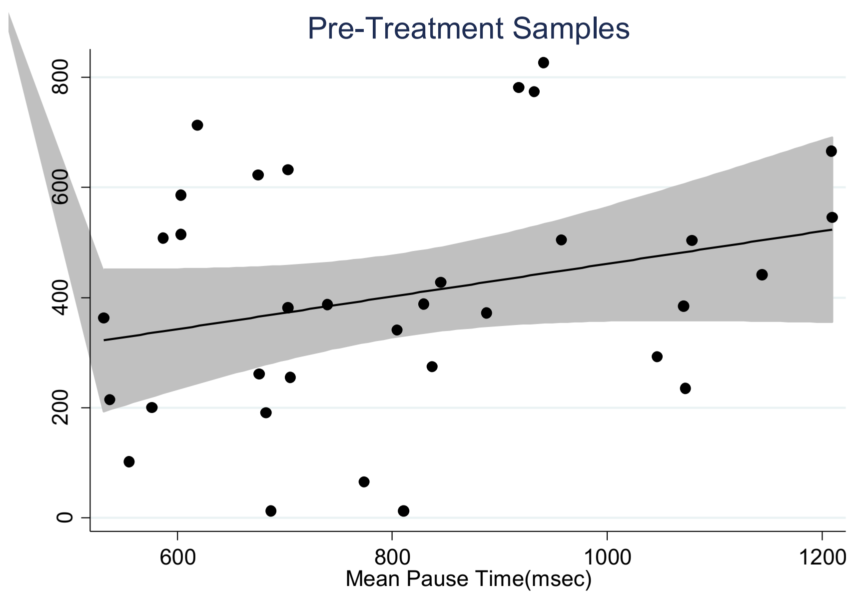

Figure 9. A scatter plot with (RAU-transformed) mean naturalness ratings on the Y-axis and mean pause duration in milliseconds on the X-axis. The solid black line represents the line of best fit and the shaded grey area shows the 95% confidence interval about the line. Note the data in this plot is from the pre-treatment condition. There appears to be a very weak positive association between mean pause duration and mean naturalness ratings for the pre-treatment condition. However, linear regression analysis indicated that this trend did not meet the criteria for statistical significance (R^2: 0.07, F(1, 32): 2.56, NS).

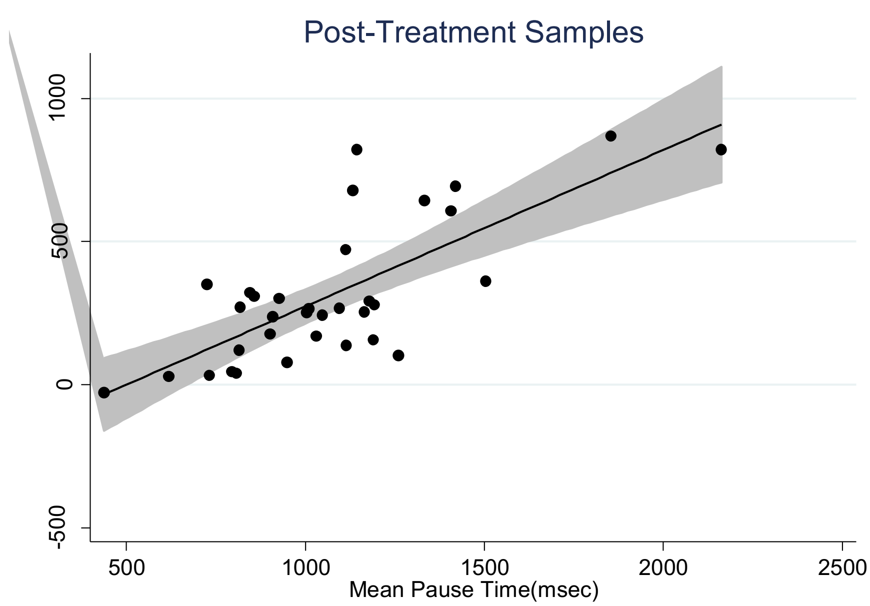

Figure 10. A scatter plot with (RAU-transformed) mean naturalness ratings on the Y-axis and mean pause duration in milliseconds on the X-axis. The solid black line represents the line of best fit and the shaded grey area shows the 95% confidence interval about the line. This plot only includes data for the post-treatment condition. A strong, positive, linear association is observed between mean pause duration and mean naturalness ratings. This result is supported by linear regression analysis (R^2: 0.55, F(1, 32): 39.44, p<0.00005).

Chapter 4: Discussion

The purpose of this study was to assess possible relationships between pause duration, fluent syllable rate, stuttering frequency, and pre-treatment and post-treatment speech naturalness ratings. As perceived speech naturalness can be influenced by several different factors, several bivariate regression models were developed and compared. The study revealed that some of these measures are correlated with speech naturalness. It is important to acknowledge that through careful examination and analysis of stuttering frequency, syllable rate and pause duration, varying relationships were discovered. Four main findings consisted in the current study.

The first key finding of the study is that there was a significant decrease in mean naturalness ratings from pre-treatment to post-treatment, indicating that AWS sound more unnatural prior to treatment as compared to after treatment. This was an expected finding and tends to correspond with findings of other studies. For example, Martin et al. (1984) found that speech naturalness ratings for AWS who had not had treatment was higher than AWS using Delayed Auditory Feedback (DAF), a fluency enhancing device used either as an adjunct or primary treatment approach. It should be highlighted that while the post-treatment samples were rated as more natural than the pre-treatment samples, the naturalness ratings were still quite high for a substantial number of participants. In other words, they continued to exhibit unnatural sounding speech. This finding is consistent with results from Martin et al., (1984) and leads us to conclude that speech naturalness ratings may improve, but not necessarily "normalize" as a consequence of stuttering treatment.

The second key finding revealed that stuttering frequency had a strong relationship with speech naturalness ratings for only the pre-treatment speech samples. This relationship is revealed in Figure 5. On the other hand, stuttering frequency did not serve as a predictive factor of speech naturalness for the post-treatment speech samples (Figure 6). Clearly, stuttering frequency greatly impacts speech naturalness prior to treatment but does not after treatment. The main reason for this is that the stuttering moments in the post-treatment samples were quite rare. As a result, the stuttering frequency was very low and with little variance. If stuttering is not present during post-treatment, then the frequency of stuttering cannot be predictive. This finding was supported by findings made by Martin, Triden and Haroldson (1984). When investigating the percentage of stuttered words in samples of AWS, a high correlation coefficient between percent stuttered words and speech naturalness was discovered (r=.81). Therefore, reinforcing the relationship between stuttering frequency and speech naturalness before treatment.

The key third finding of the study focuses on syllable rate and its effect on speech naturalness ratings. Based on previous studies, it was hypothesized that the syllable rate in AWS would be lower following stuttering treatment and that it would be negatively correlated with post-treatment naturalness ratings. Instead, there was no significant difference in syllable rate between pre-treatment and post-treatment ratings and syllable rate was not predictive of speech naturalness ratings for either pre or post-treatment samples. This finding was somewhat surprising, as previous studies have provided support that slower speech is correlated with ratings of unnatural speech (Jessen, 2016; Metz et al. 1990; Tasko et al. 2007). For example, using the same dataset to the one used for this study, Tasko et al. (2007) and Jessen (2016) found that syllable rate was a predictor of post-treatment naturalness scores. Further Metz et al., (1990) found that speech naturalness ratings were positively correlated with temporal measures of speech including voice onset and vowel duration. There may be some key methodological differences between the present study and others that could be an explanation for the significant differences between findings. First, the samples used in the current study came from a monologue task, while the samples used from Tasko et al. (2007) were extracted from recited speech, samples in Jessen (2016) came from oral reading and samples from Metz et al. (1990) came from word lists and picture descriptions. Second, the samples in the current study specifically eliminated any disfluent samples, which would affect syllable rate. This was not the case for Jessen (2016). These methodological differences could have significantly impacted the syllable rate results found in the current study.

The last finding was the relationship between pause duration and speech naturalness ratings. It was hypothesized that pause duration after treatment would be larger as compared to pause duration before treatment. The results supported this hypothesis as post-treatment mean pause duration was significantly longer than the pre-treatment mean pause duration. Further, pause duration was found to be a significant predictor of post-treatment naturalness scores. The results suggest that treatment may have influenced the duration of pauses in AWS, creating an overall higher naturalness rating (more unnatural) for post-treatment pause duration scores. It is likely that the longer pause times seen in the post-treatment samples were a consequence of employing new speech targets as well as individual proficiency of speech skills. This finding differed from that of Jessen (2016), where it was concluded that there was not a significant correlation between pause duration and naturalness ratings. Additionally, Jessen found that the duration of pauses had less of an influence of listener perception of speech naturalness than the overall speech rate did. This finding differed from that of the current study. It is important to note that the key difference between these two studies was where the current study examined monologue samples only, Jessen examined reading samples only. Therefore, leaving the question, why are people having longer pause durations in monologue samples? A possible explanation to this difference in pause duration could be the lack of script in a monologue sample. If speakers are not only concerned with "what" they have to say, but also "how" they will say it (using fluency targets), they may need to use additional planning time prior to the utterance. This would be revealed as longer pauses before each utterance. Further study would be needed with differing communication loads to better address this question.

Overall, we find that stuttering frequency is strongly correlated with speech naturalness prior to therapy, but is not correlated with speech naturalness at all after therapy. Therefore, indicating a significant reduction of stuttering from before treatment to after treatment. Pause duration displayed opposite results, where it was positively correlated after treatment as opposed to before treatment. This indicates that stuttering therapy approaches can modify pause patterns and durations observed in speech, impacting a listener's perception of naturalness. In conclusion, though speech naturalness in AWS is perceived as sounding more natural after treatment as opposed to before treatment, correlates of therapeutic approaches impact listener's ratings of speech naturalness of AWS.

Study Limitations and Future Research Directions

One limitation of this study was the little gender variability within the samples. Out of the 34 samples used in the study, the male to female ratio was 33:1. With only one female sample used in the study, an understanding of the impact of gender on speech naturalness is hard to obtain. A second limitation of this study was that the samples used were from a single stuttering treatment program. Therefore, it is unknown if naturalness scores would have been consistent across other programs because of variability in programs and methods being practiced at this specific site. And lastly, a final limitation of this study was the failure to fully assess the full data set provided. This study only assessed monologue samples, a narrow speaking task when stuttering is highly variable.

Future research directions of this study include utilizing other treatment programs. Exploring different methods and approaches may prove to be beneficial when seeking to highlight factors relating to speech naturalness in AWS. To learn how stuttering treatment approaches impact speech naturalness after receiving therapy, researching other approaches and therapies will be helpful in understanding the full effect of treatment on speech naturalness. Variability of treatment approaches will be vital for future research to determine differences between pre and post-stuttering treatment. Lastly, to investigate if stuttering generalizes across speaking activities, incorporating reading samples into the data set to assess speaking task variability would be a valuable feature to research in the future.

References

Bothe, A.K., Davidow, J.H., Bramlett, R.E., Ingham, R.J. (2006). Stuttering Treatmemmt Research 1970-2005: I. Systematic Review Incorporating Trail Quality Assessment of Behavioral, Cognitive, and Related Approaches. *American Journal of Speech-Language Pathology, 15,* 321-341.

Courmont, R.D., Kempf, J.B., Aimar, L., Francois, C. (1989). VideoLan Media Player [Computer Software]. Boston, MA: Free Software Foundation Inc.

Goebel, M. (1988). *CAFET: A computer-aided fluency establishment trainer.* Annandale, VA: Annandale Fluency Clinic.

Jessen, C. (2016). *Acoustic Correlates of Speech Naturalness in Post-Treatment Adults who Stutter: Role of Speaking Rate.* Western Michigan University Lee Honors College Thesis.

Martin, R., & Haroldson, S. (1992). Stuttering and Speech Naturalness: Audio and Audiovisual Judgments. *Journal of Speech and Hearing Research, 35,* 521-528

Martin, R. R., Haroldson, S. K., & Triden, K. A. (1984). Stuttering and Speech Naturalness. *Journal of Speech and Hearing Disorders*, *49*(1), 53.

Metz, D.E., Schiavetti, N., Sacco, P.R. (1990). Acoustic and Psychophysical Dimensions of The Perceived Speech Naturalness of Nonstutterers and Posttreatment Stutterers. *Journal of Speech and Hearing Disorders, 55,* 516-525.

Milenkovic, P. (2000). Time Frequency Analysis [Computer Software].

Novelli, A. (2018). *Speech Naturalness Before and Following Treatment in Adults Who Stutter*. Western Michigan University Lee Honors College Thesis.

Onslow, M., Adams, R., Ingham, R. (1992). Reliability of Speech Naturalness Ratings of Stuttered Speech During Treatment. *Journal of Speech and Hearing Research, 35,* 994-1001.

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.

Riley, G. (1994). Stuttering severity instrument. Austin, TX: Pro-Ed.

Studebaker, G.A. (1985). A "Rationalized" Arcsine Transform. *Journal of Speech and Hearing Research, 28* 455-462

Tasko, S.M., McClean, M.D., Runyan, C.M. (2007). Speech motor correlates of treatment-related changes in stuttering severity and speech naturalness. *Journal of Communication Disorders, 40* 42-65

Teshima, S., Langevin, M., Hagler, P., Kully, D. (2010). Post-treatment speech naturalness of Comprehensive Stuttering Program clients and differences in ratings among listener groups. *Journal of Fluency Disorders, 35,* 44-58.

Trajkovski, N., Andrews, C., Onslow, M., Packman, A., O'Brian, S., Menzies, R. (2009). Using syllable-timed speech to treat preschool children who stutter: a multiple baseline experiment. *Journal of Fluency Disorders, 34,* 1-10.

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

Yairi, E., & Seery, C. H. (2015). *Stuttering Foundations and Clinical Applications* (2nd ed.). Pearson.