4-18-2018

Speech Naturalness Before and Following Treatment in Adults Who Stutter

Alexis Novelli
Western Michigan University, lexilamb2015@gmail.com

Speech Naturalness Before and Following Treatment in Adults Who Stutter

Alexis Novelli

Western Michigan University

Abstract

Purpose: The speech of persons who stutter can often sound unnatural. This can be due to a variety of factors including the presence of stuttering itself and/or the unintended consequence of speech-altering treatment. Understanding factors that contribute to speech naturalness within the stuttering population can direct improved management approaches. This study examines associations between speech task, behavioral measures of stuttering, and selected acoustic measures on listener ratings of speech naturalness for a group of adults immediately before and after participation in an intensive fluency-shaping treatment program. A secondary aim of this study is to assess the use of a novel rating scale (i.e. a visual analog scale) to judge speech naturalness.

Methods: Participants include 50 young adult college students who passed a standard hearing screening. The participants served as judges who rated the level of perceived speech naturalness of video recordings made on adults who stutter. A computer interface allowed judges to rate speech naturalness by moving a slider on a computer screen along a continuum ranging from highly natural to highly unnatural. The video recordings that were judged consisted of one-minute excerpts drawn from 40 adults who stutter performing an oral reading and monologue before and following participation in a group-based intensive, four-week fluency shaping program. Mean naturalness ratings were compared across the speech task (oral reading vs. monologue) and time (pre-treatment vs. post-treatment). Additionally, regression analysis was used to evaluate the relationship between speech naturalness ratings and measures of stuttering frequency and selected acoustic measures.

Results: Rater agreement using the visual analog approach to speech naturalness rating was considered excellent and was highly correlated with data previously collected using a more

conventional nine-point equal appearing interval scale. A mixed-effects model was used to assess effects of speech task and treatment participation on mean speech naturalness ratings of the video samples. There was a significant main effect for treatment but not for speech task. Post-treatment samples were rated as more natural than the pre-treatment samples. Pre-treatment naturalness ratings were positively correlated with stuttering frequency, but post-treatment naturalness ratings were unrelated to stuttering frequency. Post-treatment naturalness ratings were negatively correlated with articulation rate, and positively correlated with mean pause duration.

Conclusion: Excellent rater agreement and strong correlations with more traditional rating scales suggest that a visual analog scale can be a viable scaling method for rating speech naturalness. Although stuttering frequency was a strong predictor of speech naturalness prior to treatment, it provided little prediction of speech naturalness following treatment. Instead, articulation rate and pause duration measured from the speech samples predict post-treatment naturalness ratings. Specifically, those speech samples with the slowest articulation rates and the largest speech pauses between phrases were associated with less natural sounding speech and may represent negative consequences of stuttering treatment. This study demonstrates that the combined use of stuttering frequency measures and acoustic measures of speech performance can help understand the factors that determine why some persons who stutter sound more or less natural than others.

Background

Stuttering is a speech disorder primarily characterized by intermittent, involuntary disruptions of speech, such as part-word repetitions, sound prolongations, and/or blockage of sound and airflow. The disorder typically begins in early childhood, between the ages of two and four years when speech and language skills are rapidly expanding. Although as many as 5-10 % of the preschool population will exhibit speech patterns consistent with stuttering, most will slowly resolve regardless of intervation. It has been reported that recovery rates in stuttering children is as high as eighty percent (Yairi & Ambrose, 2005). However, the longer the stuttering continues, the less likely recovery will occur (Yairi & Ambrose, 2005). As a result, for about one percent of the general population, stuttering is chronic, life-long challenge, with a male-to-female ration of about three to one (Bloodstein & Ratner, 2008). For those continue to stutter, more advanced characteristics often develop, such as excessive physical tension, facial grimacing, eye blinking when speaking, a growing avoidance of difficult words, sounds and speaking situations, and the development of strong negative thoughts and/or emotions associated with communication (Van Riper, 1982).

While a lot is known about the characteristics and natural history of stuttering, its cause remains elusive. There is compelling evidence for a genetic component to stuttering. For example, a child who stutters is more likely to have relatives who stutter, and recent biological genetic studies have revealed candidate genes for stuttering (e.g. Raza et al., 2012). Even so, the causal relationship between the known genetic anomalies and the physical manifestations of the disorder remain unclear. There is also a growing body of brain imaging studies that suggests children and adults who stutter are more likely to exhibit brain anomalies involving the circuits

associated with speech motor control as compared to fluently speaking peers (e.g. Chang, Erickson, Ambrose, Hasegawa-Johnson, & Ludlow, 2008).

Regardless of the specific etiology, stuttering can have large persistent effects on communication and can negatively influence cognitive, emotional and social well-being. Therefore, treatment approaches for adults who chronically stutter not only focus on techniques to manage or reduce the frequency of stuttering, but also address negative thoughts and feelings about communication in general and stuttering in particular. One common approach for reducing the frequency of stuttering involves learning new speaking techniques that are known to prevent the likelihood of stuttering. Techniques include the systematically slowing speech rate, or stretching vowel sounds, improving breath support, producing gradual or easy onset of voicing, producing a light contact of articulators, and continuously phonating through utterances (e.g. Gobel, 1988). These techniques must be learned through practice and are typically employed consistently every time to person speaks. While such methods, can be very successful in decreasing the frequency of stuttering, they can also have unintended consequences for the clients. While the client's speech may be fluent and free of stuttering, the speech modifications employed can sound unnatural to the listener. The presence of stuttering also sounds unnatural to listeners, creating a dilemma for the person who stutters. Do they stutter and sound unnatural, or use fluency enhancing speech targets which may also lead to unnatural sounding speech?

The use of speech naturalness ratings as a more global way to measure speech behaviors in persons who stuttering became more prominent in the early 1980s. Martin, Haroldson and Triden (1984) had raters use a one to nine equal appearing interval scale to rate speech naturalness in adults who stutter, adults who do not stutter, and adults who stutter during use of

delayed auditory feedback (DAF). The last condition was used since DAF is known to reduce stuttering frequency. The authors found that adults who stutter were generally rated as sounding less natural than adults who did not stutter. Additionally, while DAF reduced stuttering, listeners still rated the speech of this group of adults who stutter less natural than the non-stuttering control group, suggesting that factors other than stuttering may influence ratings. Finally, measures of rater reliability and agreement were generally good within and across raters (Martin et al., 1984). Since then, a number of studies have been published that have examined the speech naturalness of persons who stutter across a range of different dimensions (Martin & Haroldson, 1992; Metz et al., 1990; Onslow et al., 1992; Runyan, Bell & Prosek, 1990; Teshima et al., 2010). Naturalness ratings for persons who stutter are typically higher when using audio-visual as compared to audio-only samples (Martin and Haroldson, 1992) suggesting that visual information is important dimension to consider. Naturalness ratings tend to be more consistent for longer speech samples (one minute or more in length) compared to shorter excerpts of speech (Onslow et al, 1992). The characteristics of the judges also appears to have an influence on speech naturalness ratings. Onslow et al (1992) found that sophisticated judges tended to be more reliable in their ratings as compared to unsophisticated judges. Teshima et al. (2010) also examined the influence of three different types of judges; speech language pathology students, community members and persons who stutter. Interestingly, the speech language pathology students were the least critical judges as compared to the other two groups.

One area that has received attention is the effect of treatment participation on speech naturalness in persons who stutter. For example, Runyan et al. (1990) compared naturalness ratings of groups of adults who stutter following participation in one of six different treatment programs with a non-stuttering control group. The authors found that post-treatment adults who

stutter were rated less natural than non-stuttering controls in spite of the stuttering groups producing only fluent speech. Second, stuttering severity prior to entering therapy was not found to have any influence on the naturalness ratings following treatment. In contrast, Onslow et al. (1992) found that those who had the most severe pre-treatment stuttering severity ratings exhibited more unnatural sound speech following speech, even though the stuttering treatment had the intended effect of largely eliminating stuttering. Metz et al. (1990) also found that the speech of treated adults who stutter was judged to be less natural than the speech of non-stuttering controls, in spite of the general absence of stuttering. Furthermore, the authors found that measures of speech duration and voice onset time were correlated with speech naturalness ratings suggesting that speech naturalness may be related to temporal aspects of speech. This is consistent with studies of healthy speakers where speech samples were experimentally lengthened (Schaeffer & Eichorn, 2001; Schaeffer, 2006). It is worth noting that many of the studies used post-treatment speech samples relatively close to the conclusion of treatment making it less clear how the post-treatment speech naturalness might change over time. Teshima et al. (2010) compared a group of participants immediately following stuttering treatment and five years later. The authors found that speech was rated as more natural five years later as compared to the immediately following treatment. Overall, these studies suggest that persons who stutter can struggle with producing natural sounding speech even after treatment. These observations present a challenge to clients and clinicians, since at least one dimension of a positive treatment outcome is relatively fluent and natural sounding speech. If a client sounds unnatural to listeners following treatment, is it due to the continued presence of stuttering, a result of the treatment itself, other unidentified factors, or some combination? Understanding factors that influence speech naturalness within the stuttering population can help guide the

choice of treatment targets, and possibly enhance willingness to continue with treatment. The primary aim of this study is to attempt to better understand those factors that influence speech naturalness ratings in adults who stutter.

Most speech naturalness studies to date have used a one to nine equal appearing interval (EAI) scale to judge speech naturalness. The higher the rating, the less natural the speech. It is used with such frequency that it can be considered the standard convention for rating speech naturalness. Other methods have been developed. For example, Metz et al. (1990) compared the standard EAI scale with a direct magnitude estimation procedure and found no difference in the two approaches. Pravesh and Geetha (2012) used an approach that required listeners to make binary judgements (natural or unnatural) across a range of speech dimensions that included rate, continuity, effort, stress, intonation-rhythm, articulation and breathing pattern. The authors found that the ratings were reliable. Cronbach's alpha was found to be 0.84 across judges and 0.98 within judges. Schaeffer and colleagues (2001, 2006) used a different rating approach in their experiments that required naturalness judgements of synthetically modified speech. Judges chose "Strongly Agree", "Agree", or "Disagree" that the speech sounded natural. One approach that has not yet been used to assess speech naturalness is a visual analog scale. A visual analog scale requires judges to either make a mark or adjust a "slider" on a continuous line to a location that matches their impression of speech naturalness. Unlike EAI or binary scales, a visual analog scale allows judges to rate speech naturalness along a continuum, thus giving them a greater range of options to choose from rather than a set number of two, three or nine.

This study examines associations between speech task, behavioral measures of stuttering, and selected acoustic measures on listener ratings of speech naturalness for a large group of

adults immediately before and after participation in an intensive fluency-shaping treatment program. Speech task (monologue vs. oral reading) was included as a factor since persons who stutter commonly exhibit variable stuttering frequency across different speech tasks.

A secondary aim of this study is methodological in focus. Since Martin et al. (1984), the speech naturalness rating scale of choice has been a nine-point equal appearing interval (EAI) scale (for exceptions, see Metz et al., 1990; Schaeffer, 2006). This study implements a visual analog scale to rate speech naturalness. Rater agreement and comparisons with ratings based on the traditional EAI scale are reported.

Methods

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

Participants (Judges)

Participants were recruited from the WMU student body. Inclusion criteria required participants to be over the age of eighteen and pass a standard hearing screening (25 dB_{HL} at 1, 2, and 4 kHz). Fifty-two individuals (Gender: 51 females, one male; Mean Age: 22 years; Range: 20-29 years) volunteered to serve as participants for this study. Two participants did not pass the hearing screening resulting in fifty individuals who served as the judges in the main experiment. Although there was no monetary compensation, the judges received extra credit in a course registration in fluency disorders. The judges were speech language pathology students 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.

Video samples of adults who stutter

Video samples were drawn from the WRAMC-WMU stuttering database, a behavioral and physiological dataset that includes 43 adults who stutter (AWS). Details about the collection and processing of the dataset can be found in Tasko, McClean & Runyan (2007). The video recordings took place in a dedicated recording studio using trained technicians. Recordings were made on two occasions: (1) within a 48-hour period of beginning a four-week intensive stuttering treatment program and (2) within 48 hours of completion of the final week of the program. Three different speech tasks were recorded for each participant. First, each participant read

aloud paragraph given to them by the experimenter (Appendix A). The second task was an approximately four-minute monologue on a topic of choice. The third task included a series of telephone calls to local hotels to enquire about room rates. For this study, only the oral reading and monologue samples were used.

A video-editing program (MS Moviemaker) was used to extract one-minute samples of the oral reading and monologue tasks for the pre and post-treatment recording periods. Not all participants completed the treatment program and not all video samples were usable due to technical problems. A total of 152 samples were generated from 40 AWS (Gender: 37 males, 3 females; Mean age: 26 years; Range = 19-43 years).

Stuttering Treatment Program

All 40 subjects participated in The Walter Reed Stuttering Treatment Program (Tasko et al., 2007). This program was very intensive, requiring participation five to six hours a day, five days a week for one month. It was a group-based therapy with 3-6 members per treatment group. In general, therapy and consisted of three partially overlapping parts. The first part focused on speech production, characteristics of stuttering, and attitudes/feelings toward stuttering. The second part focused on establishing behavior target speech behaviors such as increased abdominal breathing, breath support and volume, continuity of air flow, voicing before exhalation, easy articulatory and phonatory onset, continuous phonation, and phrasing. This was done with the aid of a computer-based biofeedback system (Gobel, 1988). The third part provided settings and activities to transfer the skills learned to enhance fluency into other speaking situations.

Previously Collected Measures

One of the aims of the study is to determine associations between speech naturalness ratings and other behavioral and acoustic features of the speech samples. These other measures were drawn from previously completed studies using the WRAMC-WMU stuttering database.

Disfluency Analysis

A detailed disfluency analysis of all video samples was completed prior to the current study. Two certified speech pathologists highly experienced with stuttering independently orthographically transcribed all video recordings, counted words and syllable and identified all disfluent events within the samples using standard techniques. Following the individual analysis, the two clinicians reviewed the results together to compare results. If there was a disagreement, the two clinicians watched the video repeatedly until a consensus was reached (Tasko et al., 2007). These measures were used to determine a range of behavioral measures including percent stuttered words, type and frequency of disfluency classes.

Acoustic Measurements

Selected acoustic measures of some of the speech samples were available from a previous, related study using the WRAMC-WMU stuttering database (Jesson, 2016). This study extracted speech rate, articulation rate and speech pause information on post-treatment oral reading video recordings. A commonly used acoustic analysis package was used to complete the analysis. A synchronous waveform and broadband spectrogram was used determine the onset and offset of speech runs. Gaps in acoustical energy greater than 250 milliseconds in the sample was used to determine pauses in the sample. After acoustic segmentation was completed, syllable

counts were collected and the following measures were derived: speech rate (total number of syllables/[articulation + pause time]), articulation rate (total number of syllables/articulation time), and average pause duration (total pause time/total number of pauses) (Jessen, 2016).

Naturalness Ratings using a Nine-Point Equal Appearing Interval (EAI) Scale

In a previous study, pre and post-treatment monologue samples were also submitted to more traditional naturalness ratings using a nine-point equal-appearing-interval scale (Tasko et al., 2007). This previously reported data provided an opportunity to comparison against the data collected in the current study using a different scaling procedure.

Experimental Procedure

Each experimental session required judges to rate speech naturalness on 38 video samples. The number of judged samples was based on the need for the experimental session to be completed within one hour. For each judge, the samples were selected from the larger group of 152 samples. Sample selection was a quasi-random process so that samples assigned to a given judge was not predictable but still allowed for ensuring that each video sample was submitted to an equivalent number of judgements.

The experiment was controlled by locally developed computer software that automatically presented the video samples and recorded naturalness ratings made by the judge. The judge was seated in in quiet room in front of a computer and was provided instructions to watch each video presented and then rate each one using the visual analog scale on the computer screen (Figure 1). Judges could view each video only once. Audio levels were set to a comfortable level and judges were also instructed not to increase or decrease the volume of the

recordings. Audio levels remained constant across all sessions. After each participant session, the computer software recorded a text file that encoded the presentation order of the video samples, the response duration of the participant and the naturalness rating, which was expressed as an integer between 0 (highly natural) and 1000 (highly unnatural). It is important to note that larger values on the naturalness rating scale is associated with less natural speech.

Figure 1. A screen shot of computer interface used by judges to rate speech naturalness. The black rectangle contained the video sample. The judges used a mouse to place the 'slider icon' to a location on the sale that matched their subjective impression of speech naturalness of the stimuli presented.

Data Analysis

Prior to statistical analysis, all naturalness ratings and disfluency proportions were submitted to a rationalized arcsine transform (Studebaker, 1985), or a RAU transform. This data transformation is recommended for data such as proportions, 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. A number of statistical measures were employed. Rater reliability was assessed using a one-way random effects intraclass correlation coefficient (ICC). A mixed-effects model was used to compare speech naturalness ratings across time (pre-treament vs. post-treatment and speech task (oral reading vs. monologue). Finally, linear regression analysis was used to determine statistical relationships between speech naturalness ratings and other measures of speech performance.

Results

Rater Agreement

Each of the 152 samples received neither 12 or 13 separate ratings. No single judge rated a sample more than once. As a result, it was not possible to estimate within-rater reliability. Also, not all judges rated all samples. For this reason, it was necessary to a use one-way random effects intraclass correlation coefficient to assess rater reliability. The ICC is a value that ranges between 0 (no agreement) and 1 (perfect agreement). The ICC for a one-way random effects model was .93, with a 95% confidence interval between .89 and .95. Using conventional methods, this level of rater agreement is considered excellent, suggesting that speech naturalness can be reliably measured using a visual analog scale.

Figure 2 is a frequency histogram for the 1900 speech naturalness ratings collected in the study. Recall that ratings near zero reflect a judgement that the sample was highly natural and ratings near 1000 reflect a judgement that the sample was highly unnatural. There are a number of things to note in the figure. As a group, the judges used the entire range of the scale. Another observation is that the ratings are not evenly (or normally) distributed across the range of the scale. Instead, the distribution has a bimodal appearance, with a set of ratings clustering toward the bottom half of the scale and a second cluster in the top half of the scale. Finally, the cluster toward the bottom part of the scale has a large number of ratings at or very close to zero (highly natural).

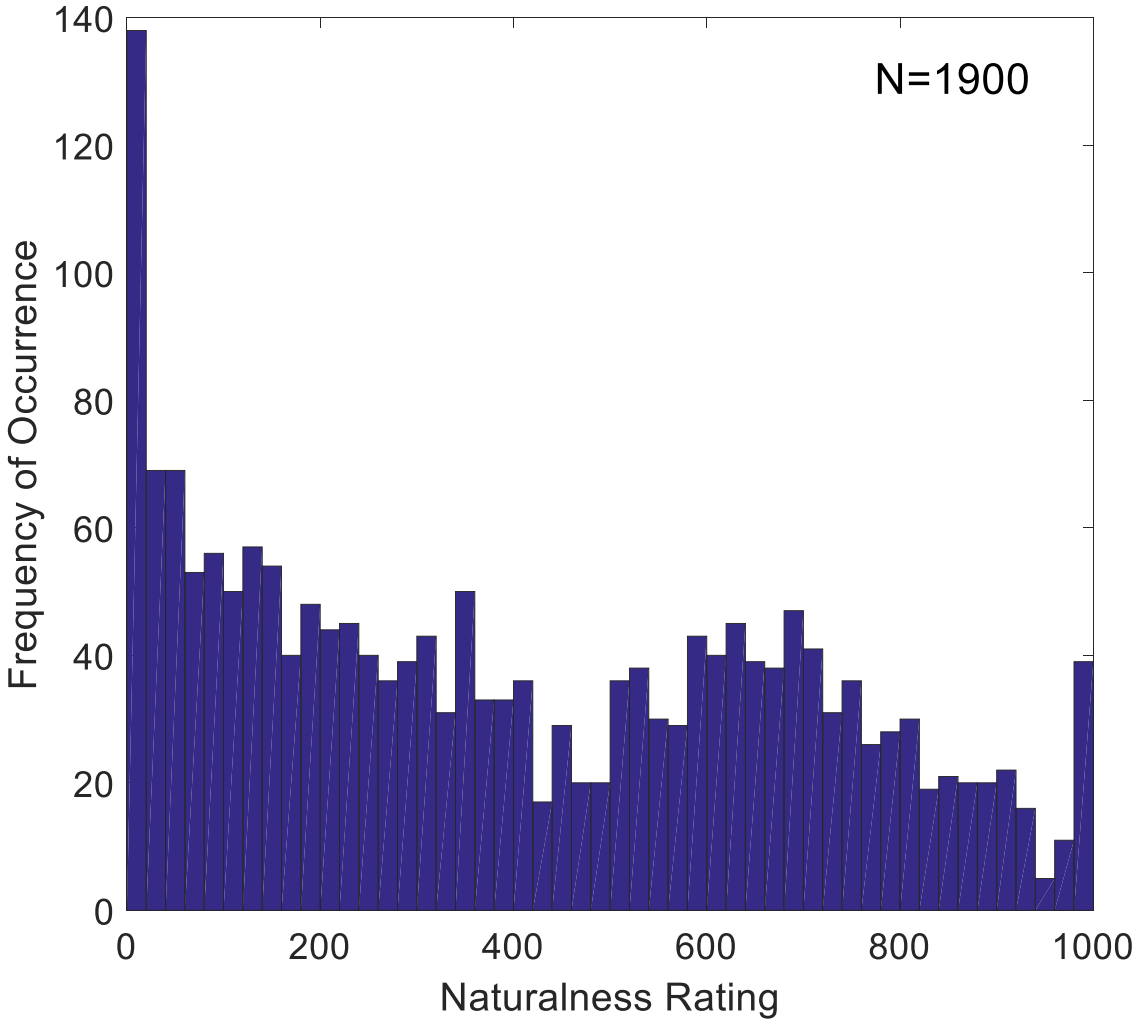

Figure 2. Frequency histogram for speech naturalness ratings. Data is collapsed across all sample videos and raters.

Most speech naturalness studies in the literature have used a nine-point EAI scale. Tasko et al. (2007) reported speech naturalness ratings on a subset of the current data (monologue samples) using the EAI scale. Therefore, it would be helpful to evaluate how naturalness ratings using the visual analog scaling procedure compared to ratings using the nine-point EAI scale. Figure 3 is a plot that shows the mean (RAU-transformed) naturalness ratings using the visual

analog scale on the horizontal axis and mean (RAU-transformed) speech naturalness ratings using the EAI scale on the vertical axis for the monologue samples where both ratings were available. A strong, positive, linear association can be observed between naturalness ratings using a visual analog scale and ratings using the nine-point EAI scale. Linear regression analysis revealed a significant linear model fit ($F(1, 71)$: 118, $p < 0.00005$) that accounts for more than 60 percent of the data variance (Adjusted R^2: 0.62).

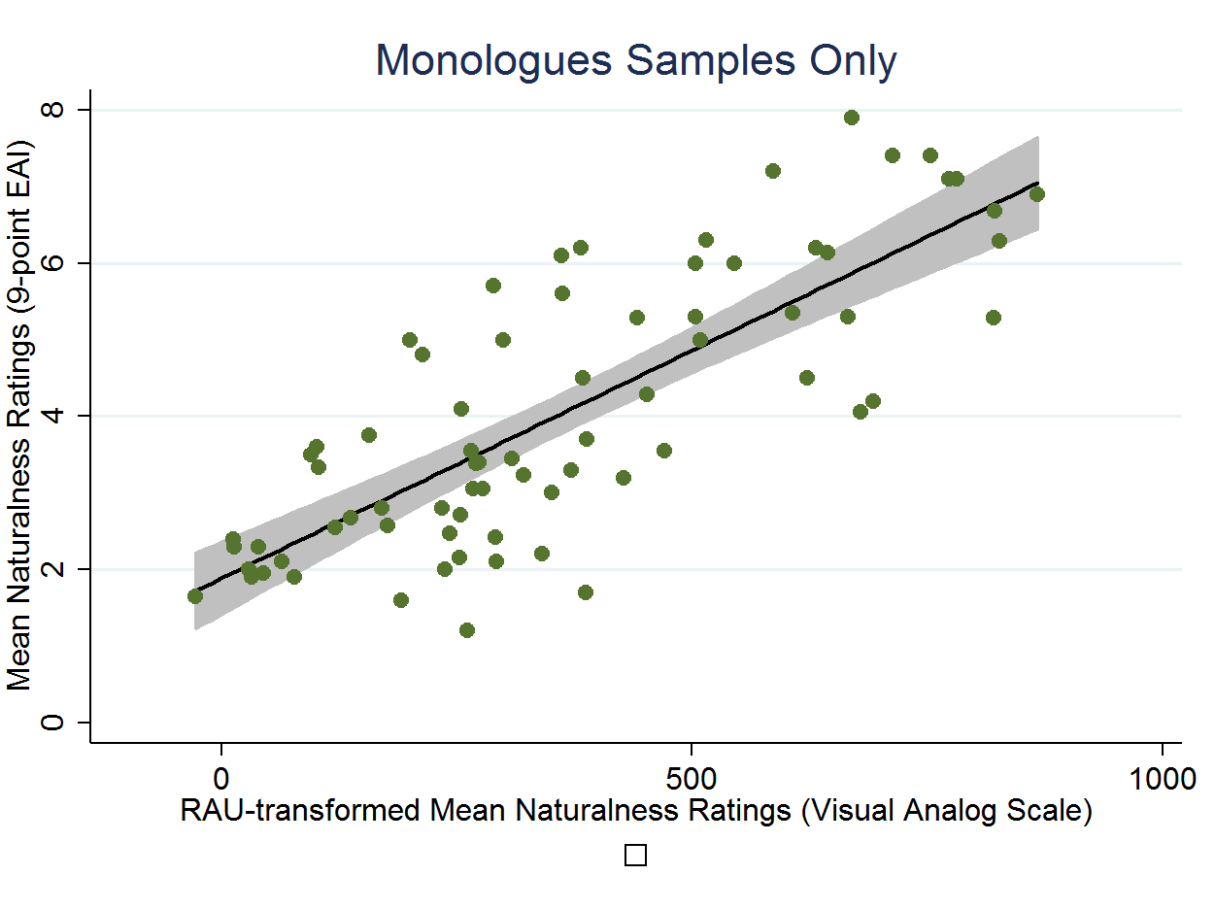

Figure 3. Scatterplot with mean naturalness ratings using the visual analog scale on the horizontal axis and the mean naturalness ratings using a nine-point EAI on the vertical axis. Note this data only includes samples from the pre and post-treatment monologue recordings. The solid black line represents the line of best fit and the shaded grey area shows the 95% confidence interval about the line. The linear regression model was significant ($F(1, 71)$: 118, $p < 0.00005$), accounting for more than 60 percent of the variance in the data (Adjusted R^2: 0.62).

Effects of Speech Task and Treatment Participation on Speech Naturalness Ratings

Table 1 and Figure 4 provide a summary of speech naturalness ratings across recording period and speech task. As previously stated, a speech naturalness score of 0 was highly natural and 1000 was highly unnatural. Therefore, higher values are associated with less natural sounding speech. Across both speech tasks, the pre-treatment speech naturalness ratings scores are higher (less natural) than for the post-treatment ratings. Second, the oral reading samples were generally rated as less natural than the monologue samples. A mixed-effects linear model was used to assess the effect of stuttering treatment and speech task on speech naturalness ratings. Treatment and speech task were considered fixed factors and subject (person who stutters) was considered a random factor. Both main effects and the interaction term were included in the model. The overall model was significant (X^2 (3) = 9.02, p = 0.029). As suggested by the mean values, the main effect for treatment was significant (p = 0.04). However, the effect of speech task was not significant (p = 0.449), nor was the treatment-by-speech task interaction (p = .857). In summary, results revealed that speech samples were rated as more natural following treatment and that speech task did not have a significant influence on speech naturalness ratings.

Table 1. Summary of mean and standard error (SE) of speech naturalness ratings across recording period and speech task.

	Monologue		**Reading**	
	Mean	SE	Mean	SE
Pre-Treatment	402	39	432	39
Post-Treatment	317	40	357	40

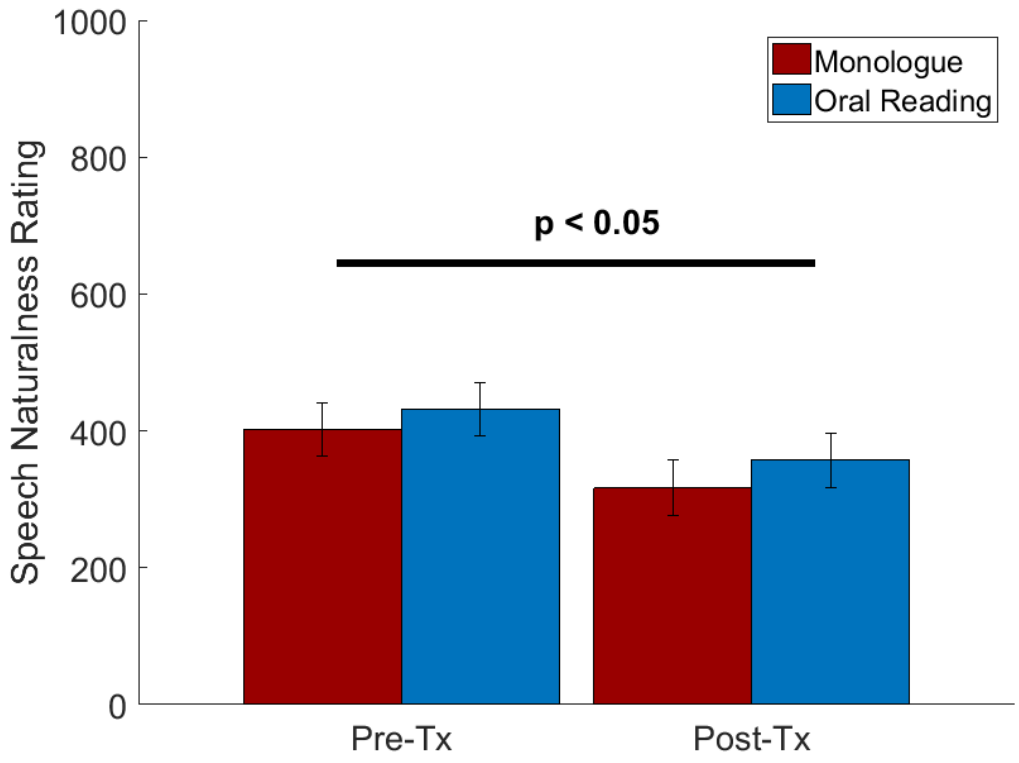

Figure 4. Bar plot of mean naturalness ratings for monologue (left bar in red) and oral reading (right bar in blue) for pre-treatment and post-treatment samples. The error bars reflect standard error. On the vertical axis is the mean speech naturalness ratings (X^2 (3) = 9.02, p=0.029). The mean ratings for post-treatment was lower than the pre-treatment mean ratings, meaning the main effect was significant (p=0.039). It was not significant for task, though, because there is virtually no difference between the pre-treatment reading and monologue scores versus the post-treatment reading and monologue scores (p=0.449).

Relations between speech naturalness and stuttering frequency

One possible correlate of speech naturalness in persons who stutter is the presence of stuttering behaviors. Therefore, linear regression was used to assess for a relationship between speech naturalness ratings and stuttering frequency. Based on the results from the mixed-effect analysis in the last section, separate analyses were performed on the pre and post-treatment data. However, due to the similarities in results for the monologue and oral reading samples, speech

tasks were collapsed together for the analyses. Figures 5 and 6 are scatterplots of speech naturalness plotted as a function of stuttering frequency for the pre-treatment and post-treatment data respectively. For the pre-treatment data (Figure 5), there is a strong, positive, linear association between stuttering frequency and speech naturalness ratings. Linear regression analysis revealed a significant linear model fit ($F(1, 70)$: 107, $p < 0.00005$) that accounts for 58 percent of the data variance (Adjusted R^2: 0.58). Clearly, for the pre-treatment samples, speech naturalness ratings are strongly influenced by the amount of stuttering in the samples.

The same cannot be said for the post-treatment data (Figure 6). There is little visual evidence that post-treatment naturalness ratings are associated in any systematic way to stuttering frequency. The linear model that included the two variables was not significant ($F(1, 70)$: 2.2, *NS*, Adjusted R^2: 0.02). The lack of association can most likely be attributed to the small amount of variance in the post-treatment stuttering frequency data. It is somewhat difficult to interpret the data following RAU transformation. The negative values suggest the frequency counts are all very small for the post-treatment samples. This is indeed the case with 90 percent of the samples exhibiting less than 3.5 percent stuttered words and almost one third of the samples having no stuttering at all. So, while post-treatment samples are often deemed unnatural (note in Figure 6 that ratings exist on the upper end of the naturalness rating scale), it does not appear to be related to the amount of stuttering in the video samples. Other factors must be in play. As Metz et al. (1990) suggested that, in their sample, post-treatment naturalness ratings were correlated with durational aspects of the speech acoustic signal, further analysis of the post-treatment data was performed.

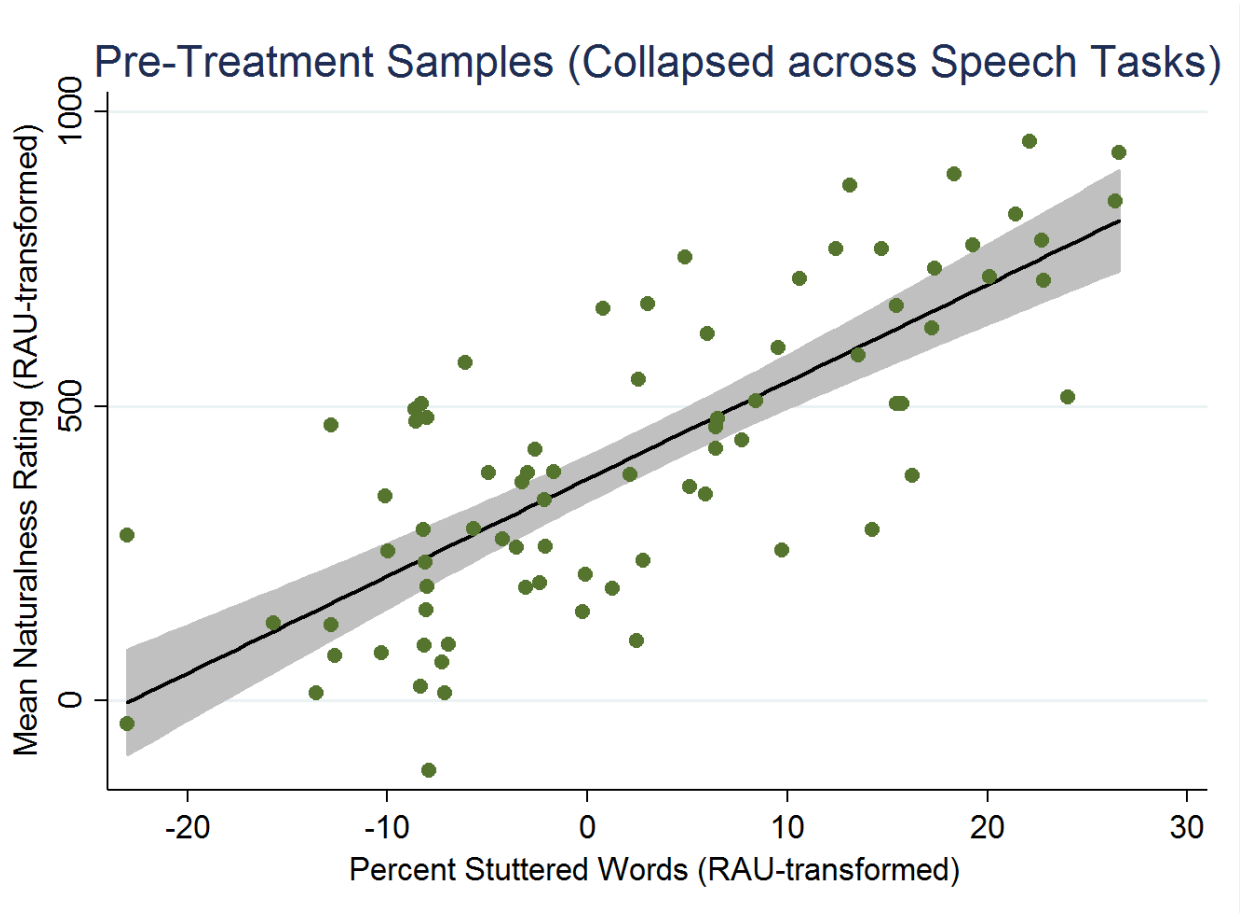

Figure 5. Scatterplot with (RAU-transformed) percent stuttered words on the horizontal axis and mean naturalness ratings (RAU-transformed) on the vertical axis. These sets of data are for pre-treatment stimuli, collapsed across speech tasks. The plot shows a very strong, positive, linear association between the two variables (Adjusted R^2: 0.58, $F(1, 70)$: 107, $p<0.00005$). The solid black line represents the line of best fit and the shaded grey area shows the 95% confidence interval about the line.

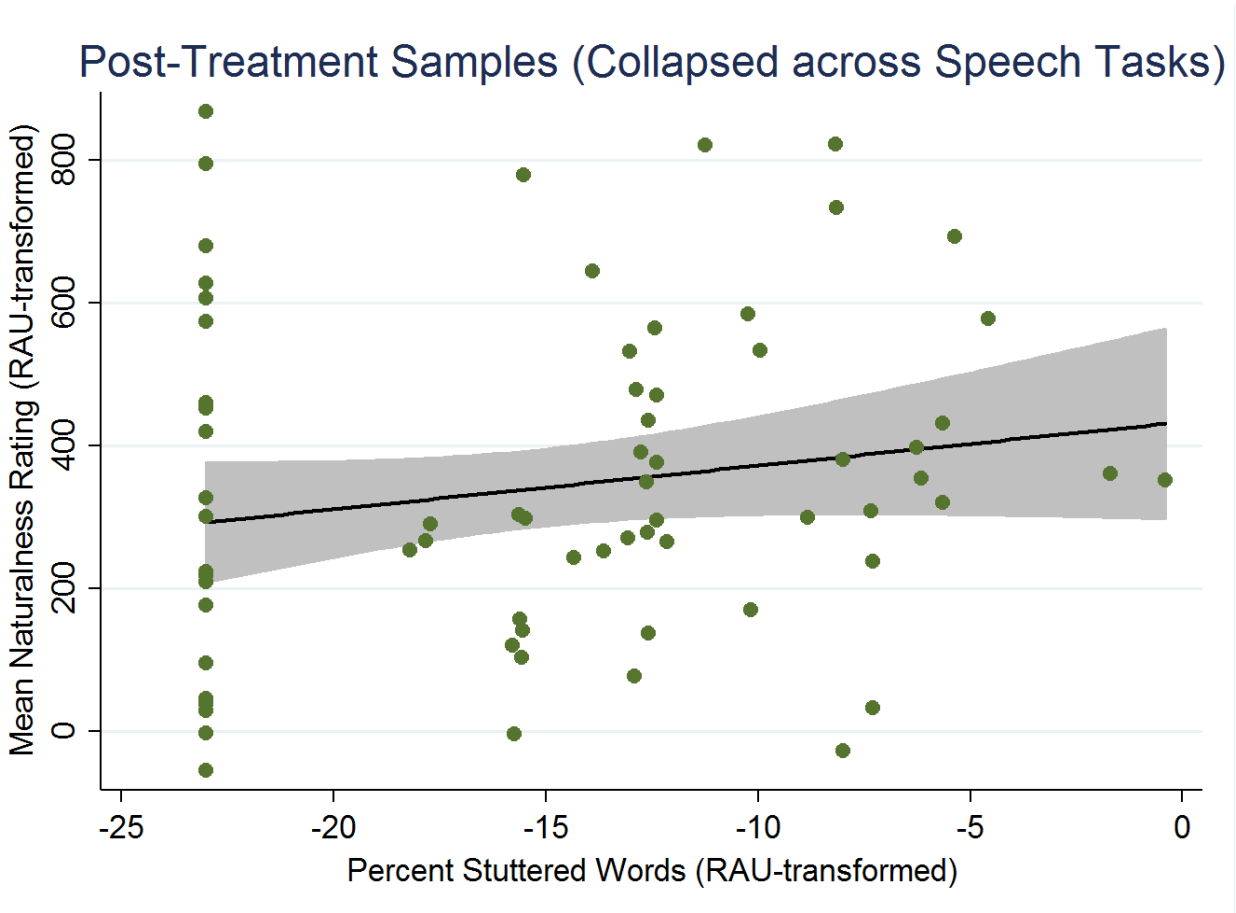

Figure 6. Scatterplot with (RAU-transformed) percent stuttered words on the horizontal axis and mean naturalness ratings (RAU-transformed) on the vertical axis. These sets of data are for post-treatment stimuli, collapsed across speech tasks. The plot shows no clear association between the two variables (Adjusted R^2: 0.02, $F(1, 70)$: 2.3, *NS)*. The solid black line represents the line of best fit and the shaded grey area shows the 95% confidence interval about the line.

Relations between speech naturalness and selected acoustics measures

As described in the Methods, Jessen (2016) performed an acoustic analysis on many of the post-treatment oral reading samples that were used in the current study. Measures of articulation rate, speech rate and mean pause time from Jessen (2016) were used to predict speech naturalness within a series of linear regression models. Scatterplots associating speech naturalness ratings with articulation rate and mean pause duration are found in Figures 7 and 8 respectively. Note in Figure 7, there is a negative, linear association between articulation rate and speech naturalness. Linear regression analysis revealed a significant linear model fit (F(1, 27): 13.25, p = 0.0011) that accounts for almost one third of the data variance (Adjusted R^2: 0.30). The negative association indicates that the less natural sounding the speech samples, the slower the articulation rate. A similar finding was observed for speech rate (which includes both articulation time and pause time). While not shown graphically, the linear model was nearly identical to the results for articulation rate (Adjusted R^2: 0.28, F(1, 27): 12.07, p = 0.0017).

Speech naturalness ratings were also plotted against the mean pause durations within the oral reading samples (Figure 8). While the data scatter is not nearly as tight as that observed in Figure 5 or even Figure 7, there is a clear positive association between the two measures. A significant linear regression model fit was also found (F(1,27): 5.78; p = 0.023). However, this speech pause duration only accounted for about 15 percent of the variance in the speech naturalness ratings. The direction of the association indicates that the most unnatural sunding speech was associated with longer pauses between utterances.

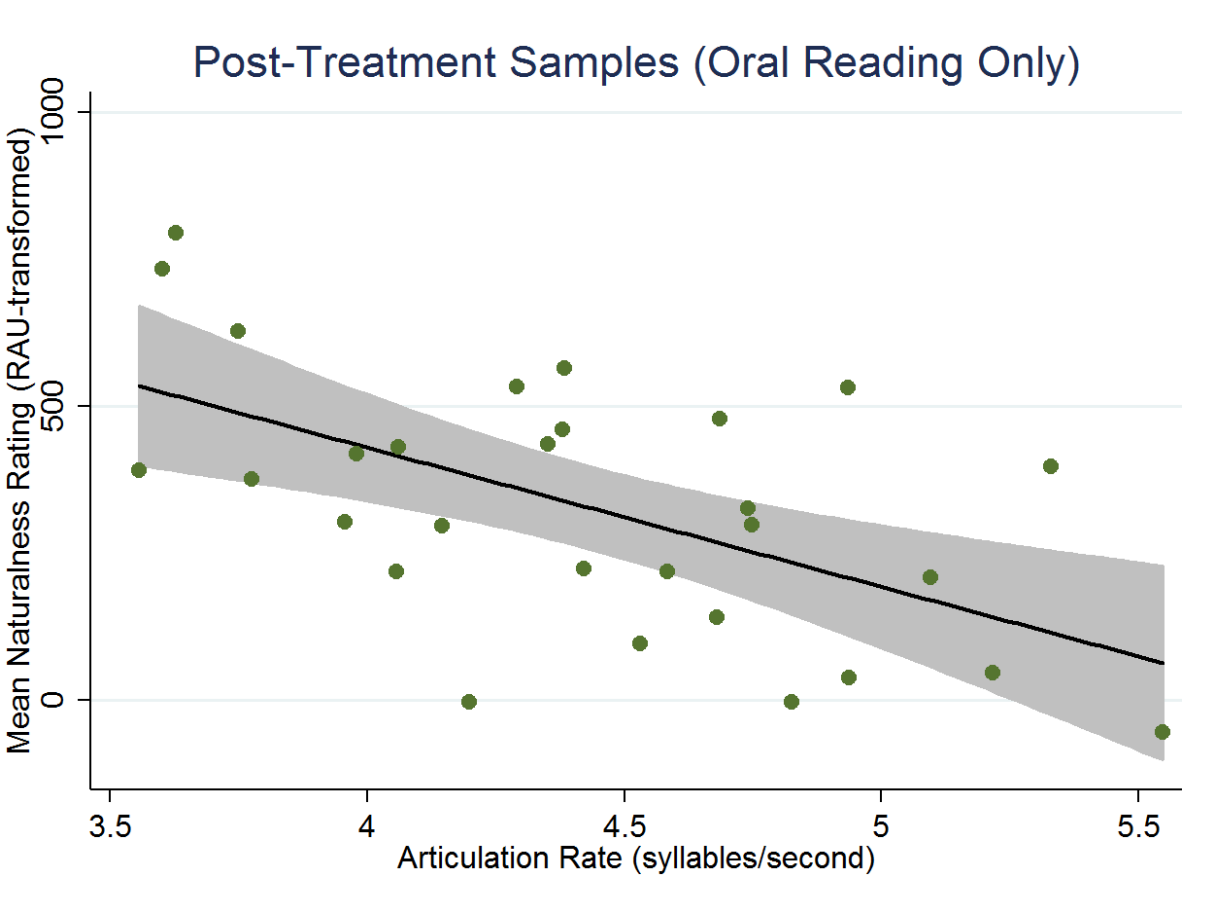

Figure 7. Scatterplot with articulation rate (syllables/second) on the horizontal axis and mean naturalness ratings (RAU-transformed) on the vertical axis. These sets of data are for post-treatment recordings of the oral reading sample. The plot shows a clear negative association between articulation rate and speech naturalness (Adjusted R^2: 0.30, $F(1, 27)$: 13.25, $p = 0.0011$). The solid black line represents the line of best fit and the shaded grey area shows the 95% confidence interval about the line.

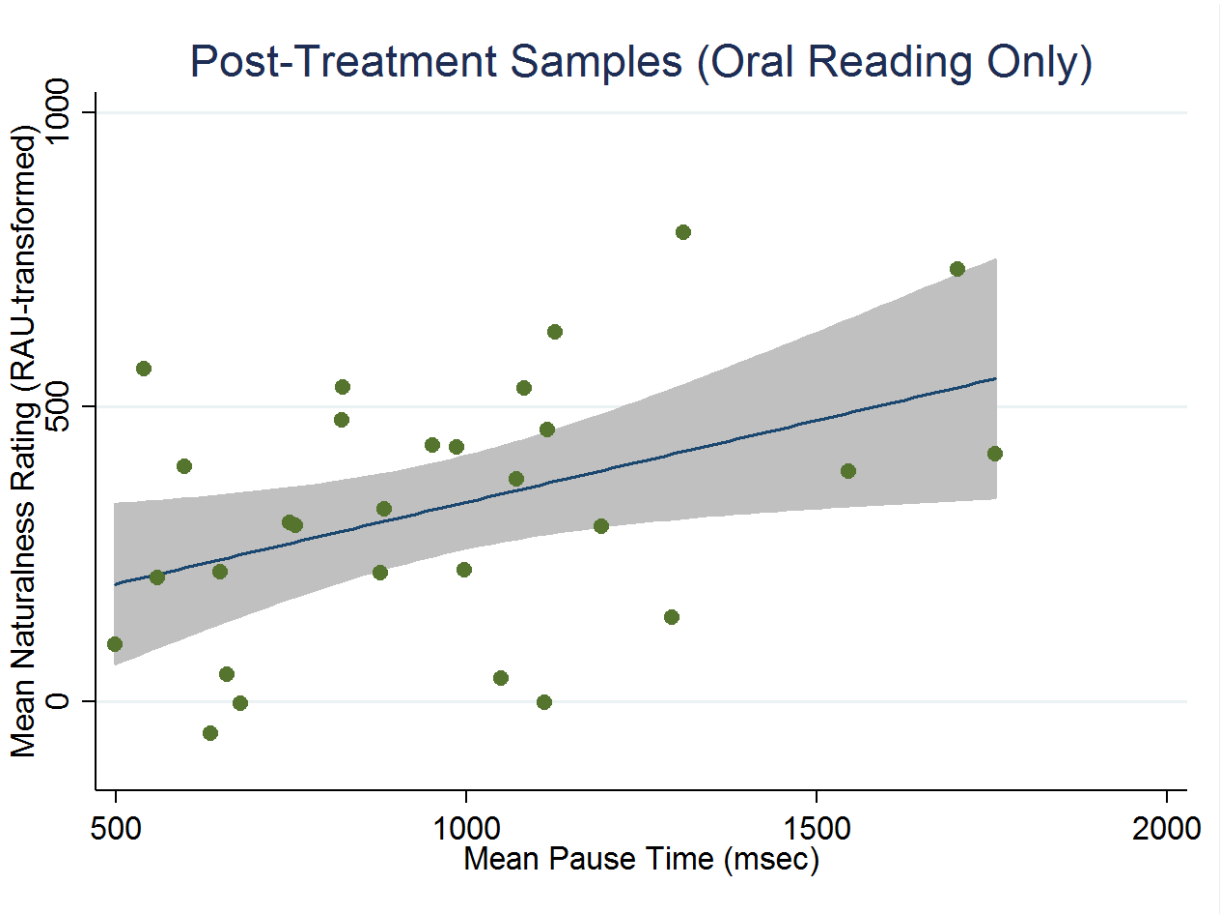

Figure 8. Scatterplot with mean speech pause time (msec) on the horizontal axis and mean naturalness ratings (RAU-transformed) on the vertical axis. These sets of data are for post-treatment recordings of the oral reading sample. The plot shows a clear negative association between articulation rate and speech naturalness (Adjusted R^2: 0.15, $F(1, 27)$: 5.78, $p = 0.023$). The solid black line represents the line of best fit and the shaded grey area shows the 95% confidence interval about the line.

Discussion

The naturalness of speech produced by persons who stutter can be influenced by a number of possible variables. This study found that correlates of speech naturalness were different for pre and post-treatment speech samples. For the pre-treatment speech samples, stuttering frequency was a very strong predictor of speech naturalness. Almost two-thirds of the variance in the speech naturalness ratings was accounted for by this single variable. For post-treatment data, stuttering frequency served no predictive value. Much of this may be attributed to the fact that stuttering was largely absent in the post-treatment speech samples. A variable cannot be predictive if it is not present. That does not mean that the post-treatment samples were generally judged as sounding natural to listeners. As a group, the post-treatment samples were rated more natural than the pre-treatment samples. But a quick examination of Figures 4, 6, 7, and 8 reveals that many individual post-treatment samples were rated as highly unnatural. This is somewhat concerning given that speech treatment programs aspire to help a person communicate in a way that does not create negative attention from the listener. While the samples contained little stuttering, listeners judged them unnatural. What was it about the post-treatment samples that negatively influenced the judges' ratings?

Articulation rate, speech rate and mean pause duration were significant predicters of post-treatment naturalness scores. This finding is consistent with Metz et al. (1990) who found that speech naturalness ratings were positively correlated with temporal measures of speech including voice onset time and vowel duration. The results suggest that the slower a person speaks, and the larger the pauses between words and phrases, the more unnatural the speech will sound. While the treatment program under study did not explicitly alter speech rate, it is likely the slower speech rate and longer pause times were a consequence of employing the new speech

targets as well as individual proficiency in the speech skills. There are stuttering treatment programs that purposely slow articulation rate to enhance fluency. It is likely that those programs could also put its participants at risk for having reduced post-treatment speech naturalness. Regardless of the details of the treatment program, the analytic approach used in the current study may be of use to help refine treatment programs to optimize both fluent and natural sounding speech.

One aim of the study was to analyze the effect of speech task on speech naturalness. In this experiment, oral reading and a monologue task were compared. Results indicated that speech naturalness ratings were minimally influenced by task. This finding has some practical significance in that the speech task choice may not have substantive effects on study results.

Finally, a secondary aim of the study was to assess the use of a visual analog scale when rating speech naturalness. There was excellent rater agreement and a strong positive correlation with the previously collected EAI scale data. This suggests that a visual analog scale can be a viable scaling method for assessing speech naturalness.

Study Limitations and Future Directions

One potential limitation of the study was the participant pool who served as judges. All were undergraduate and graduate speech language pathology students. While having some background in speech language pathology was not listed in the inclusion criteria, recruitment efforts lead to such a result. It is likely that the participants who served as judges have a level of knowledge and experience that might influence the outcomes. How that experience would affect the results is not clear. One might expect judges with less knowledge and experience would be either less reliable, less critical in their naturalness ratings, or both. While Onslow et al. (1992)

indicated differences in rater reliability between sophisticated and unsophisticated listeners, the differences were not large. With regards to rating bias by rater group, Teshima et al. (2010) noted that students studying speech language pathology were actually less critical as compared to persons who stutter or judges with little knowledge of stuttering. This study is unable to address these issues since only one listener group was used. The WRAMC-WMU stuttering database is a unique resource that could be used to assess how listener characteristics influence naturalness ratings.

A second potential limitation of the study was that speech samples were only available from a single stuttering treatment program from a specific clinical site. There are many different types of fluency shaping treatments, and we do not know if the naturalness scores reported would be consistent across programs. Further exploration of different types of fluency shaping and stuttering modification programs would be beneficial in elucidating the various factors that can contribute to unnatural sounding speech following therapy.

A third potential limitation relates to when the speech sample recordings were made. When post-treatment recordings were made, the program participants had just finished one month of a very intense therapeutic experience. It is possible that speech samples collected at a later time might be a better representation of post-treatment speech patterns. As outlined in the Introduction, Teshima et al. (2010) found that speech patterns were rated as more natural for recordings taken in long term follow up as compared to immediately after therapy. Future studies that use long term follow up should assess speech naturalness as part of their outcome measures.

A fourth limitation is that the association between speech naturalness and acoustic measures was based on very narrow subset of data. For example, when reporting on the

relationship between speech naturalness scores and articulation rate, only oral reading videos for post-treatment samples were used. Using a wider range of acoustic measures across all tasks and recording periods would allow for a more thorough analysis factors that predict speech naturalness.

Finally, clinicians may benefit from looking at the use of speech naturalness scores when implementing a treatment program. One aim of stuttering treatment programs is to reduce the stuttering severity, but it is also to make speech sound natural. One way this could be achieved is the use of speech naturalness to see if the treatment is giving the desired results.

References

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

Chang, S., Erickson, K. I., Ambrose, N. G., Hasegawa-Johnson, M. A., & Ludlow, C. L. (2008). Brain anatomy differences in childhood stuttering. *NeuroImage, 39*(3), 1333-1344.

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

Hillenbrand, J.H. (2015) Phonetics Exercises Using the Alvin Experiment-Control Software. JSLHR, 58, 171–184.

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

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

Martin, R. R., & Haroldson, S. K. (1992). Stuttering and Speech Naturalness. *Journal of Speech Language and Hearing Research, 35(3),* 521.

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*(3), 516.

Onslow, M., Hayes, B., Hutchins, L., & Newman, D. (1992). Speech Naturalness and Prolonged-Speech Treatments for Stuttering. *Journal of Speech Language and Hearing Research, 35*(2), 274.

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.

Pravesh, A. & Geetha Y. (2012). Speech Naturalness of Recovered and Relapsed Persons with Stuttering Following Treatment. *Language in India, 12*(12), 1-27.

Raza, M. H., Gertz, E. M., Mundorff, J., Lukong, J., Kuster, J., Schäffer, A. A., & Drayna, D. (2012). Linkage analysis of a large African family segregating stuttering suggests polygenic inheritance. Human Genetics, 132(4), 385-396.

Runyan, C.M., Bell, J.N. & Prosek, R.A. (1990). Speech naturalness ratings of treated stutterers. *Journal of Speech and Hearing Disorders*, 55, 434-438.

Schaeffer, N., & Eichorn, N. (2001). The effects of differential vowel prolongations on perceptions of speech naturalness. *Journal of Fluency Disorders, 26*(4), 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.

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*(1), 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*(1), 44-58.

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.: PrenticeHall.

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.