DETECTING STRESS AND DEPRESSION IN ADULTS WITH APHASIA THROUGH SPEECH ANALYSIS

Stephanie Gillespie[1], Elliot Moore[1], Jacqueline Laures-Gore[2], Matthew Farina[2], Scott Russell[3], Yash-Yee Logan[1]

[1]Georgia Institute of Technology, [2]Georgia State University, [3]Grady Memorial Hospital

ABSTRACT

Aphasia is an acquired communication disorder often resulting from stroke that can impact quality of life and may lead to high levels of stress and depression. Depression diagnosis in this population is often completed through subjective caregiver questionnaires. Stress diagnostic tests have not been modified for language difficulties. This work proposes to use speech analysis as an objective measure of stress and depression in patients with aphasia.

Preliminary analysis used linear support vector regression models to predict depression scores and stress scores for a total of 19 and 18 participants respectively. Teager Energy Operator- Amplitude Modulation features performed the best in predicting the Perceived Stress Scale score based on various measures. The complications of speech in people with aphasia are examined and indicate the need for future work on this understudied population.

Index Terms— Depression detection, stress detection, speech analysis, aphasia, stroke

1. INTRODUCTION

Aphasia is a communication disorder often resulting from stroke which can impair an individual's ability to read, write, comprehend auditory dialogue, and express himself/herself verbally. An estimated two million people in the United States suffer from aphasia, with nearly 180,000 acquiring the disorder each year [1]. Due to the difficulty with language skills, individuals living with aphasia may be under considerable stress [2, 3] which can be a factor associated with increased risk of depression. However, work by Laures-Gore and DeFife indicated that most studies of post-stroke depression exclude adults living with aphasia due to comprehension and expression disabilities that many questionnaires cannot accommodate [4].

Assessing stress and depression in persons living with aphasia is a challenging task. The Perceived Stress Scale (PSS) is a 14-item questionnaire designed to be completed by an individual to assess the degree of stress they perceive to be associated with their current life circumstances [5]. While the PSS has been used in studies on those with aphasia [4], it was not designed specifically for aphasic populations. Therefore, administration of the PSS generally requires the assistance of a caregiver or interviewer who reads the questionnaire aloud as the person with aphasia selects responses. In contrast, the Stroke Aphasia Depression Questionnaire (SADQ-10) was developed to assess depressed mood in individuals with aphasia [6]. The questionnaire is completed by a caregiver who rates 10 depression-associated behaviors based on frequency with higher scores indicating increased presence of depressive symptoms.

While there is a large body of work in speech signal processing on stress and depression, little has addressed the population living with aphasia due to limited access to this group, as well as the challenge of collecting speech. The work presented in this article builds on previous work by the authors [7] as part of ongoing research to investigate various methodologies for utilizing speech analysis in detection and classification of stress and depression in persons with aphasia. While the work in [7] presented results of a depression classification based on support vector machines (SVM), the current work constructs a linear support-vector regression (linear-SVR) model in an effort to predict SADQ-10 and PSS scores from speech in a database collected for this study.

2. APHASIA DATABASE

Between spring 2014 and summer 2015, data was collected from participants who were at least one month post-stroke and presented symptoms of aphasia. Interviews based on the Western Aphasia Battery-Revised (WAB-R) [8] were conducted at the Georgia State University Aphasia and Motor Speech Disorders Laboratory. The severity and type of aphasia for each participant is determined from the Aphasia Quotient (AQ) computed in a range from 0-100 with a score higher than 93.8 indicative of person with no discernible aphasia. At present, WAB-R interviews have been collected for a total of 26 participants. Of these, 19 participants were selected for analysis by regression of their SADQ-10 scores and 18 were selected for analysis by regression on their PSS (one participant was not able to complete the PSS and was excluded from analysis). A

Table 1: Summary of demographic and clinical information for the PSS and SADQ-10 analysis (*one of the males was excluded from the PSS study since he could not complete the PSS questionnaire)

# Males	12*
# Females	7
Age Range	31-70
AQ	31.9-99.4
# with AQ>93.8	2
SADQ-10 score	6-25
PSS score	14-40

summary of the data used in this analysis is presented in Table 1.

The aphasia quotients of the participants ranged from 31.9 (Broca's Aphasia) to 99.4 (no Aphasia). However, only 2 participants were above the threshold of 93.8 indicating no discernible aphasia. Speech was recorded using a AKG C520 headset condenser microphone and sampled at 16kHz. Speech data consisted of a variety of spontaneous speaking, sentence completion, and word/object identification tasks. The amount of speech available varied across participants, which was expected due to the nature of aphasia as a communication disorder. Therefore, the analysis for this article was restricted to approximately eighteen sentences of spontaneous speech per participant to balance the data distribution.

Detection of depression in individuals with aphasia relies on caregiver-based questionnaires to overcome the language difficulties associated with aphasia. The community stroke aphasia depression questionnaire-10 (SADQ-10) is a ten-question survey requiring caregivers to assess the frequency of specific behaviors of the participant, ranging from "never" (0) to "always" (3) [6]. Participant SADQ-10 scores ranged from 6-25. While a SADQ-10 score of 14 or above has been considered the threshold for classifying a person as depressed [9], the exact classification of depression was not a part of this regression study.

The Perceived Stress Scale (PSS) is a fourteen-question assessment to determine the degree to which situations in an individual's life were considered stressful by that individual [5]. For each item, a user can respond with never (item score of 0) to very often (4), for a total of 56 points. The PSS was not designed with severity thresholds or categories, nor can the authors find any published study validating proposed categories. However, some sources that publish the PSS-10 (a 10-question version of the test with max score of 40) advertise the severities as low stress (0-13), moderate stress (14-26), and high stress (27-40) [10, 11]. The participant PSS scores ranged from 14-40 with higher scores indicating higher levels of perceived stress. The range of values for PSS and SADQ-10 served as the targets for the linear-SVR model based on the speech features discussed in Section 3.

3. SPEECH ANALYSIS FOR STRESS AND DEPRESSION

3.1. Previous work in stress, depression, and aphasia

A recent review article by Cummins et al. [12] summarized speech analysis in depression and suicide risk over the last 10 years, including meta-analysis on vocal features as they relate to diagnosis and classification of depression. Much of the work on stress detection uses samples of short-term stress (e.g. the SUSAS database [13]) instead of clinical stress. As such, there is a lack of work understanding the detection of long-term stress in vocal acoustics. The features chosen in this work have all been traditionally used for their ability to predict stress, depression, or emotional state.

The majority of aphasia research from a speech-processing perspective has been limited to analysis at the phoneme level and is often used to diagnose aphasia itself, not the effect of stress and depression associated with living with aphasia. Le et al. developed automatic speech intelligibility tracking for patients with aphasia [14], and others have focused on diagnosis of aphasia subtypes [15].

3.2. Feature extraction and selection

Prosodic, spectral, TEO and glottal features were extracted from the voiced sections of speech, with low-level descriptors (LLD) statistics calculated at the sentence level as described in openSMILE [16]. LLDs of various prosodic and spectral features originally described in [7] were calculated, in addition to the following:

- Teager Energy Operator (TEO) features including:
    - Amplitude modulation
    - Frequency modulation
    - 16 critical band areas [17]
    - RMS-Energy
    - Log-Energy
- Glottal features including
    - H1-H2 [18]
    - Parabolic spectrum parameter (PSP) [19]
    - Harmonic richness factor (HRF) [20]
    - 18 glottal waveform time parameters based off of work by Torres et al. [21], identified as "GLTP" in this work

A total of 1596 low-level statistics of the features were extracted in MATLAB for each sentence of each participant and normalized using Z-normalization across each individual feature. Feature selection on the full data set as well as each individual feature type grouping was performed first by removing any features with a correlation greater than 0.75, and then using 10-fold cross-validation sequential feature selection to reduce the size of the feature subsets. Only those features that were selected were used to train and

Table 2: SADQ-10 regression results by feature subtype after feature selection. MAE= Mean Absolute Error with respect to SADQ-10 standard deviation, R2=R-Squared Score, P1SD= Percentage of predictions within one SADQ-10 standard deviation from the actual value.

Feature Type	MAE (SADQ-σ)	R2	P1SD (%)
All	1.24	0.04	46.0%
Pitch + Jitter	1.08	**0.34**	48.4%
RMS-Energy	1.05	0.05	52.3%
LSF + Δ	1.11	0.07	53.0%
MFCC + Δ	1.25	0.04	47.7%
HNR	**0.97**	0.15	**57.5%**
CPP	1.03	0.13	54.4%
TEO-All	1.15	0.03	48.4%
TEO-AM	1.04	0.13	56.8%
TEO-FM	**0.91**	0.00	**62.1%**
TEO-CBarea	1.06	0.03	53.7%
TEO-RMS Energy	1.04	0.00	54.0%
TEO-log Energy	1.04	0.14	50.2%
Glottal-All	1.16	0.12	49.5%
H1-H2	1.16	0.12	49.5%
PSP	1.06	**0.44**	51.9%
HRF	1.07	0.17	51.9%
GLTP	1.21	0.18	44.9%

test the feature-subset models built using the Support Vector Regression function in Matlab.

3.3. Regression as model selection

The SADQ-10 and PSS are scored on numeric scales (0-30 and 0-56 respectively) and do not have multiple thresholds representing degrees of severity. Leeds et al. [9] determined a SADQ-10 threshold of 14 as a clinical threshold for the manifestation of depressive symptoms. However, it is difficult to determine how SADQ scores within a range of 1 or 2 points should be interpreted for distinct degrees of depression. In previous work using SVM classification to detect depression in patients with aphasia [7], participants were labeled as depressed or not-depressed based on their SADQ-10 score and the SADQ-10 proposed clinical threshold [9]. Classification by Cepstral Peak Prominence and MFCC + delta feature subsets performed the best with respect to sentence-level precision, recall, and accuracy. Two of the three participants with below 50% accuracy had SADQ-10 scores near the threshold of 14 in the range of 13-16. Cummins et al. [12] recommended excluding any participants that score in the moderate categories of a depression scale due to the ordinal nature of mental state scales. While we agree that such an exclusion represents an ideal circumstance, it is our experience that a non-trivial amount of participants will fall in the "moderate" score range, as evidenced by this dataset which contains 7 of 19 participants within ± 2 of the threshold of 14. As a result, this work will explore the use of a linear support-vector regression model as a predictor of SADQ-10 and PSS values within the current dataset. A linear regression model was examined in [22] on a non-aphasia database containing ratings based on the Hamilton Depression Rating Scale (HDRS) [23], which is not designed for persons with aphasia.

To assess the proposed regression tasks, our outcomes are reported with respect to mean absolute error (MAE), the R-squared coefficient of determination (R2), and percentage of predicted scores within one standard deviation of the true score (P1SD) [22]. Mean Absolute Error is the average difference between each actual and predicted score, and represents the measure of how close the predicted scores are to the clinical score of interest (SADQ-10 or PSS). The MAE is reported with respect to the sample standard deviation of the diagnostic (SADQ-10 or PSS) in our study by dividing the absolute MAE by the standard deviation of the SADQ-10 (SADQ-σ) and PSS (PSS-σ) scores. R-Squared is used to determine to what extent the variation in the predicted values is determined linearly by the variation in the dependent variable (either SADQ-10 or PSS). P1SD is borrowed from previous work using regression with depression with the HDRS [22] to detect how many predictions were close to the true value.

3.4. Regression results for SADQ-10 scores

SADQ-10 scores were predicted for a total of 19 participants (including 2 who had WAB scores that indicated they were non-aphasic). Within our analysis, the SADQ-10 average score was 14.63 and sample standard deviation was 4.97. MAE, R2, and P1SD scores are shown in Table 2 for the various feature types considered in the linear-SVR on the SADQ-10 scores. The MAE results suggest that the average predicted score values were approximately one standard deviation from their true values with HNR and TEO-FM feature subsets providing predictions at slightly less than one standard deviation from the true values. This result is additionally noted in the P1SD scores where HNR and TEO_FM feature subsets were 57.5% and 62.1%, respectively. The R2 scores are too close to zero to indicate any significant linear dependency between the predicted values and the true SADQ-10 scores for the HNR and TEO-FM features. Pitch + Jitter and the H1-H2 glottal feature subset exhibited the highest R2 of 0.34 and 0.44, respectively.

3.5. Regression results for PSS scores

PSS scores were predicted for a total of 18 participants. Within our population, the PSS mean was 28.55 with a standard deviation of 7.31. Results in Table 3 show TEO-AM performed the best according to the measures considered in the study, with a relative MAE less than 1 standard deviation of the sample PSS, the second-highest R2

Table 3: PSS regression results by feature subtype after feature selection. MAE= Mean Absolute Error, units are with respect to PSS-σ, R2=R-Squared Score, P1SD= Percentage of predictions within one PSS-σ from the actual value

Feature Type	MAE (PSS-σ)	R2	P1SD (%)
All	1.51	0.12	36.8%
Pitch + Jitter	1.05	0.18	54.4%
RMS-Energy	**0.94**	0.12	**61.4%**
LSF + Δ	1.17	0.11	43.9%
MFCC + Δ	1.33	0.11	43.9%
HNR	1.02	0.15	55.8%
CPP	0.97	0.07	54.7%
TEO-All	1.01	0.00	55.3%
TEO-AM	**0.94**	**0.29**	**61.1%**
TEO-FM	1.06	0.25	53.2%
TEO-CBarea	0.97	0.00	55.0%
TEO-RMS Energy	1.01	**0.40**	56.1%
TEO-log Energy	0.98	0.09	59.4%
Glottal-All	1.05	0.02	52.9%
H1-H2	1.05	0.02	52.9%
PSP	1.01	0.24	52.9%
HRF	**0.89**	0.00	59.9%
GLTP	1.00	0.01	53.5%

score of 0.29, and 61.1% of sentences predicted PSS within one standard deviation of the true PSS score. Further analysis of the R2 score of the TEO-AM feature subset results indicated a negative correlation between the predicted PSS scores and the true PSS scores, as can be seen in Figure 1. These results indicate the linear-SVR model was not sufficient to capture the complexities of the PSS and SADQ scores from the aphasia database.

4. DISCUSSION AND CONCLUSION

A correlation analysis was completed to determine any correlations between participants' clinical or demographic information and the prediction median, mean, standard deviation, IQR, and/or accuracy. For the top performing PSS feature subset (TEO-AM), no statistically significant correlation was found between the calculated measures and any of the clinical or demographic information. Similarly, no other feature subset results correlated significantly between the predictions and any clinical or demographic information available for analysis. The lack of significant correlations indicates that the models do not appear to perform significantly better or worse due to any specific trait of the participants. Instead, the linear-SVR performance is likely hindered by the complexities of the diverse population from which speech was collected.

There are three main challenges to working with speech from participants with aphasia: 1) the database size is limited by eligible participants and their caregivers who are willing and able to successfully completed the entire data-collection process and clinical diagnosis procedure, 2) the

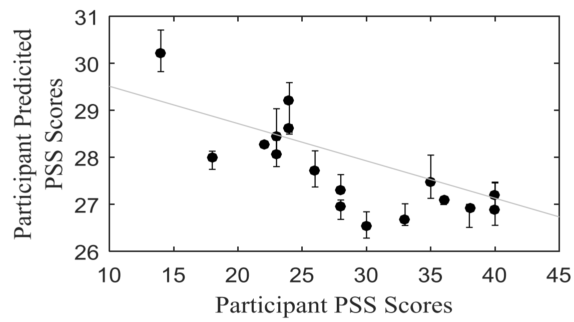

Figure 1: 1st, 2nd, and 3rd quartiles and linear regression line for prediction of PSS scores for TEO-AM

diverse clinical and demographic conditions presented by those who chose to participate, and 3) the limited "snap-shot" available of a participants mood during a single recording session. The participants in this study vary greatly with respect to age and gender. Many participants with aphasia are also diagnosed with dysarthria or apraxia, motor disorders impacting the speech produced. As an additional complication, the PSS and SADQ-10 scores are based off of the participants' emotional state throughout the past month; it is often possible for a participant who is depressed or stressed to appear or be happy or calm during the interview time period which could change and impact the emotional content of their speech.

The lack of conclusive results in this study is indicative of a need for further analysis within this complicated population that has often been excluded from larger studies due to the many uncertainties associated with how affect would manifest itself alongside a communication disorder. It is clear that while regression may be a better solution than classification due to clinical interests and current diagnostic test scores of PSS and SADQ-10, the simple linear support vector regression model is not advanced enough to handle the characteristics of affect in the aphasic population without more data taken from a multiple-interview setup. This exploratory work with aphasia will continue to look for appropriate models, features, and preprocessing strategies to handle the complexities of detecting affect from speech in those living with aphasia.

5. ACKNOWLEGEMENTS

Supported by the Emory-Georgia Institute of Technology Healthcare Innovation Program and the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR000454. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship, Grant No. DGE-1148903. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Health or the National Science Foundation.

6. REFERENCES

[1] National Aphasia Association. (September 10, 2016). *Aphasia FAQs*. Available: http://www.aphasia.org/aphasia-faqs/

[2] C. Code, G. Hemsley, and M. Haerrmann, "The Emotional Impact of Aphasia," *Seminars in Speech and Language,* vol. 20, pp. 19-31, 1999.

[3] L. Murray and H. Ray, "A Comparison of Relaxation Training and Syntax Stimulation for Chronic Nonfluent Aphasia," *Journal of Communication Disorders,* vol. 34, pp. 87-113, 2001.

[4] J. S. Laures-Gore and L. C. DeFife, "Perceived Stress and Depression in Left and Right Hemisphere Post-Stroke Patients," *Neuropsychological Rehabilitation,* vol. 23, pp. 783-797, 2013.

[5] S. Cohen, T. Kamarck, and R. Mermelstein, "A Global Measure of Perceived Stress," *Journal of Health and Social Behavior,* vol. 24, pp. 385-396, 1983.

[6] L. M. Sutcliffe and N. B. Lincoln, "The Assessment of Depression in Aphasic Stroke Patients: the Development of the Stroke Aphasic Depression Questionaire," *Clinical Rehablitation,* vol. 12, pp. 506-513, 1998.

[7] S. Gillespie, E. Moore II, J. S. Laures-Gore, and M. Farina, "Exploratory Analysis of Speech Features Related to Depression in Adults with Aphasia," in *41st IEEE International Conference on Acoustics, Speech, and Signal Processing*, Shanghai, China, 2016, pp. 5185-5189.

[8] A. Kertesz, *Western aphasia battery-revised (WAB-R)*: Pearson, 2006.

[9] L. Leeds, R. J. Meara, and J. P. Hobson, "The Utility of the Stroke Aphasia Depression Questionaire (SADQ) in a Stroke Rehabilitation Unit," *Clinical Rehablitation,* vol. 18, pp. 228-231, 2004.

[10] New York State United Teachers Union. (September 1, 2016). *Stress Assessments: Perceived Stress Scale*. Available: http://www.nysut.org/~/media/files/nysut/resources/2013/april/social-services/socialservices_stressassessments.pdf?la=en

[11] New Hampshire Department of Administrative Services. (September 1, 2016). *Perceived Stress Scale*. Available: http://das.nh.gov/wellness/Docs%5CPercieved%20Stress%20Scale.pdf

[12] N. Cummins, S. Scherer, J. Krajewski, S. Schneider, J. Epps, and T. F. Quatieri, "A Review of Depression and Suicide Risk Assessment using Speech Analysis," *Speech Communication,* vol. 71, pp. 10-49, 2015.

[13] B. D. Womack and J. H. L. Hansen, "Stress Independent Robust HMM Speech Recognition Using Neural Network Stress Classification," in *4th European Conference on Speech, Communication, and Technology (EUROSPEECH '95')*, Madrid, Spain, 1995, pp. 1999-2002.

[14] D. Le, K. Licata, C. Persad, and E. Mower Provost, "Automatic Assessment of Speech Intelligibility for Individuals with Aphasia," *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* vol. 24, pp. 2187-2199, 2016.

[15] K. C. Fraser, F. Rudzicz, and E. Rochon, "Using Text and Acoustic Features to Diagnose Progressive Aphasia and Its Subtypes," in *14th Annual Conference of the International Speech Communication Association*, Lyon, France, 2013, pp. 2177-2181.

[16] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent Developments in OpenSMILE, the Munich Open-Source Multimedia Feature Extractor," in *Proceedings of ACM Multimedia (MM)*, Barcelona, Spain, 2013, pp. 835-838.

[17] G. Zhou, J. H. L. Hansen, and J. F. Kaiser, "Nonlinear Feature Based Classification of Speech Under Stress," *IEEE Transactions on Speech and Audio Processing,* vol. 9, pp. 201-216, 2001.

[18] I. R. Titze and J. Sundberg, "Vocal Intensity in Speakers and Singers," *Journal of the Acoustical Society of America,* vol. 91, pp. 2936-2946, 1992.

[19] P. Alku, H. Strik, and E. Vilkman, "Parabolic Spectral Paramater- A New Method for Quanitification of the Glottal Flow," *Speech Communication,* vol. 22, pp. 67-79, 1997.

[20] D. G. Childers and C. K. Lee, "Vocal Quality Factors: Analysis, Synthesis, and Perception," *Journal of the Acoustical Society of America,* vol. 90, pp. 2394-2410, 1991.

[21] J. F. Torres, E. Moore II, and E. Bryant, "A study of glottal waveform features for deceptive speech classification," presented at the 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing, Las Vegas, NV, USA, 2008.

[22] N. Wahidah Hashim, M. Wilkes, R. Salomon, J. Meggs, and D. J. France, "Evaluation of Voice Acoustics as Predictors of Clinical Depression Scores," *Journal of Voice,* vol. In Press, 2016.

[23] M. Hamilton, "A Rating Scale for Depression," *Journal of Neurology, Neurosurgery and Psychiatry,* vol. 23, pp. 56-62, 1960.