EXPLORATORY ANALYSIS OF SPEECH FEATURES RELATED TO DEPRESSION IN ADULTS WITH APHASIA

Stephanie Gillespie[1], Elliot Moore[1], Jacqueline Laures-Gore[2], Matthew Farina[2]

[1] Georgia Institute of Technology, [2] Georgia State University

ABSTRACT

Aphasia is an acquired communication disorder resulting from brain damage and impairs an individual's ability to use, produce, and comprehend language. Loss of communication skills can be stressful and may result in depression, yet most depression diagnostic tools are designed for adults without aphasia. This paper discusses preliminary results from a research effort to examine acoustic profiles of adults with aphasia who have been assessed as having possible depression versus those who assessment suggests they are not depressed based on tools completed by their caretakers. This study analyzes prosodic and spectral features in 14 participants (7 assessed as having possible depression and 7 whose assessment does not suggest depression). The results showed using Cepstral Peak Prominence provided the best overall performance in separating depressed and non-depressed speech among adults with aphasia.

***Index Terms*—** aphasia, depression, speech analysis, prosodic features

1. INTRODUCTION

Aphasia is an acquired communication disorder resulting from brain damage impairing an individual's ability to use, produce, and comprehend language. Often, people with aphasia have difficulty writing and/or reading in addition to their speech difficulties. There are various types of aphasia, each with their own symptoms and impact on speech. Wernicke's aphasia is a type of fluent aphasia categorized by poor comprehension, resulting in spoken jargon or nonsensical words and phrases. Broca's aphasia is a type of non-fluent aphasia categorized by numerous pauses and effortful speech. Anomic aphasia is a milder form of aphasia, often resulting in word-finding difficulties and comprehension difficulties. Conduction aphasia results in fluent speech production but poor speech repetition.

Individuals living with aphasia may be under considerable stress related to their difficulty with language skills [1,2]. Depression is one possible behavioral response to stress. Diagnosis of post-stroke depression is a controversial issue [3]. Because of the controversy, the percentage of stroke patients identified as depressed presents a rather wide range from 10-60% of stroke patients [4,5], although a recent meta-analysis indicates that one-third of stroke patients have depression [6]. The large range of depression diagnoses is indicative of the complexity of this disorder and the potential of under- or over-diagnosis in this population. However, work by Laures-Gore [7] indicated that most studies of post-stroke depression exclude adults living with aphasia due to comprehension and expression disabilities that many questionnaires cannot accommodate. Thus, an understanding of the contribution of aphasia on post-stroke depression is limited.

This paper presents preliminary results from a study designed to collect speech from adults with various types of aphasia to identify acoustic distinctions that can be correlated to the presence of depression.

2. APHASIA AND DEPRESSION DATABASE

Speech from 26 adults who were at least one-month post-onset of stroke was collected at the Aphasia and Motor Speech Disorders Laboratory at Georgia State University over a period of approximately 1 year from spring 2014 to summer 2015. Participants in the study exhibited Broca's, Wernicke's, Conduction, and Anomic aphasia as determined by the Western Aphasia Battery (WAB) [8]. The WAB also assigns an Aphasia Quotient (AQ) that assesses the severity of the subject's aphasia. The range of values for the AQ is from 0-100 (most to least severe) with a score higher than 93.8 within normal limits indicating no aphasia. After excluding participants with missing depression questionnaire scores, participants who scored as normal on the WAB, and participants who had technical difficulties in the recording process, a total of 19 participants were available for analysis. A subset of 14 participants was chosen for analysis in this paper to balance gender and depression labels.

Different care-giver assessments were completed in order to determine which participants were considered depressed. The community stroke aphasia depression questionnaire-10 (SADQ-10) was developed to assess depressed mood in individuals with aphasia [9]. A participant's score of greater than 14 is assigned a label of high depressive symptoms [10]. As one of the few depression scales available for adults with aphasia, the SADQ-10 was used to determine the potential presence of

Participant Number	Gender	SADQ Score	Label (Depressed if SADQ>14)	Aphasia Type (AQ)	Dysarthria Score	Apraxia Score
15	Female	12	Not Depressed	Wernicke's (41)	None	Moderate-Severe
20	Female	12	Not Depressed	Conduction (54.6)	None	Moderate
24	Female	11	Not Depressed	Broca's (68.5)	Mild	Moderate
8	Female	19	Depressed	Broca's (59.6)	None	Moderate-Severe
19	Female	25	Depressed	Anomic (88.3)	Mild	Mild
5	Female	16	Depressed	Anomic (92.2)	None	Mild
25	Male	14	Not Depressed	Wernicke's (87.4)	Mild	Mild-Moderate
7	Male	10	Not Depressed	Anomic (82.1)	Mild	Mild
9	Male	8	Not Depressed	Anomic (83.2)	None	Mild
6	Male	13	Not Depressed	Anomic (87.4)	None	Mild
11	Male	19	Depressed	Anomic (78.0)	Mild-Moderate	Mild-Moderate
18	Male	20	Depressed	Anomic (83.2)	Mild	Mild
21	Male	16	Depressed	Anomic (83.3)	None	Mild
13	Male	16	Depressed	Anomic (87.4)	Mild	None

Table 1: Participants' Depression Label, Aphasia Type (Aphasia Quotient (AQ)), Dysarthria and Apraxia assessment scores

depression. In this paper, the term 'depressed' is used as shorthand to indicate high depressive symptoms while 'not depressed' indicates low depressive symptoms. Speech was recorded at 44.1 kHz (32 bit) with an AKG C520 headset condenser microphone.

2.1. Speech Segmentation

Participants were asked to complete a series of tasks as a part of the Western Aphasia Battery, including conversational questions, a picture description, auditory and verbal comprehension in the form of yes or no questions, auditory word recognition (e.g. point to the door), sequential commands, repetition, object naming and word finding, word fluency, sentence completion, and responsive speech (e.g. "Where can you buy stamps?"). Approximately 55 responses per participant were segmented, ranging from a single word answer to an extended description. Additionally, two picture descriptions were included in an attempt to collect additional spontaneous speech. Longer responses, including those of the picture descriptions, were segmented into individual utterances based on the completion of an idea. In total, the participants had at least 3.5 minutes of responses recorded after segmentation and approximately 75 utterances. This was reduced to the first 33 utterances per person in the current analysis based on the criteria of including only phrase responses as opposed to single-words based on initial analysis comparing classification accuracies of models trained on words and models trained on phrases.

2.2. Aphasic Patient Profiles

Table 1 shows the demographics for the participants chosen for this portion of the preliminary analysis. The 14 participants were chosen to be approximately balanced in gender and depression label, while still including as many participants as possible. While the primary interest in the study is the acoustic assessment of depression, adults with aphasia may also demonstrate motor speech disorders related to their stroke, such as dysarthria and apraxia of speech. Dysarthria results from impaired movements on the lips, tongue, vocal folds, and/or diaphragm. Common speech errors due to dysarthria include distortions such as slurred speech or mumbled speech. Apraxia results from the difficulties generating motor programs for speech movement. The resulting speech often includes sound distortions, substitutions and/or omissions. Both of these motor speech disorders can affect objective measures of speech and communication and subsequently impact depression assessment through acoustics alone. Table 1 shows the participants have varying levels of dysarthria per the Frenchay Dysarthria Test [11]. All except one participant had at least a mild form of apraxia as measured by the Apraxia Battery for Adults-2 [12]. The impact of motor disorders will be analyzed in future studies, but sub-scores for each test were examined to ensure that the motor disorders minimally impacted the speech in current analysis.

3. SPEECH ANALYSIS

The detection of depression in the voice of non-aphasic speech is directly related to a large foundation of research in the study of affect (emotion, stress) in the voice. Several literature surveys over the years [13-16] have provided overviews and updates on work related to the creation and analysis of speech emotion databases. From these surveys (and other more recent work, some of which will remain uncited for the sake of brevity), it is clear that prosodic (e.g. pitch, speaking rate, etc.) and spectral features (i.e. formants, cepstral coefficients, etc.) dominate the prevailing literature in terms of feature use. While there is an extensive body of literature for the recognition of stress and depression in persons without aphasia, there is an analogous lack of such research for persons with aphasia. This lack of research is not surprising considering that aphasia affects the production and/or language coherency of persons who suffer from it. Therefore, a lot of the analysis related to speech in adults with aphasia involves the choice of words and phonemes in naming tasks or connected speech with little emphasis on the acoustic properties of the speech produced [17-21].

One recent study showed that the prosody of subjects with Broca's and Wernicke's aphasia showed statistically a significant difference in prosody stability (i.e. jitter, shimmer, etc.) during the phonation of a sustained vowel compared to speech from a non-aphasic subject [22]. While the study was not correlated with any analysis of the affective state of the subject, it does highlight that there are measurable acoustic differences that can exist between different types of aphasia that must be considered in an acoustical analysis of affective state.

3.1. Features and Classification Method

Prosodic and spectral features were extracted based on the features used in the INTERSPEECH 2009 emotion challenge [23] with some additional common features and statistics from openSMILE [24]. Specifically, low-level descriptors (LLD) statistics for the following features were calculated yielding a total of 874 features:

- Cepstral Peak Prominence (CPP), computed by the VoiceSauce application using Hillenbrand's formula [25,26]
- 4 different Harmonic-to-Noise Ratios (HNR), computed by the VoiceSauce application using de Krom's method [25,27]
- 14 mel-frequency cepstral coefficients (MFCC) and deltas, computed by the VoiceBox toolbox [28]
- Pitch, based on Sun's Subharmonic-to-harmonic ratio [29] and jitter
- 8 Line Spectral Frequencies (LSF) derived from the LPC
- Root Mean Square-Energy (RMS-Energy)

Features and their statistics were extracted in MATLAB for each utterance of each participant and WEKA [30] was used for building the classifier using the SMO-SVM (Sequential Minimal Optimization-Support Vector Machine). Feature selection on the full data set as well as each individual feature type was performed using 5-fold cross validation in Weka with the Correlation-based Feature Subset Selection algorithm [31] and the Best-First search method. Only those features that were selected by at least 3 of the 5 folds were chosen to be used to train and test the feature-subset classifiers. Restricted by the small dataset, a leave-one-out approach was used in which a training model was built on 13 participants and tested on the excluded participant, yielding 14 separate classification tasks per feature subset. Analyses were conducted based on all 874 features and each reduced subset of features grouped by the following categories: pitch+jitter, RMS-Energy, Harmonic Noise Ratio (HNR), Cepstral Peak Prominence (CPP), MFCC+delta, and LSF+delta.

Features (no. of features)	Avg. Recall	Avg. Precision	Avg. Accuracy (standard dev.)
All (874)	0.359	0.411	0.422 (0.264)
Reduced (41)	0.459	0.447	0.446 (0.325)
Pitch + Jitter (7)	0.394	0.399	0.400 (0.303)
RMS-Energy (8)	0.814	0.487	0.478 (0.478)
HNR (10)	0.545	0.472	0.468 (0.311)
CPP (6)	0.563	0.634	0.619 (0.190)
MFCC+delta (19)	0.432	0.588	0.502 (0.349)
LSF+delta (20)	0.308	0.286	0.374 (0.246)

Table 2. Classification results by feature subtype in assigning the correct depression label to each utterance. All categories except 'All' are based on the reduced feature subset after feature-selection

4. RESULTS AND DISCUSSION

Table 2 shows the precision (i.e. percentage correctly marked as depression out of all observations marked as depression by the classifier), recall (i.e., percentage correctly marked as depression out of all possible true labels of depression), and average accuracy for both the original full feature set and each set of reduced features after the feature selection took place. Precision and recall balance each other as a high recall and low precision would indicate a tendency to correctly classify more of the true labels of depression at the cost of misdiagnosing a greater number of participants. A high precision and low recall would indicate a tendency for the classifier to miss more true diagnoses of depression while being more certain of the samples marked as depression. The values in Table 2 are based on assigning the appropriate label of depressed/not-depressed based on an individual's SADQ score across their 33 utterances. A total of 462 utterances were classified across all 14 participants.

Based on the results in Table 2, the Cepstral Peak Prominence feature subset classified the best overall considering recall, precision, and accuracy. Only RMS-Energy had a higher recall, but also had a lower precision and accuracy, indicating a tendency to assign the label of depression more often resulting in misdiagnoses. Cepstral Peak Prominence has been found to be indicative of breathy voices and has been extended to an evaluation of the overall periodicity of the speech [26,32]. While the exact quantities that CPP measures are unknown, it is suggested that the CPP integrates aperiodicity and other waveform features [33]. Jitter, a similar feature that measures the noise of the signal's pitch, has been found to be statistically significant in classifying between suicidal and not-depressed participants, as well as between depressed and control participants

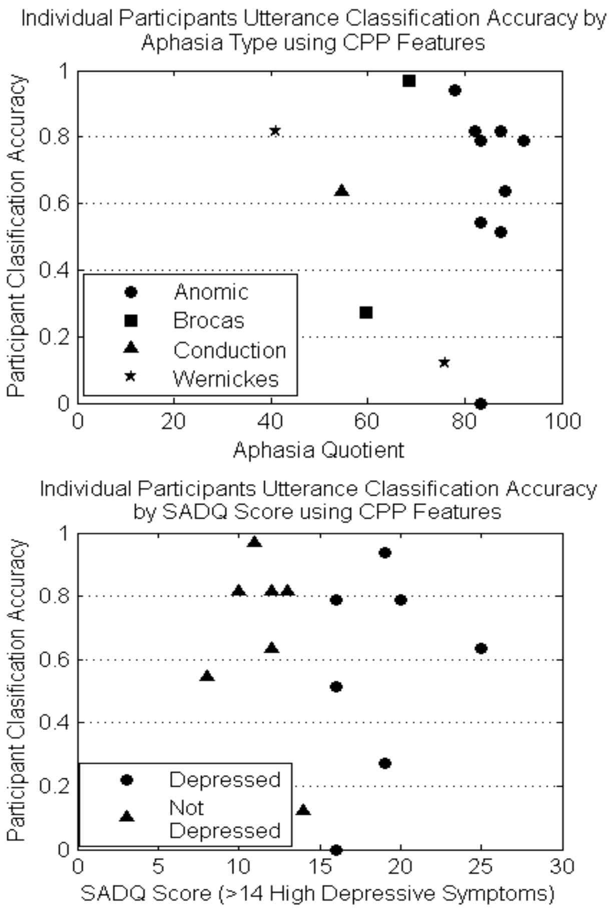

Figure 1. Classification accuracy of each participant plotted against their aphasia quotient and aphasia type (top) and against their SADQ-score and depression label (bottom).

[34,35]. Understanding the jitter and CPP of depressed aphasic speech and not-depressed aphasic speech will need to be continued in further detail to determine if this is unique to aphasic speech exhibiting depressive symptoms or if the motor disorders of dysarthria and apraxia are also contributing to these differences found in these features.

MFCC and deltas were the next highest performing group of features overall regarding accuracy and precision, while RMS-Energy had the highest recall of any feature subset. The worst performance was from the LLDs of the LSF + delta group which exhibited a classification accuracy well-below chance (~0.5). While all of the features included above have been used in the past to assess depression, they are also potentially linked to many of the communication and motor disorders experienced by adults with aphasia. Future work will continue this analysis regarding the potential impact on vocal acoustics from motor disorders, especially those measures relating to depression.

In order to determine if the classifier was actually creating a model based on depression or if it was discovering and representing the other clinical information besides depression, the classification accuracies of the Cepstral Peak Prominence subgroup were compared to the aphasia type, aphasia quotient, and SADQ numerical score of the participants. Figure 1 (top) shows the classification accuracy when assigning the label of depressed/not depressed to each of their utterances for each participant. A score of more than 50% would indicate a majority of their labels were correctly assigned according to the SADQ-depression score. 11 of the 14 participants in this study had the correct label assigned to the majority of their utterances. Figure 1 (top) shows there does not appear to be an indication that the aphasia type or aphasia quotient impacted the classifier's ability to predict the depression label.

Figure 1 (bottom) shows the classification accuracy in comparison to SADQ scores and depression label. There does not appear to be any indication that the classifier trained on the CPP features classifies either the depressed or not-depressed participants better. An interesting point is that there are three participants whose classification accuracies are below 30%. Two of these three have a SADQ score of immediately below indication of depression (14) or immediately above indication of depressed (15). As mentioned earlier in this study, the SADQ score is a caregiver's assessment from which depression may be indicated. However, it is not a precise diagnosis and it is possible that the depression label we are assuming to be a ground truth is false; the results indicate that their speech has similar features to those of the opposite class of their depression label. Future work will explore if an uncertain label can be analyzed using other clinical measures.

5. CONCLUSIONS

This research explored the difference between depressed and not-depressed speech as it relates to aphasia. While aphasic speech may manifest itself differently based on the aphasia type, depression can still be identified across aphasia types. Our best overall classifier performance utilized LLDs of Cepstral Peak Prominence. While the results of this study are limited to the relatively small dataset, they do provide an impetus for an expanded study on acoustic features that are resilient to the communication impairments exhibited by adults with aphasia, yet sensitive enough to help assess depression and other affective states.

6. ACKNOWLEDGEMENTS

A special thank you to Scott Russell. 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. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Work supported by the National Science Foundation Graduate Research Fellowship, Grant No. DGE-1148903.

7. REFERENCES

[1] C. Code *et al*., "The Emotional Impact of Aphasia," *Seminars in Speech and Language,* vol. 20, pp. 19-31, 1999.

[2] 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.

[3] S. Paolucci, "Epidemiology and Treatment of Post-Stroke Depression," *Neuropsychiatric Disease and Treatment,* vol. 4, pp. 145–154, 2008.

[4] J. Fang and Q. Cheng, "Etiological Mechanisms of Post-Stroke Depression: A Review," *Neurological Research,* vol. 31, pp. 904-909, 2009.

[5] L. Ayerbe *et al*., "Natural History, Predictors and Outcomes of Depression After Stroke: Systematic Review and Meta-Analysis," *The British Journal of Psychiatry,* vol. 202, pp. 14-21, 2013.

[6] M. L. Hackett *et al*., "Frequency of Depression After Stroke," *Stroke,* vol. 36, pp. 1330–1340, 2005.

[7] J. Laures-Gore and L. C. Defife, "Perceived Stress and Depression in Left and Right Hemisphere Stroke Patients," *Journal of Neuropsychological Rehabilitation,* to be published.

[8] A. Kertesz, *Western Aphasia Battery-Revised (WAB-R):* Pearson, 2006.

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

[10] L. Leeds *et al., "*The Utility of the Stroke Aphasia Depression Questionnaire (SADQ) in a Stroke Rehabilitation Unit," *Clinical Rehabilitation*, vol. 18, pp. 228-231, 2014.

[11] P. M. Enderby and R. Palmer, *Frenchay Dysarthria Assessment—Second Edition,* Austin, TX: Pro-Ed, 2008.

[12] B. L. Dabul, *Apraxia Battery for Adults—Second Edition,* Austin, TX: Pro-Ed, 2000.

[13] R. Cowie *et al.*, "Emotion Recognition in Human-Computer interaction," *Signal Processing Magazine, IEEE,* vol. 18, pp. 32-80, 2001.

[14] E. Douglas-Cowie *et al.*, "Emotional Speech: Towards a New Generation of Databases," *Speech Communication,* vol. 40, pp. 33-60, 2003.

[15] D. Ververidis and C. Kotropoulos, "Emotional speech Recognition: Resources, features, and methods," *Speech Communication,* vol. 48, pp. 1162-1181, 2006.

[16] Z. Zhihong *et al.*, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 31, pp. 39-58, 2009.

[17] E. Bates *et al.*, "Linguistic and Nonlinguistic Priming in Aphasia," *Brain and Language,* vol. 76, pp. 62-69, 2001.

[18] J. Druks *et al.*, "Quantitative Analysis of Connected Speech in Aphasia: Insights from Greek-speaking Patients with Fluent and Non-fluent Types of Aphasia " *Procedia - Social and Behavioral Sciences,* vol. 23, pp. 125-126, 2011.

[19] R. Goldmann *et al.*, "The Influence of Phonological Context on the Sound Errors of a Speaker with Wernicke's Aphasia," *Brain and Language,* vol. 78, pp. 279-307, 2001.

[20] J. K. Gordon., "Interpreting speech errors in aphasia," *Brain and Language,* vol. 103, pp. 122-123, 2007.

[21] G. Hickok *et al.*, "The Role of Broca's Area in Speech Perception: Evidence from Aphasia Revisited," *Brain and Language,* vol. 119, pp. 214-220, 2011.

[22] M. Vuković *et al.*, "Analysis of Voice Impairment in Aphasia After Stroke-Underlying Neuroanatomical Substrates," *Brain and Language,* vol. 123, pp. 22-29, 2012.

[23] B. Schuller *et al.,* "The INTERSPEECH 2009 Emotion Challenge," INTERSPEECH 2009, pp. 312-315.

[24] F. Eyben *et al.*, "Recent Developments in openSMILE, the Munich Open-Source Multimedia Feature Extractor", *In Proc. ACM Multimedia (MM),* Barcelona, Spain, 2013, pp. 835-838.

[25] A. Alwan. (2012) Voicesauce: A Program for Voice Analysis [Online]. Available: http://www.ee.ucla.edu/spapl/voicesauce/

[26] J. Hillenbrand *et al.*, "Acoustic Correlates of Breathy Vocal Quality," *Journal of Speech and Hearing Research*, vol. 37, pp. 769-778, 2004.

[27] G. de Krom, ''A Cepstrum-Based Techniques for Determining a Harmonics-to-Noise Ratio in Speech Signals,'' *J. Speech Hear. Res,* vol. 36, pp. 254–266, 1999.

[28] M. Brookes. (2014). VOICEBOX: Speech Processing Toolbox for Matlab. [online]. Available: http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

[29] X. Sun, "Pitch Determination and Voice Quality Analysis Using Subharmonic-To-Harmonic Ratio." ," *Acoustics, Speech and Signal Processing (ICASSP), 2002 IEEE International Conference on*, 2002, pp. 333-336.

[30] M. Hall *et al.,* "The WEKA Data Mining Software: An Update," SIGKDD Explorations, vol. 11, iss. 1, 2009.

[31] M. A. Hall. "Correlation-Based Feature Subset Selection for Machine Learning. Hamilton," New Zealand, 1998.

[32] P. R. Murphy, "Periodicity Estimation in Synthesized Phonation Signals using Cepstral Rahmonic Peaks," *Speech Communications*, vol. 48 pp. 1704-1713, 2006.

[33] R. Fraile and J. I. Godino-LLorente, "Cepstral Peak Prominence: A Comprehensive Analysis." *Biomedical Signal Processing and Control*. vol. 14, pp. 42-54, 2004.

[34] A. Ozdas *et al.*, "Investigation of Vocal Jitter and Glottal Flow Spectrum as Possible Cues for Depression and Near-Term Suicidal Risk." *IEEE Transactions on Biomedical Engineering*, vol .51, pp. 1530-1540, 2004.

[35] S. Alghowinem et al., "Detecting Depression: A Comparison Between Spontaneous and Read speech." *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 7457-7451.