JSLHR

Research Note

Identification of Affective State Change in Adults With Aphasia Using Speech Acoustics

Stephanie Gillespie,[a] Jacqueline Laures-Gore,[b] Elliot Moore,[a] Matthew Farina,[b] Scott Russell,[c] and Benjamin Haaland[d]

Purpose: The current study aimed to identify objective acoustic measures related to affective state change in the speech of adults with post-stroke aphasia.
Method: The speech of 20 post-stroke adults with aphasia was recorded during picture description and administration of the Western Aphasia Battery–Revised (Kertesz, 2006). In addition, participants completed the Self-Assessment Manikin (Bradley & Lang, 1994) and the Stress Scale (Tobii Dynavox, 1981–2016) before and after the language tasks. Speech from each participant was used to detect a change in affective state test scores between the beginning and ending speech.
Results: Machine learning revealed moderate success in classifying depression, minimal success in predicting depression and stress numeric scores, and minimal success in classifying changes in affective state class between the beginning and ending speech.
Conclusions: The results suggest the existence of objectively measurable aspects of speech that may be used to identify changes in acute affect from adults with aphasia. This work is exploratory and hypothesis-generating; more work will be needed to make conclusive claims. Further work in this area could lead to automated tools to assist clinicians with their diagnoses of stress, depression, and other forms of affect in adults with aphasia.

Accurate diagnosis of stress, depression, and determination of affective states in adults with aphasia is challenging because of the linguistic burden of many self-report measures, the potential for psychophysiological measures to be compromised by the neurological changes accompanying stroke, and problems associated with proxy-based questionnaires. The challenge of assessing stress, depression, and affective state in adults with aphasia is most obviously demonstrated by the exclusion of many adults with aphasia from post-stroke depression studies due to their inability to complete standard depression questionnaires (Kouwenhoven, Kirkevold, Engedal, & Kim, 2011). Often, post-stroke assessment of stress, depression, and affect in adults with aphasia is limited to proxy-based questionnaires or visual scale substitutions. Prior work reveals concerns that stroke survivors may have difficulty with visual analogue scales for mood (Price, Curless, & Rodgers, 1999). In addition, Bennett, Thomas, Austen, Morris, and Lincoln (2006) found some visual analogue scales may be poor tools for screening low mood in adults with stroke, even if useful over time for monitoring mood changes. Both prior studies excluded participants with aphasia. Exclusion of adults with aphasia from research as well as the clinical implications of inaccurate diagnosis of mood disorders and affective states can have negative effects on post-stroke recovery, mental and physical health, and cognitive functioning (Code & Herrmann, 2003). Development of accurate and accessible techniques that avoid the pitfalls of conventional assessment methods would enable health professionals to more effectively treat mental health problems that develop during post-stroke recovery in adults with aphasia.

The use of speech acoustics in diagnosing stress, depression, and affective states in adults with aphasia holds promise as a technique that permits avoidance of self-report questionnaires or behavioral observations and could potentially replace or augment conventional approaches to

[a]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta
[b]Communication Disorders Program, Georgia State University, Atlanta
[c]Department of Speech-Language Pathology, Grady Memorial Hospital, Atlanta, GA
[d]Department of Population Health Sciences, University of Utah, Salt Lake City
Correspondence to Stephanie Gillespie: s.gillespie812@gmail.com

Editor-in-Chief: Julie Liss
Received February 10, 2017
Revision received May 10, 2017
Accepted June 14, 2018
https://doi.org/10.1044/2018_JSLHR-S-17-0057

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

assessment. Several literature surveys over the years (Cowie et al., 2001; Ververidis & Kotropoulos, 2006; Zeng, Pantic, Roisman, & Huang, 2009) have provided updates on work related to the creation and analysis of speech emotion databases for affect analysis. Although there is an extensive body of literature for affect recognition in persons without aphasia, there is an analogous lack of such research for persons with aphasia, as they are often excluded from speech analysis studies. Much of the analysis related to speech in adults with aphasia involves word choice in naming tasks or phoneme analysis with little emphasis on the acoustic properties of continuous speech. The current study aims to identify objective acoustic measures related to affective state change in the speech of adults with aphasia while simultaneously introducing a previously unexplored method of affect labeling to aphasiology.

Affect Recognition and Digital Speech Processing

Affect can be described as the experience of feeling or emotion (Hogg, Abrams, & Martin, 2010). In general, this can be simplified to three dimensions, including valence (pleasure to displeasure), arousal (neutral to activated), and dominance (small to large representing control; Bradley & Lang, 1994). Throughout this work, "affect" will be used as a broad term encompassing the idea of emotion, and "affective state" will be used for a general feeling of an individual that could be represented along the three dimensions.

Affective computing utilizes computers for automatic emotion recognition, analysis, and synthesis (Picard, 1997). When using speech, affective speech processing usually attempts to classify or detect various emotions (e.g., joy, anger, sadness) or emotional states (e.g., stress, frustration) from analysis of the speech acoustics of the speaker. Studies often utilize speech that is acted, elicited via an emotion-prompting task, or taken from other media (e.g., video interview, TV shows). Although detection of emotion has achieved high success rates in research using controlled recordings, detection from less constrained settings with spontaneous or unprompted emotions is still challenging (Zeng et al., 2009). Sustained, strong emotional expressions are relatively simple to distinguish acoustically from lower-energy emotions (e.g., angry from calm). However, recognizing subtle changes in emotion or emotional state (e.g., calm to sadness or depression) is critical when designing automated persistent monitoring of mental health.

The detection of emotional stress is considered part of the overall efforts for affect analysis and automated recognition. However, much of the work on stress detection utilizes speech from workload-simulated stress (e.g., Kurniawan, Maslov, & Pechenizkiy, 2013) or short-term stress (e.g., the Speech Under Simulated and Actual Stress database; Womack & Hansen, 1995). As such, minimal work has focused on the detection of long-term clinical stress from speech acoustics, an area that has been identified to have a different form than that of short-term stress (Hu et al., 2015).

Depression detection through speech acoustic analysis has been an area of interest for a number of years in the speech processing community as illustrated by multiple studies (Ellgring & Scherer, 1996; Low, Maddage, Lech, Sheeber, & Allen, 2011; Ozdas, Shiavi, Silverman, Silverman, & Wilkes, 2004). Depression detection using speech acoustics is an active area of research interest due to the complications of detecting a long-term clinical diagnosis from speech acoustics that often include varying short-term emotional states as well. A recent review (Cummins et al., 2015) summarized speech analysis in depression and suicide over the last 10 years, including meta-analysis on depression-related speech features. Cummins concludes with a call to action for better research collaboration and further study of various demographic and/or clinical factors that will lead to variability in depressed speech.

Some of the recent research utilizing speech processing techniques in aphasia has dealt with challenges related to automated assessment of intelligibility in aphasia (Le, Licata, Persad, & Mower Provost, 2016). In addition, speech processing techniques in speech recognition and language processing are an area of research in designing computerized therapy and rehabilitation systems for aphasia and other language disorders (Chen et al., 2016; Pompili, Abad, & Trancoso, 2011). However, studies using speech processing to identify affective states and mood disorders in persons with aphasia are limited. Progress in detecting stress, depression, and affect in other populations via speech analysis prompts the need to expand and develop affective analysis models to accommodate adults with aphasia and other limiting language disorders who are at risk for developing depression and other serious mood disorders that can affect progress in therapy as well as overall life expectancy.

Prior Work in Depression Classification, Aphasia, and Speech Acoustics

Previous work by Gillespie, Moore, Laures-Gore, and Farina (2016) analyzed 14 participants for depression classification based on the Stroke Aphasia Depression Questionaire-10 (SADQ-10; Sutcliffe & Lincoln, 1998). Prosodic and spectral features were extracted, and a sequential minimal optimization–support vector machine (SVM) was built as a classifier in the Waikato Environment for Knowledge Analysis software (WEKA; Hall et al., 2009). A leave-one-subject-out approach was used in which a training model was built on 13 participants and tested on the excluded participant. Precision, recall, and accuracy were calculated.

The cepstral peak prominence feature subset classified the best overall considering recall, precision, and accuracy. Because of the spectrum of clinical features in the patients examined, there was concern that the depression classifier was, in fact, detecting clinical features other than depression. The analysis suggested there did not appear to be an indication that the aphasia type, aphasia quotient, or SADQ-10 score affected the classifier's ability to predict the depression label. However, two of the three participants who performed with less-than-chance accuracies had a SADQ-10 score near the threshold that was used to determine

the binary depression label based on suggested levels by Leeds, Meara, and Hobson (2004). As such, it was determined a binary label of depression may not be the best representation of depression when using machine learning for a clinical setting.

Prior Work Predicting Stress and Depression Scores in Adults With Aphasia With Speech Acoustics

The SADQ-10 and the Perceived Stress Scale (PSS; Cohen, Kamarck, & Mermelstein, 1983) are scored on numeric scales (0–30 and 0–56, respectively) and do not have multiple thresholds representing degrees of severity. Leeds et al. (2004) 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-10 scores within 1 or 2 points of each other should be interpreted for distinct degrees of depression. A survey paper on detection of depression from speech acoustics recommended excluding participants in the moderate categories of a depression scale due to the ordinal nature of mental state scales (Cummins et al., 2015). Although this may represent an ideal circumstance for machine learning, there is a need for research that uses real-world data that are likely to have moderate depression scores. As such, the next progression of the prior work was to focus on prediction of SADQ-10 and PSS scores (Gillespie et al., 2017). The use of regression avoided determining categories that sort the participants based on thresholds but instead attempted to predict the depression or stress scores as closely as possible.

Nineteen participants were selected for regression analysis of their SADQ-10 scores, and 18 were selected for regression analysis of their PSS score. Prosodic, spectral, Teager energy operator (TEO), and glottal features were extracted from the voiced sections of speech, with low-level descriptors statistics calculated at the sentence level as described previously. After feature selection, the selected feature subsets were then used to build models using support vector regression (SVR) in MATLAB. None of the feature groups analyzed performed significantly better than the others or at a level that would be considered significantly successful. It was concluded that the mediocre performance of the linear SVR was likely hindered by the small amount of data available for analysis. These "snapshot" speech acoustics were taken from a single recording and would likely be dominated by the short-term affective states, including current emotional state, instead of the long-term stress or depression states.

Motivation

From the literature and our previous work, it is clear there is a need to better understand and explore the translation of existing practices for speech-based assessment of stress, depression, and general affective state to the population of adults affected by aphasia. The acute nature of affect and emotional states makes building a predictive model for long-term states (such as clinical depression) challenging due to the temporal nature of affective expression and mood. As a result, the work presented in this note seeks to find features of speech that may show sensitivity to subtle changes in affective states over the course of a single recording session. First, we analyzed the correlation between various assessment scores, determining the consistency of the self-reported scores and the potential implications of aphasia on affect. Then, we determined if relevant speech measures could indicate subtle changes in affect, motivated by the idea that short-term affective states may be more prevalent in the vocal acoustics than long-term stress or depression in adults with post-stroke aphasia during single-session interviews. As the models proposed in this work analyze the differences within an individual's speech over a single recording session, the impact of motor speech disorders (including dysarthria and apraxia of speech) as they relate to affect was not addressed in this work and is left for future analysis.

Method

Speech from 26 adults who were at least 1-month post-onset of stroke was collected in single sessions at the Aphasia and Motor Speech Disorders Laboratory at Georgia State University from spring 2014 to summer 2015. Approval from the institutional review boards at Georgia State University and Georgia Institute of Technology was collected. Grady Memorial Hospital approved the research through the Grady Research Oversight Committee.

Table 1 shows the demographic and clinical data of the participants included in this study, a subset of those previously studied by Laures-Gore, Farina, Moore, and Russell (2016). Participants exhibited Broca's, Wernicke's, conduction, and anomic aphasia as determined by the Western Aphasia Battery–Revised (WAB-R; Kertesz, 2006). The WAB-R also assigns an Aphasia Quotient that assesses the severity of the aphasia, ranging from 0 to 100 (*most* to *least severe*) with a score higher than 93.8 within normal limits indicating no aphasia. Two participants scored above the WAB-R cutoff but were included as they self-identified as having aphasia and were referred to the study due to their aphasia. Five participants had technical difficulties during the recording process, and one participant did not confirm stroke. As such, 20 participants with complete recordings and confirmed history of stroke were available for analysis. Although dysarthria and apraxia of speech were determined by the Frenchay Dysarthria Assessment–Second Edition (Enderby & Palmer, 2008) and the Apraxia Battery for Adults–Second Edition (Dabul, 2000), this study did not consider their impact on affective speech.

In order to elicit and record speech, participants were asked to complete a series of tasks as a part of the WAB-R. Prior to the WAB-R, two additional picture descriptions, the cookie theft picture from the Boston Diagnostic Aphasia Examination–Third Edition (Goodglass, Kaplan, & Barresi, 2000) and the "Cat in Tree" sketch (Nicholas & Brookshire, 1993), were provided in an attempt to

Table 1. Select clinical and demographic information for selected[a] participants included in this work.

ID	Gender	Age	Aphasia type (AQ)	Dysarthria score	Apraxia score	SAM–Valence Pre	SAM–Valence Post	SAM–Arousal Pre	SAM–Arousal Post	Stress Scale Pre	Stress Scale Post
6	M	61	Anomic (87.4)	None	Mild	3	2	3	1	1	1
7	M	49	Anomic (82.1)	Mild	Mild	2	2	1	1	2	2
8	F	46	Broca's (59.6)	None	Moderate–severe	1	1	1	1	1	1
9	M	70	Anomic (83.2)	None	Mild	1	1	1	1	1	1
10	M	32	Anomic (99.4)	None	None	2	1	1	1	1	1
11	M	55	Anomic (78.0)	Mild–Moderate	Mild–moderate	5	5	3	3	7	4
12	M	52	Broca's (58.3)	Mild	Moderate–severe	1	2	1	2	3	5
13	M	57	Anomic (87.4)	Mild	None	3	2	3	3	2	5
14	M	67	Anomic (98.4)	Mild	None	2	2	3	3	4	3
15	F	51	Wernicke's (41)	None	Moderate–severe	1	5	1	5	1	7
16	F	39	Anomic (93.2)	Mild	Mild	2	3	2	3	3	4
17	F	33	Anomic (92.4)	Mild	None	1	1	1	1	1	1
18	M	63	Anomic (83.2)	Mild	Mild	1	1	2	1	4	1
19	F	52	Anomic (88.3)	Mild	Mild	4	4	4	1	5	1
20	F	36	Conduction (54.6)	None	Moderate	3	3	3	4	3	3
21	M	49	Anomic (83.3)	None	Mild	2	2	1	1	1	1
23	M	31	Broca's (31.9)	Mild	Mild–moderate	1	1	1	1	1	1
24	F	64	Broca's (68.5)	Mild	Moderate	3	3	3	3	5	5
25	M	40	Wernicke's (87.4)	Mild	Mild–moderate	1	5	1	1	1	2
26	M	25	Anomic (92.3)	None	None	3	3	3	2	4	3

Note. AQ = Aphasia Quotient; SAM = Self-Assessment Manikin; M = male; F = female.

[a]Selected participants were those who completed recordings without technical difficulties and confirmed stroke.

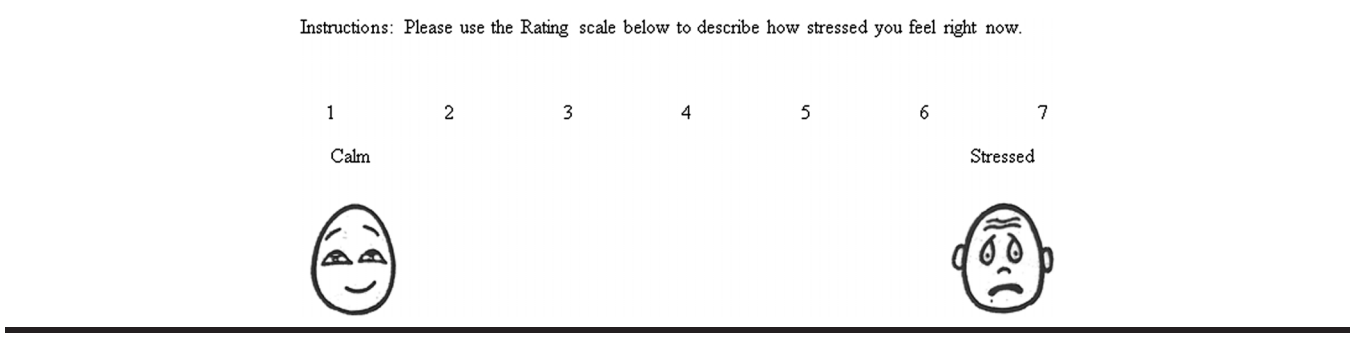
elicit additional spontaneous speech. Speech was recorded during a single session with an AKG C520 headset condenser microphone and sampled at 16 kHz in Audacity (SourceForge, n.d.).

Acute assessments of affective state were collected at the start and end of the testing session. Participants were asked to complete valence and arousal Self-Assessment Manikin (SAM; Bradley & Lang, 1994). The nonverbal assessment consists of two sets of line-drawn figures expressing different emotional scales for a total of five figures in each set. The participants in this study were limited to selecting the picture that best represented their mood, limiting the numeric outputs to a 5-point scale instead of the original 9-point scale. An additional acute pictorial assessment, identified in this work as the "Stress Scale," was used and is shown in Figure 1. The assessment output was a number between 1 and 7, identified by the authors as *calm* and *stressed*, respectively. The Stress Scale has been used in previous work by the second author (Laures-Gore, Heim, & Hsu, 2007). Both the SAM and the Stress Scale were used in this work as acute measures of affect as they allowed feedback from the participants with linguistic challenges due to aphasia. For all three affective assessments, each participant was asked to point to the picture that best represented how he or she felt at that moment with the investigator providing the word fitting each extreme on the scale. Each participant received the same instructions with repetition of instructions if requested.

Using Version 2.0.5 of Audacity (SourceForge, n.d.), approximately 55 responses per participant were segmented from the recorded speech. The segments ranged from a single-word answer to an extended description based on the individual task of the WAB-R. Longer responses, including those of the picture descriptions, were segmented into individual sentences or phrases based on the completion of an idea or task. In total, the participants had at least 3.5 min of responses recorded after segmentation and approximately 75 segments. Each segment excluded the interviewer's speech as well as any incidental noise.

Features are mathematical quantities from equations that have been created to highlight specific patterns or trends in the speech acoustics. Often, the numeric values on their own are not meaningful until looked at with comparisons to other data. Prosodic features are designed to be features related to rhythmic or intonational properties of speech. Spectral features are those related to the content in the frequency spectrum of speech. The prosodic and spectral features extracted in this work include cepstral peak prominence, four different harmonic-to-noise ratios for various frequency bands, 14 mel-frequency cepstral coefficients (MFCCs), and their deltas, pitch, jitter, eight line spectral frequencies, and root mean square (RMS) energy. Additional features include those based on the TEO and glottal waveform, which assist in quantifying airflow through the vocal folds related to voice quality. Specific TEO features extracted in this work include amplitude modulation, frequency modulation, 16 critical band areas, RMS energy, and log energy. The glottal features are based on prior work detecting depression and are detailed in previous literature (Moore, Clements, Peifer, & Weisser, 2003, 2004, 2008). Full methodologies for feature extraction techniques and justification of selection are described

Figure 1. Stress Scale. The Picture Communication Symbols 1981–2016 by Tobii Dynavox. All Rights Reserved Worldwide. Used with permission from Tobii Dynavox (1981–2016). Boardmaker® is a trademark of Tobii Dynavox. Tobii Dynavox, 2100 Wharton Street, Suite 400, Pittsburgh, PA 15203. Phone: 1 (800) 588-4548.



in prior work (Gillespie et al., 2017). For each feature extracted, various low-level descriptors (e.g., mean, interquartile range [IQR], skew) were calculated based on those described in openSMILE (Eyben, Weninger, Gross, & Schuller, 2013).

Analysis of Correlation of Affective Labels

Because of the variety of labels collected from self-reports from the participant, the correlation between similar affective scores was necessary to validate the consistency in responses between the various affective assessments. Laures-Gore et al. (2016) recently published the results comparing the Aphasia Depression Rating Scale (Benaim, Cailly, Perennou, & Pelissier, 2004), SADQ-10, and PSS scores for 25 of the participants included in the original data collection. Laures-Gore et al. did not find any significant correlations between the Aphasia Depression Rating Scale or SADQ-10 scores and the other demographic information considered in their analysis (age, WAB-R–Aphasia Quotient, or time post-onset from stroke). In the current work, further correlation analysis using Pearson product–moment correlation was completed with a larger subset of the available clinical scores for the 20 participants analyzed for affect.

Machine Learning to Predict Changes in Affect

The underlying premise of this study is to measure and analyze speech features that can be correlated to changes in affective state during a single-session interview. It is important to note that this study does not (at this time) attempt to make an identification of a specific affective state. Instead, the features are examined to determine those that correlate with changes (positive or negative) in the affective ratings provided during the pre– and post–affective assessments. The expectation would be that participants who recorded changes in their affective condition between the pre- and post-assessments should also exhibit measures that correlate respectively. Participants with no change in their assessments should exhibit little or no relevant changes in certain speech measures. Statistics between the two sets of measurements were calculated and used with machine learning to detect potential changes in affective state scores.

The first and last 10 sentences from the 20 participants who had complete recordings and confirmed stroke were selected for analysis in MATLAB. Limiting the speech considered at the beginning and end of the interview to 10 samples ensured minimal time elapsed between the SAM and Stress Scale assessments and when the sentences were recorded, minimizing the likelihood that the self-reported affective labels would have changed. The first 10 samples were mostly spontaneous speech elicited from the picture descriptions, whereas the last 10 samples were mostly directed speech of either word or phrase length. One thousand five hundred ninety-five statistics from the aforementioned prosodic, spectral, glottal, and TEO features were extracted for each sentence and individually normalized with respect to the features' mean and variance across participants. In order to compare the change in feature distributions between the first and last 10 samples, six statistics were calculated on each dataset to represent the distribution: mean, median, min, max, IQR, and variance. The change in the first and last sets was then calculated by determining the change of the statistic values for each feature, which created 9,570 features to analyze.

Given the number of statistics under consideration in the study, it was necessary to employ a means to reduce the dimensionality through a feature selection strategy. Feature selection is the process of selecting the most important features so as to reduce the number analyzed while also retaining as much class-discriminating information as possible (Theodoridis & Koutroumbas, 2003). Multiple methods and algorithms exist, often using separate training and testing subsets of the data (Bagherzadeh-Khiabani et al., 2016). With only 20 participants available for analysis, reserving some of the participants as an exclusive test set would not be practical, as the model would then have even fewer samples from which to learn and build the machine learning model. Instead, the feature selection and machine learning process were conducted using a leave-one-subject-out strategy, creating 20 different train/test sets. The feature selection was conducted for 19 participants in MATLAB using a bootstrap-aggregated (bagged) decision classification tree model with 100 individual trees. Each individual tree utilized randomized initial parameter

selection, which helped to avoid overfitting. The reduced feature sets were then selected to include the top 10 features based on the tree predictor importance for the machine learning tasks. This feature selection process was completed for each of the individual feature types, as well as in groupings including the prosodics and spectral features, TEO features, and glottal feature sets. The leave-one-out strategy was continued into the machine learning analysis where a subset of selected features from the 19 participants was analyzed to predict the results for the individual held out. As such, the testing individual was never seen or considered by the model until the test phase of the machine learning task.

The three different self-assessments of interest in this study were the SAM–Valence, SAM–Arousal, and Stress Scale assessments. The change in affective score was calculated by subtracting the pre-recording score from the post-recording score. Because the reported changes in self-assessment scores were not evenly distributed (many individuals reported no change in scores between their pre- and post-assessments), two sets of experiments were conducted. A two-class experiment grouped individuals by whether they reported a change (positive or negative) in their scores or not (no change), which will be referred to as the presence-of-change (PC) analysis. A three-class experiment analyzed individuals by whether they reported a negative change, a positive change, or no change in scores, referred to as the sign-change (SC) analysis.

A multiclass SVM classification model using a linear kernel in MATLAB was built. The initialization of the hyperparameters of the SVM was optimized with an internal fivefold cross-validation system within MATLAB. Following a leave-one-subject-out methodology, 20 individual experiments were conducted in which there were 19 training points used to build the model using the appropriately reduced feature subset and predict the change-in-affect response for the remaining participant for each of the six affective state tasks.

Results

Selected results of the extended correlation analysis are shown in Table 2. The pre-tests for the SAM–Valence and SAM–Arousal scores had a high correlation of .82, whereas the post-test score correlation was only .54. High correlation values between the SAM–Arousal and Stress Scale scores were found between both pre-recording scores and both post-recording scores (pre: $r = .712$, post: $r = .873$). The SAM–Valence and the Stress Scale were highly correlated for the pre-scores, but not for the post-scores (pre: $r = .70$, post: $r = .54$).

The most common method of judging the performance of a machine learning model is with accuracy. Guessing the correct class in a two- or three-class case with equal class distributions would occur by chance approximately 50% or 33% of the time, respectively. In this study, the class distributions were not equal. As such, the best results without learning a model would be achieved by always predicting the class of the majority label, which would be the correct class at the same rate as the majority class membership. In both the PC and SC analyses, the class majority label was that of no-change (class = 0), with membership probabilities of 60%, 65%, and 55% for the SAM–Valence, the SAM–Arousal, and the Stress Scale, respectively. Using these values, 15.8% of the SC models and 29.8% of the PC models resulted in equal or better accuracies than the best-guess accuracies.

Machine learning models often use measures in addition to accuracy, especially when there are an uneven number of individuals in each class that were used to train the model. In these cases, the Matthew correlation coefficient (MCC) value can be calculated to understand the quality of a model. The MCC ranges from −1 to +1, with 0 representing random prediction, negatives scores representing disagreement between predictions and observations, and positive scores representing agreement. A perfect model would result in MCC = 1.

None of the models for PC or SC of any of the affect labels resulted in a large positive MCC score. Because of the small number of training samples, there were multiple models across the various feature set sizes, feature set types, and affective state scales that resulted in negative MCC values, indicating poorly fitting models. The analysis presented in this work utilized machine learning from features that represented the change in two distributions over time. In general, the most selected statistics across most of the feature types included change in maximum, change in minimum, and change in IQR. Change in mean, change in median, and change in variance together were only selected as approximately 10% of the features.

Discussion

The work presented here sought to investigate statistical trends in speech acoustics that correlated with changes in reported affective scores. As an initial inquiry, the consistency of the SAM, Stress Scale, and WAB-R labels was analyzed to validate the use of self-report scores for detecting acute affective states in adults with aphasia. The current study utilized the valence and arousal dimensions of the SAM, which, as Bradley and Lang (1994) describe, "ranges from a smiling, happy figure to a frowning, unhappy figure when representing the pleasure dimension, and ranges from an excited, wide eyed figure to a relaxed, sleepy figure for the arousal dimension" (p. 50). Similarly, the Stress Scale uses pictures of faces ranging from a calm, relaxed appearance to one that is wide-eyed and not appearing relaxed. In addition, the Stress Scale utilizes the words "calm" and "stressed" to further anchor the scale and label the pictures of the faces. Consistently, the SAM–arousal dimension and the Stress Scale were strongly related across pre-recording, post-recording, and change scores. Arousal has been a long-standing component of the human stress response (Sanders, 1983, 1998), and our current findings seem to support this connection. SAM–valence is strongly linked with SAM–arousal and Stress Scale pre-recording

Table 2. Pearson correlation coefficients between various affective assessment scores and the WAB-R.

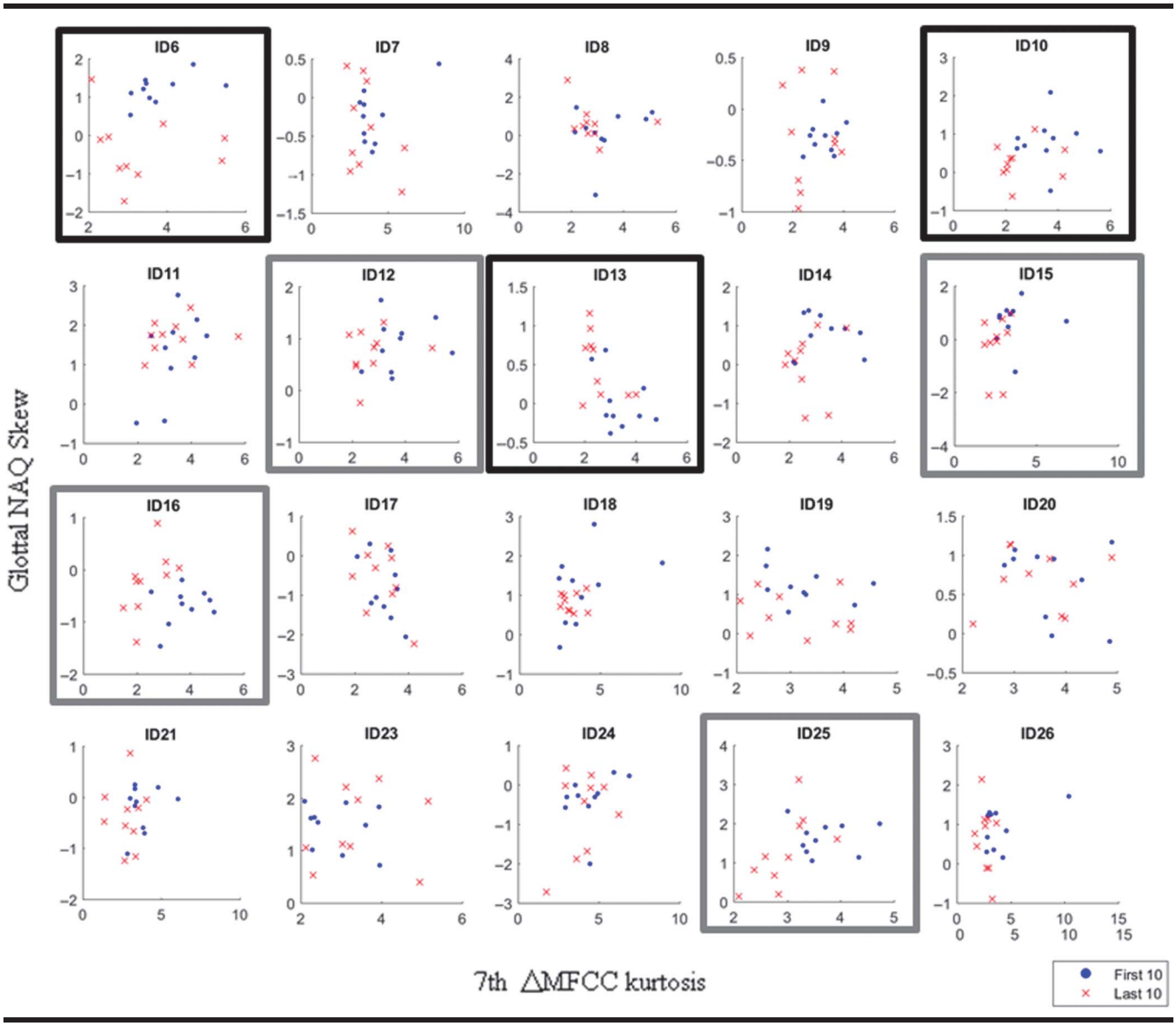
Measure	SAM–Arousal Post	ΔSAM–Arousal	SAM–Valence Pre	SAM–Valence Post	ΔSAM–Valence	Stress Scale Pre	Stress Scale Post	ΔStress Scale	WAB-R Subscore–Spontaneous	WAB-R Subscore–Word Finding	WAB-R–Aphasia Quotient
SAM–Arousal Pre	.35	–.45	**.82**	.33	–.37	**.71**	.20	–.42	.24	.16	.22
SAM–Arousal Post		.67	.29	.54	.30	.35	**.87**	.47	–.31	–.23	–.33
ΔSAM–Arousal			–.38	.25	.58	–.23	.67	**.79**	–.49	–.42	–.49
SAM–Valence Pre				.45	–.40	**.70**	.19	–.42	.20	.31	.22
SAM–Valence Post					.64	.42	.54	.12	–.08	.01	–.18
ΔSAM–Valence						–.18	.39	.50	–.25	–.25	–.37
Stress Scale Pre							.33	–.37	.05	.23	.11
Stress Scale Post								.61	–.36	–.18	–.34
ΔStress Scale									–.36	–.35	–.39
WAB-R Subscore–Spontaneous										**.88**	**.93**
WAB-R Subscore–Word Finding											**.93**

Note. Bold scores indicate strong correlation with $|r| > .7$. SAM = Self-Assessment Manikin; WAB-R = Western Aphasia Battery–Revised; Δ = change in, calculated by subtracting pre-score from post-score.

scores. Valence, or pleasure, and arousal have endured a long history of cohabitation within the emotion literature (e.g., Russell, 1980). Within aphasia, the relation between valence, arousal, and stress should be further explored to deepen our understanding of these emotional constructs on the well-being of adults with aphasia. Interestingly, the aphasia severity (as measured by the Aphasia Quotient) had low to moderate correlations with some of the affective labels, indicating that the short-term affective state of the individual appears to have some association with the severity of the aphasia presenting in the individual. This is unlike previous results in which there appears to be no relation between aphasia severity and both physiological and self-report measures of stress in the aphasia population (Laures-Gore, 2012).

Figure 2 shows a visualization of the calculated feature values for the first and last 10 samples of speech across all 20 participants with two arbitrary features for the SAM–Valence label, selected for visualization because of its simplicity with just two features, represented in two dimensions. The plots boxed with solid black boxes, light gray boxes, and no boxes identify participants in the negative-change group, the positive-change group, and the no-change group, respectively, for the SC analysis. Visually, it can be seen that certain participants (e.g., 16 and 21) had a clear change in their sample distribution between the first and last 10 samples.

Figure 2. Distributions of coverage method feature values for the first 10 and last 10 sentences of each person for the Self-Assessment Manikin (SAM)–Valence test scores. Plots with a solid black outline displayed a negative change in SAM–Valence scores, plots with a solid gray outline displayed a positive change in SAM–Valence scores, and plots with no outline did not report a change in SAM–Valence scores. MFCC = mel-frequency cepstral coefficient; NAQ = normalized amplitude quotient.



However, not every participant with a change in SAM–Valence score (indicated by either a gray or black box around the plot) had clear acoustic changes that confirmed their self-reported change in affect. In addition, some participants appear to have changes in speech acoustics but did not report a change in affect. Changes in the speech acoustics extracted in this work do not perfectly predict nor create a perfect model relating the self-reported affect scores. The small number of models with promising results based on accuracy values suggests the possibility of detecting changes in affective states in individuals with aphasia. Because an individual's long-term states (e.g., depression) or demographics do not change during the recording process, consistent trends in the changes of affective state are a likely explanation for the calculated changes in vocal acoustics. However, small fluctuation in vocal acoustics would be expected due to sentence intonation, type of speech, or even the abilities of the individual to pronounce the specific phonemes in the required phrases. This can be seen by lack of a consistent pattern between the first and last 10 feature samples for individuals with the same type of affective change seen in Figure 2. With more speech from a larger number of individuals, especially those indicating a change in affective states, we would expect to see the trends more clearly.

The statistics analyzed in this work were selected for their ability to represent a summary of data. Because the first 10 and last 10 feature sets for each individual were suspected to change significantly only with a reported change in affective state, documenting the location and spread was essential. The prominence of the feature selection algorithm's selection of changes in maximum, minimum, and IQR was repeatedly shown across multiple prediction models. Changes in mean, median, and variance together were only selected as 10% of the features, even though they could be argued to be more commonly used measures for simple distribution representation. Although the results here are exploratory with only 20 individuals used for training and testing, these preliminary findings suggest that trends in the "extreme values" of speech features may be useful when comparing time-series trends for reported changes in affective states as compared to information about the centralities.

The utility of the prosodic, spectral, glottal, and TEO features in multiple areas of signal processing analysis, including the detection of stress, depression, and emotion recognition, was a driving factor for analyzing the change in feature distributions related to acute affective states. MFCCs are often used in affective research and are considered one of the most versatile features in speech analysis (Cummins et al., 2015; Rabiner & Schafer, 2011). Although MFCC features were found to be one of the subsets that produced higher accuracies in the PC experiment, the feature set did not perform well within the SC experiment, potentially suggesting that MFCCs are more useful in detecting the presence of change rather than identifying a change as positive or negative. Though often indicative of emotion, it is possible the pitch already fluctuated from sentence to sentence to such an extent that the potential affective changes were neither consistent nor large enough to create appropriately predictive models with high MCC scores. These results suggest the importance of expanding the scope of features analyzed beyond those of the prosodic nature; initiating collaborations between the fields of clinical speech research and digital signal processing could facilitate this transition to include other promising feature types.

The presented work on affect analysis was not easy due to multiple challenges, many of which can be attributed to the data collection process. The small database was limited by eligible participants who were willing and able to successfully complete the entire data collection process and clinical diagnosis procedure, as well as the extensive time commitment to collect the data. As such, there was a large range in clinical and demographic information for those participants who chose to participate, making generalizations of models complicated and challenging. The analysis results must thus be considered exploratory, and we encourage future studies to analyze changes in affective state instead of the traditional models of classification or prediction for replication of the presented results. Furthermore, accurate self-report of affective state can be complicated, especially given the linguistic challenges associated with aphasia. In the current study, we attempted to collect pre– and post–self-report of affective states in a short time frame. Despite the scale directing each participant to report feelings at that moment, it is possible that some of the participants did not change their rating in the post–self-report, attempting to be consistent with the pre–self-report. Subsequent work should explore the use of different self-report intervals and the effects on consistency of ratings in adults with aphasia. In addition, because this work was prompted due to the clinical need to objectively detect stress and depression, it is clear acute assessments should not be a primary diagnostic tool for long-term stress and depression. Future studies of affect in adults with aphasia, especially stress and depression, should strongly consider a multisession recording and assessment process to allow the speech collected to extend over a longer time period, reducing the effect of short-term affective states on the speech collected.

Conclusion

The work presented identified objective measures of vocal acoustics suggestive of changes in affective state using speech analysis for adults with aphasia. Reported changes in affect and vocal acoustics were used to predict state changes, resulting in better-than-chance accuracies but low MCC values. Future work for detection of stress and/or depression in adults with aphasia should consider a data collection process over multiple sessions for each individual to compensate for changes in short-term affective state. Although participants in this study presented with dysarthria and apraxia of speech, the impact of these motor speech disorders was not considered in this analysis and will need to be studied further with respect to their relation to affect in speech. The results of this exploratory

study are promising and support the possibility of developing a clinical tool in the future to assist clinicians with the diagnosis of stress, depression, and affect in adults with aphasia using speech acoustics. We will continue to look for appropriate models and features to handle the complexities of detecting affect from speech in those living with aphasia.

Acknowledgments

This study was 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 UL1TR000454, awarded to Elliot Moore, Jacqueline Laures-Gore, and Scott Russell. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship (Grant DGE-1148903) awarded to Stephanie Gillespie. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation.

References

Bagherzadeh-Khiabani, F., Ramezankhani, A., Azizi, F., Hadaegh, F., Steyerberg, E., & Khalili, D. (2016). A tutorial on variable selection for clinical prediction models: Feature selection methods in data mining could improve the results. *Journal of Clinical Epidemiology, 71,* 76–85.

Benaim, C., Cailly, B., Perennou, D., & Pelissier, J. (2004). Validation of the Aphasic Depression Rating Scale. *Stroke, 40*(2), 523–529.

Bennett, H. E., Thomas, S. A., Austen, R., Morris, A. M. S., & Lincoln, N. B. (2006). Validation of screening measures for assessing mood in stroke patients. *British Journal of Clinical Psychology, 45,* 367–376.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49–59.

Chen, Y.-P. P., Johnson, C., Lalbakhsh, P., Caelli, T., Deng, G., Tay, D., . . . Morris, M. E. (2016). Systematic review of virtual speech therapists for speech disorders. *Computer Speech & Language, 37,* 98–128.

Code, C., & Herrmann, M. (2003). The relevance of emotional and psychosocial factors in aphasia to rehabilitation. *Neuropsychological Rehabilitation, 13*(1–2), 109–132.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24,* 385–396.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine, 18,* 32–80.

Cummins, N., Scherer, S., Krajewski, J., Schneider, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication, 71,* 10–49.

Dabul, B. L. (2000). *Apraxia Battery for Adults–Second Edition*. Austin, TX: Pro-Ed.

Ellgring, H., & Scherer, K. R. (1996). Vocal indicators of mood change in depression. *Journal of Nonverbal Behavior, 20*(2), 83–110.

Enderby, P., & Palmer, R. (2008). *Frenchay Dysarthria Assessment–Second Edition (FDA-2)*. Austin, TX: Pro-Ed.

Eyben, F., Weninger, F., Gross, F., & Schuller, B. (2013). *Recent developments in OpenSMILE, the Munich open-source multimedia feature extractor*. Paper presented at the Proceedings of ACM Multimedia (MM), Barcelona, Spain.

Gillespie, S., Moore, E., II, Laures-Gore, J. S., & Farina, M. (2016). *Exploratory analysis of speech features related to depression in adults with aphasia*. Paper presented at the 41st IEEE International Conference on Acoustics, Speech, and Signal Processing, Shanghai, China.

Gillespie, S., Moore, E., II, Laures-Gore, J. S., Farina, M., Russell, S., & Logan, Y.-Y. (2017). *Detecting stress and depression in adults with aphasia through speech analysis*. Paper presented at the 42nd IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2017), New Orleans, LA.

Goodglass, H., Kaplan, E., & Barresi, B. (2000). *Boston Diagnostic Aphasia Examination–Third Edition (BDAE-3)*. San Antonio, TX: Pearson.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Peutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter, 11*(1), 10–18.

Hogg, M. A., Abrams, D., & Martin, G. N. (2010). Social cognition and attitudes. In G. N. Martin, N. R. Carlson, & W. Buskist (Eds.), *Psychology* (pp. 646–677). Harlow, United Kingdom: Pearson Education.

Hu, B., Liu, Z., Yan, L., Wang, T., Liu, F., Li, Z., & Kang, H. (2015). *Feature selection and classification of speech under long-term stress*. Paper presented at the 2015 IEEE International Conference on Bioinformatics and Biomedicine, Washington, DC.

Kertesz, A. (2006). *Western Aphasia Battery–Revised (WAB-R)*. San Antonio, TX: Pearson.

Kouwenhoven, S. E., Kirkevold, M., Engedal, K., & Kim, H. S. (2011). Depression in acute stroke: Prevalence, dominant symptoms, and associated factors. A systematic literature review. *Disability and Rehabilitation, 33*(7), 539–556.

Kurniawan, H., Maslov, A. V., & Pechenizkiy, M. (2013). *Stress detection of speech and galvanic skin response signals*. Paper presented at the 2014 IEEE 27th International Symposium on Computer-Based Medical Systems, New York, NY.

Laures-Gore, J. S. (2012). Aphasia severity and salivary cortisol over time. *Journal of Clinical and Experimental Neuropsychology, 34*(5), 489–496.

Laures-Gore, J. S, Farina, M., Moore, E., & Russell, S. (2016). Stress and depression scales in aphasia: Relation between the Aphasia Depression Rating Scale, Stroke Aphasia Depression Questionnaire-10, and the Perceived Stress Scale. *Topics in Stroke Rehabilitation, 24,* 114–118. https://doi.org/10.1080/10749357.2016.1198528

Laures-Gore, J. S., Heim, C. M., & Hsu, Y.-S. (2007). Assessing cortisol reactivity to a lingtuistic task as a marker of stress in individuals with left-hemisphere stroke and aphasia. *Journal of Speech, Language, and Hearing Research, 50*(2), 493–507.

Le, D., Licata, K., Persad, C., & Mower Provost, E. (2016). Automatic assessment of speech intelligibility for individuals with aphasia. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 24*(11), 2187–2199.

Leeds, L., Meara, R. J., & Hobson, J. P. (2004). The utility of the Stroke Aphasia Depression Questionaire (SADQ) in a stroke rehabilitation unit. *Clinical Rehabilitation, 18,* 228–231.

Low, L.-S. A., Maddage, N. C., Lech, M., Sheeber, L. B., & Allen, N. B. (2011). Detection of clinical depression in adolescents'

speech during family interactions. *IEEE Transactions on Biomedical Engineering, 58*(3), 574–586.

Moore, E., II, Clements, M., Peifer, J., & Weisser, L. (2003). *Investigating the role of glottal features in classifying clinical depression*. Paper presented at the Proceedings of the 25th Annual International Conference of the IEEE-EMBS, Cancun, Mexico.

Moore, E., II, Clements, M., Peifer, J., & Weisser, L. (2004). *Comparing objective feature statistics of speech for classifying clinical depression*. Paper presented at the Proceedings of the 26th Annual International Conference of the IEEE-EMBS, San Francisco, CA.

Moore, E., II, Clements, M., Peifer, J., & Weisser, L. (2008). Critical analysis of the impact of glottal features in the classification of clinical depression of speech. *IEEE Transactions on Biomedical Engineering, 55*(1), 96–1070.

Nicholas, L. E., & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research, 36,* 338–350.

Ozdas, A., Shiavi, R. G., Silverman, S. E., Silverman, M. K., & Wilkes, D. M. (2004). Investigation of vocal jitter and glottal flow spectrum as possible cues for depression and near-term suicidal risk. *IEEE Transactions on Biomedical Engineering, 51*(9), 1530–1540.

Picard, R. W. (1997). *Affective computing*. Cambridge, MA: MIT Press.

Pompili, A., Abad, A., & Trancoso, I. (2011). *Virtual therapist for aphasia treatment*. Rome, Italy: Rome University of Tor Vergata.

Price, C. I. M., Curless, R. H., & Rodgers, H. (1999). Can stroke patients use visual analogue scales? *Stroke, 30,* 1357–1361.

Rabiner, L. R., & Schafer, R. W. (2011). *Theory and applications of digital speech processing* (1st ed.). Upper Saddle River, NJ: Pearson.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Sanders, A. F. (1983). Towards a model of stress and human performance. *Acta Psychologica, 53*(1), 61–97.

Sanders, A. F. (1998). *Elements of human performance: Reaction processes and attention in human skill*. Mahwan, NJ: Erlbaum.

SourceForge. (n.d.). Audacity (Version 2.0.5). Retrieved from http://audacity.sourceforge.net/

Sutcliffe, L. M., & Lincoln, N. B. (1998). The assessment of depression in aphasic stroke patients: The development of the stroke aphasic depression questionaire. *Clinical Rehabilitation, 12*(6), 506–513.

Theodoridis, S., & Koutroumbas, K. (2003). *Feature selection pattern recognition* (2nd ed.). London, United Kingdom: Elsevier Academic Press.

Tobii Dynavox. (1981–2016). Picture communication symbols. Retrieved from http://www.mayer-johnson.com

Ververidis, D., & Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech Communication, 48,* 1162–1181.

Womack, B. D., & Hansen, J. H. L. (1995). *Stress independent robust HMM speech recognition using neural network stress classification*. Paper presented at the 4th European Conference on Speech, Communication, and Technology, Madrid, Spain.

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(1), 39–58.