

Characterization of Divergence in Impaired Speech of ALS Patients

Archna Bhatia¹, Bonnie J. Dorr¹, Kristy Hollingshead¹, Samuel L. Phillips², and Barbara McKenzie²

¹Florida Institute for Human and Machine Cognition, 15 SE Osceola Ave, Ocala, FL 34471

²James A. Haley VA Hospital, 13000 Bruce B. Downs, Tampa, FL 33612

¹{abhatia, bdorr, kseitz}@ihmc.us

²{samuel.phillips, barbara.mckenzie}@va.gov

Abstract

Approximately 80% to 95% of patients with Amyotrophic Lateral Sclerosis (ALS) eventually develop speech impairments (Beukelman et al., 2011), such as defective articulation, slow laborious speech and hypernasality (Duffy, 2013). The relationship between impaired speech and asymptomatic speech may be seen as a *divergence* from a baseline. This relationship can be characterized in terms of measurable combinations of phonological characteristics that are indicative of the degree to which the two diverge. We demonstrate that divergence measurements based on phonological characteristics of speech correlate with physiological assessments of ALS. Speech-based assessments offer benefits over commonly-used physiological assessments in that they are inexpensive, non-intrusive, and do not require trained clinical personnel for administering and interpreting the results.

1 Introduction

Amyotrophic lateral sclerosis (ALS) or Lou Gehrig’s Disease, the most common form of motor neuron disease, is a rapidly progressive, neurodegenerative condition. It is characterized by muscle atrophy, muscle weakness, muscle spasticity, hyperreflexia, difficulty speaking (dysarthria), difficulty swallowing (dysphagia), and difficulty breathing (dyspnea). Mean survival time for ALS patients is three to five years from the time it is diagnosed; however, death may occur within months, or survival may last decades.

Most physiological assessments used to determine the functional status of patients with ALS are invasive, involving the use of expensive equipment

and requiring trained clinical personnel to administer the tests and interpret the results. This is the case for a number of standardized objective assessments of bulbar function in ALS patients (Green et al., 2013), for example: breathing patterns, articulatory patterns, and voice loudness. These are generally measured by technologies that record chest wall movements, oral pressures and flows, oral movement and strength, and speech acoustics.

This paper lays the foundation for the development of less invasive phonologically-inspired measures that correlate strongly with (more invasive) physiological measures of ALS. Speech impairments eventually affect 80% to 95% of patients with ALS (Beukelman et al., 2011). In fact, Yorkston et al. (1993) noted that speech impairments may be present up to 33 months prior to diagnosis of ALS. Several previous studies (Yunusova et al., 2016) have shown that speech impairments correlate with physiological changes associated with ALS. Thus, we focus on correlating measures based on phonological features with standard physiological measures, thus enabling new, non-invasive measures for assessing the functionality of an ALS patient without significant overhead for personnel training and administration.

To bring this about, we determine the degree of *divergence* of symptomatic speech from asymptomatic speech taken as a baseline.¹ This determination is based on phonological features in speech, most of which have been previously identified in the literature as being associated with ALS, e.g., *monoloudness*, *hypernasality* and *distorted vowels*, see (Duffy, 2013). These are annotated, for the current study, by specialists, i.e., a phonolo-

¹As part of a longitudinal study, we are exploring individual baselines for each ALS speaker: speech from each speaker’s first visit is taken as an individual baseline for the speaker.

gist and a speech therapist experienced in working with ALS patients. The degree of divergence is correlated with physiological assessments of ALS, namely %FVC (Forced Vital Capacity) in sitting (%FVC-SIT) as well as supine (%FVC-SUP) positions.²

The rest of the paper is organized as follows: In Section 2 we discuss related work that motivates and informs our research. Section 3 describes data used for our experiments. A discussion of speech divergence is presented in Section 4. Section 5 presents an assessment of the degree to which divergent characteristics in the speech match the level of progress of the ALS condition. This is followed by a discussion of future work and conclusions in Section 6.

2 Related Work

A number of past studies have investigated the utility of measuring the “voice signal” in order to answer questions about a speaker’s state from their speech (Schuller et al., 2015, 2011). One such study attempts to distinguish classes of individuals with various speech impairments, such as stuttering (Nöth et al., 2000), aphasia (Fraser et al., 2014), and developmental language disorders (Gorman et al., 2016). The recognition of impaired speech has been employed to detect Alzheimer’s (Rudzicz et al., 2014). Various speech-related features have been employed to detect whether the speech is affected by Parkinson’s Disease (Bocklet et al., 2011). Relatedly, variations in speech properties under intoxicated and sober conditions have also been conducted (Beadsy et al., 2011).

Our work differs from prior approaches in that we explore perceivable phonological characteristics through the analysis of language divergences. One of the motivations for using phonological features exclusively rather than also using other features employed in prior studies was that phonological features did not require expensive equipment to collect data from speakers as e.g., a feature like *maximum subglottal pressure* would require. Since the goal of this work is to develop a measure that is completely based on speech features that can be identified with a simple click on a

²%FVC-SUP refers to the percent value of the Forced Vital Capacity while the person is in supine position, and %FVC-SIT refers to the percent value of the Forced Vital Capacity while the person is in sitting position. See (Brinkmann et al., 1997; Czaplinski et al., 2006) for additional information about use of FVC in ALS assessments.

device such as a phone, we focused on phonological features on which a machine can be trained to analyze automatically. Our focus on correlations with phonological features—tied to the notion of divergence from a baseline—is a significant contribution beyond what has been investigated previously.

The notion of *divergence* itself is not a new one in natural language processing. The characterization of divergence classes (Dorr, 1994) has been at the heart of solutions to many different problems ranging from word alignment (Dorr et al., 2002) to machine translation (Habash and Dorr, 2002) to acquisition of semantic lexicons (Olsen et al., 1998). Finding the minimal primitive units—and determining their possible combinations—was the foundation for this earlier work. However, in these earlier studies, primitives consisted of properties that were syntactic, lexical, or semantic in nature, whereas the primitives for the current work consist of properties that are phonological in nature.

Beukelman et al. (2011), Duffy (2013), Green et al. (2013), and Orimaye et al. (2014) have established that pronunciation varies systematically within categories of speech impairment. (Silbergleit et al., 1997; Carrow et al., 1974) have shown that ALS speech shows deviant characteristics. For example, (Ball et al., 2001) observe that ALS speakers manifest altered voice quality. A number of speaker-level characteristics associated with impaired speech studied in prior work have been leveraged for our speech-related divergence detection. For example, Duffy (2013) specifically has enumerated speaker-level characteristics, such as *monopitch* and *monoloudness*.

Rong et al. (2015; 2016), and Yunusova et al. (2016) have previously attempted to identify measures of speech motor function for ALS speech. While certain components of speech such as speaking rate, breathing patterns, and voice loudness have proven too variable to provide a reliable marker (Green et al., 2013), we demonstrate that divergence measurements based on phonological characteristics of speech correlate with physiological assessments of ALS.

In addition to speaker-level characteristics and associated properties, our work defines divergence in terms of speech/span-level characteristics, as described in Section 3. Smaller vowel space areas have been found in ALS speech (Turner et al., 1995; Weismer et al., 2001) which suggests that

vowels may be distorted in ALS speech. Similarly, Kent et al (Kent et al., 1990) found place and manner of articulation for some consonants, and regulation of tongue height for vowels to diverge from asymptomatic speech; these were expected to result in imprecise consonants and distorted vowels. Caruso and Burton (1987) observed that ALS speakers and asymptomatic speakers exhibited significant differences in stop-gap durations as well as in vowel durations.

Yunusova et al. (2016) have also previously shown a correlation between speaking rate and physiological measures of ALS, specifically ALS Functional Rating Scale (ALSFRS).³ Our own work differs from this prior work in that we define *divergence* in terms of a wider range of speech characteristics—beyond speaking rate—and then demonstrate that divergence measures correlate with physiological measures of ALS.

3 Data: Transcriptions and Phonological Annotations

The data for our experiments consist of recorded speech of 16 recruited subjects with ALS in a clinical setting, collected quarterly for each subject. The subjects range between 35-74 years of age. Their age distribution is as follows: one subject in the 30s, two subjects in their 40s, one subject in their 50s, five subjects in their 60s, and seven subjects in their 70s. Out of the 16 subjects, only one of them is female, the other 15 subjects are male. In terms of race of the subjects, we have the following distribution: White (12), Asian (1), African-American (1), Not reported (2).

The criteria for the recruitment of a particular subject are that the subject: (1) has been diagnosed with ALS; (2) is a native monolingual speaker of American English; (3) has bulbar involvement identified during initial ALS inpatient evaluation; (4) has a forced vital capacity (FVC) of greater than 50% of the expected value for age; and (5) has an ALSFRS-R score⁴ of 40 or greater. Excluded from the study are those who have received a diagnosis of dementia, FVC of less than 50%, inability to speak, or inability to follow directions.

As part of their standard clinical care, each ALS-diagnosed subject reports for a quarterly

evaluation during which the following measures are collected: Forced Vital Capacity (FVC; (Brinkmann et al., 1997; Czaplinski et al., 2006)), Penetration-Aspiration Scale (a paper-pencil screen; (Rosenbek et al., 1996)), ALSFRS-R (Cedarbaum and Stambler, 1997; Cedarbaum et al., 1999), and Speech Intelligibility Test (SIT (Beukelman et al., 2011; Yorkston et al., 2007)).

- | |
|---|
| <ol style="list-style-type: none"> 1. She held your dark suit in greasy wash water all year. 2. Don't ask me to hold an oily rag like that. 3. The big dog loved to chew on the old rag doll. 4. Chocolate and roses never fail as a romantic gift. |
|---|

Table 1: Four TIMIT Sentences

Speech recordings of the same four sentences, that have been preselected, are made during each (quarterly) visit of each of the patients.⁵ The four sentences, presented in Table 1, are selected from the Texas-Instrument/MIT (TIMIT) corpus (Garofolo et al., 1993) and were designed to be phonetically rich, thus providing solid coverage of the phonetic space from each subject.⁶

The data also include recordings of four control (asymptomatic) subjects, two of whom are female and two are male, reading the same four TIMIT sentences in the same setting as the symptomatic subjects. These are used as the baseline speech against which divergence scores (defined in the next section) are calculated for the ALS symptomatic speech.

Our hypothesis is that a higher divergence is indicative of the progression of the ALS condition. This study focuses on divergence with respect to asymptomatic speech—taken as a baseline—to determine whether the divergence is speaker dependent or whether it is more generally indicative of ALS progression. If the latter, this would help diagnose patients for which no previous/longitudinal data is available.⁷

ALS speech data for the 16 subjects was transcribed and annotated via speech-analysis software called Praat (Boersma and van Heuven, 2001) for the 14 phonological characteristics enumerated in Table 2. These characteristics

⁵All uses of these data as reported in this paper have been approved by the relevant Institutional Review Board (IRB).

⁶Note the TIMIT sentences 1 and 2 are slightly different from the original TIMIT sentences; the original TIMIT sentences are as follows: (1) She **had** your dark suit in greasy wash water all year; (2) Don't ask me to **carry** an oily rag like that.

⁷However, as longitudinal data becomes available to us in our future work, we will also look at divergence based on speaker dependent baselines.

³ALSFRS is a standard questionnaire-based scale to measure functionality of a person in carrying out daily activities, see (Cedarbaum and Stambler, 1997; Brooks, 1997).

⁴ALSFRS-R is a revised ALSFRS that incorporates assessments of respiratory function (Cedarbaum et al., 1999).

Speaker level characteristics	
<i>monopitch</i>	Voice lacks inflectional changes; pitch does not change much.
<i>monoloudness</i>	Voice for which the volume/loudness does not change, hence lacking normal variations in loudness.
Speech/span related characteristics	
<i>harshness</i>	Voice seems harsh, rough and raspy—sometimes referred to as pressed voice—similar to what happens when a person talks while lifting a heavy load.
<i>imprecise consonants</i>	Consonant sounds lack precision. There may be slurring, inadequate sharpness, distortions, lack of crispness, and clumsiness in transitioning from one consonant to another. For example, a “w” may be produced instead of a “b”.
<i>distorted vowels</i>	Vowel sounds distorted throughout their total duration. For example, a “a” may be produced instead of “i”.
<i>prolonged phonemes</i>	A phoneme (i.e., a consonant or a vowel) is prolonged, i.e., its sound (when it is produced) continues over an unusual period of time.
<i>inappropriate silences</i>	Pauses that are produced not at syntactic or prosodic boundaries.
<i>hypernasality</i>	Vowels that are supposed to be non-nasalized are instead nasalized in speech.
<i>strained or strangled quality</i>	Tenseness in voice (as with overall muscular tension). Perceived as increased effort, may seem tense or harsh as if talking and lifting at the same time or as if talking with breath held.
<i>breathiness</i>	Voice seems breathy, weak and thin. May seem like a sighing sound. There may be non-modulated turbulence noise in the produced sound, i.e., audible air escape in voice or bursts of breathiness.
<i>audible inspiration/stridor</i>	Noisy breathing and wheezing may accompany inhaling. There may be a harsh, crowing, or vibratory sound of variable pitch resulting from turbulent air flow caused by partial obstruction of the respiratory passages.
<i>unusual stress</i>	Speech sounds where most important parts of a sentence may be de-stressed or all parts of a sentence are stressed as if all are important or speech sounds may be perceived as robotic, with the same stress—where there is no variation in stress throughout sentence/phrase/word/syllable.
<i>hoarseness</i>	Abnormal voice changes, where voice may sound breathy, raspy, strained, or there may be changes in volume (loudness) or pitch (how high or low the voice is).
<i>vocal fry</i>	Popping or rattling sound of a very low frequency—also known as a creaky voice.

Table 2: Phonological characteristics annotated in symptomatic speech

were pre-identified mostly based on the clinical research literature on ALS speech, e.g., see Duffy (2013): p248. The phonological annotations were made by two specialists: one of whom was a phonologist and the other was a speech therapist with experience working with ALS speakers.

Two classes of phonological characteristics served as the basis of annotations, each with a set of primitive phonological features: *speaker level characteristics* and *speech/span related characteristics*. *Speaker level characteristics* refer to features in speech that are more characteristic of a specific speaker’s voice—independent of individual sounds/spans, e.g., *monopitch* which indicates the lack of inflectional changes in voice. These were annotated only once for each speaker.

Speech/span related characteristics, on the other hand, refer to features in speech that are characteristic of a specific sound or are observed for a portion of speech—as opposed to features that are characteristic of the voice itself. For example, the feature *imprecise consonants* refers to the portion of speech where a specific consonant is produced imprecisely, it may involve slurring or inadequate sharpness, e.g., producing a “w” instead of a “b”. For these annotations, spans in speech were marked over which these features

were observed.

For each of these characteristics, the annotators also assigned a 1-10 Likert scale (Likert, 1932) rating to indicate the severity of the characteristic when it is observed, where 10 indicates “very severe” and 1 indicates “negligible.”

4 Divergence in Speech

Understanding the relationship between impaired speech and asymptomatic speech is facilitated by measuring the degree to which symptomatic speech diverges from a baseline. For the current study, asymptomatic speech—which serves as a baseline—was created from a combination of recordings from asymptomatic speakers as described in Section 3. Simplistically, the degree of divergence is defined as the sum of the changes in a speech utterance from its asymptomatic equivalent. For a correlation to be supported, a large number of changes in speech (i.e., a strong divergence from asymptomatic speech) would correspond to advanced progression of the disease. The relationship between impaired speech and asymptomatic speech is characterized in terms of measurable combinations of phonological characteristics that are indicative of the degree to which the two diverge. The degree of divergence can be used

as a diagnostic tool at regular intervals for checking the severity of physiological changes.

Multiple methods have been applied in order to calculate divergence scores:

1. **Feature count based divergence score:** *Feature count* refers to the number of characteristics observed in the speech samples.⁸ The *Feature count based divergence score* for each ALS speaker is the difference between the *feature count* for the ALS speech and the *feature count* for the control asymptomatic speech.

Four variations of this score are used based on how the *feature count* for the control asymptomatic speech is obtained:

(a) **Average feature count for controls:** It is assumed that asymptomatic speakers may display characteristics identified in Table 2 but to a much smaller extent. Thus, taking a simple average of the *feature count* for control speakers is taken to be most representative of all asymptomatic speakers. The *feature count* for the control asymptomatic speech is the average of the *feature count* for all the control speakers.

(b) **Minimum feature count for controls:** The control speaker with the minimum number of characteristics in his/her speech is assumed to be the most asymptomatic. Hence, the *feature count* for the control asymptomatic speech is the minimum of the *feature counts* for all control speakers.

(c) **Gender dependent average feature count for controls:** The presence (or absence) of characteristics may be dependent on the gender of the speaker. To calculate divergence scores, it is best if speakers of the same gender are compared. Hence, the *feature count* for the control asymptomatic speech is the average of the *feature count* for all the control speakers of the same gender as the ALS speaker.

(d) **Gender dependent minimum feature count for controls:** It is assumed the control speaker with the minimum number of characteristics in his/her speech is the most asymptomatic, but the presence (or absence) of characteristics may be gender dependent. To calculate divergence scores, it is best if speakers of the same gender

are compared. Hence, the *feature count* for the control asymptomatic speech is the minimum of the *feature count* for all the control speakers of the same gender as the ALS speaker.

2. **Total frequency based divergence score:** For each speaker, an *observed frequency score* is computed as an aggregate of the frequencies of all observed characteristics in the speech of the speaker.⁹ The average of the *observed frequency score* of both the annotators for a given speaker is taken as the *frequency score* for the speaker. The *divergence score* for an ALS speaker is the difference between the *frequency score* for the ALS speech and the *frequency score* for the control asymptomatic speech.

The same four variations of this score are examined as described in the case of *Feature count based divergence score*, depending on how the *frequency score* for the control asymptomatic speech is obtained.

3. **Likert Scale rating based divergence score:** It is assumed that each of the characteristics may be as indicative of the condition as other characteristics in various ALS speakers. It is also assumed that the severity of the characteristics indicates progression of ALS. An *observed Likert score of the speech samples from a speaker* is taken to be an aggregate of the multiples of Likert Scale rating assigned by an annotator for each occurrence of a characteristic with the weight of the characteristic (which is uniformly taken to be 1 for all the characteristics in the current analysis). A *Likert score for a speaker* is calculated as an average of the two annotators' *observed Likert scores of the speech samples from the speaker*. A *Likert Scale rating based divergence score* for each ALS speaker is then taken to be the difference between the *Likert score* for the ALS speech and the *Likert score* for the control asymptomatic speech. The same four variations of this score are examined as described in the case of *Feature count based divergence score*, depending on how the *Likert score* for the control asymptomatic speech is obtained.

For each of the three divergence measures defined above, a higher score indicates that the patients speech diverges from an asymptomatic

⁸The counts from the two annotators were combined together in five different ways described in Section 5.1.

⁹For example, if a characteristic, say *distorted vowels*, is observed 6 times, the frequency for *distorted vowels* is 6.

Spkr	%FVC-SUP	%FVC-SIT	Feature count	Total frequency	Likert scale
18	79	88	-0.75	-5	52.5
9	77	78	4.25	7	53.25
1	77	77	0.25	-2.5	37.75
6	75	79	0.5	0.5	29.5
14	66	41	4.5	10.25	83.5
10	53	64	7	13.75	98.75
5	52	56	7.5	22	127.25
4	51	44	8.75	29.5	189.75
8	50	56	5.75	13.75	53.25
17	50	52	4.25	16	150.25
2	38	78	9.25	87	593
11	32	26	6.25	7	82.75
13	29	46	9.25	36	263.46
19	29	40	3.5	1.25	52.5
7	29	29	6.75	36.5	297.25
12	25	26	4.75	12	123.25

Table 3: Physiological Scores (%FVC) and Divergence Scores (D.S.) variant (d) for average of all four utterances for ALS speakers. The feature count was based on a union of features across the two annotators. Total frequency and Likert scale values were computed as a maximum across the two annotators.

speech baseline more than would be indicated by a lower score. Divergence scores are expected to correlate with physiological measures of changes associated with ALS. Increasing divergence scores would thus serve as an indicator of the disease progression, analogous to decreasing physiological outputs (lower scores) associated with ALS—thus, the two measures are expected to be negatively correlated.

5 Results and Discussion

Table 3 presents two physiological assessment scores (%FVC-SUP and %FVC-SIT) and three divergence scores (defined above) for the 16 ALS speakers.¹⁰ The scores are sorted by %FVC-SUP. The table indicates that the %FVC scores tend to drop as the divergence scores go up. As expected, a decrease in %FVC scores indicates disease progression, and similarly, a higher divergence score indicates disease progression.

5.1 Dealing with differences in Annotations to Calculate Divergence Scores

Since the nature of the annotated phonological characteristics was such that multiple characteristics might share various aspects of speech, annotators were asked to mark all characteristics that

¹⁰Only the variant (d) for each of the divergence scores computed using the three methods is presented in the table to maintain clarity. Note variant (d) refers to the divergence scores calculated with *Gender dependent minimum feature count for controls* setting, as described in Section 4 above.

Divergence Score Type	Correlation	p-value
<i>Feature ct D.S.(d)-Union</i>	0.65	0.007
<i>Feature ct D.S.(d)-Avg</i>	0.58	0.017
<i>Feature ct D.S.(d)-Max</i>	0.58	0.018
<i>Feature ct D.S.(d)-Min</i>	0.58	0.019
<i>Feature ct D.S.(d)-Intersection</i>	0.51	0.045
<i>Likert D.S.(a)-Max</i>	0.49	0.055

Table 4: Correlations between the Physiological Scores (%FVC) and Divergence Scores (D.S.) for all four variants

seemed relevant to them. The general descriptions provided in Table 2 were used as heuristics by the annotators, providing additional help in identifying the characteristics.¹¹ In order to resolve differences across annotations, we used five different methods to combine the two sets of annotations. Table 3 shows a representative combination of the first case below for the feature count measure and the third case below for both total frequency measure and Likert scale measure:

1. *Union*: The characteristics identified by both the annotators were considered only once.
2. *Intersection*: Only the features annotated by both the annotators were considered.
3. *Max*: The maximum of the two annotators’ feature counts was used.
4. *Min*: Minimum of the two annotators’ feature counts was used.
5. *Avg*: An average of the two annotators’ feature counts was used.

5.2 Association between Divergence Scores and Physiological Scores

To determine whether there was an association between any or all of the divergence scores and the physiological measures of ALS, we correlated the divergence scores with the physiological assessment scores, %FVC-SUP and %FVC-SIT, using Pearson’s correlation coefficient. The results are presented in Table 4. For simplicity, we report the correlations in the table as $-1 * \langle correlation \rangle$. Refer Section A for correlations with all the divergence scores.

We observe that while divergence scores do not seem to correlate with the %FVC-SIT score, they do show a moderate correlation with the %FVC-

¹¹Although these descriptions were somewhat coarse-grained, the idea was to start at this level and to learn more precise features associated with acoustic inputs corresponding to these characteristics. These precise features are expected to be critical for automatic classification of speech samples with respect to ALS progression and, correspondingly, predictive of the physiological scores for patients.

SUP score ($0.49 < r < 0.66$) with moderate p -values ($p < 0.05$). The stronger correlation effect we observe with %FVC-SUP than with %FVC-SIT may be due to higher difficulty in breathing that a patient may experience when (s)he is in supine position than in sitting position.

Consistent with the point above, patients with other pulmonary conditions have also been reported to experience higher difficulty in breathing when in supine position than in sitting or standing positions. Since the patients need to exert higher effort to achieve the same result in supine position than in sitting position, they may not be physiologically able to perform the same in the two positions, i.e., %FVC-SUP may be more sensitive than %FVC-SIT to the condition's progression. Since speech symptoms have also been found to be more readily apparent than other physiological symptoms (Yorkston et al., 1993), this results in a stronger correlation of the speech divergence scores with %FVC-SUP than with %FVC-SIT.

The table also indicates that divergence scores based on a simple measure—counts of features observed in ALS speech—correlate even better with %FVC-SUP scores than divergence scores that are based on slightly more complicated measures such as features' frequency or the Likert Scale ratings.

6 Conclusion and Future Work

This paper has presented a case for viewing the relationship between impaired speech and asymptomatic speech as a *divergence* from a baseline. Novel divergence measures have been developed for distinguishing asymptomatic speech from symptomatic speech, and these have been tested for correlations with physiological measures of ALS progression.

These speech divergence measures are a first step toward developing automated speech-based assessments of progression of the ALS condition that are both less expensive and less intrusive than their physiological counterparts. The current approach has enabled the identification speech-based measures that correlate well with other physiological measures currently used to monitor the progression of the ALS condition. The next step is to test if these measures can be used to predict the values for the currently used physiological measures including %FVC.

Also, the current study is based on manual annotations provided by human specialist annotators. Future research will involve exploration of

approaches that can be trained to produce such annotations automatically. These could, in turn, be used to calculate divergence scores and eventually to predict values for other physiological measures.

The theoretical groundwork for developing speech-based measures defines speech divergence in terms of clinically-informed phonological speech characteristics associated with ALS symptomatic speech. We presented three methods, with four variants apiece, to compute speech divergence scores for symptomatic speech. We also showed that speech divergence scores are indeed correlated with physiological assessment scores for the progression of the disease.

Future research will investigate other methods to compute divergence between the symptomatic and asymptomatic speech that yield even stronger correlations with the physiological assessments measures. For example, it would be useful to explore whether the proportion of speech that is affected by the characteristics listed in Table 2 has any relation to the progression of the disease. Divergence scores that incorporate characteristics related to a proportion of the span are expected to be strongly correlated with the progression of ALS.

Two possible variants of how one may compute divergence scores based on such proportion-related information are as follows:

(1) Take *proportion* to be the proportion of speech that is affected by any of the characteristics.¹² One may calculate a divergence score for each ALS speaker as the difference between the *proportion* of speech of the ALS speaker affected by these characteristics and the *proportion* of controls' speech affected by these characteristics. An average of the *proportion* in the annotators' annotations may be used for the calculation of the divergence score.

(2) As a simple analytic, one may also consider proportion-based divergence scores corresponding to each of the characteristics for each ALS speaker. This analytic may be useful for providing a direct relation between a specific characteristic and the progression of the condition. However, it may also be useful to explore divergence classes based on groupings of characteristics that are similarly affected due to the progression of the condition, if any.¹³

¹²Note there may be overlapping spans for more than one characteristics.

¹³For the calculation of this variant, an average across the portions of speech for which a characteristic is annotated may

Some characteristics may be grouped to further explore divergences. Green et al. (2013) grouped features according to the speech subsystem involved (e.g., respiratory, phonatory, resonatory and articulatory). A reviewer also mentioned that gender-specific degree of severity of certain features would be interesting to explore. For example, there seems to be evidence that voicing control is more vulnerable in male patients (Kent et al., 1994). Such findings suggest that characteristics such as gender and possibly age may also need to be considered while developing speech divergence-based measures.

In addition, for the current study, each of the characteristics was treated uniformly with respect to ALS. Future work will explore the hypothesis that certain characteristics are more indicative than others with respect to the progression of ALS.

Finally, while prior studies indicate that prosodic recognition is not affected in ALS speakers (Zimmerman et al., 2007), articulatory or phonatory deficits might alter the correct production of interrogative, imperative, or declarative sentences (Congia et al., 1987). These may be found to be useful in the development of speech-based measures of ALS. Thus, future work will investigate the extent to which these variables would be more or less difficult to analyze automatically.

Acknowledgments

This work is supported in part by the VA Office of Rural Health Program “Increasing access to pulmonary function testing for rural veterans with ALS through at home testing” (N08-FY15Q1-S1-P01346) and Award Number 1I21RX001902 titled “DESIPHER: Speech Degradation as an Indicator of Physiological Degeneration in ALS.” Additionally, the work is supported by resources and the use of facilities at the James A. Haley Veterans’ Hospital. The contents of this paper represents solely the views of the authors and not the views of the Department of Veterans Affairs or the United States Government. The authors are also grateful for constructive feedback from anonymous reviewers. An additional thank you is extended to the two specialists for their time and expertise in providing us with annotations of the phonological features in ALS speech that enabled us to conduct the analyses reported herein.

be used to determine the *proportion* of speech for a specific speaker that is affected by the characteristics.

References

- L.J. Ball, A. Willis, D.R. Beukelman, and G.L. Pattee. 2001. A protocol for identification of early bulbar signs in als. *Journal of the Neurological Sciences* 191:43–53.
- D. Beukelman, S. Fager, and A. Nordness. 2011. Communication support for people with ALS. *Neurology Research International*.
- F. Biadsy, W.Y. Wang, and A. Rosenberg. 2011. Intoxication detection using phonetic, phonotactic and prosodic cues. In *Proceedings of INTERSPEECH*.
- T. Bocklet, E. Nöth, G. Stemmer, H. Ruzickova, and J. Rusz. 2011. Detection of persons with Parkinson’s disease by acoustic, vocal, and prosodic analysis. In *Proceedings of ASRU*.
- P. Boersma and V. van Heuven. 2001. Praat, a system for doing phonetics by computer. *Glott International* 5(9/10):341–345.
- J.R. Brinkmann, P. Andres, M. Mendoza, and M. Sanjak. 1997. Guidelines for the use and performance of quantitative outcome measures in als clinical trials. *Journal of the Neurological Sciences* 147(1):97–111.
- B.R. Brooks. 1997. Amyotrophic lateral sclerosis clinimetric scales guidelines for administration and scoring. In Herndon R., editor, *Handbook of clinical neurologic scales*, Demos Vermande, pages 27–80.
- E. Carrow, V. Rivera, M. Mauldin, and Shamblin L. 1974. Deviant speech characteristic in motor neuron disease. *Archives of Otolaryngology - Head and Neck Surgery Journal* 100:212–218.
- A.J. Caruso and E.K. Burton. 1987. Temporal acoustic measures of dysarthria associated with amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research* 30(1):80–87.
- J.M. Cedarbaum and N. Stambler. 1997. Performance of the amyotrophic lateral sclerosis functional rating scale (ALSFERS) in multicenter clinical trials. *Journal of Neurological Sciences* 152(Suppl 1):S1–S9.
- J.M. Cedarbaum, N. Stambler, E. Malta, C. Fuller, D. Hilt, B. Thurmond, and A. Nakanishi. 1999. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. *Journal of the Neurological Sciences* 169(1-2):13–21.
- S. Congia, S. Di Ninni, F. Zonza, and M. Melis. 1987. Prosody alterations in amyotrophic lateral sclerosis. In *Amyotrophic Lateral Sclerosis*, Springer US, pages 121–123.
- A. Czaplinski, A.A. Yen, and S.H. Appel. 2006. Forced vital capacity (fvc) as an indicator of survival and disease progression in an als clinic population. *Journal of Neurology, Neurosurgery, and Psychiatry* 77(3):390–392.
- B.J. Dorr. 1994. Machine translation divergences: A formal description and proposed solution. *Computational Linguistics* 20(4):597–633.
- B.J. Dorr, L. Pearl, R. Hwa, and N. Habash. 2002. DUSTER: A method for unraveling cross-language

- divergences for statistical word-level alignment. In *Proceedings of the Fifth Conference of the Association for Machine Translation in the Americas (AMTA)*. pages 31–43.
- J. Duffy. 2013. *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. Elsevier Health Sciences, 3rd edition.
- K.C. Fraser, G. Hirst, N.L. Graham, J.A. Meltzer, S.E. Black, and E. Rochon. 2014. Comparison of different feature sets for identification of variants in progressive aphasia. In *Proceedings of the ACL CLPsych Workshop*. pages 17–26.
- J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren, and V. Zue. 1993. TIMIT acoustic-phonetic continuous speech corpus. http://web.mit.edu/course/6/6.863/share/nltk_lite/timit/allsenlist.txt.
- K. Gorman, L. Olson, A.P. Hill, R. Lunsford, P.A. Heeman, and J.P.H. van Santen. 2016. Uh and um in children with autism spectrum disorders or language impairment. *Autism Research* 9(8):854–865.
- J.R. Green, Y. Yunusova, M.S. Kuruvilla, J. Wang, G.L. Pattee, L. Synhorst, L. Zinman, and J.D. Berry. 2013. Bulbar and speech motor assessment in ALS: Challenges and future directions. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* 14:494–500.
- N. Habash and B.J. Dorr. 2002. Handling translation divergences: Combining statistical and symbolic techniques in generation-heavy machine translation. In *Proceedings of the Fifth Conference of the Association for Machine Translation in the Americas, AMTA-2002*. pages 84–93.
- R.D. Kent, J.F. Kent, G. Weismer, R.L. Sufit, J.C. Rosenbek, R.E. Martin, and B.R. Brooks. 1990. Impairment of speech intelligibility in men with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Disorders* 55:721–728.
- R.D. Kent, H. Kim, G. Weismer, J.F. Kent, J.C. Rosenbek, B.R. Brooks, and M. Workinger. 1994. Laryngeal dysfunction in neurological disease: Amyotrophic lateral sclerosis, parkinson disease, and stroke. *Journal of Medical Speech-Language Pathology* 2(3).
- R. Likert. 1932. A technique for the measurement of attitudes. *Archives of Psychology* 140:1–55.
- E. Nöth, H. Niemann, T. Haderlein, M. Decher, U. Eysholgt, F. Rosanowski, and T. Wittenberg. 2000. Automatic stuttering recognition using hidden Markov models. In *Proceedings of ICSLP*.
- M.B. Olsen, B.J. Dorr, and S. Thomas. 1998. Enhancing automatic acquisition of thematic structure in a large-scale lexicon for mandarin chinese. In *Proceedings of the Fifth Conference of the Association for Machine Translation in the Americas, AMTA-2002*. Langhorne, PA, pages 41–50.
- S.O. Orimaye, J. Wong, and K.J. Golden. 2014. Learning predictive linguistic features for Alzheimer’s disease and related dementias using verbal utterances. In *Proceedings of the ACL Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. pages 78–87.
- P. Rong, Y. Yunusova, J. Wang, and J.R. Green. 2015. Predicting early bulbar decline in amyotrophic lateral sclerosis: A speech subsystem approach. *Behavioral Neurology*.
- P. Rong, Y. Yunusova, J. Wang, L. Zinman, G.L. Pattee, J.D. Berry, B. Perry, and J.R. Green. 2016. Predicting speech intelligibility decline in amyotrophic lateral sclerosis based on the deterioration of individual speech subsystems. *PLoS ONE* 11(5).
- J.C. Rosenbek, J.A. Robbins, E.B. Roecker, J.L. Coyle, and J.L. Wood. 1996. A penetration-aspiration scale. *Dysphagia* 11(2):93–98.
- F. Rudzicz, L. Chan Currie, A. Danks, T. Mehta, and S. Zhao. 2014. Automatically identifying trouble-indicating speech behaviors in Alzheimer’s disease. In *Proceedings of ASSETS*. ACM, pages 241–242.
- B. Schuller, S. Steidl, A. Batliner, S. Hankte, F. Hönl, J.R. Orozco-Aroyave, E. Nöth, Y. Zhang, and F. Weninger. 2015. The interspeech 2015 computational paralinguistics challenge: nativeness, parkinson’s & eating condition.
- B. Schuller, S. Steidl, A. Batliner, F. Schiel, and J. Krajewski. 2011. The INTERSPEECH 2011 speaker state challenge. In *Proceedings of INTERSPEECH*. pages 3201–3204.
- A.K. Silbergleit, A.F. Johnson, and B.H. Jacobson. 1997. Acousticanalysis of voice in individuals with amyotrophic lateral sclerosis and perceptually normal vocal quality. *Journal of Voice* 11:222–231.
- G.S. Turner, K. Tjaden, and G. Weismer. 1995. The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research* 38:1001–1013.
- G. Weismer, J.-Y. Jeng, J.S. Laues, R.D. Kent, and J.F. Kent. 2001. Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. *Folia Phoniatr Logop* 53:1–18.
- K. Yorkston, D. Beukelman, M. Hakel, and M. Dorsey. 2007. Sentence intelligibility test, speech intelligibility test.
- K.M. Yorkston, E. Strand, R. Miller, A. Hillel, and K. Smith. 1993. Speech deterioration in amyotrophic lateral sclerosis: implications for the timing of intervention. *Journal of medical speech-language pathology* 1:35–46.
- Y. Yunusova, N.L. Graham, S. Shellikeri, K. Phuong, M. Kulkarni, E. Rochon, D.F. Tang-Wai, T.W. Chow, S.E. Black, L.H. Zinman, and J.R. Green. 2016. Profiling speech and pausing in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). *PLoS ONE* 11(1).
- E.K. Zimmerman, P.J. Eslinger, Z. Simmons, and A.M. Barrett. 2007. Emotional perception deficits in amyotrophic lateral sclerosis. *Cognitive and Behavioral Neurology* 20(2):70–82.

A Supplemental Material

The correlation results between each of the three types of speech divergence scores with all of their variants and the %FVC-SUP are presented in Tables 5, 6, and 7. As mentioned before, for simplicity, we report the correlations in the tables as $-1 * \langle correlation \rangle$.

Divergence Score Type	Correlation	p-value
<i>Feature ct D.S.(a)-Max</i>	0.58	0.018
<i>Feature ct D.S.(a)-Min</i>	0.58	0.019
<i>Feature ct D.S.(a)-Ave</i>	0.58	0.017
<i>Feature ct D.S.(a)-Union</i>	0.65	0.007
<i>Feature ct D.S.(a)-Intersection</i>	0.51	0.045
<i>Feature ct D.S.(b)-Max</i>	0.58	0.018
<i>Feature ct D.S.(b)-Min</i>	0.58	0.019
<i>Feature ct D.S.(b)-Ave</i>	0.58	0.017
<i>Feature ct D.S.(b)-Union</i>	0.65	0.007
<i>Feature ct D.S.(b)-Intersection</i>	0.51	0.045
<i>Feature ct D.S.(c)-Max</i>	0.58	0.018
<i>Feature ct D.S.(c)-Min</i>	0.58	0.019
<i>Feature ct D.S.(c)-Ave</i>	0.58	0.017
<i>Feature ct D.S.(c)-Union</i>	0.64	0.007
<i>Feature ct D.S.(c)-Intersection</i>	0.50	0.045
<i>Feature ct D.S.(d)-Max</i>	0.58	0.018
<i>Feature ct D.S.(d)-Min</i>	0.58	0.019
<i>Feature ct D.S.(d)-Avg</i>	0.58	0.017
<i>Feature ct D.S.(d)-Union</i>	0.65	0.007
<i>Feature ct D.S.(d)-Intersection</i>	0.51	0.045

Table 5: Correlations between the Physiological Scores (%FVC) and Feature Count Based Divergence Scores (D.S.) for all four variants

Divergence Score Type	Correlation	p-value
<i>Feature freq D.S.(a)-Max</i>	0.45	0.077
<i>Feature freq D.S.(a)-Min</i>	0.42	0.103
<i>Feature freq D.S.(a)-Ave</i>	0.44	0.085
<i>Feature freq D.S.(b)-Max</i>	0.45	0.077
<i>Feature freq D.S.(b)-Min</i>	0.42	0.103
<i>Feature freq D.S.(b)-Ave</i>	0.44	0.085
<i>Feature freq D.S.(c)-Max</i>	0.45	0.083
<i>Feature freq D.S.(c)-Min</i>	0.42	0.106
<i>Feature freq D.S.(c)-Ave</i>	0.44	0.09
<i>Feature freq D.S.(d)-Max</i>	0.46	0.075
<i>Feature freq D.S.(d)-Min</i>	0.42	0.103
<i>Feature freq D.S.(d)-Avg</i>	0.45	0.083

Table 6: Correlations between the Physiological Scores (%FVC) and Feature Frequency Based Divergence Scores (D.S.) for all four variants

Divergence Score Type	Correlation	p-value
<i>Likert Scale D.S.(a)-Max</i>	0.49	0.055
<i>Likert Scale D.S.(a)-Min</i>	0.44	0.089
<i>Likert Scale D.S.(a)-Ave</i>	0.47	0.068
<i>Likert Scale D.S.(b)-Max</i>	0.48	0.061
<i>Likert Scale D.S.(b)-Min</i>	0.43	0.095
<i>Likert Scale D.S.(b)-Ave</i>	0.46	0.074
<i>Likert Scale D.S.(c)-Max</i>	0.48	0.059
<i>Likert Scale D.S.(c)-Min</i>	0.43	0.094
<i>Likert Scale D.S.(c)-Ave</i>	0.46	0.071
<i>Likert Scale D.S.(d)-Max</i>	0.47	0.067
<i>Likert Scale D.S.(d)-Min</i>	0.43	0.094
<i>Likert Scale D.S.(d)-Avg</i>	0.46	0.076

Table 7: Correlations between the Physiological Scores (%FVC) and Likert Scale Based Divergence Scores (D.S.) for all four variants