CARNEGIE MELLON UNIVERSITY

SCS Honors Undergraduate Research Thesis Final Report

Vocal Fold Dynamics for Automatic Detection of Amyotrophic Lateral Sclerosis from Voice

by

Jiayi (Maggie) Zhang
Computational Biology Department

supervised by

Dr. Rita Singh
Language Technology Institute

April 2022

(This is page is intentionally left blank.)

During the course of this project, I received suggestions and support from many people, so I would like to dedicate a space to express my sincere appreciation. I would like to first express my deep gratitude to Dr. Rita Singh, my thesis mentor, who shared her expertise and provided me with extensive mentorship. I would also like to thank Dr. Marcelo Magnasco (the Rockefeller University) for his guidance on a precursory project as well as for offering access to the dataset of the current project. In addition, I am thankful to Dr. Christina Bjorndahl (CMU DC, Department of Philosophy), Dr. Phillip Compeau (CMU SCS, Computational Biology Department), Dr. Or Alus (the Rockefeller University) and Dr. Liat Shenhav (the Rockefeller University) for their comments, advice, and help. Finally, I would like to mention my friends and family, who inspired and motivated me with their constant, warm encouragements. Thank you all–I would not have been able to finish this project without you:)

Abstract

Amyotrophic Lateral Sclerosis (ALS) is a progressive neurodegenerative disease. Current diagnostic methods for ALS are complicated and rely on subjective judgements from physicians. This situation motivates the development of an expedient and objective diagnostic aid. Since ALS affects motor neurons and causes dysfunctions in speech and respiration, we hypothesized that analyses of features that capture the essential characteristics of the biomechanical process of voice production can successfully distinguish ALS patients from non-ALS controls. We focus on representing voices with algorithmically estimated vocal fold dynamics from physical models of phonation and aim to validate our hypothesis by identifying a set of features that are effective for our desired separation. To achieve our goal, we have explored 2 main sets of features: simple statistical measurements (Set 1) and phase-space characterizations (Set 2) of estimated vocal fold displacements and range of displacements. Random Forest Classifiers based on these features were used to differentiate the voices of ALS and non-ALS individuals. In 10-way cross-validation experiments, classifiers with Set 1 and Set 2 features yielded average AUC-ROC of 99.6% and 82.3%, respectively. These results demonstrate the potential of the use of vocal fold dynamics in detecting ALS from voice recordings.

Contents

1

Problem Introduction

Voice is the result of complex interactions between cognitive and biomechanical processes that control the physical structures and airflow dynamics of the vocal tract. It carries enormous information about the physical, physiological, medical and other states of the speaker. We might guess that someone is tired by hearing an increase of "vocal fry" (linguistically known as "creak") in their speech; we might notice if someone has just contracted the flu or a common cold by detecting an increase of nasality. Researchers and clinicians, knowing how changes in physical states or properties could potentially interact with voice production, have been exploring a wide range of vocal acoustic biomarkers for diagnoses or progression-tracking of several diseases that are connected to speech or motor control (3; 35; 25; 22). Examples of such features include fundamental frequency (F0), jitter, shimmer, nasality, breathiness, etc. (36; 22) and the repertoire is still growing over time.

The current project falls well within the field of disease detection through vocal acoustic biomarkers. Specifically, we are interested in exploring the use of estimated vocal fold dynamics for automatically detecting Amyotrophic Lateral Sclerosis (ALS). ALS is a progressive, neurodegenerative disease (23). Currently, the diagnosis of ALS is time-consuming and complicated, requiring clinical tests conducted by experienced medical practitioners (23; 30). In addition, the existence of definitive diagnostic biomarkers has not been indicated in the

current literature (23). These situations suggest a need for an efficient, objective, and reliable diagnostic aid for ALS. A diagnostic tool based on vocal acoustic biomarkers fits these criteria—the raw data (usually voice recordings) can be acquired cheaply and non-intrusively. The analyses of such recordings would involve automatic extractions of features that are both highly informative and computationally efficient. Upon examining the case studies and reports of common speech pathology in ALS patients (37; 5; 33) as well as prior studies on automatic ALS detection through voice, we decided to explore this open problem and expand on previous efforts. We hypothesized that 1) *ALS, as a motor-neuron disease, affects the motion and range of motion of the vocal fold during phonation* 2) *any analysis that captures the essential characteristics of the biomechanical process of voice production can effectively and reliably differentiate the voices of ALS patients from non-ALS individuals.* In this work, we set out to investigate features that are based on algorithmically estimated vocal fold dynamics data from speech recordings of ALS patients and non-ALS controls. To the best of our knowledge, features regarding vocal fold dynamics have not been used in the literature for ALS detection from voice. We aim to fill this gap by testing the viability of such a method as well as devising features that are especially relevant for ALS detection. Since our analysis is based on physical models of voice production and deals with estimated vocal fold dynamics directly, our results would naturally have high physiological interpretability, which is essential for clinical diagnosis (but sometimes hard to achieve), as mentioned in the existing literature.

2

Background

2.1 Human voice & disease detection through voices

It's often overlooked that the production of voice (or speech sounds) is the result of a fine choreography of the respiratory functions and the motion of the articulators, with high spatial and temporal precision. Air is pumped out of the lungs of the speaker and subsequently passes through the vocal folds, generating the source signal. Undoubtedly, the vibration of the vocal folds is the fastest motion in the entire human body. Though the vibration itself is self-sustained and driven by myoelastic aerodynamic forces, it can be and is almost always modulated by the laryngeal muscles that control the states (i.e., abduction vs. adduction) of the vocal folds. The source signal is then filtered by the cavities of the vocal tract and shaped by articulators (the tongue, the teeth, the palate, etc.). Even slight alterations in the configuration of the vocal tract can yield perceptually noticeable changes in the output, filtered signal.

Naturally, voices carry enormous information about the physical states of the speakers (36). In the context of daily life, there are utterances like "you sound tired" or "you sound like you've caught a cold," which suggest that we humans are capable of detecting information regarding physical states in voices simultaneously when we parse the content information of the speech. Historically, voices were used as one of the diagnostic metrics in traditional Chinese medicine.

In more recent years, with the advancements in signal analyses and computational methods, a diverse body of literature on automatic disease detection through voices has emerged. The diseases of interest usually involve *respiratory pathology*, e.g., COVID-19 (12; 3; 13), chronic pharyngitis (25), asthma (35); or affects *the motor system*, e.g., Parkinson's Disease (44; 20) and Alzheimer's Disease (35; 27). Some other studies have also explored the possibility of detecting coronary heart disease (31; 18) as well as Type II diabetes (9) through vocal biomarkers.

With the advancements in fields such as bio-acoustic, signal processing, linguistics, etc., most of the biomarkers in these diagnosis studies are automatically extracted and analyzed. Automatic methods support more objective and efficient measures, which is beneficial especially when the medical tests that are necessary to arrive at a diagnosis are hard to conduct or require substantial experience from clinicians. In addition, voice recordings—the raw data from which the biomarkers are extracted—can be obtained reasonably easily, without discomforting the patients. This implies that we might be able to repeatedly take measurements at the point of care; we might even be able to track disease progression through these biomarkers. Though we are yet to see the popularity and rigorous use of voice biomarkers in clinical practice, their potential is increasingly recognized.

2.2 Amyotrophic Lateral Sclerosis (ALS)

Amyotrophic Lateral Sclerosis (ALS) is a fatal, idiopathic, progressive neurodegenerative disease (23). It affects the upper and lower motor neurons (UMN and LMN) (23), resulting in an array of symptoms, such as limb atrophies, dysarthria and respiratory difficulties, dementia (in some cases), etc. Though there is considerable clinical heterogeneity with regard to symptoms and disease progression, most patients are presented with limb weakening or bulbar associated pathologies (such as "slurred speech," difficulty in swallowing) at onset (23), and the wasting and paralysis will gradually spread to other regions of their bodies, leading to respiratory failure and eventual death. The prognoses for ALS patients are bleak—despite continuous efforts in novel drug and therapy development since the documentation of the disease some 150 years ago (42), the present life expectancy for 50% of patients is within 30 months after the onset of symptoms (23).

Given the lack of clear mechanisms for parthenogenesis, heterogeneity across patients, and the existence of multiple confusing diseases (such as Parkinson's disease), diagnostics (and therapeutics) of ALS have proven to be difficult. *Currently, there is no definitive diagnostic test or biomarker for ALS* (23; 30). Medical practitioners resort to clinal diagnostic criteria—the El Escorial criteria (23)—and assess diagnostic certainty based on identification of certain

UMN and LMN signs (23; 1; 10). Though the El Escorial criteria is specific and has helped to standardize the diagnostic process, its reported sensitivity is not ideal, with a 30% of false-negative rate even after revision (1). Other tests (mostly based on electromyography results) have been reported to increase diagnostic sensitivity (10) but are still waiting to be formally incorporated into the criteria. Overall, current diagnostics of ALS is complicated and relies on various clinical tests that require physicians' subjective judgements (thus experience). As a result, it is time-consuming (median time to diagnosis=14 months (23)), with a relatively low sensitivity, which prevents or delays valuable early treatment and intervention that can improve patients' prognoses and life quality. *Naturally, clinicians are in the search for more efficient, relatively objective, and standardized methods (23) to advance the diagnostic process.*

2.3 Prior studies on ALS detection through voice

In this section, I will first provide an overview of prior studies on ALS detection through voice (and or speech[1]). Readers that are unfamiliar with vocal signal and/or acoustic analysis might find some of the features and terms mentioned in the overview confusing, so I will present a list of common features used in these studies that are most germane to this current project and comment on their (physiological) relevance. Finally, I will briefly discuss methods and results from my summer project, which utilized features regarding hyper-resolution spectrograms of voices as the basis of classification.

2.3.1 Published studies and literature reviews

Previous studies have explored many features derived from voices and speech, including but not limited to physical properties of sounds (acoustic measures), pausing pattern analyses (measures of respiratory function) (4; 34), word choices in question-answering tasks (semantic measures), the intelligibility of utterances (2), etc. For example, Tomik et al. (2015) (37), a case-control study (N_{case}=17, $N_{control}$=60), examined features regarding "perceptual assessments of voice," measurements with videolaryngostroboscopy, percentage jitter (%jitter), percentage shimmer (%shimmer), and noise-to-harmonic ratio (NHR) (37). They found that most of the patients exhibit hoarseness, roughness, and breathiness in their voices as well as abnormalities in the range of vibration of fold folds, glottal closure, etc. (37). They reported a significant increase in jitter in patients compared to the control group (37). Allison et al.

[1]In daily conversations, "voice" and "speech" are usually interchangeable concepts. Here, by "voice," I am referring to phonated segments that are not necessarily meaningful, (e.g. prolonged phonation of a vowel); by "speech," I am referring to segments of voice that contain semantic meaning.

(2017) (4), a case-control study (N_{case}=36, $N_{control}$=17), also examined both perceptual and instrumental speech analyses. They found that the percent of pause time has the highest AUC-ROC for separating patients from non-patients, compared to other acoustic measurements such as maximum fundamental frequency, nasality, maximum lip opening velocity, etc. (4). Norel et al. (2018) (29), a case-control study (N_{case}=67, $N_{control}$=56), utilized an existing, open-source speech analysis toolkit (openSMILE) and extracted Mel-frequency Cepstral Coefficients (MFCC), Linear Predictive Coding Coefficients (LPCC), spectral entropy, features measuring randomness/irregularity in speech signal spectra, etc. (29). While some MFCCs were found to be most indicative of ALS in females, LPCC features reflecting spectral entropy, variance ranked top for males (29). A linear SVM achieved 79% and 83% accuracy for males and females respectively in a "leave- five-subjects-out" cross-validation scheme (29). For more information on prior studies and compilation of features, readers can refer to review articles such as Chiaramonte and Bonfiglio (2020) (7) as well as Vieira et al (2019) (38).

Some of the measures, e.g. pausing pattern analyses, though having been reported to show high specificity in detecting bulbar symptoms of ALS (34), require specific instruments for measurement. Such data and equipment were not readily accessible to us. Similarly, analyses of semantic features require certain contextual richness in speech datasets as well as in-depth topic-specific knowledge. While these are potential features of interest for future studies, I do not consider them in this project. In the following paragraphs, I will briefly describe and discuss some broad acoustic feature types that I have noticed in the papers I am surveyed. By doing so, I wish to provide readers of this report with some ideas about the measures, and more importantly, the types of speech abnormalities that have interested researchers working on the same problem.

One type of commonly used features is measurements of *jitter*. Jitter reflects variations in the time intervals between glottal pulses (glottal vibration cycles). Healthy speech (especially in modal phonation) usually shows a very low %jitter while pathological speech might show a drastic increase in jitter. Perhaps because they are fundamental and easy to obtain, jitter measurements are broadly used in the detection and diagnosis of voice-related diseases. Jitter measurements appeared in an array of ALS detection studies such as Chiaramonte et al. (2019) (8), Tomik et al. (2015) (37), Wang et al. (2016) (39), Xie et al. (2014) (41), etc. Similarly, measurements of *shimmer* are also widely used. Shimmer is variations in the amplitude between glottal pulses. Since jitter and shimmer require virtually the same analyses (namely finding glottal pulses from waveforms), most studies that utilize jitter measurements would also explore shimmer measurements.

Both jitter and shimmer measurements can capture the pulse-to-pulse variations as a result of abnormal vocal fold motions. However, they do not provide us insights into the physiological

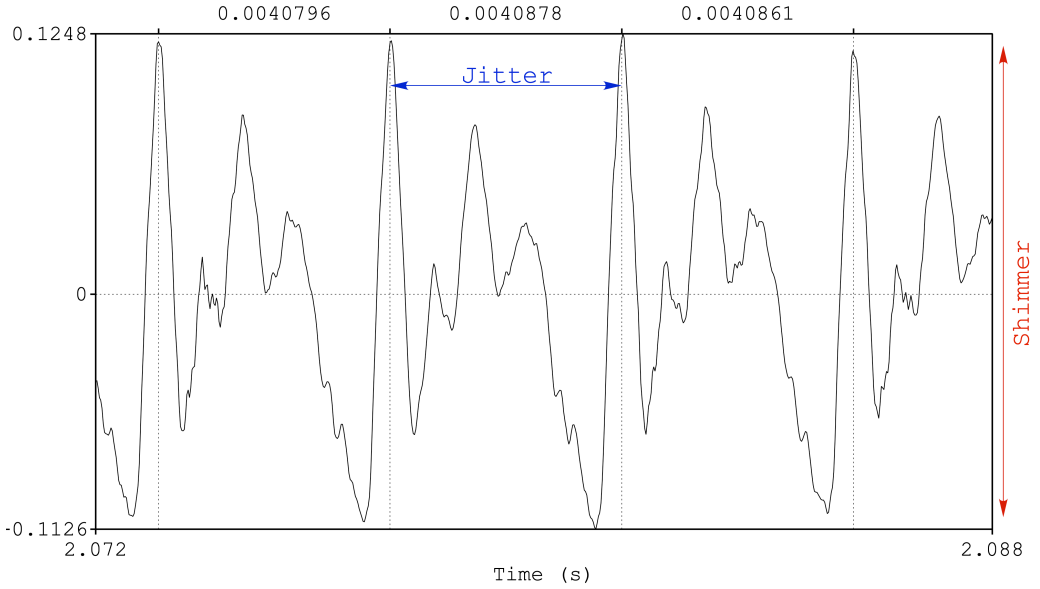

FIGURE 2.1: A portion of the waveform of a recording of prolonged phonation of [a] ("ah"), generated and annotated with *Praat* (6). The peaks of glottal pulses are indicated. As described before, jitter concerns time interval differences while shimmer concerns amplitude differences across subsequent glottal pulses.

mechanism behind the abnormality.

Many other studies also looked at measurements of *Fundamental Frequency (F0)*. F0 is the rate of vocal fold vibration. Its acoustic correlate to "pitch" (how "high" or "low" a sound is). Features regarding F0 can be F0 itself (21), maximum F0 in an utterance, range of F0 over time (29), variation of F0 (37), etc. Lowered or heightened F0, as well as high variation of F0 in declarative sentences or phonation tasks, could indicate pathology of the underlying biomechanical system. Interestingly, deviant, lowered F0 of ALS patients, noted in Agurto et al. (2019) (2), is related to the observation made in my summer project on the same subject (see section 3.2.2 below).

Features that are adopted from normal speech analyses are also used. For example, *Mel-Frequency Cepstral Coefficients (MFCCs)* and *Linear Predictive Coding Coefficients (LPCCs)* were used in Norel et al. (2018) (29) and Agurto et al. (2019) (2). These features are well-studied in the context of speech analyses and coding and are convenient to researchers since implementations can be found in many open-source vocal analysis toolkits. I do want to point out a potential limitation of MFCCs. Mel-scale is a non-linear scale derived from human listeners' perception of equal distance of pitch. Since the previous studies are focused on the automatic separation of speech between ALS and non-ALS individuals, representing speech in Mel-scale might (unnecessarily) mask important cepstral information of the original signal.

The goal of the discussion above is to familiarize the readers with the existing studies on the automatic detection of ALS through voice. These studies each investigated (a combination of) features and showed the relevance and effectiveness of vocal biomarkers with promising results and statistically significant features. Still, we do not have a standardized protocol and there are a number of gaps that needs to be filled (38), one of them being the continuation of the discovery of effective features. This situation prompted me to explore and test novel features, preferably ones that have high physiological interpretability. My first attempt, prior to conducting thesis research, is briefly documented in the next section.

2.3.2 Prior summer project

In the summer of 2021, the student researcher conducted a short research project on automatic ALS detection through hyper-resolution spectrograms in the Magnasco Lab at the Rockefeller University. The goal of this project is to explore features, automatically extracted from hyper-resolution spectrograms that can reliably separate ALS and non-ALS voice recordings. Spectrograms is one common way to represent audio signals that show the frequency compositions (along with their amplitudes) of the signal over time. Hyper-resolution spectrograms, developed in the Magnasco Lab, is a type of re-assigned spectrograms that offer greater temporal and spatial resolution (17) compared to conventional spectrograms generated through Short-Time Fourier Transform alone.

The dataset used consisted of 52 samples of prolonged phonation of [a] ("ah"), 39 of which belonged to confirmed patients of ALS and 13 from non-ALS controls who reported no existing respiratory nor neuro-motor diseases or conditions. The voice recordings of cases are a subset from a larger collection of ALS patient recordings, accessed through the Magnasco Lab. Hyper-resolution spectrograms with 3 different levels of time-resolution were generated for each recording.

A total of 17 features were used: 15 of them were extracted directly from the spectrograms (5 distinct features $\times$ 3 time-resolution) and another 2 were biological gender of the speaker and length of phonation. The 5 distinct acoustic features were:

1. average number of high energy pixels across columns (time steps)
2. count of vertical connected components (above a size threshold)
3. average 1st spectral moment ("center of gravity" of energy distribution)
4. median of 1st spectral moments
5. variance of 1st spectral moments.

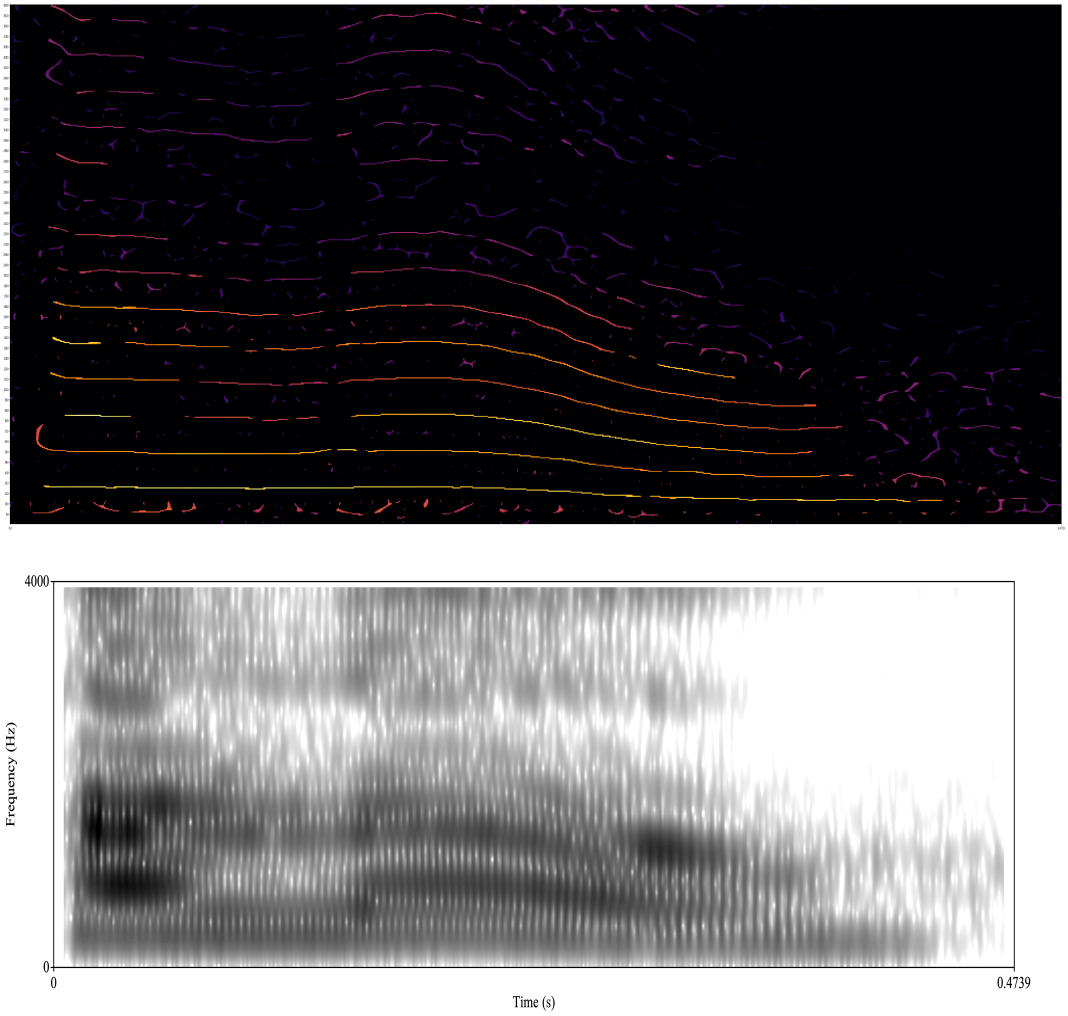

FIGURE 2.2: A hyper-resolution spectrogram (top) and a conventional broad-band spectrogram generated with *Praat* (6) (bottom) of the same recording of "Hello."

Features 1 and 2 were designed to capture the amount of "creakiness" in voices. Upon listening to the recordings, I noticed common and persistent creakiness in the case recordings. Creak is usually produced when vocal folds vibrate irregularly at lower frequencies. Typically, when speakers reach and maintain a pitch that is lower than their natural register, they would sound "creaky." Though creaks might appear in any person's speech, persistent and pronounced creaks tend not to happen in the phonation of lax vowels (for example, [a] used in our dataset). Therefore, I hypothesized measures reflecting "creakiness" would be effective markers for separation. In the hyper-resolution spectrograms, creak is represented by a series of pulse trains, which are large, vertical connected components (smooth, more tonal voices, on the other hand, are represented by horizontally oriented connected components that span the time domain). Thus, the average number of high-energy pixels across columns and the count of large vertical connected components were devised as features. Features 3-5 reflect

the distribution of energy as well as the stability (or consistency) of the distribution. The feature regarding speaker gender was included because previous studies (2; 37) noted potential heterogeneity between female and male patients in terms of acoustic features.

A Random Forest Classifier trained on 70% of the data yielded an average ROC-AUC of 87.5% in a 10-fold cross-validation trial, which is comparable to many of the existing studies. Variance and median of 1st spectral moments, as well as the average number of light pixels, ranked the highest among all the features.

2.4 Summary

With the background information described above, it should be clear that the human voice carries important information about the physical state of the speaker and can serve as indicators and aids in the diagnosis of certain diseases. Studies regarding the automatic detection of ALS through voice have explored many vocal features, and the field is still actively searching for better biomarkers for this complicated and devastating neuro-degenerative disease. As such, this current project aims to investigate the effectiveness of novel features and apply techniques developed for vocal analyses to ALS.

3 Method

3.1 Estimated Vocal Fold Dynamics

Among the many possible representations of voices, this project focuses on *vocal fold dynamics that are algorithmically estimated using physical models of voice production.* Specifically in this project, we represent voices as vocal fold oscillation trajectories (displacements and velocities) derived from physical models, with parameters fitted for each voice. As mentioned in previous sections, this representation is not employed in the ALS detection literature. However, novelty is not the sole reason that this method is chosen. Most of the previous features, such as measurements of jitter and shimmer, give us some information about the variations of the glottal pulses as a result of vocal fold oscillations. With estimated vocal fold oscillation data, we can directly infer the actual motion of the vocal folds and yield results with greater physiological relevance, which is crucial to our disease-diagnosis-oriented goal.

3.1.1 Physical models of phonation

The high-level problem for generating estimated vocal fold dynamics data is to solve a well-chosen physical model with appropriate parameters. In bio-acoustics studies, glottal (i.e.,

vocal fold) dynamics are usually characterized by "mass-(damper)-string" oscillator models. These models might be 1-mass (16; 26), 2-mass (19; 28; 11), or multi-mass (43) and parameterize them to capture the biophysical properties of the vocal folds—mass, elasticity, volume, etc. Some models assume symmetry in motion, meaning the left and right vocal folds are in phase with each other and oscillate in synchrony. Others incorporate parameters to allow asymmetry in motion. For this project, a single-mass model by Lucero et al. (2015) (26) that is *capable of capturing asymmetric vocal fold oscillation* is chosen. Lucero et al. proposed a set of coupled, nonlinear equations based on van der Pol oscillators (26):

$$\ddot{x}_r + \beta(1 + x_r^2)\dot{x}_r + x_r - (\Delta/2)x_r = \alpha(\dot{x}_r + \dot{x}_l)$$
$$\ddot{x}_l + \beta(1 + x_l^2)\dot{x}_l + x_l - (\Delta/2)x_l = \alpha(\dot{x}_r + \dot{x}_l)$$

(26)

Where x_r and x_l correspond to the displacements of right and left vocal folds (from the center of the larynx, or the glottis) respectively. It then follows that $\dot{x}_i$ corresponds to velocity, and $\ddot{x}_i$ to acceleration. β describes the "damping" effect of the oscillator (is proportional to an underlying damping coefficient); α relates to subglottal pressure; Δ (with $|\Delta| < 2$) captures "asymmetry" (26).

The ability to characterize asymmetry is especially important for our study (and studies with similar purposes) since we hypothesized that ALS patients have affected vocal fold dynamics and that we would be able to detect the pathology from their voices. Asymmetric oscillation is one of the most common forms of irregularity in vocal fold oscillations. It can underlie abnormal voice properties such as hoarseness (15), which is reported in ALS patients' voices (37).

After choosing a model, the next step is to acquire a set of necessary parameters for each of the speakers—since we want to be able to differentiate speaker groups from their voices, the parameters need to be reasonably accurate to the true values.

3.1.2 Adjoint Least-Squares (ADLES) Algorithm

Typically, parameters to the phonation models are clinically measured or observed (45) to ensure accuracy. These measurements need to be made for each individual since using averaged parameters will yield vocal fold dynamics estimations that are not specific to each individual—the direct opposite of our goal. However, we do not have access to this type of data—our dataset consists of voice recordings from different speakers (as will be discussed

in the following section). In addition, the traditional practice requires manual measurements with specific equipment, which might cause discomfort to patient larynx and thus not be sustainable for ALS diagnostics in reality. Considering these situations, this project utilizes an *Adjoint Least-Squares (ADLES) algorithm* (45) proposed in the Singh Lab to *infer optimal parameters directly from voice recordings*. A code repository containing an implementation of the ADLES algorithm authored by other members of the Singh Lab is used, with only minor adjustments made to suit the dataset of this project.

Given a short segment of recording (usual window length=50ms), the ADLES algorithm works by minimizing the squared error between (airflow) volume velocities derived from 2 separate methods: one based on inversely filtered vocal signal (i.e. voice recordings), while the other based on displacements of the vocal folds, modeled with a chosen physical model (in this case, one presented in Lucero et al. (2015) (26)). As analytical (close form) solutions would be infeasible for the objective, the algorithm employs gradient descent and iteratively finds more and more optimal parameters (the aforementioned β, α and Δ) to the model for each vocal signal. We can then supply the parameters estimated by the ADLES algorithm to Lucero et al. (2015) (26)'s model. By iteratively solving the coupled, non-linear system of equations with these parameter settings, we obtain phase-space trajectory of variables in the model which corresponds to vocal fold dynamics (including information on vocal fold oscillation trajectory, velocity and acceleration) (45). See Figures 3.1 and 3.2 for visual representations of some outputs of the analysis described above.

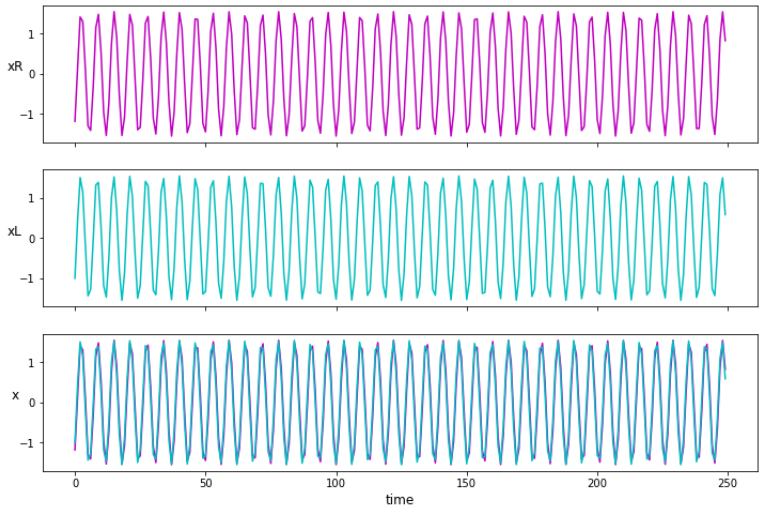

FIGURE 3.1: Estimated vocal folds oscillation by time. The top plot shows the oscillation of the right fold; the middle shows that of the left fold. The bottom plot is created by superimposing the top and middle plots.

For healthy voices, their phase-portraits (behavior in the system's phase space) appear regular (in a qualitative sense, since some small irregularity and asymmetry can exist in healthy

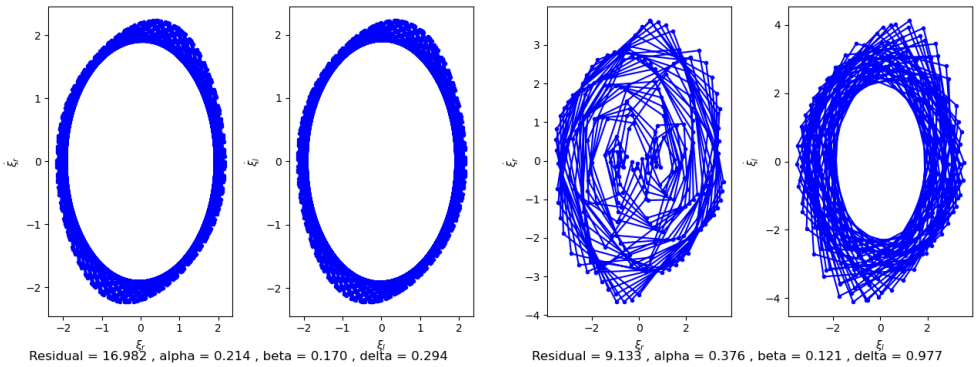

FIGURE 3.2: Estimated vocal fold velocity is plotted against estimated displacements of the left and right vocal folds for 2 distinct analysis windows, obtained by solving Lucero et al.'s model (26), with parameters supplied by the ADLES algorithm. The left panel shows highly periodic and in-sync vocal fold oscillation while the right shows aperiodic and asymmetric oscillations.

voices) and symmetrical between the left and right vocal folds. For pathological voices, we expect to observe highly irregular and non-symmetrical phase portraits.

We can come up with an intuition about the algorithm by considering the nature of speech sounds and voice recordings. As air is passed through the oscillating vocal folds, a source signal is produced and then modulated by the supra-glottal vocal tract. Finally, the filtered signal (essentially pressure differences over time) is released, captured by microphones and finally transformed into digital signals. In the ADLES algorithm, on one hand, we "inverse" the filtered signal to estimate the source signal, and on the other, we describe the source signal based on a bio-physical model of phonation. Naturally, our objective is to solve for parameters of the model that could bring the 2 representations of the source signal as close as possible.

3.2 Building the Feature Sets

Given estimated vocal fold dynamics data for a speech recording, we can extract relevant features to test our hypothesized separation. In this project, we explored 2 overall sets of features: 1) *simple statistical measurements* and 2) *Phase-space characterizations.*

3.2.1 Set 1: Simple Statistical Measurements

Statistical measurements (e.g., mean, maximum, variance, etc.) were first taken for the vocal fold displacements time series data. Though these features are conceptually simple, they are

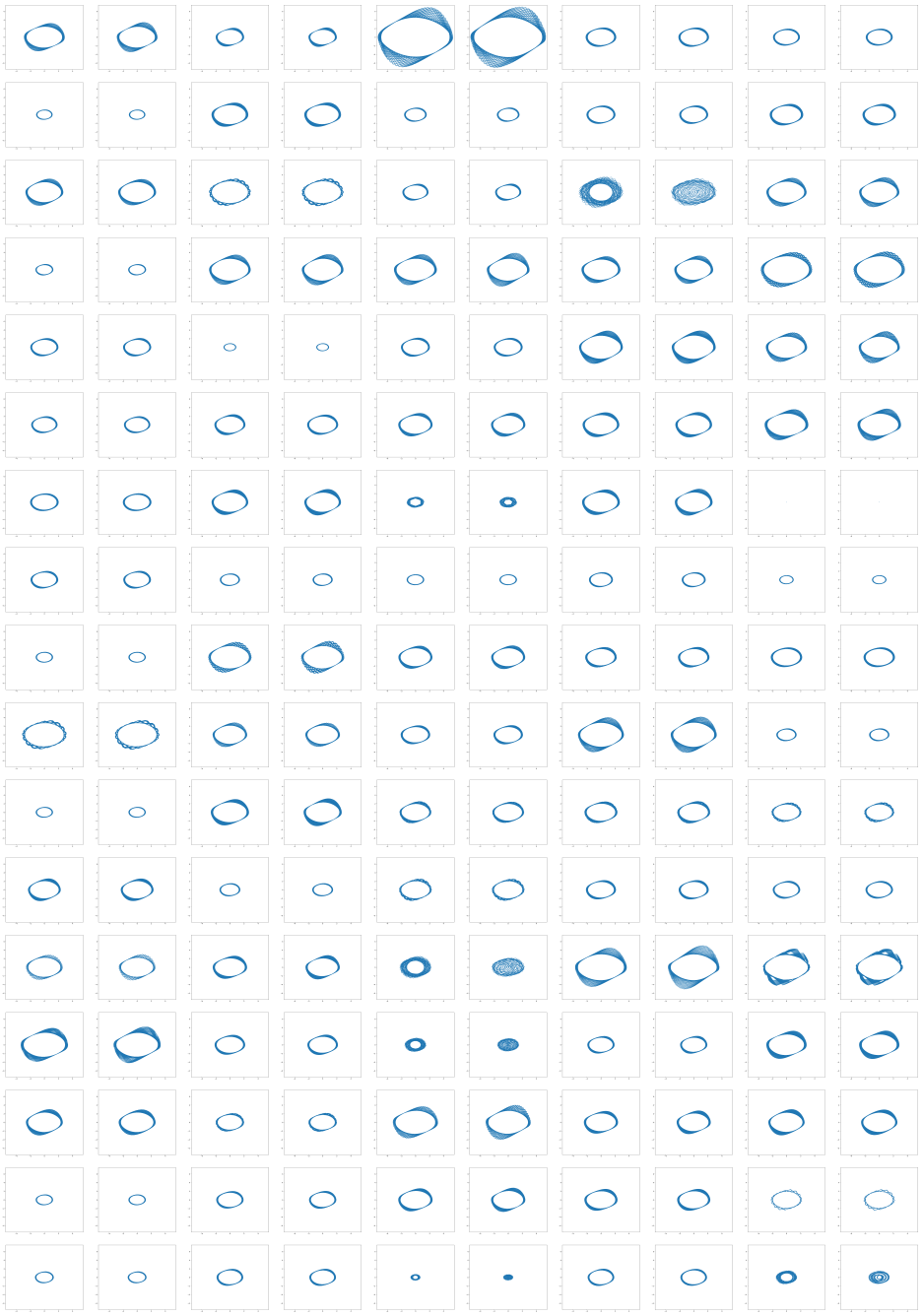

FIGURE 3.3: A montage of vocal fold oscillation trajectories in successive analysis windows for a recording. The x- and y-axis are normalized across all frames. Every 2 panels correspond to the estimation for the left and right (in order) vocal fold in each frame.

suitable starting points for exploration: they are easy to extract and can yield preliminary insights (about the distribution of data) that could in turn guide our later explorations. A total of 10 measurements were implemented and could be further divided into 2 groups (according to the kind of data that they are generated from):

Group 1 ($n = 4$): *average*, *max*, *min*, and *variance* of vocal fold displacements
Group 2 ($n = 6$): *average*, *max*, *min*, *variance*, *center* and *skew* of max range of displacement (MROD)[1]

Since the motion of the left and right vocal folds are generated separately, each of the above measurements is calculated for each side respectively, resulting in a total of $10 \times 2 = 20$ features for each recording. Among these measurements, *average*, *max*, *min*, and *variance* are calculated based on the generic definition of these terms. *center* and *skew* are calculated based on 10-bin histograms generate from MRODs from all analysis windows of a given recording. To clarify the construction of these 10-bin histograms, recall that in the previous section, we described that the analysis window length is set to 50ms. A recording of 5s would then consists of 100 analysis windows and thus 100 trajectories of vocal fold oscillation and dynamics. The 10-bin histogram for said 5s-recording would be a histogram for 100 MROD values, each calculated from a displacement trajectory of an analysis window.

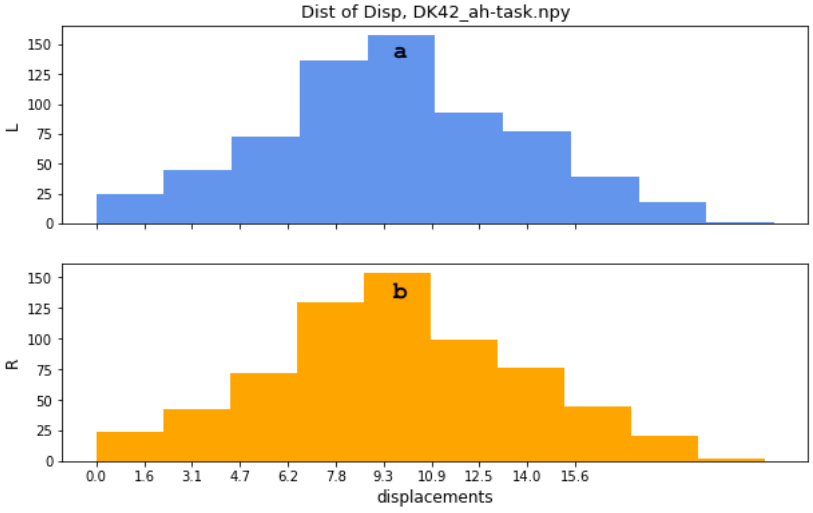

FIGURE 3.4: 10-bin histograms of MRODs for a particular frame of the recording `DK42_ah_task`. The top (in blue) is generated for left vocal fold displacement data and the bottom (in orange) for those of the right side. The highest frequencies bins are labeled **a** and **b** respectively.

Give such a histogram, *center* refers to the average of all MRODs of the highest frequency bin (the bin with the largest count); *skew* reflects the skewness of the histogram and is implemented based on the definition of Pearson's Second Coefficient for Skewness:

$$3 \cdot (\text{average} - \text{median}) / \text{standard deviation} \quad (40)$$

[1]Given a time series of displacements (D) of a trajectory (belonging to a single analysis window), it MROD is defined as: `D.max()-D.min()`.

The larger the magnitude of *skew*, the more skewed or asymmetrical the histogram is.

It should be clear to the readers that this set of features, though simplistic, can help us investigate if the vocal fold motion is affected in ALS patients and if we can distinguish their motion from that of a non-ALS individual. With maximum displacements as well as the maximum range of displacements, we will be able to infer if patients' have increased/reduced range of vocal fold motion; with measurements such as *variance* and *skew*, we can explore if there is a higher degree of irregularity in the voice of ALS patients. This increased irregularity, should it exist, will then lead us to consider the physiological causes and thus a deeper understanding of the mechanisms and symptoms of ALS.

3.2.2 Set 2: Phase-space Characterization

Previously, we discussed that estimated vocal fold dynamics data is the output of a non-linear dynamical system; therefore, we naturally availed ourselves of established methods for analyzing trajectories in phase-space from the field of dynamics. The second set of features are developed with regard to these existing methods. In particular, we explored the Lyapunov and Hurst coefficient spectra. The following sections outline the features derived from these spectra respectively.

3.2.2.1 Features based on Lyapunov Spectra

Lyapunov exponents is a measurement of "speed of divergence" of a potentially chaotic system (14). For a discrete-time system, such as ours, the Lyapunov exponent λ is defined by:

$$\lambda \approx \frac{1}{n} ln \left| \frac{f^n(x_0 + \epsilon) - f^n(x_0)}{\epsilon} \right| \textit{(Eq1)} \quad (14)$$

$f(x)$ denotes a function (or time series), x_0 is the starting point and ϵ is a small perturbation. $f^n(x_0)$ is the value of the function after n steps, starting from x_0. Essentially, we want to measure how "far" the system diverges if we start with 2 points, x_0 and $x_0 + \epsilon$, that are in close proximity. Intuitively, with a regular vocal fold oscillation trajectory (whose phase-portrait is a torus), its Lyapunov exponent(s) will be low compared to that (or those) of an irregular, pathological oscillation trajectory. We can thus capture the "stability" of vocal fold dynamics of ALS and non-ALS individuals for comparison.

For this project, we implemented an algorithm for calculating Lyapunov exponents so that we can maximally customize the variables (or parameters) involved. The implementation takes

a trajectory, a step constant n (that specifies time steps of divergence), a distance constant d [2](based on ϵ and specifies the radius of search of "neighbors" of the starting points) and an index offset i [3]. For every point $X = (\dot{x}, \ddot{x})$ (velocity and displacement) in a trajectory, we find all of its neighbors bounded by $[\dot{x} \pm d, \ddot{x} \pm d]$ and are at least some i indices apart from X. We then take k steps for both X and its neighbor(s). If a pair of points exist after k steps, we can solve for an exponent λ with the formula specified by *Eq1*. We return the maximum of all the λ's acquired for a trajectory and finalize to the highest 5 values over all trajectories (of a single recording). The highest 5 exponents are chosen to provide a description of the fastest diverging behavior in the system.

3.2.2.2 Features based on Hurst Spectra

Additionally, we extracted features that are derived from Hurst exponents. In complexity theory, Hurst exponents are used for estimating attractors in dynamical systems and can quantify the strengths of such attractors, ranging from 0 to 1 (32). An exponent of value 0.5 indicates a complete lack of correlation (i.e., "perfect random motion") (32); a value above 0.5 indicates "trending" behavior (value is proportional to the strength of the trend) (24); a value below 0.5 indicates "mean-reverting" (24), meaning the times series tend to "regress" to its mean value. In sum, these values describe the level of autocorrelation or "long-term memory" in a system (24). With these features, we aim to further characterize the regularity and irregularity in the voices of ALS and non-ALS and explore this established analysis method for dynamics.

Given a delay time τ, the corresponding Hurst exponents ($H(\tau)$) can be found with the traditional rescaled range (R/S) analysis:

$$H(\tau) = \frac{log(\frac{R(\tau)}{S(\tau)})}{log(\tau)} \quad (32)$$

where

$$R(\tau) = \max_{1 \le t \le \tau} X(t, \tau) - \min_{1 \le t \le \tau} X(t, \tau)$$

[3]By default, k is set to 50, $d = 0.2$, $i = 25$ for trajectories of length 250. Empirically, varying k values does not have much impact on the output exponents.

[3]Until the last k points, since there wouldn't be a valid value for these starting points after k steps.

$$S(\tau) = \sqrt{\frac{1}{\tau}\sum_{i=1}^{\tau}(x_i - \bar{x}(\tau))^2} \quad (32)$$

For convenience and feasibility (considering the number of time series in our dataset), I adopted the simplified method presented by Lewinson (24) for my own implementation. The intuition behind the method is to find the exponents as a measure of the correlation between variance in data with some lag time τ and τ itself.

Mathematically, we want to find H such that:

$$Var(\tau) \propto \tau^{2H} \quad (24)$$

In the code implementation, given a trajectory time series, we consider all τ from 2 (steps) to a `max_lag` constant that can be specified. For this project, `max_lag` is set to 50. We record {average, maximum, variance} $\times$ {x, $\dot{x}$} $\times$ {left, right}, a total of 12 exponents for each time series.

3.3 Data Collection & Experiments Setup

In order to test if our proposed analysis method and features are effective at detecting ALS through voice, we utilized the dataset that was used for the summer project: a case-control dataset that consists of recordings of prolonged phonation of the vowel [a] (as in "ah"). The original dataset contains 39 case recordings (N_{male}=24, N_{female}=15) and 13 control recordings (N_{male}=1, N_{female}=12). Age and disease progression information are not provided as a part of the dataset and are therefore unknown and impossible to control for. Indeed, there is a speaker gender imbalance in our set, but since 1) the proposed method does not rely on fundamental frequency (F0, which is often significantly different among biological male and female speakers), 2) from low feature weight reported in the summer study, we consider this imbalance minor to the current project. Upon applying the pipeline for estimating vocal fold dynamics, 7 case recordings and 3 control recordings were ruled out because the algorithm timed out before converging on a set of parameters. As a result, a total of 42 recordings (32 cases and 10 controls) are used for this project.

Subsequently, all 37 features (20 statistical + 17 phase-space) discussed above are obtained from the vocal fold dynamics estimated from each voice recording.

3.3.1 Classification trials

A 10-way cross-validation trial with 70:30 train-test split[4] was arranged for Set 1 and Set 2 features, respectively. Random Forest Classifiers (with max-depth=2) are chosen for the classification tasks because of their simplicity (to avoid overfitting). In addition, RFCs provide information about relative feature importance and enable us to compare the effectiveness of features in and across sets as well as those used in the summer study.

[4] There is no overlap between training and test data.
[4] Ditto.

4

Results & Analyses

In this section, we will first present distributions of each of the afore-described features. In this way, we want to offer the readers some initial ideas of the ranges and separation of the features on their own. Then, we will examine the results regarding the performances of the classifiers based on Set 1 and Set 2 features. Given the amount of data and results, we will embed some analyses and interpretations (of figures and tables) within each subsection (instead of culminating all the observations in a separate section). Towards the end of the section, we will also include a summary of the most essential observations and tie the results back to our hypotheses and project objectives.

4.1 Distributions of features

4.1.1 *disp* features

Based on Figure 4.1, we can make a few observations. First, the levels of separation of the 4 features are not consistent. On one hand, for *disp: {avg, max, min}*, case and control seem to overlap in range. However, we see most control values are centered (with regard to the distribution of all values) while case values occupy a much larger range. On the other hand,

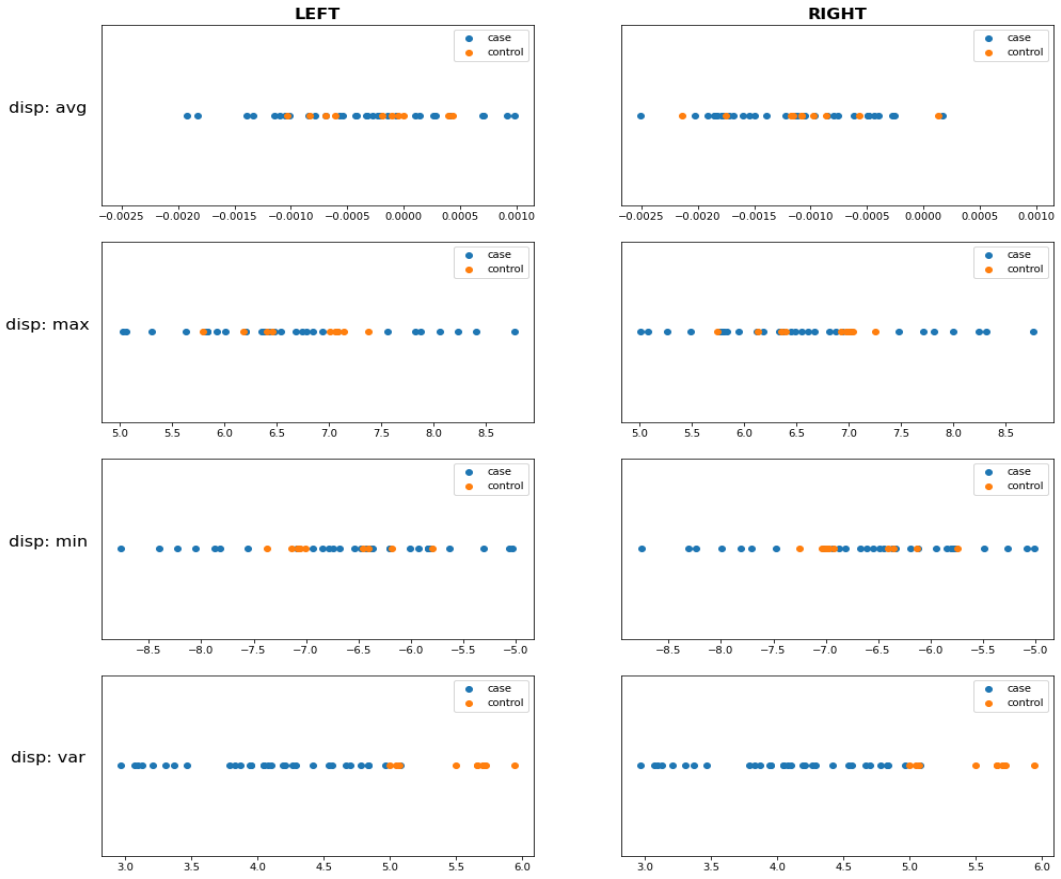

FIGURE 4.1: The distribution of *disp* features ($n = 4 \times 2 = 8$) among case and control voices. The left column shows measurements for the estimated displacements for the left vocal fold and the right for the right fold.

there appears to be high separation for *disp: var* for both left and right vocal folds. Second, we observe asymmetry of distribution in the left and right *disp: avg* in case data, which is in line with our prediction of asymmetry in the dynamics of cases—this supports our decision of using a physical model that can capture the asymmetry.

The centering of *disp: avg* might be interesting because it could reflect the level of "symmetry" of vocal fold motion: if a vocal fold oscillates consistently and periodically, we would expect to find the average displacements around 0 (which is what we observe for the controls). The distribution of case *disp:avg* values also seem to show sidedness: the values for the left seem slightly skewed towards the left and the right to the right (essentially the folds stay across the glottis more often). The biological implication of this phenomenon is not obvious.

Mann-Whitney U tests are applied and show that the distributions of *disp: var* for both sides are significantly different between the case and control groups. The significance remains after

[1] "**" denotes statistically significant after correction.

TABLE 4.1: p-values of *disp* features. (Corrected threshold $= 0.000625$)

Feature	(Raw) p-value
disp: avg_L	0.54491987
disp: avg_R	0.84777907
disp: max_L	0.35225022
disp: max_R	0.36774665
disp: min_L	0.35225022
disp: min_R	0.36774665
disp: var_L	3.81E-06**[1]
disp: var_R	3.81E-06**

applying Bonferroni correction.

4.1.2 *MROD* features

From Figure 4.2, we can likewise observe differences in distribution among features. In particular, there appear to be high separations in both left and right *MROD: {avg, var, (min)}* data. Notice most of the control values for both sides of *MROD: {avg, var}* are higher than those for cases. Higher *MROD: var* values are likely related to the higher *disp: var*. Higher *MROD: avg* values indicate that controls have a larger range of movements than cases. In other words, ALS might have some negative impact on the degree of movement of vocal folds.

Also, notice that most of the control values for *MROD: skew* within ± 0.1 while most values for cases are high in magnitude and in the positive range. As we described before (see 3.2.1), *MROD: skew* is a measure of skewness of the 10-bin histograms of MROD data, with larger magnitudes denoting higher skewness and its sign denoting the direction of skew. Thus, the distribution observed means that the histograms for cases tend to have more counts in higher MROD ranges (with regard to each individual) instead of showing more symmetry.

Mann-Whitney U tests reveal that the distributions of *MROD: {avg, var, center, skew}* for both sides between case and control are significantly different from each other. After applying Bonferroni correction, differences in both sides of *MROD: {avg, var}* and the left side of MROD: skew remain significant. It seems although the extremes of MROD (i.e. the maxima and minima) are not informative, most other *MROD* features provide some basis for separation. From the distributions, we might expect higher relative importance of these features in the classification trials.

[2] "*" denotes significance before correction.

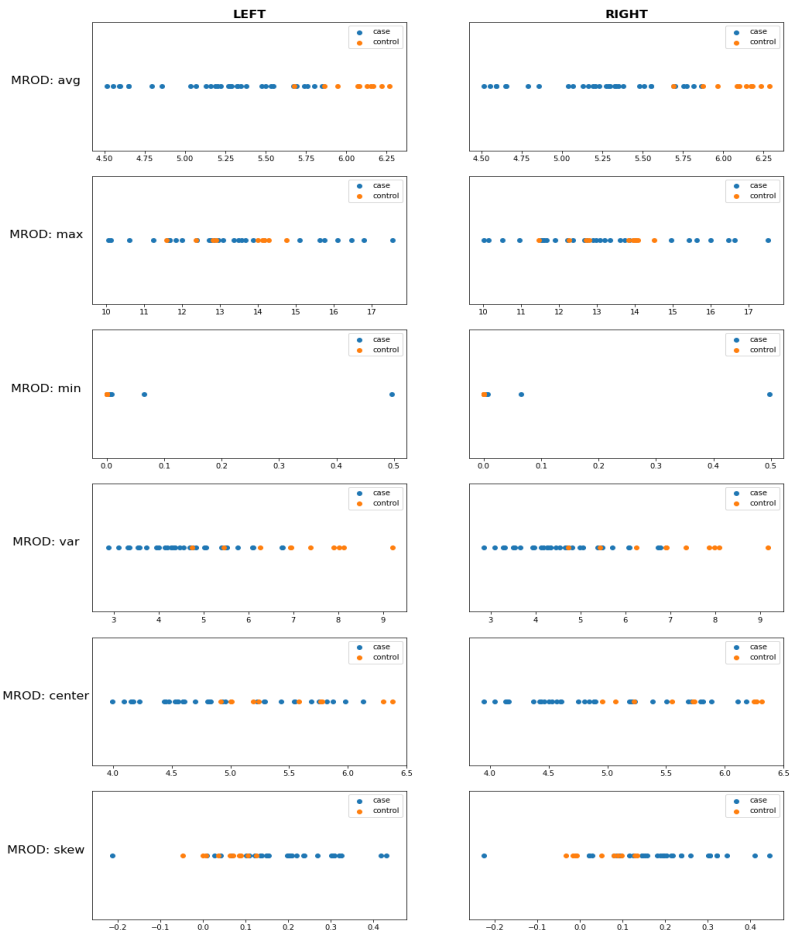

FIGURE 4.2: The distribution of *MROD* features ($n = 6 \times 2 = 12$) among case and control voices. The left column shows measurements of MROD for the left vocal fold and the right for the right fold.

TABLE 4.2: p-values of *MROD* features. (Corrected threshold ≈ 0.000417)

Feature	(Raw) p-value
MROD: avg_L	5.05E-06**
MROD: avg_R	5.81E-06**
MROD: max_L	0.35225022
MROD: max_R	0.36774665
MROD: min_L	0.16968963
MROD: min_R	0.15206822
MROD: var_L	4.90E-05**
MROD: var_R	4.90E-05**
MROD: center_L	0.01258194*[2]
MROD: center_R	0.00166044*
MROD: skew_L	0.00037296**
MROD: skew_R	0.00089291*

4.1.3 *Lyapunov* features

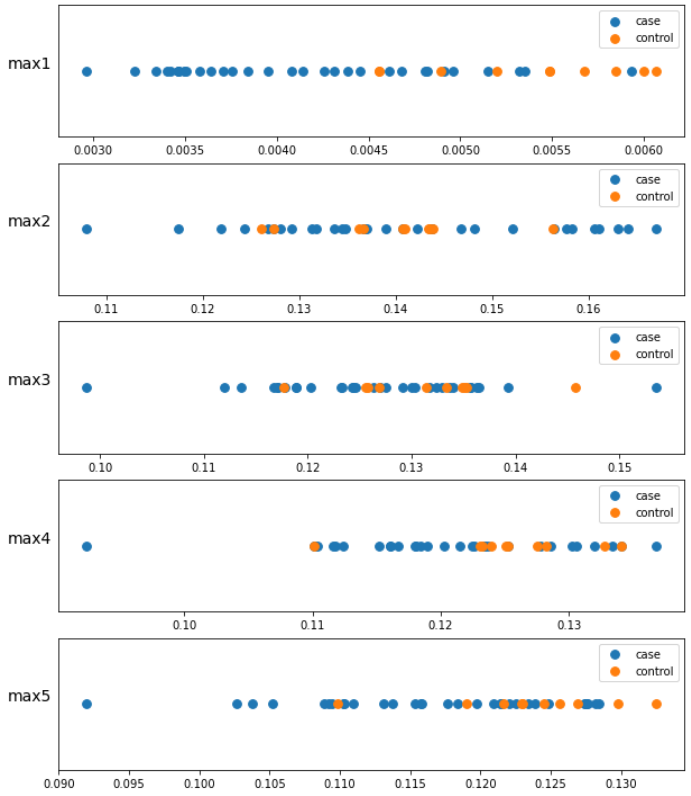

FIGURE 4.3: The distribution of *Lyapunov* features ($n = 5$) among case and control voices.

TABLE 4.3: p-vales of *Lyapunov* features. (Corrected threshold $= 0.00100$)

Feature	(Raw) p-value
Lyapunov: max1	0.00018719**
Lyapunov: max2	0.87097377
Lyapunov: max3	0.17905494
Lyapunov: max4	0.08949923
Lyapunov: max5	0.03473088*

From Figure 4.3 (below) it seems controls overall have higher *Lyapunov: max1* (which is the largest Lyapunov exponent for a given voice). This might seem somewhat counter-intuitive, considering the previous discussion on Lyapunov exponents. By inspection, there are no apparent trends or differences in the distributions of the other measurements between cases and controls. Mann-Whitney U tests reveal significant differences in the distributions of both *Lyapunov: {max1, max5}* in terms of raw p-values, though the significance for *Lyapunov: max5* disappears after applying Bonferroni correction.

4.1.4 *Hurst* features

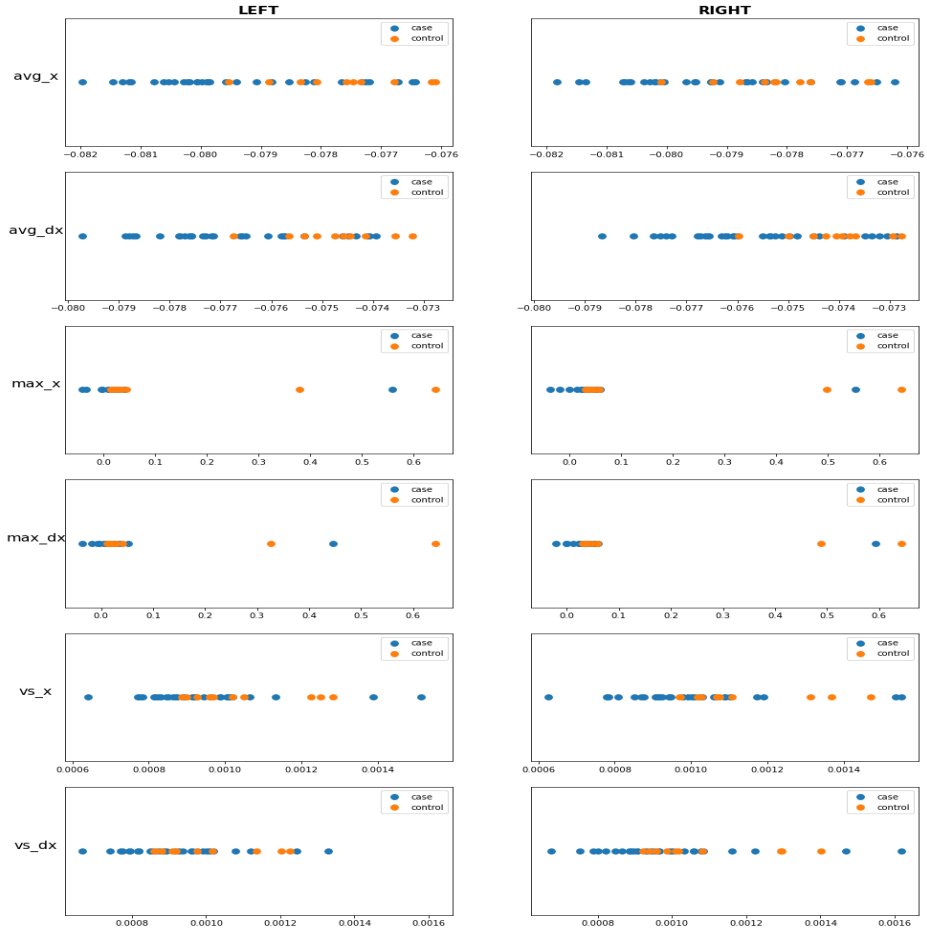

FIGURE 4.4: The distribution of *Hurst* features ($n = 6 \times 2 = 12$) among case and control voices. The left column shows the Hurst-exponent-related features for the left vocal fold and the right for the right fold.

As shown in Figure 4.4 below, controls appear to have slightly higher *Hurst: avg_dx* values for both sides. Such a trend might exist but is not obvious for *Hurst: avg_x*. Some differences in distribution between case and control values can also be observed for *Hurst: {max_x, max_dx}* (the distributions are highly similar between displacements and velocity data). Results from Mann-Whitney U tests partially corroborate with these observations—significant differences are detected for *Hurst: {avg, max, var}* for both displacement and velocity data, of both sides. However, these distinctions all disappear after applying Bonferroni correction.

TABLE 4.4: p-values of *Hurst* features. (Corrected threshold ≈ 0.000417)

Feature	(Raw) p-value
Hurst: avg_x_L	0.0036279*
Hurst: avg_x_R	0.02209855*
Hurst: avg_dx_L	0.0015004*
Hurst: avg_dx_R	0.00398567*
Hurst: max_x_L	0.01744163*
Hurst: max_x_R	0.02780151*
Hurst: max_dx_L	0.01258194*
Hurst: max_dx_R	0.02577353*
Hurst: var_x_L	0.0299656*
Hurst: var_x_R	0.02043841*
Hurst: var_dx_L	0.09521374
Hurst: var_dx_R	0.05307687

4.2 Classification results with Set 1 features

Random Forest Classifiers (max-depth=2) with Set 1 features ($\{disp\} \cup \{MROD\}, n = 20$) yielded an **average AUC-ROC of 99.6%** in the afore-described 10-way cross-validation setting. The minimum AUC-ROC across these 10 trials was 96.3%; the median and maximum AUC-ROC were both 100%. The overall variance was 0.00001.

We report the relative feature importance in the following figure. (For a report of the numerical values, see Appendix A.)

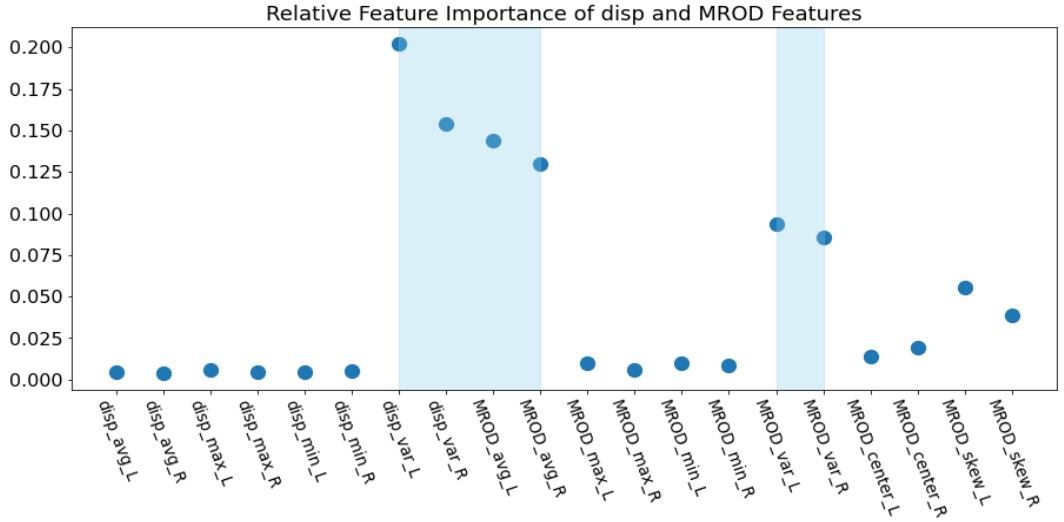

FIGURE 4.5: The relative importance of Set 1 features in the RFC. The highest-ranking features are highlighted.

The highest-ranking features in Set 1 are *disp: var* and *MROD: {avg, var}*, both sides. Together, they account for 81.0% of feature weight. Interestingly, we observe asymmetry of relative ranking between left and right measurements: for *disp* features, the left measurements seem to be slightly more important than those of the right, but the opposite is observed for *MROD* features; though the significance of such differences is unknown and could possibly be ignored.

4.3 Classification results with Set 2 features

Random Forest Classifiers (max-depth=2) with Set 2 features ($\{Lyapunov\} \cup \{Hurst\}$, $n = 17$) yielded an *average AUC-ROC of 82.3%* in the afore-described 10-way cross-validation setting. The minimum AUC-ROC across these 10 trials was 63.3%; the median and maximum AUC-ROC were 86.7% and 96.7%, respectively. The overall variance was 0.01312.

Again, we report the relative feature importance in the following figure. (For a report of the numerical values, see Appendix A.)

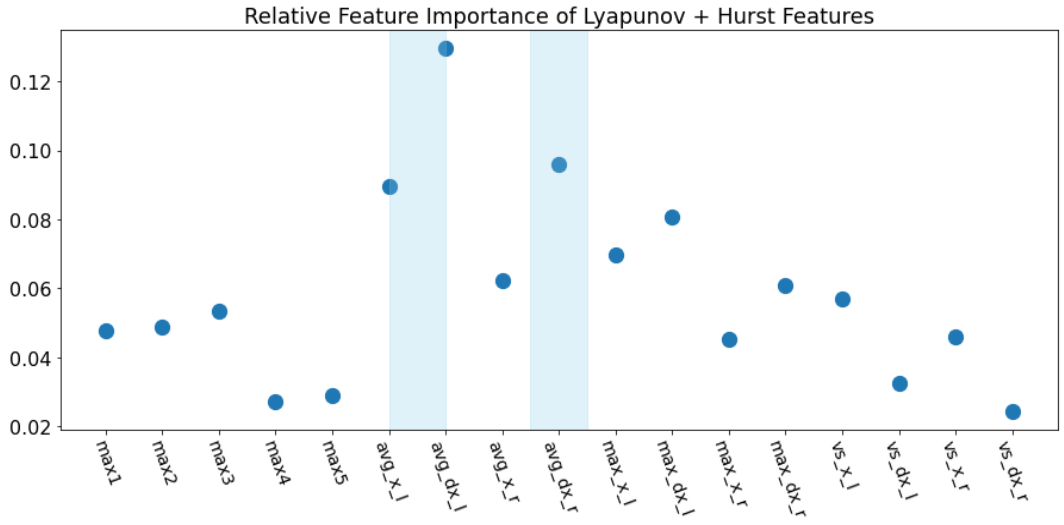

FIGURE 4.6: The relative importance of Set 2 features in the RFC.

The 3 highest-ranking features in Set 2 are *Hurst: {avg_dx_L, avg_dx_R, avg_x_L}*. Together, they account for 31.5% of relative feature weight. Features derived from Lyapunov exponents, in comparison, do not appear to be as effective in current classification trials.

4.4 Discussion & Summary

In this section, we presented the distributions of each of the features and the performances of classifiers built with Set 1 and 2 features, respectively. Here, we would like to highlight some of the observations.

First, the distributions of a subset of *disp* and *MROD* features indicate the effects of ALS on voice production and more fundamentally, vocal fold motion. In our dataset, cases show significantly lower *disp: var* and *MROD: {avg, var}* (in both the left and right vocal folds), suggesting a reduced range of oscillation as well as (capability of) motion variety. Including these features, we found a total of 8 features (out of 37) that distribute significantly differently between cases and controls after applying Bonferroni correction (22 before correction). Overall, more separation is seen in simple statistical features (Set 1). These findings support our hypothesis that ALS affects vocal fold motion and dynamics and we are capable of separating the 2 groups based on a subset of the features that we proposed.

Second, the classifiers based on our features yield promising results (comparable to those reported in previous studies, from a numerical perspective). In particular, RFC with Set 1 features is capable of distinguishing voices from ALS patients and non-ALS individuals with very high accuracy across runs. Features that had statistically significant differences in distributions (*MROD: {avg, var}* and *disp: var*) were ranked the most important among all Set 1 features. Classifiers with Set 2 features, comparatively, are not as accurate and stable (though the accuracy is still much higher above chance). We observe a much larger variance in the accuracy across runs. Interestingly, *Hurst: {avg, max}* turned out to be the most important for separation. Recall that *Hurst: {avg, max}* were not reported to show statistically significant differences in distribution across groups (after correction). On the other hand, *Lyapunov: max1* and (especially) *Lyapunov: max5* contributed less to effective separation, despite significant differences in distribution. This might be a result of the strict False Discovery correction in the analysis. It might also result from the simplicity of the physical model of phonation employed in this study. Again, these results show that features regarding vocal fold dynamics can be used to distinguish ALS voices. **Overall, we conclude that it is surprisingly easy to detect the presence of ALS with high accuracy based on features derived from vocal fold dynamics.**

5 Discussion

5.1 Comparison with previous studies

This project is similar to published studies in several ways. Like prior work, we adopted a case-control structure and explored the separability of the two groups in the feature space that we devised. Though the overall structure is the same, we introduced the use of vocal fold dynamics and novel features to the existing body of work. This is to fulfill our initial goal of innovation and exploration.

Our results corroborate with prior findings that ALS patients experience decreased range of vocal fold motion as well as abnormalities in vocal fold dynamics (38; 37). In particular, our results show a decreased variability in vocal fold motion through decreased variances in vocal fold displacements and maximum range of displacements in ALS patients compared to non-ALS controls. These observations likely echo the existing qualitative description that ALS patients show a limited ability to "make acoustic contrasts," especially as the disease progresses (22).

While direct comparisons of numerical values of classification accuracy are not always appropriate because of the differences in datasets and implementation, it is safe to say that our

method yield high to relatively high performances. Compared to the summer project [1], we see a marked improvement in accuracy for classifiers with Set 1 features. Thus, like previous work, we show that voice is a feasible source of data for the detection of ALS and that there is a wealth of features that could be extracted for analysis. We hope our findings will facilitate the use of vocal fold dynamics in the field of automatic detection of ALS (and ALS-like) diseases from voice.

5.2 Limitations

Though we demonstrated promising results, we want to point out a number of limitations of the current project. First of all, similar to many of the previous studies, our dataset is small (38), which impacts the generalizability and significance of our results, especially given the heterogeneity of this disease reported in the literature. ALS is inherently rare (comparatively speaking); in addition, voice recording data of ALS patients are not readily available. These reasons contribute to the small cohort sizes seen across our type of studies. However, as we noted, voice recording is relatively easy and cheap to acquire—we are hopeful that larger cohort sizes will be possible as medical centers or physicians take recordings alongside procedural diagnostic examinations and publish some of their data. Our dataset also suffers from imbalances between case and control numbers as well as with regard to gender. Though in the precursory study we did not find differences due to gender in the current dataset and our current analysis method does not rely on speaker gender information, the imbalance might still have impacted the results implicitly. Furthermore, the current recordings in the dataset are uniform, as all of them only contain a sustained phonation of an [a] (“ah”) vowel. This kind of recording is simplistic to collect but might have limited the range of vocal fold motion that is involved.

Second, our method is based on a single-mass model of phonation proposed by Lucero et al. (2015) (26). Though this model is capable of capturing the asymmetry of motion of the left and right vocal folds, there are many other factors and complexity that are not taken into consideration. For example, the vocal folds are not uniform (in terms of thickness, tissue properties, etc.) along the sagittal and transverse planes, so modeling them as single mas objects might have been an oversimplification.

Third, there is some lacking in our data analyses. In this project, we performed statistical tests on features and further examined their effectiveness in the context of classification trials. For

[1] Direct comparisons made since the summer project utilized the same structure, dataset and classification trial setup

some of the features, we also noted their physical or physiological relevance. However, we are yet to construct a more systematic characterization or to relate the estimated physical behavior of the folds and trends of such behavior back to more detailed underlying physiological and neurological mechanisms. Finding potential neurological explanations is especially relevant considering ALS is a neurodegenerative disease.

5.3 Future Work

Since our ultimate aim is to develop an "efficient and reliable" diagnostic tool for ALS, there are many enhancements that can be made. First, it is of paramount importance that we increase the size of both case and control groups, as well as improve sample balance with regard to gender, age, etc. In addition, we should try to include other patient information such as stage of disease progression, onset subtypes, relevant disease history, ethnicity, etc. ALS is a complicated and highly idiosyncratic disease; by collecting the outlined information, we can better control for potential confounding factors, match case and control groups, as well as achieve higher generalizability with our classifiers. At the same time, knowing more patient information might enable us to investigate how our proposed features change over the course of disease progression. This might confer useful insights as we search for underlying mechanistic causes of the observed abnormalities in voice dynamics. Second, as noted in the previous section, we can incorporate voice recordings of other phonation tasks and explore if vocal fold motion elicited in qualitatively different conditions could provide us with a richer dataset for analysis.

To address the limitation regarding the physical model we employed, we can use the ADLES algorithm in conjunction with other, more specific or sophisticated models of voice production. As mentioned in 3.1.1, there exists a long line of research on these models. For example, we could utilize a double or multi-mass model to more accurately capture the anatomical properties of the vocal folds. We could alternatively consider vocal fold motion in a 3D space—currently, we only characterize vocal fold motion in terms of displacements from the glottis (i.e. motion in 2D space).

As briefly noted in 2.2, ALS has a number of confusing diseases, such as Parkinson's disease (PD) and bulbar palsy. In this project, we merely distinguished the voices of ALS patients from non-ALS individuals. Would we see a clear distinction along our proposed features between the voices of an early-stage ALS patient and a PD patient? Differentiating between these diseases is difficult under existing criteria and procedures. It will be interesting to explore if

examining vocal fold dynamics can provide a high enough resolution for separation between these conditions, thus creating a footing in solving this clinical challenge.

Table 1: Relative feature importance of Set 1 features.

disp			
avg_L	*avg_R*	*max_L*	*max_R*
0.00399241	0.0111772	0.0069379	0.0037697
min_L	*min_R*	*var_L*	*var_R*
0.00353045	0.00546683	0.16725762	0.13985571

MROD			
avg_L	*avg_R*	*max_L*	*max_R*
0.17273227	0.15276863	0.01004331	0.0135628
min_L	*min_R*	*var_L*	*var_R*
0.00815652	0.00477268	0.09232693	0.08291031
center_L	*center_R*	*skew_L*	*skew_R*
0.0256859	0.0296961	0.03224492	0.0331118

Table 2: Relatvie feature importance of Set 2 features.

Lyapunov				
max1	*max2*	*max3*	*max4*	*max5*
0.04791625	0.0487193	0.05344315	0.02718908	0.028799

Hurst			
avg_x_L	*avg_dx_L*	*avg_x_R*	*avg_dx_R*
0.08974696	0.12958645	0.06224641	0.09607327
max_x_L	*max_dx_L*	*max_x_R*	*max_dx_R*
0.06975146	0.08069075	0.04526309	0.06076164
var_x_L	*var_dx_L*	*var_x_R*	*var_dx_R*
0.05712626	0.03245048	0.04582491	0.02441151

Bibliography

[1] Agosta, F., Al-Chalabi, A., Filippi, M., Hardiman, O., Kaji, R., Meininger, V., Nakano, I., Shaw, P., Shefner, J., Van Den Berg, L. H., et al. The el escorial criteria: strengths and weaknesses. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration 16*, 1-2 (2015), 1–7.

[2] Agurto, C., Pietrowicz, M., Eyigoz, E. K., Mosmiller, E., Baxi, E., Rothstein, J. D., Roy, P., Berry, J., Maragakis, N. J., Ahmad, O., et al. Analyzing progression of motor and speech impairment in als. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (2019), IEEE, pp. 6097–6102.

[3] Al Ismail, M., Deshmukh, S., and Singh, R. Detection of covid-19 through the analysis of vocal fold oscillations. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2021), IEEE, pp. 1035–1039.

[4] Allison, K. M., Yunusova, Y., Campbell, T. F., Wang, J., Berry, J. D., and Green, J. R. The diagnostic utility of patient-report and speech-language pathologists' ratings for detecting the early onset of bulbar symptoms due to als. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration 18*, 5-6 (2017), 358–366.

[5] Aronson, A. E., Winholtz, W. S., Ramig, L. O., and Silber, S. R. Rapid voice tremor, or "flutter," in amyotrophic lateral sclerosis. *Annals of Otology, Rhinology & Laryngology 101*, 6 (1992), 511–518.

[6] Boersma, P. Praat, a system for doing phonetics by computer. *Glot. Int. 5*, 9 (2001), 341–345.

[7] Chiaramonte, R., and Bonfiglio, M. Acoustic analysis of voice in bulbar amyotrophic lateral sclerosis: a systematic review and meta-analysis of studies. *Logopedics Phoniatrics Vocology 45*, 4 (2020), 151–163.

[8] Chiaramonte, R., Di Luciano, C., Chiaramonte, I., Serra, A., and Bonfiglio, M. Multi-disciplinary clinical protocol for the diagnosis of bulbar amyotrophic lateral sclerosis. *Acta Otorrinolaringologica (English Edition) 70*, 1 (2019), 25–31.

[9] Chitkara, D., and Sharma, R. Voice based detection of type 2 diabetes mellitus. In *2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)* (2016), pp. 83–87.

[10] Costa, J., Swash, M., and de Carvalho, M. Awaji criteria for the diagnosis of amyotrophic lateral sclerosis: a systematic review. *Archives of neurology 69*, 11 (2012), 1410–1416.

[11] De Vries, M., Schutte, H., Veldman, A., and Verkerke, G. Glottal flow through a two-mass model: comparison of navier–stokes solutions with simplified models. *The Journal of the Acoustical Society of America 111*, 4 (2002), 1847–1853.

[12] Deshmukh, S., Al Ismail, M., and Singh, R. Interpreting glottal flow dynamics for detecting covid-19 from voice. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2021), IEEE, pp. 1055–1059.

[13] Despotovic, V., Ismael, M., Cornil, M., Mc Call, R., and Fagherazzi, G. Detection of covid-19 from voice, cough and breathing patterns: Dataset and preliminary results. *Computers in Biology and Medicine 138* (2021), 104944.

[14] Dingwell, J. B. Lyapunov exponents. *Wiley encyclopedia of biomedical engineering* (2006).

[15] Eysholdt, U., Rosanowski, F., and Hoppe, U. Vocal fold vibration irregularities caused by different types of laryngeal asymmetry. *European Archives of Oto-rhino-laryngology 260*, 8 (2003), 412–417.

[16] Flanagan, J., and Landgraf, L. Self-oscillating source for vocal-tract synthesizers. *IEEE Transactions on Audio and Electroacoustics 16*, 1 (1968), 57–64.

[17] Gardner, T. J., and Magnasco, M. O. Sparse time-frequency representations. *Proceedings of the National Academy of Sciences 103*, 16 (2006), 6094–6099.

[18] Gouda, P., Ganni, E., Chung, P., Randhawa, V. K., Marquis-Gravel, G., Avram, R., Ezekowitz, J. A., and Sharma, A. Feasibility of incorporating voice technology and virtual assistants in cardiovascular care and clinical trials. *Current Cardiovascular Risk Reports 15*, 8 (2021), 1–8.

[19] Ishizaka, K., and Flanagan, J. L. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell system technical journal 51*, 6 (1972), 1233–1268.

[20] Johri, A., Tripathi, A., et al. Parkinson disease detection using deep neural networks. In *2019 Twelfth International Conference on Contemporary Computing (IC3)* (2019), IEEE, pp. 1–4.

[21] Kent, J. F., Kent, R. D., Rosenbek, J. C., Weismer, G., Martin, R., Sufit, R., and Brooks, B. R. Quantitative description of the dysarthria in women with amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research 35*, 4 (1992), 723–733.

[22] Kent, R. D. Vocal tract acoustics. *Journal of Voice 7*, 2 (1993), 97–117.

[23] Kiernan, M. C., Vucic, S., Cheah, B. C., Turner, M. R., Eisen, A., Hardiman, O., Burrell, J. R., and Zoing, M. C. Amyotrophic lateral sclerosis. *The lancet 377*, 9769 (2011), 942–955.

[24] Lewinson, E. Introduction to the hurst exponent- with code in python, May 2021.

[25] Li, Z., Huang, J., and Hu, Z. Screening and diagnosis of chronic pharyngitis based on deep learning. *International Journal of Environmental Research and Public Health 16*, 10 (2019), 1688.

[26] Lucero, J. C., Schoentgen, J., Haas, J., Luizard, P., and Pelorson, X. Self-entrainment of the right and left vocal fold oscillators. *The Journal of the Acoustical Society of America 137*, 4 (2015), 2036–2046.

[27] Martínez-Sánchez, F., Meilán, J. J. G., Carro, J., and Ivanova, O. A prototype for the voice analysis diagnosis of alzheimer's disease. *Journal of Alzheimer's Disease 64*, 2 (2018), 473–481.

[28] Mittal, R., Erath, B. D., and Plesniak, M. W. Fluid dynamics of human phonation and speech. *Annual review of fluid mechanics 45* (2013), 437–467.

[29] Norel, R., Pietrowicz, M., Agurto, C., Rishoni, S., and Cecchi, G. Detection of amyotrophic lateral sclerosis (als) via acoustic analysis. *bioRxiv* (2018), 383414.

[30] of Neurological Disorders, N. I., of Neuroscience Communications, S.-O., and Engagement. Amyotrophic lateral sclerosis (als) fact sheet, 2022.

[31] Pareek, V., and Sharma, R. K. Coronary heart disease detection from voice analysis. In *2016 IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS)* (2016), pp. 1–6.

[32] Piovesan, D., Carlson, J. N., et al. The hurst exponent-a novel approach for assessing focus during trauma resuscitation. In *2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)* (2018), IEEE, pp. 1–5.

[33] Ramig, L. O., Scherer, R. C., Klasner, E. R., Titze, I. R., and Horii, Y. Acoustic analysis of voice in amyotrophic lateral sclerosis: A longitudinal case study. *Journal of Speech and Hearing Disorders 55*, 1 (1990), 2–14.

[34] Rong, P., Yunusova, Y., Wang, J., and Green, J. R. Predicting early bulbar decline in amyotrophic lateral sclerosis: A speech subsystem approach. *Behavioural neurology 2015* (2015).

[35] Shubhangi, D., and Pratibha, A. Asthma, alzheimer's and dementia disease detection based on voice recognition using multi-layer perceptron algorithm. In *2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)* (2021), IEEE, pp. 1–7.

[36] Singh, R. *Profiling humans from their voice*. Springer, 2019.

[37] Tomik, J., Tomik, B., Wiatr, M., Składzień, J., Stręk, P., and Szczudlik, A. The evaluation of abnormal voice qualities in patients with amyotrophic lateral sclerosis. *Neurodegenerative Diseases 15*, 4 (2015), 225–232.

[38] Vieira, H., Costa, N., Sousa, T., Reis, S., and Coelho, L. Voice-based classification of amyotrophic lateral sclerosis: where are we and where are we going? a systematic review. *Neurodegenerative Diseases 19*, 5-6 (2019), 163–170.

[39] Wang, J., Kothalkar, P. V., Cao, B., and Heitzman, D. Towards automatic detection of amyotrophic lateral sclerosis from speech acoustic and articulatory samples. In *Interspeech* (2016), pp. 1195–1199.

[40] Weisstein, E. W. Pearson's skewness coefficients, n.d.

[41] Xie, H., Ma, F., Fan, D., Wang, L., Yan, Y., and Lu, P. Acoustic analysis for 21 patients with amyotrophic lateral sclerosis complaining of dysarthria. *Beijing da xue xue bao. Yi xue ban= Journal of Peking University. Health Sciences 46*, 5 (2014), 751–755.

[42] Xu, X., Shen, D., Gao, Y., Zhou, Q., Ni, Y., Meng, H., Shi, H., Le, W., Chen, S., and Chen, S. A perspective on therapies for amyotrophic lateral sclerosis: can disease progression be curbed? *Translational Neurodegeneration 10*, 1 (2021), 1–18.

[43] Yang, A., Stingl, M., Berry, D. A., Lohscheller, J., Voigt, D., Eysholdt, U., and Döllinger, M. Computation of physiological human vocal fold parameters by mathematical optimization of a biomechanical model. *The Journal of the Acoustical Society of America 130*, 2 (2011), 948–964.

[44] Zhang, T., Zhang, Y., Sun, H., and Shan, H. Parkinson disease detection using energy direction features based on emd from voice signal. *Biocybernetics and Biomedical Engineering 41*, 1 (2021), 127–141.

[45] Zhao, W., and Singh, R. Speech-based parameter estimation of an asymmetric vocal fold oscillation model and its application in discriminating vocal fold pathologies. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2020), IEEE, pp. 7344–7348.