Fractal Features for Automatic Detection of Dysarthria

Taylor Spangler[1] and N. V. Vinodchandran[1] and Ashok Samal[1] and Jordan R. Green[2]

Abstract— **Amytrophic lateral sclerosis (ALS) is an incurable neurodegenerative disease. Difficulty articulating speech, dysarthria, is a common early symptom of ALS. Detecting dysarthria currently requires manual analysis of several different speech tasks by pathology experts. This is time consuming and can lead to misdiagnosis. Many existing automatic classification approaches require manually preprocessing recordings, separating individual spoken utterances from a repetitive task. In this paper, we propose a fully automated approach which does not rely on manual preprocessing. The proposed method uses novel features based on fractal analysis. Acoustic and associated articulatory recordings of a standard speech diagnostic task, the diadochokinetic test (DDK), are used for classification. This study's experiments show that this approach attains 90.2% accuracy with 94.2% sensitivity and 85.1% specificity.**

I. INTRODUCTION

The assessment of speech is relevant to the diagnosis of a variety of progressive neurologic conditions, such as amyotrophic lateral sclerosis (ALS) because it can be the first presenting symptom [1]. Bulbar motor changes, (i.e., difficulty with speech or swallowing) are the first symptoms in approximately 30% of persons with ALS [2]. The detection of speech motor involvement in ALS and other neurologic disorders, however, currently relies on subjective measures based on clinicians' auditory perceptions. Clinicians' perceptions of speech can be inconsistent [3] and some symptoms of bulbar ALS cannot be easily detected without instrumentation [4]. Because the late detection of dysarthria can lead to the late detection of ALS, imprecision in speech assessment can significantly delay diagnosis.

The need for improved speech diagnostic tools has motivated recent work on the development of automated speech analyses. One obstacle to developing an automated system has been the need to manually pre-segment the relevant speech units prior to feature extraction. For example, recent attempts to detect abnormal speech movements in persons with Parkinson's required the manual parsing of vowel consonant vowel (VCV) recordings prior to analysis [5]. Manually segmenting these VCV recordings, each containing 10 or more individual utterances, is a time consuming process.

Rather than analyze tasks which require manual preprocessing, our approach extracts measurements from unsegmented, rapid repetitions of syllables. During this speaking task, which is often referred to as a diadochokinetic task (DDK), participants are asked to produce the maximum number of syllable (e.g., "tah" and "pah") as rapidly and accurately as possible in a single breath. This task is widely used in differential diagnosis and the identification of speech muscle system impairments, because DDK irregularities are often seen in speakers with dysarthria [6].

II. RELATED WORK

While a wide range of features for speech analysis has been developed, little research has been conducted on using fractal analysis to detect or describe dysarthric speech. Fractal analysis has been increasingly used to understand disorder in several aspects of human physiology, which behave like complex systems. For example, it has been shown that a fractal measure called the *fractal scaling exponent* (FSE) can be used in differentiating between healthy and disordered populations. For heart rate time series, FSE has been used to quantify the difference between healthy hearts and hearts with severe congestive heart failure [7].

In Neuroscience, computing the FSE of neuronal oscillations was shown to be correlated with the severity of depression. It has also been used to describe differences in brain activity between early stage Alzheimer's and healthy brain activity. The FSE has also been used as a biomarker indicating that a patient is responding to treatment. For example, to show that a medication altered the brain activity of patients with epilepsy to be more similar to the brain activity of a healthy control group [8].

In regards to biomechanical systems like gait analysis, the FSE has been used to distinguish between healthy subjects and subjects with Parkinson's disease [9], and to differentiate between old and young subjects [10]. The FSE has also been applied to physical therapy. For example, altering the FSE of an audio recording played for a patient was shown to alter the fractal structure of a person's gait, in both diseased and healthy subjects [11], [12]. Such studies have led to a broader theory describing healthy human movement as that which exhibits an appropriate amount of self-similar variability. This theory suggests that healthy movement demonstrates stability, while allowing for variability due to adaptation to changes in the environment. Thus a change in FSE could indicate a change in the ability to adapt to the environment [13], [14], [15].

A. Detrended Fluctuation Analysis

One existing method for approximating the FSE of a time series is called *detrended fluctuation analysis* (DFA). The DFA algorithm consists of 6 steps: profiling, segmentation,

*This research was partially supported by the National Institute on Deafness and Other Communication Disorders (NIH-NIDCD) R01 DC009890, and R01 DC0135470

[1]Department of Computer Science and Engineering, University of Nebraska - Lincoln, Lincoln, Nebraska, USA {tspangle/vinod/samal}@cse.unl.edu

[2]MGH Institute of Health Professions, Boston, Massachusetts, USA jgreen2@mghihp.edu

978-1-5090-4179-4/17/$31.00 ©2017 IEEE

trending, detrending, root mean square error (RMSE), and regression. Profiling is comprised of integrating over the entire length of the time series. Segmentation consists of partitioning the time series uniformly to a certain scale. Next, the trending step fits a polynomial to the partition, constructing a local trend. To detrend each partition, the local trend is subtracted, leaving the local deviations from the trend. Then, the (RMSE) of the average deviation across all partitions is computed. The first five computations are repeated for increasing scales, typically doubling each round. Finally, a linear regression is fit to the log-log plot of each scale and it's corresponding RMSE. The slope of the regression line is an approximation of the FSE. A more detailed description of this algorithm can be found in [7].

III. FRACTAL SPEECH FEATURES

While DFA has been used in speech analysis, it is typically applied directly to the acoustic wave. One such result combines DFA with recurrence period density entropy (RPDE) for classification of different speech pathologies [16]. The mel frequency cepstral coefficients (MFCCs) have also been used in conjunction with DFA to detect Parkinson's disease [17]. As the FSE is really analyzing the variability, this forces the tool to measure both the variability of the frequency and amplitude of the signal. This is unlike how DFA is used in other applications.

A. Fractal Jitter

To adapt DFA for use in dysarthria detection, we first looked at an existing measure, *jitter*. Jitter measures the variation in the fundamental frequency of an acoustic waveform, and has been shown to correlate with dysarthria [18]. In this paper we have developed a *fractal jitter* feature which looks at the self-similarity of the variations in the fundamental frequency over time.

The first step in extracting the fractal jitter is to ignore the non-speech portions of the signal. Because this measure is developed for DDK analysis, the recordings have little to no interruption and the amplitude of speech tends to significantly overpower any background noise. Thus, a simple voice activity detection (VAD) approach can be used to remove any non-speech portions of the signal. Next, the detected speech is analyzed to extract the consecutive zero-crossing intervals. The fractal scaling exponent of the consecutive zero crossing intervals is the fractal jitter.

The fractal jitter shows a significant correlation with the health of the speaker, having a correlation coefficient of -0.402. Using the Mann-Whitney U test to determine whether the fractal jitter of ALS speakers is greater than that of the healthy group resulted in a $p = 1.376 \times 10^{-6}$, indicating significantly different FSE distributions across the two groups. The distributions can be seen in Fig. 1

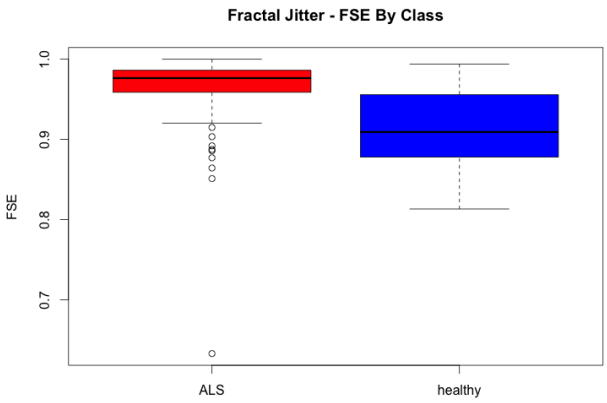

Fig. 1. A box and whisker plot showing the distribution of the fractal jitter for the ALS and Healthy classes

B. Fractal Articulatory Analysis

In addition to acoustic speech, articulatory recordings have been increasingly used to augment acoustic speech in clinical analysis. Articulatory recordings capture the positions $x, y,$ and z positions of the tongue tip, tongue back, upper lip, lower lip, and jaw articulators across the entire length of a DDK recording.

In order to normalize the sensor data to remove variance of physiological differences between speakers, procrustes normalization was used. This technique essentially rotates the three dimensional space in which the sensor points lie, so that the upper lip is directly above the lower lip. The points are then scaled to a unit size [19]. Here we used only the y (up/down) and z (front/back) coordinates of the sensors. The x (left/right) coordinate has been shown to carry minimal information for speech classification tasks [20].

Unlike the acoustic time series, the articulatory time series is both multidimensional and multivariate. Therefore, we implemented a recent extension of DFA to multivariate data, *multivariate detrended fluctuation analysis* (MVDFA) [21]. Our fractal articulatory measure uses the MVDFA on all sensors and dimensions, treating the different dimensions of a sensor as different variables. Thus MVDFA was applied to a 10 dimensional multivariate time series. This results in a *multivariate fractal scaling exponent* (MVFSE) for the articulatory recording. The distribution of this analysis for speakers in the two classes can be seen in Fig. 2

IV. EXPERIMENTS

A. Data Collection

To validate these new features, real world clinical samples were used. A total of 83 speakers were recorded, with 34 healthy speakers, and 49 speakers having been diagnosed with ALS. The ALS speakers were each evaluated and given a sentence intelligibility score, using the Sentence Intelligibility Test software [22]. The average intelligibility among ALS speakers is 96.0% (SD 8.3). More than 80% having $\geq$ 99% intelligibility, as most of the speakers are early in their diagnosis. The average age in the healthy group was 61.7 (SD 8.8) and the average age in the ALS group was 59.5 (SD 9.2), allowing for analysis of nearly identical age groups across classes.

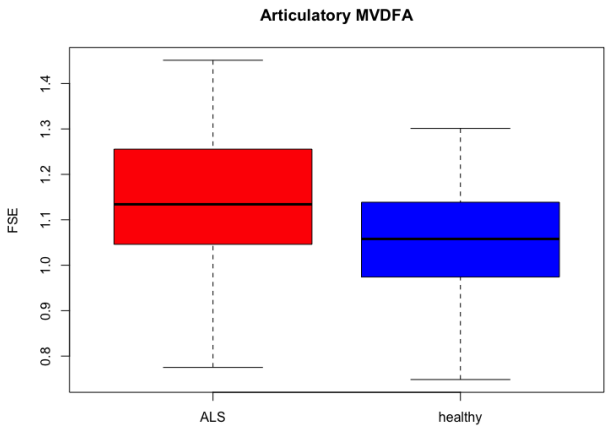

Fig. 2. A box and whisker plot showing the distribution of the MVFSE for the ALS and Healthy classes

The articulatory data was collected using the NDI Wave Speech Research System [23]. The tongue tip sensor was placed on the tongue within 1cm of the tongue tip. The tongue posterior sensor was placed 4cm from the tip. The upper and lower lip sensors were placed on the center of each lip. The jaw sensor was placed on the jaw in line with the corner of the right lip. Additionally, a reference sensor was placed in the center of the forehead. All sensors are measured relative to the position of the reference sensor. These sensors have millimeter precision, and capture positions at a 100hz sample rate.

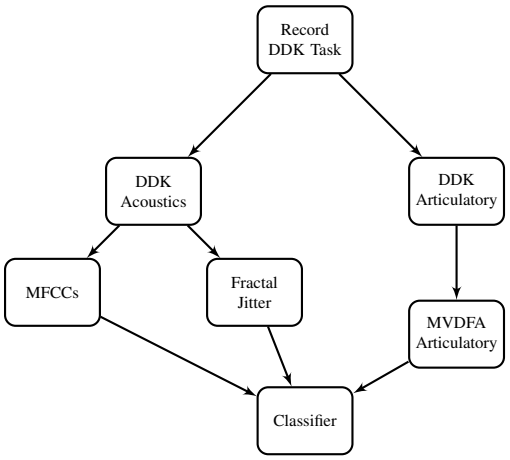

Fig. 3. An outline of the algorithm for classifying dysarthria.

B. Classification

For the purposes of computing the fractal jitter, an R program was written to extract the sequence of zero crossing intervals, and an existing R implementation was used to compute the FSE [24]. A new implementation of the MVDFA algorithm was written in python for computing the MVFSE of the articulatory recordings [21].

In addition to the two new fractal features presented in Section III, three existing feature sets widely used for detecting disordered speech were used: jitter, MFCCs, and the RPDE. Because the MFCCs of a timeseries are also a timeseries, we averaged the MFCCs across the time domain to condense them to a single set of coefficients for each recording.

The classification algorithm used for this task is Extreme Gradient Boosting (XGBoost). XGBoost has been shown to work very well for a wide variety of problems, from high energy physics event detection to customer behavior prediction [25]. It has also been applied to clinical data sets [26].

A schematic for our dysarthria detection algorithm using the fractal features presented in Section III is shown in Fig. 3. The algorithm takes in an acoustic recording and associated articulatory recording, taken from sensors placed on the 5 different articulators, featurizes the different time series, and classifies the sample.

Several different feature configurations are explored. First, as a baseline, we used the RPDE, MFCCs, and jitter as input to XGBoost. Next, we augmented these features by adding the fractal jitter and MVFSE.

V. RESULTS AND ANALYSIS

To validate our model with the given feature sets, leave one out cross validation was used. Unfortunately, the data set is significantly unbalanced in favor of ALS samples. To minimize this, we first randomly selected a balanced subset of the two classes (e.g. 30 randomly chosen samples from each class). We then performed leave one out cross validation on this randomly subsampled set. This was repeated with 100 random subsampled training and validation sets. The average accuracy, sensitivity, and specificity can be seen in TABLE I.

While the baseline (BASE) gives 82.6% accuracy, the disparity between the sensitivity and specificity clearly indicates that it is much more accurately classifying dysarthric speech than healthy speech. Combining the baseline features with fractal jitter (BASE+FJ) does give a significant improvement in the accuracy, but still leaves a significant gap between ALS and healthy classification. As the distributions of the articulatory MVDFAs suggest, adding the articulatory MVDFA feature to the baseline (BASE+MV) results in a less significant increase in accuracy than fractal jitter. Given that the articulatory recordings are collected at a much lower sample rate, it is possible that this is because there is less variability resolution. Still, combining both of these fractal features with the baseline, we observed improvement over only using one or the other. Thus, it is clear that the variability in the fractal jitter is not completely characterized by the variability in the articulatory and vice versa.

Finally, we utilized XGBoost's ability to rank the features based on how frequently they are used to separate the two classes. This allowed us to remove both the RPDE and jitter features from our model, as they were routinely not used to differentiate between the two classes by the classifier. Therefore using only the MFCCs, fractal jitter, and articulatory MVDFA features (MFCCs+FJ+MV), we were able to achieve the highest accuracy of 90.2%.

TABLE I

THE AVERAGE ACCURACY, SENSITIVITY, AND SPECIFICITY OF THE DIFFERENT FEATURE SETS OVER ALL VALIDATION SETS (BEST IN BOLD FONT). HERE BASE IS THE SET OF RPDE, JITTER, AND MFCCS, +FJ MEANS INCLUDING FRACTAL JITTER, +MV MEANS INCLUDING MULTIVARIATE, AND MFCCS IS THE SET OF MFCCS

Features	Avg. Accuracy	Avg. Sensitivity	Avg. Specificity
BASE	82.6%	88.6%	73.7%
BASE+FJ	84.5%	91.8%	79.5%
BASE+MV	82.9%	90.3%	77.6%
BASE+FJ+MV.	85.4%	93.0%	82.3%
MFCCs+FJ+MV	**90.2%**	**94.2%**	**85.1%**

Overall, we have shown that DDK sessions can be used to accurately detect dysarthria in ALS patients who still have very high intelligibility. We have also shown that this can be done in a completely automated fashion, not requiring any manual preprocessing of speech samples.

VI. CONCLUSIONS

This study presented two novel features for detecting speech dysarthria in ALS speakers who still have high sentence intelligibility. These features along with several existing features were leveraged into a fully automated process for detection of dysarthria utilizing the XGBoost classification algorithm. TABLE I clearly indicates that adding either fractal feature significantly improved the accuracy. Combining these features further improves the accuracy, achieving best results when used solely in conjunction with the MFCCs, ignoring the RPDE and traditional jitter measure, 90.2% accuracy (94.2% sensitivity, 85.1% specificity).

Unlike other existing automatic diagnostic tools, which require manual preprocessing of the samples for classification, this approach is fully automated. While the sample size used in this research is not large by machine learning standards, the number of patients is significant for this kind of clinical application.

This approach may also be useful in discerning different levels of ALS severity. In future work we plan to investigate how this algorithm performs on later stages of diagnosis, using longitudinal data.

REFERENCES

[1] B. Tomik and R. J. G. Professor, "Dysarthria in amyotrophic lateral sclerosis: A review," *Amyotrophic Lateral Sclerosis*, vol. 11, no. 1-2, pp. 4–15, 2010.

[2] L. J. Haverkamp, V. Appel, and S. H. Appel, "Natural history of amyotrophic lateral sclerosis in a database population. validation of a scoring system and a model for survival prediction." *Brain : a journal of neurology*, vol. 118 (Pt 3), pp. 707–19, 1995.

[3] R. D. Kent, "Hearing and believingsome limits to the auditory-perceptual assessment of speech and voice disorders," *American Journal of Speech-Language Pathology*, vol. 5, no. 3, pp. 7–23, 1996. [Online]. Available: + http://dx.doi.org/10.1044/1058-0360.0503.07

[4] J. R. Green, Y. Yunusova, M. S. Kuruvilla, J. Wang, G. L. Pattee, L. Synhorst, L. Zinman, and J. D. Berry, "Bulbar and speech motor assessment in als: Challenges and future directions," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 14, no. 7-8, pp. 494–500, 2013, pMID: 23898888. [Online]. Available: http://dx.doi.org/10.3109/21678421.2013.817585

[5] J. Wang, P. V. Kothalkar, B. Cao, and D. Heitzman, "Parkinson's condition estimation using speech acoustic and inversely mapped articulatory data," in *in Interspeech*, 2015.

[6] Y.-T. Wang, R. D. Kent, J. R. Duffy, and J. E. Thomas, "Analysis of diadochokinesis in ataxic dysarthria using the motor speech profile program™," *Folia phoniatrica et logopaedica : official organ of the International Association of Logopedics and Phoniatrics (IALP)*, vol. 61, no. 1, pp. 1–11, 04 2009.

[7] C. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger, "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series," *Chaos*, vol. 5, no. 1, 1995.

[8] R. Hardstone, S.-S. Poil, G. Schiavone, R. Jansen, V. Nikulin, H. Mansvelder, and K. Linkenkaer-Hansen, "Detrended fluctuation analysis: A scale-free view on neuronal oscillations," *Frontiers in Physiology*, vol. 3, p. 450, 2012.

[9] M. Kirchner, P. Schubert, M. Liebherr, and C. T. Haas, "Detrended fluctuation analysis and adaptive fractal analysis of stride time data in parkinson's disease: Stitching together short gait trials," *PLoS ONE*, vol. 9, no. 1, p. e85787, 2014.

[10] J. M. Hausdorff, "Gait dynamics, fractals and falls: Finding meaning in the stride-to-stride fluctuations of human walking," *Human movement science*, vol. 26, no. 4, pp. 555–589, 08 2007.

[11] J. P. Kaipust, D. McGrath, M. Mukherjee, and N. Stergiou, "Gait variability is altered in older adults when listening to auditory stimuli with differing temporal structures," *Annals of Biomedical Engineering*, vol. 41, no. 8, pp. 1595–1603, 2013.

[12] N. Hunt, D. McGrath, and N. Stergiou, "The influence of auditory-motor coupling on fractal dynamics in human gait," *Scientific Reports*, vol. 4, pp. 5879 EP –, Aug 2014, article.

[13] N. Stergiou, Y. Yu, and A. Kyvelidou, "A perspective on human movement variability with applications in infancy motor development," *Kinesiology Review*, vol. 2, no. 1, pp. 93–102, 2013.

[14] N. Stergiou, R. T. Harbourne, and J. T. Cavanaugh, "Optimal movement variability: A new theoretical perspective for neurologic physical therapy," *Journal of Neurologic Physical Therapy*, vol. 30, no. 3, 2006.

[15] L. M. Decker, C. Moraiti, N. Stergiou, and A. D. Georgoulis, "New insights into anterior cruciate ligament deficiency and reconstruction through the assessment of knee kinematic variability in terms of nonlinear dynamics," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 19, no. 10, pp. 1620–1633, 2011.

[16] M. Little, P. McSharry, I. Moroz, and S. Roberts, "Nonlinear, biophysically-informed speech pathology detection," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 2, May 2006, pp. II–II.

[17] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1264–1271, May 2012.

[18] M. Vieira, F. McInnes, and M. Jack, "On the influence of laryngeal pathologies on acoustic and electroglottographic jitter measures," *Journal of the Acoustical Society of America*, vol. 111, no. 2, pp. 1045–1055, 2002.

[19] J. Wang, J. R. Green, and A. Samal, "Across-speaker articulatory normalization for speaker-independent silent speech recognition," in *2014 International Speech Communication Association INTERSPEECH Proceedings*, 2014, pp. 1179–1183.

[20] J. Wang, A. Samal, J. R. Green, and F. Rudzicz, "Sentence recognition from articulatory movements for silent speech interfaces," in *ICASSP*, 2012.

[21] H. Xiong and P. Shang, "Detrended fluctuation analysis of multivariate time series," *Communications in Nonlinear Science and Numerical Simulation*, vol. 42, pp. 12 – 21, 2017.

[22] K. Yorkston, D. Beukelman, M. Hakel, and M. Dorsey, "Speech intelligibility test," Madonna Rehabilitation Hospital, Lincoln, Neb, USA, 2007.

[23] J. J. Berry, "Accuracy of the ndi wave speech research system," *Journal of Speech, Language, and Hearing Research*, vol. 54, no. 5, pp. 1295–1301, 2011.

[24] W. Constantine and D. Percival, *Fractal Time Series Modeling and Analysis*, 2016.

[25] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," *CoRR*, vol. abs/1603.02754, 2016.

[26] P. Hahn, E. Cenik, K.-J. Prommersberger, and M. Muehldorfer-Fodor, "Machine learning as a tool for predicting insincere effort in power grips," *bioRxiv*, 2016.